

The Early History of Probability Forecasts: Some Extensions and Clarifications

ALLAN H. MURPHY*

Prediction and Evaluation Systems, Corvallis, Oregon

(Manuscript received 12 May 1997, in final form 20 October 1997)

ABSTRACT

Heretofore it has been widely accepted that the contributions of W. E. Cooke in 1906 represented the first works related to the explicit treatment of uncertainty in weather forecasts. Recently, however, it has come to light that at least some aspects of the rationale for quantifying the uncertainty in forecasts were discussed prior to 1900 and that probabilities and odds were included in some weather forecasts formulated more than 200 years ago. An effort to summarize these new historical insights, as well as to clarify the precise nature of the contributions made by various individuals to early developments in this area, appears warranted.

The overall purpose of this paper is to extend and clarify the early history of probability forecasts. Highlights of the historical review include 1) various examples of the use of qualitative and quantitative probabilities or odds in forecasts during the eighteenth and nineteenth centuries, 2) a brief discussion in 1890 of the economic component of the rationale for quantifying the uncertainty in forecasts, 3) further refinement of the rationale for probability forecasts and the presentation of the results of experiments involving the formulation of quasi-probabilistic and probabilistic forecasts during the period 1900–25 (in reviewing developments during this early twentieth century period, the noteworthy contributions made by W. E. Cooke, C. Hallenbeck, and A. K. Ångström are described and clarified), and 4) a very concise overview of activities and developments in this area since 1925.

The early treatment of some basic issues related to probability forecasts is discussed and, in some cases, compared to their treatment in more recent times. These issues include 1) the underlying rationale for probability forecasts, 2) the feasibility of making probability forecasts, and 3) alternative interpretations of probability in the context of weather forecasts. A brief examination of factors related to the acceptance of—and resistance to—probability forecasts in the meteorological and user communities is also included.

1. Introduction

When reference is made to early advocates of the explicit treatment of uncertainty in weather forecasts or to the results of the first attempt to express forecasts in terms of probabilities, it is the work of the Australian meteorologist W. E. Cooke in 1906 that is most often cited (e.g., Murphy 1985, 337). Cooke's pioneering contributions in this area are indeed noteworthy (see section 4a). Recently, however, it has come to light that at least some issues related to the treatment of uncertainty in weather forecasts were raised prior to 1900 and that probabilities (or odds) were used in weather forecasts much earlier—actually, more than 200 years ago. Moreover, the precise nature of the contributions made by various individuals remains somewhat unclear. In view of the current interest in the quantification of uncertainty

in forecasts on all space and time scales, with ensemble forecasting representing a recent manifestation of this interest, extension and clarification of the early history of probability forecasts seems warranted.

The overall purpose of this paper is to trace the early history of the recognition and treatment of uncertainty in weather forecasts in somewhat greater detail than has been attempted heretofore. In describing these early developments, an effort is made to distinguish between different approaches to the treatment of uncertainty as well as different interpretations of the forecasts themselves. This historical study helps to clarify the nature of Cooke's contribution in this area and highlights important contributions made by several other individuals. The study focuses on the period from the late eighteenth century to approximately 1925.

For convenience the overall period is divided into two primary subperiods: (a) before 1900 and (b) after 1900. The first period is further subdivided into the presynoptic era (before 1860) and the postsynoptic era (after 1860). The treatment of uncertainty in forecasts in the late eighteenth and nineteenth centuries is illustrated in section 2, and this section also summarizes an early discussion of the rationale for probability forecasts. Section 3 highlights relevant work in the 1900–

* Deceased.

Corresponding author address: Barbara G. Brown, Research Applications Program, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: bgb@ncar.ucar.edu

25 period, with particular reference to the noteworthy and multifaceted contributions of W. E. Cooke, C. Hal-lenbeck, and A. K. Ångström. Section 4 consists of a very brief overview of developments in this area since 1925. A discussion of various issues related to the rationale for probability forecasts and the formulation and interpretation of such forecasts is presented in section 5. The issues examined include the rationale for probability forecasts, the feasibility of making probability forecasts, and interpretations of probabilities in the context of weather forecasting, as well as the acceptance of—and resistance to—probability forecasts by the meteorological and user communities.

2. Before 1900

a. *Presynoptic era (<1860)*

In his historical review of statistical concepts and methods in meteorology in the eighteenth and nineteenth centuries, Sheynin (1985) describes the use of qualitative, and, in some cases, quantitative, expressions of uncertainty in weather forecasts produced by J. Dalton (Dalton 1793) in the United Kingdom and J. B. Lamarck (Lamarck 1800–11) in France. The forecasting methods employed by Dalton and Lamarck are not identified. Nevertheless, Dalton's forecasts include statements such as "the probability of rain was much smaller than at other times" and "the probability of a fair day to that of a wet one is as ten to one." In the latter case, we encounter an example of forecast uncertainty expressed in terms of odds.

Lamarck evidently produced forecasts for various weather elements over 6–9-day periods up to a few months in advance during the first decade of the nineteenth century. Considerable use was made of the term "probable" in these forecasts, which according to general usage at the time implied a probability of occurrence of somewhat greater than 0.5. Moreover, in at least one instance Lamarck introduced a scale of qualitative probabilities; namely, very great, great, average, and indeterminable (close to 0.5). The extent to which this scale was actually used is unclear. According to Sheynin, Lamarck reported that readers of his annual reports indicated a preference for forecasts expressed in a less uncertain manner.

A more extensive search of the forecasting literature in the eighteenth and nineteenth centuries undoubtedly would uncover other examples of such forecasts. Although the use of probabilities or odds in weather forecasts during this era may not have been common practice, it is of considerable interest to see that at least some individuals felt a need to characterize the uncertainty in their forecasts explicitly. By today's standards, of course, the forecasting methods available during the presynoptic era were indeed quite primitive and subject to very substantial uncertainties. In any case, the use of probabilities, odds, etc., in their forecasts suggests that

individuals such as Dalton and Lamarck were quite comfortable with these probabilistic or quasi-probabilistic forecasts.

b. *Postsynoptic era (>1860)*

1) R. H. SCOTT

The creation of networks of weather stations in the United States and western Europe during the period 1850–70, and the transmission of the observations taken at these stations to central locations, greatly facilitated the preparation of synoptic charts and the formulation of real-time forecasts based on these charts. By the late 1860s efforts were undertaken to combine information gleaned from such charts with various physical principles (e.g., Buys Ballot's law) in order to produce short-range forecasts of weather conditions, especially storms. Scott (1869/1971, 340–341; see also Nebeker 1995, 44) reports an interesting example of the results derived from such efforts:

If we take the area from Valencia to Helder, and from Nairn to Rochefort, we find that whenever the difference of barometrical readings between any two stations is 0.6 in. on any morning, the chance is 7:3 that there will be a storm before next morning, whose direction will have been indicated by the [Buys Ballot's] law, somewhere within the area covered by our network of stations. On the other hand, the chance is 9:1 that no storm will come on without having given unmistakable signs of its approach by means of barometrical readings, even though the absolute difference observed may not have reached 0.6 in.

Once again, odds are used to describe the uncertainty in early weather forecasts.

2) FORECASTS AS "PROBABILITIES"

The newly created Signal Service in the United States issued its first forecasts and warnings in 1871, under the direction of C. Abbe (Whitnah 1961). It is very interesting to note that these warnings were labeled as "probabilities;" see Fig. 1. The warnings themselves contained no numerical probabilities; nevertheless, use of this label explicitly recognized the uncertainty inherent in such information. Evidently, it was viewed as desirable (in at least some quarters) to draw the attention of recipients of weather information to this characteristic of forecasts. A few years later the term "indications" was substituted for probabilities. However, it was not until 1889 that the term "forecasts" received official sanction (see Lorenz 1970). [Note: It may also of interest to point out that in the early 1870s the Chief Signal Office in Washington, D.C., was popularly known by the name "Old Probabilities" (Scott 1873/1971).]

Somewhat similar terminology was used in warnings issued by the Meteorological Department of the Board

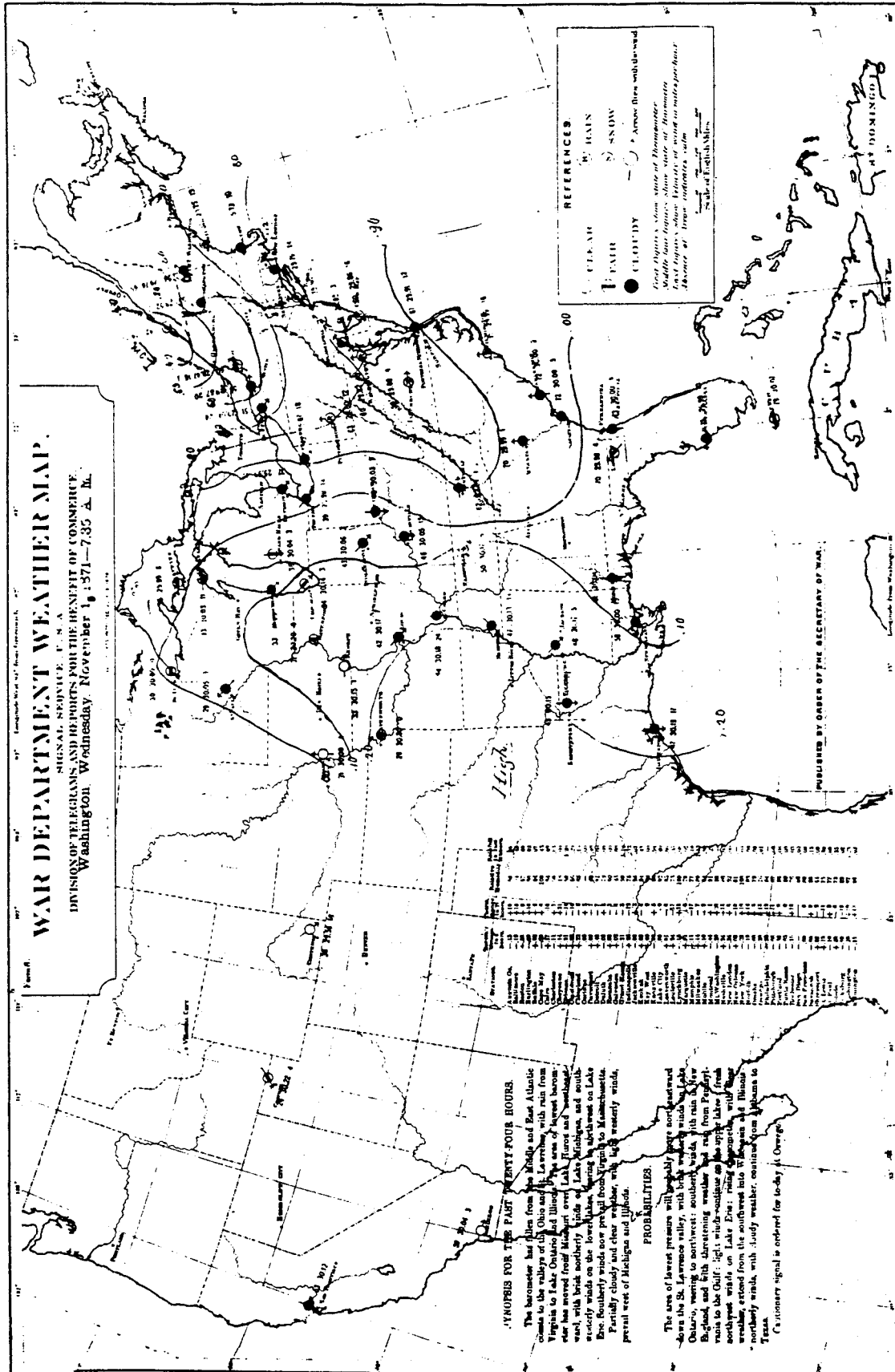


FIG. 1. The first weather map in the United States produced by telegraphic transmission of local observations (see Murphy and Williamson 1976, vii).

of Trade in the United Kingdom during the early days of its existence. To quote Scott (1869/1971, 336–337), “Admiral Fitzroy, in 1861, devised a code of meteorological telegraphy in cipher, and instituted a regular service, by means of which information of weather was received from stations on the coast, and issued to the public. This information consisted of occasional warnings of storms, and of forecasts of *probable* weather, which were published in the daily papers” (emphasis added).

3) W. S. NICHOLS

The paper by Nichols (1890) is remarkable in several respects. It focused primarily on issues related to forecast accuracy, forecast value, and the accuracy–value relationship (see Murphy 1996). However, questions related to uncertainty in forecasts and its proper treatment were also addressed. Interestingly, it is clear from the discussion in his paper that forecast uncertainty is an issue that Nichols believed arises quite naturally in the context of forecast use and value.

Nichols was concerned (inter alia) with event importance in an economic sense, its likelihood of occurrence, and the role that these two factors play in deciding whether or not to issue a forecast or warning. In this regard, he stated that, “(w)here the event is unimportant the probability should be relatively strong to justify predictions which may tend to discredit the reliability of the reports. But the greater the importance the smaller need be the probability involved, while to avoid the sacrifice of accuracy and confidence the problematic character of such predictions should, as far as possible, be indicated” (p. 391). To summarize his perspective on this issue, Nichols indicated that “(a) knowledge of the degree of certainty with which an event may be expected, increases the value of the information” (p. 391).

3. Early twentieth century: 1900–25

a. W. E. Cooke

As indicated in section 1, the work of Cooke (1906a,b) is usually cited as the first attempt to treat the problem of uncertainty in weather forecasts explicitly. Hopefully, it is now clear that this topic was actually addressed, explicitly or implicitly, by several individuals prior to 1900. Here we summarize Cooke’s papers, with the overall aim of providing a clearer picture of his contributions to this topic.

Cooke (1906a) approached the problem of uncertainty in forecasts from the perspective of forecaster confidence. Specifically, he stated that, “(a)ll those whose duty it is to issue regular daily forecasts know that there are times when they feel very confident and other times when they are doubtful as to the coming weather. It seems to me that the condition of confidence or oth-

TABLE 1. Weights assigned to each item in forecasts for Southwest and Goldfields districts in western Australia in 1905 (Cooke 1906a).

Weight	Definition
5	We may rely upon this with almost absolute certainty.
4	We may rely upon this with tolerable certainty, but may be wrong about once in 10 times.
3	Very doubtful. More likely right than wrong, but probably wrong about 4 times out of 10.
2	Just possible, but not likely. If showers are indicated, for example, they will not be heavy even if they occur at all.
1	The barest possibility. Not at all likely.

erwise forms a very important part of the prediction, and ought to find expression. It is not fair to the forecaster that equal weight should be assigned to all his predictions and the usual method tends to retard that public confidence which all practical meteorologists desire to foster. It is more scientific and honest to be allowed occasionally to say ‘I feel very doubtful about the weather for to-morrow’ . . . and it must be . . . useful to the public if one is allowed occasionally to say ‘It is practically certain that the weather will be so-and-so to-morrow’ ” (p. 23).

Cooke’s solution to the problem of variability in forecaster confidence was to define a set of weights (see Table 1) that express various degrees of confidence (or “states of doubt or uncertainty”). The larger the weight, the greater the confidence expressed in the forecast. In practice, an appropriate weight was assigned to each component of the forecast. As an example of a weighted forecast for the Southwest district (Geraldton to Esperance): “Fine weather throughout (5) except in the extreme southwest where a few light coastal showers are possible (2). Warm or sultry for the present inland (4), but a cool change is expected on the west and southwest coast (4), which will gradually extend throughout (4).”

Cooke reported some results of applying his weighting scheme to 24-h forecasts formulated twice daily for two districts in western Australia during 1905; see Table 2. Cooke’s original table included only the frequencies of right and wrong forecasts; relative frequencies were added here to facilitate interpretation of the results. Note that the relative frequency of right (wrong) forecasts decreases (increases) monotonically as weight decreases (with the exception of the forecasts involving the smallest weights for the Goldfields district). Under the assumption that degree of correctness measures forecaster confidence, it is clear that the forecasters were able to assign weights that reflected their degree of confidence. Keeping in mind the definitions of the weights in Table 1, the fact that only 1%–2% of the weight 5 forecasts were wrong and that approximately 6% of the weight 4 forecasts were wrong is especially noteworthy.

Cooke’s ideas were pronounced impracticable by Gariott (1906). The response by Cooke (1906b) merits careful attention even today. It concluded as follows: “Now this is the point I wish to make clear. Those

TABLE 2. Frequency and relative frequency of right and wrong forecasts as a function of weight for (a) Southwest district and (b) Goldfields district (Cooke 1906a).

	Right		Wrong		Total
	Weight	Frequency	Relative frequency	Frequency	
(a) Southwest district					
5	435	0.989	5	0.011	440
4	573	0.938	38	0.062	611
3	131	0.775	38	0.225	169
2	24	0.571	18	0.429	42
1	6	0.545	5	0.455	11
(b) Goldfields district					
5	240	0.980	5	0.020	245
4	337	0.939	22	0.061	359
3	102	0.803	25	0.197	127
2	13	0.542	11	0.458	24
1	5	0.625	3	0.375	8

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. But if, on the other hand, I attach a figure which practically says ‘I’m sorry, but this is the best I can do for you to-day—do not attach too much importance to it,’ I eliminate beforehand the adverse opinion which a great number of incorrect forecasts must produce, and I raise the bulk of the predictions to their true value. In particular, I create a series, marked with the maximum figure, which the public finds to be almost invariably correct, and thus raise the value of this particular series enormously.”

Recently, Jeffrey (1992) has argued that Cooke’s scheme falls short of “true” probability forecasts. Specifically, he pointed out that Cooke made definite or unqualified statements (e.g., “fine weather” or “rain”) and then indicated his confidence in such statements by attaching weights to them. As a result, Jeffrey (1992, 47) refers to Cooke’s approach as “a halfway house between dogmatism and probabilism.”

The probabilistic nature of Cooke’s forecasts can be tested by considering complementary events. In the case of “true” probability forecasts, the probability of an event (e.g., rain) and its complement (no rain) must sum to one. Cooke’s weighting scheme does not always satisfy the conditions of this test, in the sense that it is not always clear which of the available weights should be attached to an event complementary (no rain) to the basic event (rain). As a result, it may be most appropriate to refer to Cooke’s weighted forecasts as *proto-probability* forecasts.

b. Probabilities from “objective” methods

The probabilistic or quasi-probabilistic forecasts identified in earlier sections of this paper were based largely on the judgments of forecasters (although un-

conditional and/or conditional climatological relative frequencies undoubtedly played an important role in some situations). By the late nineteenth and early twentieth centuries some attention was directed toward the development of so-called objective methods of weather forecasting (Kramer 1952). In some cases, these methods produced forecasts expressed directly in terms of the probabilities of specific weather events.

For example, Besson (1905) applied methods based on the use of contingency tables and scatter diagrams to produce short-range forecasts of precipitation occurrence. In this process, he explored the relationship between precipitation occurrence and various predictors (barometric pressure, wind speed and direction, etc.), individually and in pairs. The methods specified the probability of precipitation given the predictor value(s). Although Besson evidently recognized the advantage of expressing forecasts in terms of probabilities, he judged forecasting performance solely in terms of the fraction of correct forecasts (after the probability forecasts had been converted into rainfall–no rainfall forecasts). Similar studies were undertaken in Sweden by Rolf (1917) and in France by Dunoyer and Reboul (1921).

c. Other relevant contributions

1) O. VON MYRBACH

Von Myrbach (1913) identified a basic deficiency in weather forecasts, namely, their degree of incorrectness or lack of reliability, from the perspective of users of the forecasts. He then made a strong appeal for the inclusion in forecasts of information concerning their reliability. Interestingly, von Myrbach states that “it is almost incomprehensible that this increase in reliability has still not been realized, although in the present state of meteorology it is very easy to attain. The forecaster, of course, merely needs to communicate something to the public which he already knows and that is the ap-

proximate probability of success of the individual forecast.”

As a first step in this direction, von Myrbach suggested discriminating four grades of probability: 1) very great probability (almost certainty), 2) great probability, 3) less great probability, and 4) small probability. Quantitative interpretations of these qualitative probabilities were not discussed. Although it is unclear, this kind of information may have been included in some forecasts issued in Austria around this time.

Von Myrbach concluded his paper as follows: “I believe that the very purpose of weather forecasting would be furthered extraordinarily by this importation if the possibility were offered to the public to separate those cases in which it is actually able to rely on the forecast, first from those in which it is only obliged to bear the forecast in mind and further from those in which it is better to rely on the particular local practical knowledge of the weather.”

2) CHIEF SIGNAL OFFICER

The Signal Corps Meteorological Service of the American Expeditionary Forces in France in World War I issued forecasts that included a statement as to the probable accuracy of the forecast (Chief Signal Officer 1919). Probable accuracy was expressed in terms of the odds in favor of the forecast. As an example, “odds of ‘five to one’ indicated that in the opinion of the forecaster there were five chances to one in favor of the forecast being correct.” It is reported that the inclusion of this information made it possible “to make the forecast absolutely definite and such qualifications as ‘probable’ or ‘possibly’ have never been used.” No results of this innovative experimental forecasting program were reported.

d. C. Hallenbeck

An important event in the early history of probability forecasts occurred in 1920 when a paper by C. Hallenbeck appeared in *Monthly Weather Review*. This paper (Hallenbeck 1920) reported the results of a forecasting experiment in which forecasts of the occurrence of rainfall in a 36-h period were expressed in terms of numerical probabilities. The experiment was motivated by farmers’ inquiries to the U.S. Weather Bureau Office in Roswell, New Mexico, for information regarding “the chances of rain.” These farmers were concerned with rainfall events during periods in which they faced decisions related to irrigation and alfalfa harvesting.

The forecasts were based in part on a series of composite weather maps that specified the probability of rainfall as a function of various synoptic weather types, where the latter were defined in terms of surface pressure distributions over the southwestern United States. The composite charts served as a starting point for the forecasting process, with the final forecast involving

TABLE 3. Results of a rainfall probability forecasting experiment in 1919 for the Pecos Valley of New Mexico (Hallenbeck 1920).

Forecast probability	No. of forecasts	No. of rainfall occurrences	Percentage of rainfall occurrences
90–100	1	1	100
80–89	1	1	100
70–79	7	6	86
60–69	15	11	73
50–59	13	8	62
40–49	15	8	53
30–39	18	6	33
20–29	31	7	23
10–19	22	4	18
0–9	0	0	—

careful consideration, by the forecaster, of other factors such as local topography and prevailing weather conditions. Hallenbeck (p. 647) notes that in the final analysis, “it [i.e., the forecast] is the forecaster’s personal opinion regarding the chances of verification of a rain forecast or the non-verification of a fair weather forecast.”

Some results of this probability forecasting experiment for a 6-month period in 1919 are summarized in Table 3. First, it should be noted that a monotonic relationship exists between forecast probability (in intervals) and percentage of rainfall occurrence. That is, the relative frequency of rainfall occurrence decreases as the forecast probability decreases. Moreover, the forecast probabilities and percentages of rainfall occurrences are in fairly close agreement, indicating that the probability forecasts are quite reliable.

From the data in Table 3 it is also possible to determine the skill of Hallenbeck’s probability forecasts. As a measure of skill, we computed a skill score (SS) based on the Brier score (BS) (Brier 1950), a quadratic measure of accuracy, using as a standard of reference (indicative of a no-skill forecast) a constant forecast of the sample climatological probability of rainfall occurrence. Given that this latter probability is 0.423 (= 52/123) for these data, and that the values of BS for the probability forecasts and sample climatology are 0.191 and 0.244, respectively, it follows that the SS is 0.217. Expressed as a percentage, Hallenbeck’s probability forecasts exhibit a positive skill score value of 21.7% (between one-fifth and one-fourth of the skill of perfect forecasts). In summary, the degree of reliability and level of skill achieved in this very early experiment strongly supports the thesis that forecasters—with experience—should be able to quantify the uncertainty in weather forecasts in a reliable and skillful manner.

Whereas Jeffrey (1992) referred to Cooke’s forecasts as protoprobability forecasts, Hallenbeck’s forecasts were considered by Jeffrey to be “fully probabilistic.” The latter’s forecasts pass the complementary event test. That is, the probability of no rain and the probability of rain sum to one.

e. A. K. Ångström

Noteworthy papers on probability forecasting and the use/value of forecasts by A. K. Ångström appeared in 1919 (Ångström 1919) and 1922 (Ångström 1922). The contents of these papers were summarized recently by Liljas and Murphy (1994, hereafter LM94). To minimize the overlap between this review and LM94, attention is restricted here to Ångström's discussion of uncertainty in forecasts. Readers interested in the use/value of forecasts are referred to LM94 or Ångström's original papers for more detailed discussions of these issues.

It was Ångström's belief that forecasts of weather events "will always be afflicted with a certain degree of uncertainty." After an interesting discussion of the sources of this uncertainty, Ångström (1919) presented two examples in which statistical methods were used to estimate the probability of weather events (namely, rainfall and frost occurrence). In considering the issuance of warnings for storms (or other events), Ångström recommended that warnings be issued "according to a certain scale in which a specific formulation of the warning corresponds to a specific probability range for the occurrence."

In connection with an investigation of the use/value of forecasts in a prototypical warning problem, Ångström (1922) discussed issues related to the assessment (or estimation) of the probability of a storm or other relevant event. First, he argued that an idealized forecaster with unlimited experience could estimate the probability of the event by means of the relative frequency of its occurrence on those occasions in the past when similar antecedent weather conditions prevailed. According to this approach, probability is interpreted as a relative frequency (e.g., Winkler 1972, 10–15; Morgan and Henrion 1990, 48–50).

Recognizing the difficulties associated with applying this relatively simple and idealized approach in real-world situations, Ångström presented the following argument for a more pragmatic approach: "(I)t seems scarcely necessary to give P (the limiting relative frequency in the previous paragraph) the narrow definition previously indicated. We may define P simply as resulting from the combined experience, judgement, intuition, computation, and knowledge of the forecaster. In fact, the evaluation of P enters as a fundamental, but perhaps often too less conscious condition for forecasting in general." Here, Ångström appeared to be arguing for a subjective interpretation of probability, according to which the probability represents the forecaster's degree of belief in the event "storm (frost, . . .) today" (Winkler 1972, 15–18; Morgan and Henrion 1990, 48–50).

Ångström then discussed the issuance of warnings, in the course of which he made several insightful remarks related to the rationale for probability forecasts. First, Ångström pointed out the difficulties encountered when a forecaster is constrained to issue a warning in

categorical (i.e., nonprobabilistic) terms. This difficulty relates to the fact that the forecaster's knowledge of the user's loss function (e.g., cost–loss ratio) is generally insufficient to determine whether "issue warning" or "do not issue warning" is the best possible decision. Moreover, Ångström noted that "(w)hat makes the matter still more complicated is the fact that . . . [the loss function] . . . has very different values in different cases."

As a consequence of these considerations, Ångström arrived at the following conclusion: "The most appropriate system seems therefore to be to leave to the clients concerned by the warning to form an idea of the ratio [cost–loss ratio] . . . and to issue the warnings in such a form that the larger or smaller probability of the event gets clear from the formulation. The client may then himself consider if it is worthwhile to make arrangements of protection or to disregard a given warning." In these few words, Ångström provided a clear statement of the economic component of the rationale for probability forecasts (see section 5).

4. After 1925

Relatively few if any developments related to probability forecasts can be identified during the period between the mid-1920s and the early 1940s. However, a rebirth of interest and activity in this area occurred in the mid-1940s, stimulated in part perhaps by the demands for specialized weather forecasts imposed by military and other activities during World War II. In the United States the work of G. W. Brier in the mid-1940s is particularly noteworthy. Brier (1944, 1946) picked up the threads of many aspects of the earlier work, including the discussion of the rationale for probability forecasts and the formulation of probability forecasts by objective and subjective methods.

Brier's pioneering work on objective methods of probability forecasting led to many statistical weather forecasting studies in the late 1940s and early 1950s. Further experimental work involving both objective and subjective probability forecasts in the 1950s and early 1960s culminated in the initiation of a nationwide program of operational precipitation probability forecasting in the United States in 1965. In the 1970s statistical forecasting methods capable of producing probability forecasts were developed, and/or expanded extensively, in several countries in conjunction with the implementation of operational systems of numerical–statistical forecasting. Moreover, the use of probabilities in weather forecasts throughout the world greatly increased during this period. For a bibliography containing a relatively complete set of references to work on probability forecasting (and closely related topics) up through the mid-1960s, the reader is referred to Murphy and Allen (1970).

Probability forecasts are now made on a regular basis in many countries, objectively and/or subjectively, for

a variety of weather variables, ranging from daily precipitation occurrence, cloud amount, etc., to monthly and seasonal precipitation/temperature anomalies and the landfall of hurricanes. For relatively recent reviews of probability forecasting and probability forecasts, the reader is referred to Ehrendorfer (1989), Murphy (1985), and Murphy and Winkler (1984).

In recent years numerical weather prediction models have been integrated forward in time, starting from an ensemble of different (but plausible) initial conditions, on an experimental or quasi-operational basis at several major forecast centers. Ensemble prediction allows the formulation of probability forecasts for a wide variety of variables ranging from weather regimes over regions to weather events at specific locations. For a recent review of the use of numerical models to quantify uncertainty in forecasts, see Ehrendorfer (1997).

5. Discussion and conclusions

The primary purposes of this paper were to (a) trace the early history of the use of probabilities, and other modes of expression of uncertainty, in weather forecasts and (b) clarify the nature of the contributions made by the first workers in this area. Contributions reviewed here included arguments related to the rationale for the explicit treatment of uncertainty in forecasts as well as results of experimental forecasting activities involving probability forecasts. From this historical study, it is now clear that the use of probabilities in weather forecasts can be traced back to the late eighteenth or early nineteenth centuries and that the rationale for probability forecasts was discussed as early as the late nineteenth century. Evidently, probability forecasts, in one form or another, were not “invented” in the twentieth century.

In this section we briefly discuss the early treatment of some basic issues related to probability forecasts and, in some cases, compare this treatment with their treatment in more recent times. These issues include (a) the rationale for probability forecasts, (b) the feasibility of making probability forecasts, and (c) alternative interpretations of probabilities in the context of weather forecasting. The paper concludes with a brief examination of issues concerning the acceptance of—and resistance to—probability forecasts in the meteorological and user communities.

The rationale for probability forecasts possesses two components, namely, a “scientific” component and an “economic” component (Murphy 1985). First, weather forecasts must be expressed in terms of probabilities (or equivalent modes of expression) to accommodate the uncertainty inherent in the forecasting process. In general, forecasts expressed in a nonprobabilistic format do not accurately reflect a forecasting system’s (i.e., a forecaster’s or a model’s) true state of knowledge concerning future conditions. Second, weather forecasts must be expressed in probabilistic terms to enable users of the forecasts to make the best possible decisions, as reflect-

ed by their levels of economic and/or social welfare. The welfare of users who make decisions on the basis of forecasts expressed in a nonprobabilistic format is unlikely to attain these same relatively high levels.

These components of the rationale for probability forecasts were identified in some of the very earliest contributions reviewed in sections 2 and 3. Uncertainty in the forecasting process—or, more precisely, the variability in the degree of certainty from occasion to occasion—was emphasized by Cooke and von Myrbach. These authors also argued, albeit qualitatively, that distinguishing between these degrees of certainty is of benefit to users. Even earlier, Nichols used a model of a simple decision-making problem to “demonstrate” these economic benefits. The two components of the rationale were joined, and supported by quantitative analyses, in the works of Ångström. It is in Ångström’s papers that we see, for the first time, what might be described as a full, modern treatment of the underlying rationale for probability forecasts.

It may be of some interest to compare the rationale for probability forecasts (or their equivalent) discussed by early workers and that set forth in more recent times in a particular context. For example, Cooke (1906a,b) emphasized the need to distinguish between forecasting occasions on which forecaster confidence is high and forecasting occasions on which it is low. Cooke’s weights provided a (relatively crude) scale of distinguishing levels of forecaster confidence. In relatively recent discussions related to the uncertainty in numerical weather prediction (NWP) forecasts, Tennekes et al. (1987) (and others) underlined the need to distinguish between those occasions on which NWP forecasts deteriorated rather slowly with lead time (relatively skillful forecasts) and those occasions on which they deteriorated rather rapidly with lead time (relatively unskillful forecasts). When the very substantial differences between the forecasting processes in 1905 and 1987 (e.g., the differences in the roles of forecasters and models) are taken into account, it is evident that these two arguments are really the same argument in different guises. In this regard, Cooke’s argument relates to the uncertainties in the knowledge base of weather forecasters, whereas Tennekes et al.’s argument relates to the uncertainties in the knowledge or information base that constitutes NWP models.

Another issue that must have arisen in this early era, and still arises today in connection with the formulation of probability forecasts in new contexts, is the feasibility of quantifying the uncertainty in forecasts. That is, is it possible to assess (i.e., estimate) the uncertainty in forecasts in a quantitative manner and in such a way that the predictions possess desirable characteristics as probability forecasts? In general, the dimensionality of verification problems involving probability forecasts is relatively large, implying that is necessary to examine numerous characteristics of performance in order to assess forecast quality in a reasonably complete manner (e.g.,

Murphy 1997; Murphy and Winkler 1992). For the purposes of this discussion, it is convenient to restrict the set of desirable characteristics to reliability and skill.

Although the experimental results presented by Cooke and Hallenbeck possess obvious limitations (e.g., the range of “probabilities” in Cooke’s case, the sample size in Hallenbeck’s case), these results appear to provide at least a tentative answer to the feasibility question. First, the relationship between observed relative frequency (fraction correct) and forecast probability (weight) in the case of Hallenbeck’s (Cooke’s) forecasts is monotonic. Moreover, in the case of Hallenbeck’s forecasts the correspondence between observed relative frequency and forecast probability is relatively good, especially when the small sample sizes and other considerations (e.g., lack of forecaster experience in quantifying uncertainty) are taken into account. The skill score calculation (see section 3d) shows that these forecasts also possess positive skill. Thus, notwithstanding the obvious need for further studies involving experimental probability forecasts, it appears that an affirmative answer could have been given to the question of the feasibility of making probability forecasts as early as 1920.

Probabilities in weather forecasts (e.g., the probability of precipitation occurrence) can be accorded at least two different interpretations. First, the probability can be interpreted as a (limiting) relative frequency. For example, “the relative frequency of precipitation under these meteorological conditions is 0.30.” In this approach, the relative frequency 0.30 is an estimate of the true but unknown probability of precipitation. Alternatively, the probability 0.30 can be interpreted as the forecaster’s degree of belief in the statement “precipitation today” [in the context of a simple lottery, it implies that the forecaster is indifferent between receiving \$.30 for sure or receiving \$1.00 (\$0.00) if precipitation occurs (does not occur)]. In this framework, it is not necessary to postulate the existence of true probabilities or to invoke the concept of limiting relative frequencies. It may be of interest to note that uncertainty is a property of the system under consideration (i.e., the atmosphere) in the case of the relative frequency interpretation, whereas uncertainty is a property of the individual making the forecast (i.e., the forecaster) in the case of the degree-of-belief interpretation. For further discussion, and comparison, of the relative frequency and degree-of-belief interpretations of probability in a weather forecasting context, see Murphy and Winkler (1971).

With regard to the interpretations given to probabilities by the early contributors, it is of considerable interest to see that Ångström in 1922 discussed both interpretations of probability. Of course, it is quite natural to interpret the probabilities derived from objective forecasting methods as relative frequencies. In the context of operational weather forecasting, the distinction between these two interpretations of probability may not be of great practical significance. Nevertheless, it is im-

portant for weather forecasters to recognize that a sound theoretical basis exists for a degree-of-belief interpretation of their subjective probabilities (see Winkler 1972; Morgan and Henrion 1990).

Recognizing that the “case” for probability was made by the early 1920s, and keeping in mind the substantial developments in this area over the last 75 years, it is natural to ask why probability forecasts are still the exception rather than the rule in weather forecasting in the late 1990s? Obviously, an in-depth discussion of this multifaceted question is beyond the scope of this historical paper. Nevertheless, a brief examination of some aspects of the question may be instructive here.

An important overall consideration related to the current acceptance of—or resistance to—probability forecasts is the way in which the problem of weather forecasting is viewed by the meteorological community at large and by those segments of the community directly and indirectly involved in the forecasting process. When weather forecasting was in its infancy in the nineteenth century, the forecasting process was dominated by very large uncertainties, and the evidence presented in section 2 suggests that the meteorological community at that time was reasonably comfortable with forecasts expressed in probabilistic terms. In concert with the evolving view of meteorology in the late nineteenth and early twentieth centuries as an exact physical science, key segments of the community adopted the view that the weather forecasting problem was best understood as a deterministic problem. According to this view, the uncertainties in the forecasting process would eventually fall victim to ever-improving models and more and better observations. In accordance with this view, uncertainty in forecasts was downplayed, and nonprobabilistic (or categorical) modes of expression of forecasts became the rule. The view of the prediction problem as a deterministic problem was further reinforced, and had its lifetime extended, by the advent of numerical weather prediction in the 1950s. It is interesting to note that this view was not significantly affected by the work on probability forecasting during the period 1945–85 or by the early work related to chaos in nonlinear models (Lorenz 1963; see also Ehrendorfer 1997). To some degree, then, the meteorological community’s lack of acceptance of probability forecasts—and its resistance to the enhanced use of this mode of expression of forecasts—over the last 50–100 years can be attributed to its view of the forecasting problem as a deterministic problem.

In the past, arguments have been advanced from time to time that particular methods of forecasting could not produce reliable and/or skillful probability forecasts. However, a vast body of evidence now exists to the effect that the uncertainty in forecasts can be quantified by a variety of methods, including subjective methods, statistical methods, numerical–statistical methods, and stochastic–dynamic methods (i.e., ensemble forecasting) (e.g., see Murphy 1985; Ehrendorfer 1997). It is certainly true that not all probability forecasts produced

by these methods are perfectly reliable and highly skillful; for example, in difficult forecasting problems involving relatively rare events, probability forecasts frequently reveal substantial overforecasting (forecast probabilities larger than observed relative frequencies) and marginal skill (see Murphy 1991). Nevertheless, it can be stated without equivocation that probability forecasts exhibiting reasonable reliability (and reliability considerably in excess of that achieved by the corresponding nonprobabilistic forecasts) and at least modest skill can be produced for most if not all weather conditions of interest.

It may be argued that producing probability forecasts requires more effort on the part of forecasters, and thus would require greater investment in the forecasting process. In addition, frequently it is argued that users of forecasts (especially the general public) will not accept—or will not be able to understand—probability forecasts. However, the “evidence” on which these arguments are based is largely, if not entirely, anecdotal in nature. Only a modest amount of credible evidence concerning this issue exists, but it all suggests that users prefer forecasts with reliable characterizations of forecast uncertainty (see Murphy 1985). Moreover, use of explicit probabilities minimizes the need for specific formats for particular users (e.g., in the case of forecasts issued by a private firm).

In conclusion, perhaps the widespread interest in the development and application of ensemble prediction is a sign that the meteorological community is at last ready to acknowledge explicitly the uncertainty inherent in the forecasting process and to face up to the practical implications of its existence. These implications can and should be viewed in a positive rather than a negative light. An opportunity now exists to provide the full spectrum of users of forecasts—from weather forecasters, as users of model output, to specialized endusers and members of the general public—with reliable information concerning the likelihood of occurrence of future weather conditions, information that could substantially improve the decisions made by *all* such users. In this regard, the results of some studies of forecast use and value (e.g., Thompson 1962; Gandin et al. 1992) suggest that the benefits of such “operational” improvements are the same order of magnitude as those that can be expected from advances in the state of the art of weather forecasting.

REFERENCES

- Ångström, A. K., 1919: Probability and practical weather forecasting (in Swedish). Centraltryckeriet, Teknologföreningens Förlag, 11 pp.
- , 1922: On the effectivity of weather warnings. *Nord. Statistisk Tidskrift*, **1**, 394–408.
- Besson, L., 1905: Attempts at methodical weather forecasting (in French). *Observatoire Municipal de Montsouris, Annals*, **6**, 473–495.
- Brier, G. W., 1944: Verification of a forecaster’s confidence and the use of probability statements in weather forecasting. U.S. Weather Bureau Research Paper 16, Washington, DC, 10 pp.
- , 1946: A study of quantitative precipitation forecasting in the TVA basin. U.S. Weather Bureau Research Paper 26, Washington, DC, 40 pp.
- , 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3.
- Chief Signal Officer, 1919: The Signal Corps Meteorological Service. A. E. F. *Mon. Wea. Rev.*, **47**, 870–871.
- Cooke, W. E., 1906a: Forecasts and verifications in western Australia. *Mon. Wea. Rev.*, **34**, 23–24.
- , 1906b: Weighting forecasts. *Mon. Wea. Rev.*, **34**, 274–275.
- Dalton, J., 1793: *Meteorological Essays*. Richardson, Phillips and Pennington, 208 pp.
- Dunoyer, L., and G. Reboul, 1921: The problem of weather forecasting (in French). *J. Phys. Radium, Ser. 6*, **2**, 129–141.
- Ehrendorfer, M., 1989: Meteorological probability forecasts: Some aspects (in German). *Wetter Leben*, **41**, 123–139.
- , 1997: Predicting the uncertainty of numerical weather forecasts: A review. *Meteor. Z., Neue Folge*, **6**, 147–183.
- Gandin, L. S., A. H. Murphy, and E. E. Zhukovsky, 1992: Economically optimal decisions and the value of meteorological information. Preprints, *12th Conf. on Probability and Statistics in the Atmospheric Sciences*, Toronto, ON, Canada, Amer. Meteor. Soc., J64–J71.
- Garriott, E. B., 1906: Note. *Mon. Wea. Rev.*, **34**, 24.
- Hallenbeck, C., 1920: Forecasting precipitation in percentages of probability. *Mon. Wea. Rev.*, **48**, 645–647.
- Jeffrey, R., 1992: *Probability and the Art of Judgment*. Cambridge University Press, 244 pp.
- Kramer, M. P., 1952: Annotated bibliography on objective forecasting. *Meteor. Abstr. Bibliogr.*, **3**, 471–517.
- Lamarck, J. B., 1800–11: *Annuaire Meteorologique* (in French). Vols. 1–11. Chez l’Auteur.
- Liljas, E., and A. H. Murphy, 1994: Anders Ångström and his early papers on probability forecasting and the use/value of weather forecasts. *Bull. Amer. Meteor. Soc.*, **75**, 1227–1236.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.
- , 1970: Forecast for another century of weather progress. *A Century of Weather Progress*, J. E. Caskey, Ed., Amer. Meteor. Soc., 18–24.
- Morgan, M. G., and M. Henrion, 1990: *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, 332 pp.
- Murphy, A. H., 1985: Probabilistic weather forecasting. *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, A. H. Murphy and R. W. Katz, Eds., Westview Press, 337–377.
- , 1991: Probabilities, odds, and forecasts of rare events. *Wea. Forecasting*, **6**, 302–307.
- , 1996: The Finley affair: A signal event in the history of forecast verification. *Wea. Forecasting*, **11**, 3–20.
- , 1997: Forecast verification. *Economic Value of Weather and Climate Forecasts*, R. W. Katz and A. H. Murphy, Eds., Cambridge University Press, 19–74.
- , and R. A. Allen, 1970: Probabilistic prediction in meteorology: A bibliography. Weather Bureau Tech. Memo. WBTM TDL 35, ESSA, Silver Spring, MD, 60 pp.
- , and R. L. Winkler, 1971: Forecasters and probability forecasts: Some current problems. *Bull. Amer. Meteor. Soc.*, **52**, 239–247.
- , and D. L. Williamson, Eds., 1976: *Weather Forecasting and Weather Forecasts: Models, Systems, and Users*. Vol. 1. National Center for Atmospheric Research, 446 pp.
- , and R. L. Winkler, 1984: Probability forecasting in meteorology. *J. Amer. Stat. Assoc.*, **79**, 489–500.
- , and —, 1992: Diagnostic verification of probability forecasts. *Int. J. Forecasting*, **7**, 435–455.
- Nebeker, F., 1995: *Calculating the Weather: Meteorology in the 20th Century*. Academic Press, 251 pp.

- Nichols, W. S., 1890: The mathematical elements in the estimation of the Signal Service reports. *Amer. Meteor. J.*, **6**, 386–392.
- Rolf, B., 1917: Probability and forecasting of summer rains (in French). Ph.D. thesis, University of Uppsala, Uppsala, Sweden, 61 pp.
- Scott, R. H., 1869/1971: On the work of the Meteorological Office, past and present. *The Royal Institution Library of Science, Earth Science*, S. K. Runcorn, Ed., Vol. 1, Applied Science Publishers, 333–345.
- , 1873/1971: On recent progress in weather knowledge. *The Royal Institution Library of Science, Earth Science*, S. K. Runcorn, Ed., Vol. 2, Applied Science Publishers, 378–387.
- Sheynin, O. B., 1985: On the history of the statistical method in meteorology. *Arch. Hist. Exact Sci.*, **31**, 53–95.
- Tennekes, H., A. P. M. Baede, and J. D. Opsteegh, 1987: Forecasting forecast skill. *Proc. Workshop on Predictability in the Medium and Extended Range*, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 277–302.
- Thompson, J. C., 1962: Economic gains from scientific advances and operational improvements in meteorological prediction. *J. Appl. Meteor.*, **1**, 13–17.
- von Myrbaach, O., 1913: Reliability of weather forecasts (a proposal) (in German). *Meteor. Z.*, **30**, 496–497.
- Whitnah, D. R., 1961: *A History of the United States Weather Bureau*. University of Illinois Press, 267 pp.
- Winkler, R. L., 1972: *Introduction to Bayesian Inference and Decision*. Holt, Rinehart and Winston, 563 pp.