**1.1**

# STATISTICS IN HYDROMETEOROLOGICAL FORECASTING– CONTRIBUTIONS AND CHALLENGES

Bob Glahn[1]
Meteorological Development Laboratory
Office of Science and Technology
National Weather Service, NOAA
Silver Spring, Maryland

## 1. INTRODUCTION

It's easy to think of contributions of statistics to hydrometeorological forecasting; statistics are embedded in the professional lives of all scientists and forecasters. However, we probably don't fully appreciate the degree of contribution, but pick a paper in a journal of the AMS and note whether or not statistics are used. Of course, the boundary between statistics and other branches of mathematics is not clean, as with the boundary between some other sciences. Wilks (2011) quotes, "Statistics is the discipline concerned with the study of variability, with the study of uncertainly, and with the study of decision-making in the face of uncertainty (Lindsay et al. 2004)."

Statistics as a subject has been around essentially forever in mankind's lives, but like other sciences underwent formal development in the last few centuries. Early achievements, such as those surrounding the Gaussain distribution and Bayes theorem are the cornerstones of so much in modern statistics. Our prediction capabilities have advanced tremendously in the past century, especially in the last 60 years with Numerical Weather Prediction (NWP), capabilities that have married statistics with other sciences including physics, mathematics, and chemistry. This subject could, and does, fill books.

The theme of this annual meeting is "Taking Predictions to the Next Level: Expanding Beyond Today's Weather, Water, and Climate Forecasting and Projections." This is an especially appropriate time to consider the contributions of statistics because several international organizations have declared 2013 as the International Year of Statistics (Statistics 2013) (http://www.statistics2013.org/about-us/). I will mention only a few of the major **contributions** of statistics, and then discuss **probabilistic forecasting** and **decision making**, as I view them as major ways statistics can help in moving toward the conference goal.

## 2. CONTRIBUTIONS OF STATISTICS TO FORECASTING

Forecasting in the United States dates back to about the time the U.S. weather service was formed as a duty of the Signal Service in 1870. Observations were scarce, but were analyzed, and members of the Signal Service and later the U.S. Weather Bureau manually issued forecasts. Even then, patterns were deduced and used in making the forecasts. For instance in 1910, Weather Bureau Chief Willis Moore (1910) stated, " . . . forecasts of a week or 10 days in advance have been issued from time to time when certain well defined weather types were shown by reports from selected stations throughout the Northern Hemisphere." Such patterns were, of course, statistically determined.

### 2.1 Objective Forecasting

Allen and Vernon in the *Compendium of Meteorology* (1951) defined objective forecasting as ". . . a forecast which does not depend for its accuracy upon the forecasting experience or subjective judgment of the meteorologist using it." Even earlier, Gringorten (1949) had described an objective forecast as one that ". . . is made without recourse to the personal judgment of the forecaster." It is in that publication that he defined "predictor" and "predictand," likely the first use of those terms. Of necessity, the first such methods were graphical or some other method of showing relationships, scatter diagrams and histograms being examples. Such techniques hit their stride in the 1950's, and the *Monthly Weather Review* has numerous examples, such as the classics by Vernon (1947) and Thompson (1950).

#### 2.1.1 The Classical Era

Shortly thereafter, digital computers arrived, and tremendously expanded the horizon for such work. Two main centers of activity for early statistical forecasting were the Travelers Research Center (TRC) and the Air Force Cambridge Research Laboratory (AFCRL). For example, at AFCRL, Iver Lund (1955) and Irving Gringorten (1950) were investigating methods of objective weather prediction. Their methods were right at the cusp of digital computer arrival, and were generally based on rather small amounts of data. TRC had government contracts, and could process large quantities of data, albeit they were on low density magnetic tapes. Bob Miller (1962) brought multiple discriminant analysis (MDA) and multiple regression into the meteorological world, including Bob's attaching the name REEP (Regression Estimation of Event Probabilities) to regression with binary predictands (Miller 1958).

Much of the TRC work was using the "classical" technique where the prediction was based only on past observations. Being done on aviation-related contracts, the studies were heavily oriented toward predicting the probability of each of several categories of ceiling height

[1] Bob Glahn, National Weather Service Scientist Emeritus, 1325 East West Highway, Silver Spring, MD. 20910; email: harry.glahn@noaa.gov

and visibility. The contributions of the TRC and AFCRL work were mainly in the popularizing of statistical prediction and the specific techniques brought into the meteorological literature. None of the results of specific studies were implemented into main-stream forecasting, but rather they furnished the foundation of the statistical work that was to grow in the Weather Bureau.

#### 2.1.2 The Perfect Prog Era

While observations, or analyses of observations, are quite useful for what used to be called short range (which now may be more appropriately termed nowcasting) and are indispensable for projections of an hour or so, it was quite apparent that longer range forecasts would need NWP. But NWP wasn't producing "street level" weather, or even "county size" weather above the surface. As a result, when operational NWP results were judged good enough to be useful in prediction, Bill Klein et al.(1967) and others substituted the NWP upper air predictions in multiple regression equations to predict maximum and minimum temperature at cities where observations existed. This not only brought the forecasts down to the ground, but they also applied to specific locations. This so-called "perfect prog" technique was used first in the Extended Forecast Section of the Weather Bureau in formulating their official forecasts. Later, the results were directly distributed to Weather Bureau forecasters and others who received the teletypewriter bulletins. The first such transmissions, called experimental because the forecasts had not been evaluated by forecasters, started September 17, 1968 (WB 1968a). An improved operational product replaced the subjectively-derived one on March 18, 1970 (WB 1970). Now, statistics were playing a major role in weather forecasting for the United States.

Early studies with this technique were also being conducted in Canada (e.g., Yacowar 1968). Also, considerable work was being done at the National Hurricane Center in predicting the tracks and intensity of hurricanes by regression methods (e.g., Neuman 1979); work along these lines has continued (e.g., DeMaria and Kaplan 1994).

#### 2.1.3 Model Output Statistics

The perfect prog[2] method used upper air observations as predictors in developing the relationships, and then applied the relationships to the NWP predictions. It was soon realized that the models had biases, and the relationships developed with "perfect" predictors did not apply as well as it had been hoped to NWP predictions. Especially, the probabilities of, say, precipitation did not fair into climatological relative frequencies at very long projections, but had about the same variance at long projections as at short projections.

[2] For more complete definitions of the classical, perfect prog, and MOS techniques, see Glahn (1985) or Wilks (2011).

The first application of using model predictors in developing the relationships was with a simple advective model over the eastern United States. The first such operational product was issued to forecasters starting June 10, 1968 (WB[3] 1968b; Glahn and Lowry 1967; 1972). The first national MOS product was of probability of precipitation and started January 1, 1972 (NWS 1971; Glahn and Lowry 1972). This ushered in the modern era of statistical weather forecasting. Many countries have adopted this method of using statistics in weather forecasting, although perfect prog and classical methods are used in some circumstances, especially where sufficient samples of NWP data are not available.

MOS products were expanded to other weather elements by the Techniques Development Laboratory (TDL, now the Meteorological Development Laboratory, MDL). These products, disseminated in the early days by teletypewriter and facsimile, became a mainstay in the forecasting paradigm at forecast offices, not only in the National Weather Service (NWS) but also at military installations, and in the private sector.

MOS systems have been developed in many other countries too numerous to mention, sometimes replacing an older perfect prog system. For instance, Wilson and Vallee (2003) describe Canada's system.

The production of statistical forecasts should be viewed as a system, flowing from the taking of observations, the analysis (assimilation) of them for NWP, the development and operation of the models, the postprocessing, and whatever human activity is required to produce the final product. The modeling community should be working hand-in-glove with the postprocessing community. This would insure the right information is available for postprocessing, and the verification that is part of postprocessing ought to help modelers in their development and in deciding what new or modified models should be implemented. The modeling community needs to recognize the necessity and advantages of providing hindcasts (reforecasts; retrospective runs) for postprocessing; statistics cannot deduce the error characteristics of the model without a sample of data from that model encompassing weather conditions of many kinds. It is possible that well designed experiments involving these two groups might lead to a better understanding, for instance, of whether more ensemble members or fewer members at higher resolution would provide better forecasts. It is also possible the results might depend on whether the prime metrics were of the upper level flow, or whether the daily

[3] These WB references are to the Technical Procedures Bulletins (TPB) started by Charles Roberts in July 1967 to inform users, principally field forecasters, of centrally produced Weather Bureau products. They comprise a continuous record of new, modified, and dropped products from 1967 until shortly after 2000. The Weather Bureau under ESSA became the National Weather Service under NOAA on October 3, 1970, and the references here are "NWS" rather than "WB" after that date. Copies of the TPBs exist in the NOAA Library in Silver Spring, Maryland.

2-m temperature or precipitation relative frequency was deemed more important.

### 2.1.4 New Methods

Besides the linear regression workhorse, and to a much lesser extent discriminant analysis, other methods that can be applied to any of the classes of techniques above (classical, perfect prog, MOS) have been developed. While so-called linear regression and discriminant analysis <u>are</u> linear in their predictors, most applications define a system which is highly nonlinear. That is, predictors are defined that the developer believes have a linear relationship to the predictand. The innovations are unlimited.

The Empirical Orthogonal Functions that Lorenz (1956) brought into the meteorological literature have been used to some extent, especially in the last couple of decades in diagnosing the state of climate and its change. In addition, canonical correlation has been used to some extent. The degrees of freedom used by these techniques are usually large, and great care must be taken in assigning significance or physical meaning to the patterns associated with the coefficients.

Regression has been used to estimate the probability of an event, by using the predictand event as a “1" if it occurred in the sample, and a “0" if it didn’t. Not being constrained, regression may give a value > 1 or < 0, both not reasonable for a probability forecast. This has been dealt with by just assigning the values > 1 to 1 and the values < 0 to 0. A model that does restrain the predicted values correctly is the Logit model (Brelsford and Jones 1967; Wilks 2011). However, the solution cannot be obtained analytically (except in a very restricted case), and therefore iteration is required. As with most iterative methods, problems may arise in certain situations, and more computer time is required for large data sets and relationships.

The area of neural networks applied to weather forecasting has grown from its modest beginnings in the early ‘60's (Hu and Root 1964; Glahn 1964b);[4] a good exposition is given by Marsban (2003). Here, the intent is to let the model define the degree of non-linearity that exists among the predictors and predictand(s) in the data set. As with defining non-linear predictors for regression, the variations in building the models, devising the training algorithms, and defining the predictors are limitless.

This is not an exhaustive list of current methods, but a few that will persist.

## 2.2 Verification

It is imperative that the forecasts we make be verified, at least a representative sample of those made. This has been another rich area for the contribution of statistics to weather prediction. Any method of verification must necessarily make use of statistics. At first blush, one would think this to be a simple problem. A statement that has been attributed to various persons is, “Prediction is difficult, especially about the future.” I say, “Verification is difficult, especially about the past.”

The interest in verification grew along with the interest in objective weather prediction, although the subject had been addressed much earlier, and U.S. weather service forecasts issued by the Signal Service had been verified from their inception. Scores then were simple, usually being mean absolute error (MAE), mean square error, or root mean square error for quasi-continuous variables like temperature, and percent correct for dichotomous variables such as occurrence of precipitation. The difficulties with the latter are well recognized, and an early debate about Finley (1884) verifying his tornado forecasts in this way has been thoroughly examined by Murphy (1996). In that same paper Murphy states, “The burst of verification-related activities during the period 1884–1893 is referred to here as the ‘Finley affair.’ It marked the beginning of substantive conceptual and methodological developments and discussions in the important subdiscipline of forecast verification.” This paper by Murphy is very complete and an interesting read concerning especially the pre-1900 verification activities. It was during this period that Gilbert (1884) devised a score that has been “rediscovered” twice and is now generally called the Threat Score (Palmer and Allen 1949) or the Critical Success Index (Donaldson et al. 1975).

Glenn Brier and Roger Allen of the U.S. Weather Bureau were devoting time to verification in the late 1940's and early 1950's. They contributed a frequently quoted chapter to the *Compendium of Meteorology* (Brier and Allen 1951) in which they define three purposes of verification and discuss each: (1) economic, (2) administrative, and (3) scientific. They also state, “. . . the verification scheme should influence the forecaster in no undesirable way.” About the same time, Brier (1950) published a paper defining a score that has come to be called the Brier Score. While this is nothing more than a mean square error of a probability forecast (e.g., 0.1) where the verifying variable is binary, either 1 if the event occurred and 0 if it did not, he proved that it would encourage the forecaster to issue the forecast that he/she really believed; that is, it would not influence the forecaster in an undesirable way. In this paper, Brier also suggested that after a set of forecasts had been made, the relative frequency of the event could be computed for each value of probability forecast to determine whether the “forecast probabilities are related to the relative frequency of the events' occurrence.” By so doing, it might be concluded the forecast process should be modified. This correspondence between probability forecasts and the relative frequencies is what is now generally called “reliability.” While Brier did not use these words, he indicated a good score would be achieved when the forecasts were as sharp as possible within the restriction of reliability.

[4] Early uses were for binary predictands and the method was called “adaptive logic,” but the basic model is essentially the same.

The interest in verification paralleled to a great extent the interest in objective forecasting, both by numerical (dynamic) and statistical means. Both objective forecasting and verification were made operationally possible by the development of digital computers. The computer era also allowed the collection and processing of large volumes of data. The availability of orders of magnitude more forecasts and the ability to verify them, gave rise to renewed interest in methods of verification and in the establishment of systems to collect and verify the forecasts.

It would be impossible here to adequately review the development of verification of forecasts that has occurred since 1950. Rather I will only mention some of the important scores and concepts that have been developed.

Fred Sanders (1963) decomposed the Brier Score into validity (i.e., reliability) and sharpness components. Other partitions have been introduced, (e.g. Murphy 1972). These breakdowns indicate how the probability forecasts can be improved.

In 1969, Ed Epstein (1969) published the Ranked Probability Score (RPS), and Allan Murphy (1969) showed it to be "proper." RPS, originally described for categorical forecasts, has been extended to the Continuous RPS (CRPS), and has been decomposed into components similar to those for the Brier Score (Herbach 2000).

The ROC (Relative Operating Characteristic) brought into the meteorological literature by Mason (1982) continues to increase in use. It shows the discriminating power of a set of probability forecasts and is a good way of visualizing on a diagram the relative discriminating power of two or more sets of forecasts. However, it is probably not yet fully appreciated that the ROC does not at all address the issue of calibration.

Murphy and Winkler (1987), in an attempt to unify the many disparate verification metrics, published "A General Framework for Forecast Verification." It is based on the joint distribution of forecasts and observations, and two factorization are described, the calibration-refinement factorization and the likelihood-base rate factorization. They demonstrate the richness of information that is contained, in the case of categorical forecasts, in the contingency tables that were one of the early ways of displaying forecasts and matching observations.

Gandin and Murphy (1992) and Gerrity (1992) introduced and discuss the concept of equitability. Equitable scores ". . . discourage forecasters from exhibiting inappropriate preferences for some events at the expense of other events. In particular, constant forecasts of any particular event–as well as forecasts in which events are chosen at random–achieve the same expected score. . ." (Gandin and Murphy 1992).

Most verification has been, and continues to be, done at points. However, there has always been recognition for the need to consider the spatial aspects of forecasts. It was imperative to evaluate the upper air forecasts produced by NWP. The S1 score introduced by Teweles and Wobus (1954) was used for years and is still used to furnish a continuous record from the inception of NWP and before. The anomaly correlation[5] is heavily used today in verifying spatial fields, although it ignores biases and is more appropriately considered a measure of potential performance, as is the ROC. However useful the S1 score is in measuring the accuracy of gradients of pressure or geopotential height, or the anomaly correlation is in measuring the correspondence of patterns, these measures do not address the aspects of placement of significant features, such as rain bands or severe weather. Early work in this area has been done by Ebert and McBride (2000), and Brown et al. (2004) discuss an object-oriented approach. An intercomparison of approaches has been reported by Gilleland et al. (2010), and they cite experimental use at some forecast centers. This important area will receive much attention in the future.

The expanding use of ensembles in the past decade or so provides the mechanism of estimating probabilities with no postprocessing other than determining the relative frequency of "events," sometimes defined with thresholds, and has brought into focus verification techniques specific to probability forecasts. The Rank Histogram, evidently brought into the published literature by Hamill and Colucci (1997, 1998), is now an indispensable evaluation tool.

An important characteristic of forecasts is consistency from one forecast projection to another. For instance, a forecast for the same calendar day and time may be made multiple times, once at 7 days, once at 6 days, . . . , and finally 1 day ahead, or even oftener (e.g., and 12 h intervals). The tendency of numerical model forecasts and MOS to exhibit unwanted "jumpiness" has been recognized for many years. Such a tendency can lower the confidence of a user of the information, either a field forecaster or some other user. This aspect of forecast evaluation is beginning to get some attention through use of the Ruth-Glahn Convergence Score (Ruth et al. 2009; Lashley et al. 2008).

An excellent book on verification by Jolliffe and Stephenson (2012) is in its second edition. This is an important read for anyone interested in verification of weather forecasts. To my knowledge, it is the first book dealing exclusively with weather forecast verification. In addition, the bible on meteorological statistics has an excellent chapter on forecast verification (Wilks 2011).

[5] Wilks (2011, p. 364) points out there are two versions of the anomaly correlation.

## 3. PROBABILISTIC FORECASTING

### 3.1 Status

If we discount the general impressions formed by observing natural phenomenon including the weather, probabilistic forecasting came into serious consideration in the early 20th century. A bibliography and an early history on the subject put together by Allan Murphy (1966, 1998) lists a couple of publications in the late 1800's, then many more in the 1920's and forward. Early examples are Dalton (1793), Dines (1902), and Cook (1906) who discussed probabilities. Early in the U.S. national weather service as it began in the Signal Service, Cleveland Abbe started issuing forecasts that were called “Weather Synopses and Probabilities” in 1871 (Meyer 1871); this early history has been traced by Glahn (2012).

Probabilistic forecasting and the verification of probabilistic forecasts began to have serious attention in the U.S. Weather Bureau in the 1940's and 50's as the graphical, pre-computer, methods of objective forecasting (Allen and Vernon 1951) came into play, especially with the Research Forecaster Program, a program not unlike that of the Science and Operations Officer (SOO) in National Weather Service of today.[6] And, of course, hand in glove with these techniques came Glenn Brier’s classic paper on verification (Brier 1950). Largely due to the perseverance of Charles Roberts,[7] the Bureau started a Probability of Precipitation (PoP) national forecasting program in 1965[8] (Hughes 1980). Verification of the PoP forecasts soon started and this program remains today. Unfortunately, the systematic expansion of the production and use of probabilistic forecasts did not take place as had been envisioned by some meteorologists.

Following the graphical methods of developing forecast relationships (e.g., Thompson 1950), and as operational NWP at the National Meteorological Center (NMC) became a reality in the mid and late 1950's, the Techniques Development Laboratory (TDL) began using NMC’s IBM 704 computer to experiment with producing probabilistic forecasts for use as guidance for Weather Bureau forecasters. The first such PoP product debuted in 1969 and covered the eastern United States (WB 1969). As mentioned earlier, on January 1, 1972, an automated PoP product covering the conterminous United States replaced a manually produced one being disseminated by NMC (NWS 1971). This transition to an automated product was largely due to Harlan Saylor, a senior manager and forecaster at NMC, who recognized that statistics had a role to play alongside NWP, and was here to stay. That was a major breakthrough for statistical postprocessing.

Probabilistic guidance expanded into other weather elements, including ceiling height, visibility, and conditional probability of frozen precipitation. The products were transmitted to forecasters by teletypewriter and facsimile, along with categorical forecasts derived from the probabilities and thresholds based on desirable accuracy and/or skill metrics. However, the probabilities, even though reliable, and furnishing the path whereby the categorical forecasts were made, were not heavily used, and were largely discontinued for lack of pull from the user community, the field forecasters being the primary user for TDL.

Probabilistic forecasting has been strongly supported by certain persons and groups from its beginning, and especially during the last decade. The AMS has issued two Information Statements in strong support (2003; 2008). The National Research Council (NRC 2006) published *Completing the Forecast* with the subtitle “Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts” in 2006, following up at the request of the NWS on Recommendation 8 in NRC’s report *Fair Weather* (NRC 2003), stating “The NWS should continue to adopt and improve probabilistic methods for communicating uncertainties in the data and forecasts where such methods are accepted as scientifically valid.” Yet, probabilistic forecasting has not seen the expansion desired. Why is that?

There are undoubtedly many factors, among them, I believe, are the following:

1) While the economic benefit of using probabilistic information of weather, water, and climate forecasts and from other sources is huge (Hirschberg and Abrams 2011), a large part of the weather enterprise still services the general public, and there has not been an outcry from that public for probability forecasts. As long as there is not a strong pull for the information, many forecasters and organizations are confident in their ability to provide the non-probabilistic service they are used to, and if the user does not desire probabilities, why produce them?

2) Some persons in the weather enterprise question whether the concept of probability is understood by the general public. I firmly believe that most people appreciate what “chance of” and “probability of” means (see for instance Feller 1957, p. 2). After all, people deal with that concept every day–what are the chances of my finding a parking

---

[6] There were on the order of 10 Research Forecasters, one each at the larger forecast offices such as Los Angeles and Chicago. Today, there is a SOO position at each of the 122 Weather Forecast Offices.

[7] Memorandum from Charles Roberts, Head of the Technical Procedures Branch in the U.S. Weather Bureau, dated July 27, 1965, transmits a draft proposed circular letter to formally initiate the Weather Bureau Probability Forecast Program.

[8] An interesting article was written by Myron Tribus (1970) not long after the PoP program started in which he expressed much hope for probabilistic forecasts. Tribus, at the time Assistant Secretary of Commerce for Science and Technology, was an expert in the theory and practice, having written a book on the subject (Tribus 1969).

space, what is the probability supper will be ready when I get home, what is the likelihood my hay will get wet if I mow? Twenty or 30 is less than 80, right? Can I ever be 100% sure? While this terminology may not be scientifically precise, I believe it is sufficient for the lay public, and that it is understood. As long as scientists believe the concept of probability is too difficult for the masses to understand, they will not be prone to promote and provide such information.

3) Furnishing probabilistic information for the full range of weather elements is a daunting challenge, both in the production and dissemination. For weather elements of a semi-continuous nature, such as temperature, it takes about an order of magnitude increase in the number of numbers necessary to furnish a reasonably satisfactory description of the probability distribution (PDF) over a single number like 60 degrees, or even a range 58 to 61 degrees. A probability must be of an "event." That is, what is it the probability of? A characterization of a PDF can be a series of probabilities, each being of an event, that can be used together to form an approximation to the PDF.

One very simple probability forecast is of the probability of precipitation (PoP), the first probability forecast issued by the Weather Bureau/NWS. The event was carefully defined–liquid precipitation of $\geq$ 0.01 inch in a 12-h period at a specific point. Yet, studies have shown that this venerable product is not well understood. However, the misunderstanding of the PoP forecast, if there is one, may be due to not understanding the definition of the event rather than not understanding the concept of probability (Murphy et al. 1980). Windy tomorrow afternoon is an easy term to appreciate, but the definitional, production, educational, and dissemination challenges to go into the probabilistic realm are not small.

### 3.2 Progress

While probabilistic forecasting has not moved at the rapid pace we would have liked, statistics have provided notable progress, including:

1) The production of reliable and skillful probabilistic information has been repeatedly demonstrated with centralized operational products, starting 4 decades ago (e.g., Glahn and Lowry 1972; Glahn and Bocchieri 1975; Bocchieri and Glahn 1976), most of the progress being tied closely to the growth of NWP and computer capacity. It was feared that forecasters could not make reliable probability forecasts. This fear was largely unjustified as shown by the verification system of the Weather Bureau.[9] For many years, single model runs were made from one analysis, and the output was postprocessed into a variety of probabilistic products. Climate forecasts are furnished in terms of probabilities of broad categories. Stream flow is now characterized by probabilities (Fresch and Roe 2013). Hurricane track forecasts have uncertainty information. More recently, ensembles are furnishing the foundation for better products, although ensembles alone do not provide probabilities, rather they provide a discrete set of single value forecasts that can be processed into probabilities. When the ensemble members forecast for a specific event, the probability of the event can be estimated by counting the number of events that were and were not forecast. However, to date, the reliability of such relative frequencies is lacking, and statistical postprocessing is necessary (see for example Gneiting et al. 2009).

2) Awareness of the utility of probabilistic forecasts has been greatly increased, even if this awareness has not contributed to progress as much as desired or expected. Strong support has been voiced. The NWS has produced training material, and COMET courses are available. The National Oceanic and Atmospheric Administration (NOAA) has issued a "Fact Sheet" on weather forecast uncertainty for public consumption [available at: 'www.nrc.noaa.gov/stateofsciencefactsheets'. Short courses have been conducted at AMS meetings. The AMS Probabilistic and Statistics Committee holds conferences every 2 years, at which probabilistic methods of forecasting and verification are prime. The AMS created an Ad Hoc Committee on Uncertainty in Forecasts (ACUF); committee members produced a Weather and Climate Enterprise Strategic Implementation Plan for Generating and Communicating Forecast Uncertainty (Hirschberg and Abrams 2011), and a condensed version appeared in BAMS (Hirschberg et al. 2011). The World Meteorological Society has issued guidelines on communicating forecast uncertainty (WMO 2008). Aspects of probabilistic forecasting are now being studied and documented by social scientists (e.g., Demuth et al. 2012).

3) The science of verification of forecasts in general, and of probabilistic forecasts specifically, has grown tremendously, and is still growing. When the first PoP product was produced, we essentially had the Brier Score (Brier 1950), a tremendous

[9] The results of the Weather Bureau verification system at the beginning of the probability forecasting program is documented in a series of Technical Memoranda in the FCST series. National records were not being kept on individual forecasters, and it would be difficult to amass sufficient cases for one forecaster to document reliability reliably. Also, the emphasis was on the Brier Score and the Brier Skill Score, rather than reliability. However, Technical Memorandum WBTM FCST 11, "Report on Weather Bureau Forecast Performance 1967-68 and Comparison with Previous Years," dated March 1969 contains tables on pages 33-38 that indicate nationwide and by Weather Bureau Region the forecasts were quite reliable, except at the high end, where there was overforecasting (Roberts, et al. 1969). Diagrams in Murphy (1985, pp. 357-359) indicate high reliability for the 1980-81 seasons.

achievement of Brier at the time in realizing that it could not be "played," but nevertheless is just a mean square error.[10] Now, verification has been sliced and diced in many ingenious ways, even to different sets of components of the Brier Score itself.[11] Many of these methods are elegantly described and exampled by Wilks (2011) and Jolliffe and Stephenson (2012).

### 3.3 Challenges and Opportunities

There are many challenges and opportunities for statistics in probabilistic forecasting; the importance of statistics is becoming more recognized as the use of ensembles has grown. Some to mention are:

1) Except for the very first hours, the foundation of weather, water, and climate forecasting is in numerical models. Especially with the realization that the initialization for a model cannot be perfect, and that errors of specification will eventually overwhelm the numerical solution (Lorenz 1965), the stochastic-dynamic approach (Epstein 1969) implemented through ensembles (Lewis 2005) solidified the tie of statistics and dynamics in weather forecasting. Ensembles are composed of individual runs of a model, each initialized with a plausible but different representation of the current and recent past data. However, the three-dimensional fields that represent the correct error distribution are not known, and various methods have been devised to produce the different starting points. Initialization methods in use today produce forecasts that are underdispersive. Data assimilation is very important and is attracting much attention. Statistical methods are bound to play an increasing role, as opposed to purely mathematical methods. Data assimilation is crucial to the advance of forecasting, especially probabilistic forecasting, and deserves the highest attention.

2) The blending of the several results from the members of an ensemble is another area deserving, and receiving, considerable attention. Numerous methods have been proposed, and some are, in fact, in use. One method getting much play is Bayesian Model Averaging (Raftery et al. 2005). This is a method whereby the different members, or some grouping of members perhaps composed of members of similar model physics, can be weighted differently to arrive at the "best mean" forecast and the distribution around it Another method in use in the NWS has been called EKDMOS for Ensemble Kernel Density MOS (Glahn et al. 2009). Others (e.g., Woodcock and Engel 2005) have blended forecasts from different sources to produce an improved product. Much work is still needed in this area.

3) Not everyone in the weather enterprise is convinced of the importance of producing and providing uncertainty information. A challenge is changing the paradigm of providing only deterministic, single-value forecasts. The value of probabilistic information has to be accepted by meteorologists and users alike. This is not an insignificant challenge.

4) The general area of evaluation, even though much progress has been made in the past few decades, needs more work. The desires for good probability forecasts seem simple enough–the forecasts should be as sharp as possible within the constraint of reliability. Even so, some probability ranges may be more important to a user than others, and many times it is in the tails of the distribution that the forecasts are important. We have the RPS (Epstein 1969) and its extension the CRPS (Matheson and Winkler 1976; Unger 1985), but the latter is more sensitive to the placement of the median than the shape of the tails.

   Meteorological centers and private firms issuing statistical postprocessed products have to decide on the method/technique to use. Many papers on different techniques appear in refereed journals, but provide little help in deciding on an operational technique to implement. The reason is that there is no systematic method of comparing methods on an even playing field. Usually there is some comparison presented; it may be the author's new highly developed method versus a known technique, but the comparison technique has not been implemented optimally by the author. Or the comparison may be to some default forecast such as persistence for a 24-h forecast or to a simple climatic relative frequency for a 48-h forecast. These latter may have been adequate comparisons 40 years ago, but not today. The operational entity needs to know how results from the new method compare to what is already operational; usually these operational forecasts are readily available and could furnish a standard to beat.

5) There is inadequate recognition for the need for education and training for meteorologists in statistics, not only in dealing with probabilities. It has been my experience over the past years that meteorologists entering the work force may have <u>at most</u> one course in statistics. Applicants with significant statistics courses under their belt are the exception.

   The WMO issues "Guidelines for the Education and Training of Personnel in Meteorology and Operational Hydrology" (WMO 2001). I find <u>no</u> mention of statistics whatsoever. To secure a job in the U.S. Government as a meteorologist, no specific qualifications in statistics are required (U.S. Govt. 1998); statistics is mentioned, but is one of 1l sciences for which a total of 9 semester hours (essentially three courses) are required.

[10] Roberts (1968, p. 139) presents an interesting discussion on not reporting the probability that is your true belief.

[11] Incidently, Brier called it the "Score P."

Furthermore, the importance of statistics seems to not be recognized by those mentoring students in college. The AMS periodically publishes an Information Statement "Bachelor's Degree in Atmospheric Science," the last being in December 2010 (AMS 2010). The Statement is issued for the ". . . primary purpose . . . to provide guidance to university faculty and administrators responsible for undergraduate programs in atmospheric science," meteorology and atmospheric science being considered equivalent in the Statement. The Statement was drafted by a committee of 10 members chosen by the AMS Board on Higher Education. While the membership was capped at 10, others who had volunteered but not selected were included informally in the email discussions. Of the 16 who were regularly on the addressee list, one was an AMS staff member and of the other 15, 13 had the email domain name extension "edu." While there was one government representative and one from private industry, it seems the committee was overpoweringly weighted to university professors and had inadequate representation of those entities that actually hire meteorologists and know what the requirements are. Only after lobbying on my part, the terms probability and statistics were included in a couple of places, whereas the previous such Statement had none. For preparation for graduate school, the Statement does not mention statistics. The AMS has available a scholarship dedicated to students showing background or interest in statistics. The requirements are not stringent. However, this past year, there was no qualified applicant; this was also the case for a previous year. My conclusion is that there is not a lot of emphasis or interest in statistics in the departments training our future meteorologists. This needs to change.

## 4. DECISION MAKING

### 4.1 Status

Decision making under uncertainty has been practiced as long as decisions have been made. Even as a science, the concept is decades old (Chernoff and Moses 1959; Miller and Starr 1960), and attempts have been made to formally marry decision models and forecasts (see Glahn 1964a for an early example). Decisions based partly on weather information are made by a wide variety of individuals and organizations and are of vital importance to the saving of lives and property and to the economy of the country. While probabilistic forecasts can be made, and are for some weather elements, much of the information furnished to users is still non-probabilistic in nature. There are sophisticated users who access probabilistic information, either from the NWS or from private entities, and use that information in decision models. It is difficult to assess the degree to which this enterprise has matured because the companies involved generally treat the process as proprietary and do not publish or otherwise make it available.

The introduction into the literature of the Cost/Loss (C/L) model by Thompson (1962) raised awareness of the importance of economic value in assessing forecasts. This model has been studied and used extensively, thereby bringing a better understanding of the decision making process. However, real decisions are usually based on many factors, weather being only one of them. The decision for an action 2 days from now might depend on the forecast not only for 2 days from now, but also before that and after that. Such decision models become very complicated, and models presented in the literature generally depend only on one forecast, but discussions by Murphy et al. (1985), Krzysztofowicz (1986), and Epstein and Murphy (1988) are exceptions and paved the way for more realistic models

### 4.2 Progress

The weather enterprise has grown tremendously over the past couple of decades. Much of the progress has been not only in making more skillful forecasts, but in getting them into the hands of users who then make decisions. In fact, in many cases the users are consulted as to their needs, even to the extent they are assisted in making their decisions. This can happen at any stage in the producer/user chain.

At the end of the chain where forecasters are providing the final product, be it from the NWS or private companies, the tie has become stronger. While hard information is difficulty to come by, the tie is likely very close in the private sector–profit depends on it. In the NWS, one of the goals in its Weather Ready Nation Strategic Plan (NWS 2011) is to provide forecasts that will "compare weather risk to tolerance levels based on societal or economic impacts, communicating the potential social, economic, and environmental impacts." Close ties exist between NWS forecast producers and governmental agencies involved in protection of life and property. The contribution of statistics is to provide processed data and analyses to help the user understand how to make better decisions.

### 4.3 Challenges and Opportunities

The science and practice of decision making under uncertainty are in their infancy. The challenges below relate to understanding probability science, understanding the customer, and better decision models.

1) The meteorologist who desires to help a user in actual decision making under uncertainty needs to understand probability and decision theory and how decision models are built. Models may need to be "second order" where probabilistic weather forecasts are used to make probabilistic forecasts of environmental hazards, such as harmful algal bloom. While most everyone, I believe, understands the basic concepts of probability, the understanding may not go much farther than that. For instance, the probability of two separate

events may be available and the understanding of how to make a decision based on each one individually is known, how does one make a decision based on both together? Is that possible? What other information is needed? What should the development community be tasked to provide in order to make the decision? This is probably the simplest step up from a basic decision algorithm based on one weather variable. It is my perception that the knowledge needed here is generally lacking. The problem stems from lack of statistical training discussed above. If meteorologists are to assist in making critical economic or life and death decisions, then they should not be novices in the theory of probability and decision theory.

2) The success in helping users of weather information make decisions lies with understanding what the needs are. A major step is being able to convince the user that probability forecasts can actually help in his/her operation. The details are very important; for instance, when does the user need to be concerned about the weather, what forecasts are needed, and what are the risks? What other variables come into play in the decision? Is weather a major or minor player? Is timing of a weather event of critical importance? Can a mathematical model be built to accommodate the predominant factors, and if so, can the weather information be better tailored to assist? For instance, it might be the joint probability of two or more weather variables is needed. If so, can the developers be tasked to provide it? Statistics can provide user relevant information once it is known what is needed. The forecaster needs to work closely with the user down the chain and also the developer up the chain. This interactive aspect is not well developed in the weather enterprise and deserves much attention.

3) Decisions concerning whether one forecasting method is better than another [in terms of some metric(s)] may involve statistical significance tests. The probability of Type 1 error is usually set at some low level like 1% or 5%. This implies that the decision maker wants to be quite sure that the difference calculated is real and a difference of the same sign would also be found in another sample from the same population. Unfortunately, the probability of Type 2 error is not always considered, and the decision may not require a low probability of Type 1 error. Placing error bars can furnish more flexibility, but the same difficulty can exist. An inherent difficulty with applying significance tests is that the sample points are highly redundant, and the degrees of freedom in the sample are not known. Temporal non-independence can be dealt with in what is probably a satisfactory manner by assuming a Markov model and computing an "effective sample size" with which the variance of the statistic being used can be adjusted (e.g., Wilks 2011, p. 147, 422) . Dealing with spatial non-independence is much more difficult. Many times, data are aggregated across stations, or many grid points are involved. How is this dealt with? DeSole and Shukla (2009) discuss the issue of spatial non-independence and reference other work, but the issue is far from settled. Likely this is best dealt with by some form of block randomization such as described in Wilks (2011 ,Ch. 5; 1997).

Tests may be made for several projections, separated by 1 or more hours; typically this might be MAE for temperature every 12 hours out to, say, 11 days. Tests of the results for a particular technique may not reject the null hypothesis at the 5% level for any projection, but does for every protection at the 20% level. How does one come to an overall conclusion as to whether to implement or not. Certainly, the errors and the scores are correlated, but there is also considerable independent information among the 22 scores. Methods to deal with this situation need to be devised. A unified approach that deals with both temporal and spatial non-independence and across strata, such as projections, should be a goal.

## 5. SUMMARY

The use of statistics, models for postprocessing, and metrics for verification have come a long way in the last 60 years, marching hand in hand with NWP and computer technology. I have concentrated on forecasting because of the theme of the meeting, but the use of statistics is much broader than that. I have also concentrated on weather at the expense of water and climate, because there are summary talks to follow on those subjects (Fresch and Roe 2013; Collins 2013). As we learn more, we also realize how much more there is to know. More details and a broader scope are contained in the ACUF report mentioned above (Hirschberg and Abrams 2011).

While status and progress are important, what is of most interest and concern is the way forward–the challenges and opportunities. I have presented some of the major challenges and areas of growth involving statistics as I see them for taking predictions to the next level. The other presentations in this Symposium are also rich in details in moving forward toward that objective.

*Acknowledgments:* Several persons provided ideas for this paper; I sincerely thank them. Special thanks goes to Barbara Brown who reviewed an early draft which resulted in major revisions, and to Tamarah Curtis for assisting with document preparation. However, the views expressed here are those of the author and do not necessarily represent those of the NWS or of any other government agency.

## 6. REFERENCES


Allen, R. A., and E. M. Vernon, 1951: Objective weather forecasting. *Compendium of Meteorology*, Amer. Meteor. Soc., 796-801.

AMS, 2003: Enhancing weather information with probability forecasts. *Bull. Amer. Meteor. Soc.* **83**, 450-452.

_____, 2008: Enhancing weather information with probability forecasts. *Bull. Amer. Meteor. Soc*. **89**, 1049-1053.

_____2010: Bachelor's degree in atmospheric science. *Bul. Amer. Meteor. Soc*., **91**, 1730-1734.

Bocchieri, J. R., and H. R. Glahn, 1976: Verification and further development of an operational model for forecasting the probability of frozen precipitation*.* *Mon. Wea. Rev.*, **104**, 691-701.

Brelsford, W. M., and R. H. Jones, 1967: Estimating probabilities. *Mon. Wea. Rev*., **85**, 570-576.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev*., **78**, 1-3.

_____, and R. A. Allen, 1951: Verification of weather forecasts. *Compendium of Meteorology*, Amer. Meteor. Soc., 841-884.

Brown, B. G., R. R. Bullock, C. A. Davis, J. H. Gotway, M. B. Chapman, A. Takacs, E. Gilleland, K. Manning, and J. L. Mahoney, 2004: New verification approaches for convective weather forecasts. Preprints, *11th Conf. on Aviation, Range, and Aerospace*, Hyannis, Massachusetts, Amer, Meteor. Soc., **9.4.**

Chernoff, H., and L. E. Moses, 1959: *Elementary Decision Theory*. John Wiley & Sons, 364 pp.

Collins, D. C., 2013: Statistical methods in climate and seasonal prediction. Preprints, *Symposium on the Role of Statistical Methods in Weather and Climate Prediction*. Austin, Texas, Amer. Meteor. Soc., in press.

Cook, W. E., 1906: Weighing forecasts. *Mon. Wea. Rev*., **34**, 23-24.

Dalton, J., 1793: *Meteorological Essays*. Richardson, Phillips and Pennington, 208 pp.

DeMaria, M., and J. Kaplan: 1994: A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. *Wea. Forecasting*, **9**, 209-220.

Demuth, J. L., R. E. Morss, B. H. Morrow, and J. K. Lazo, 2012: Creation and communication of hurricane risk information. *Bul. Amer. Meteor. Soc.*, **93**, 1133-1145.

Dines, W. H., 1902: The element of chance applied to various meteorological problems. *Quart. J. Roy. Meteor. Soc*., **28(122),** 53-68.

Donaldson, R. J., R. M. Dyer, and M. J. Kraus, 1975: An objective evaluator of techniques for predicting severe weather events. Preprints, *Ninth Conf. on Severe Local Storms*, Norman, Oklahoma, Amer. Meteor. Soc., 321–326.

Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrology*, **239**, 179-202.

Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor*., **8**, 985-987.

_____, 1969: Stochastic dynamic prediction. *Tellus XXI*, **6**, 739-759.

_____,, and A. H. Murphy, 1988: Use and value of multiple-period forecasts in a dynamic model of the cost-loss ratio situation. *Mon. Wea. Rev*., **116**, 746-761.

Feller, W., 1957: A*n Introduction to Probability Theory and Its Applications*, Vol. 1. John Wiley & Sons, Inc., 461 pp.

Finley, J. P., 1884: Tornado predictions. *Amer. Meteor. J*., **1**, 85–88

Fresch, M., and J. Roe, 2013: Operational Implementation of the U.S. National Weather Service's Hydrologic Ensemble Forecast Service. Preprints, *Symposium on the Role of Statistical Methods in Weather and Climate Prediction*. Austin, Texas, Amer. Meteor. Soc. in press.

Gandin, L., and A. H. Murphy, 1992: Equitable skill scores for categorical forecasts. *Mon. Wea. Rev*., **120**, 361-370.

Gerrity, J. P., Jr., 1992: A note on Gandin and Murphy's Equitable Skill Score. *Mon. Wea. Rev*., **120**, 2709-2712.

Gilbert, G. K., 1884: Finley's tornado predictions. *Amer. Meteor. J*., **1**, 166–172.

Gilleland, E., D. A. Ahjevych, B. G. Brown, and E. E. Ebert, 1010: Verifying forecasts spatially. *Bul. Amer. Soc.*, **91**, Amer. Meteor. Soc., 1365-1373.

Glahn, H. R., 1964a: The use of decision theory in meteorology with an application to aviation weather. *Mon. Wea. Rev*., **92**, 383-388.

_____, 1964b: An application of adaptive logic to meteorological prediction. *J. Appl. Meteor*. **3**, 718-725.

_____, 1985: Statistical weather forecasting. *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, A. H. Murphy and R. W. Katz, eds., Westview Press, 289-335.

_____, 2012: *The United States Weather Service: The First 100 Years*. 116 pp. [Available from the NOAA library, Silver Spring, Maryland.]

_____, and D. A. Lowry, 1967: Short range, subsynoptic surface weather prediction. *ESSA Tech. Memo*. **WBTM TDL-11**, Techniques Development Laboratory, U.S. Weather Bureau, ESSA, U.S. Department of Commerce, 10 pp.

_____, and _____, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor*. **11**, 1203-1211.

_____, and J. R. Bocchieri, 1975: Objective estimation of the conditional probability of frozen precipitation. *Mon. Wea. Rev*., **103**, 3-15.

_____, M. Peroutka, J. Weidenfeld, J. Wagner, G. Zylstra, and B. Schuknecht, 2009: MOS uncertainty estimates in an ensemble framework. *Mon. Wea. Rev*. **137**, 246-268.

Gneiting, T., A. E. Raftery, A. H. Westveld III, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev*., **133**, 1098-1118.

Gringorten, I. I., 1949: A study in objective forecasting. *Bul. Amer. Meteor. Soc*., **30**, 10-15.

_____, 1950: Forecasting by statistical inference. *J. Meteor*. **7**, 388-394.

Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta-RSM short range ensemble forecasts. *Mon. Wea. Rev*., **125**, 1312-1327.

_____, and _____, 1998: Evaluation of Eta-RSM ensemble probabilistic precipitation forecasts. *Mon. Wea. Rev*., **126**, 711-724.

Herbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559-570.

Hirschberg, P. A., and E. Abrams, eds., 2011: Weather and climate enterprise strategic implementation plan for generating and communicating forecast uncertainty. Amer. Meteor. Soc. Rep. **99**. [Available online at: www.ametsoc.org/boardpages/cwce/docs/BEC/ACUF/2011-02-20-acuf-FINAL-REPORT.pdf.]

_____, and others, 2011: Integrating uncertainty into the forecast. *Bul. Amer. Meteor. Soc*., **92**, 1651-1666.

Hu, M. J. C., and H. E. Root, 1964: An adaptive data processing system for weather forecasting. *J. Appl. Meteor*. **3**, 513-523.

Hughes, L. A., 1980: Probability forecasting–Reasons, procedures, problems. *NOAA Tech. Memo*. NWS **FCST 24**, p. 5.

Jolliffe I. T., and D. B. Stephenson, eds. 2012: *Forecast Verification*, 2nd edition. Wiley-Blackwell, 274 pp.

Klein, W. H., F. Lewis, and G. P. Casely, 1967: Automated nationwide forecasts of maximum and minimum temperature. *J. App. Meteor*, **6**, 215-228.

Krzysztofowicz, R., 1986: Optimal water supply planning based on seasonal runoff forecasts. *Water Resour. Res.*, **22**, 313-321.

Lashley, S. L., L. Fisher, B. J. Simpson, J. Taylor, S. Weisser, J. A. Logsdon, and A. M. Lammers, 2008: Observing verification trends and applying a methodology to probabilistic precipitation forecasts at a National Weather Service forecast office. Preprints, *19th Conf. On Probability and Statistics*, New Orleans, Louisiana, Amer. Meteor. Soc., **9.4**.

Lewis, J. M., 2005: Roots of ensemble forecasting. *Mon. Wea. Rev*.,**133**, 1865-1885.

Lindsay, B. G., J. Kettenring, and D. O. Siegmund, 2004: A report on the future of statistics. *Statistical Science*, **19**, 387-413.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. *Scientific Report* No. **1**, Statistical Forecasting Project, Massachusetts Institute of Technology, Cambridge, Massachusetts, 49 pp.

_____, 1965: A study of the predictability of a 28-variable numerical model. *Tellus*, **17**, 321-333.

Lund, I. A., 1955: Estimating the probability of a future event from dichotomously classified predictors. *Bul. Amer. Meteor. Soc*., **36**, 325-328.

Marsban, C., 2003: Neural networks for postprocessing model output: ARPS. *Mon. Wea. Rev*., **131**, 1103-1111.

Mason, I. B., 1982: A model for assessment of weather forecasts. *Australian Meteor Mag*., **30**, 291-303.

Matheson, J. E., and R. L. Winkler, 1976: Scoring rules for continuous probability distributions. *Management Science*, **22**, 1087-1095.

Meyer, A. J., 1871: *Annual report of the Chief Signal Officer*, Government Printing Office, Washington, D.C., p. 8 [available on line from the NOAA library].

Michaelsen, J., 1987: Cross-validation statistical climate forecast models. *J. Climate Appl. Meteor*. **26**, 1589-1600.

Miller, D. W., and M. K. Starr, 1960: *Executive Decisions and Operations Research*. Prentice-Hall, 446 pp.

Miller, R. G., 1958: Regression estimation of event probabilities, U.S. Weather Bureau, Contract Cwb-10704, *Tech. Rep.* No. **1**, The Travelers Research Center, Inc., Hartford, Connecticut.

_____, 1962: Statistical prediction by discriminant analysis. *Meteor. Monographs*, **4**, Amer. Meteor. Soc., 54 pp.

Moore, W. L., 1910: *Report of the Chief of the Weather Bureau for 1908-1910, Part I*, U.S. Department of Agriculture, Government Printing Office, Washington, D.C., p. 15 [available on line from the NOAA library].

Murphy, A. H., 1966: Probabilistic Predictions in the atmospheric sciences–an annotated bibliography. University of Michigan, *Tech. Rep.* ORA Project 06603, 17 pp.

_____, 1969: On the "Ranked Probability Score." *J. Appl. Meteor*., **8**, 988-989.

_____, 1972: Scalar and vector partitions of the probability score: Part I. Two state situation. *J. Appl. Meteor*., **11**, 273-282.

_____, 1980: Misinterpretations of precipitation probability forecasts. *Bul. Amer. Meteor. Soc.,* **61**, 695-701.

_____, 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.

_____, 1996: The Finley affair: A signal event in the history of forecast verification. *Wea. Forecasting*, **11**, 3-20.

_____, 1998: The early history of probability forecasts: Some extensions and clarifications. *Wea. Forecasting*, **13**, 5-15.

_____, R. W. Katz, R. L. Winkler, and W.-R. Hsu, 1985: Decision making and the value of forecasts in a generalized model of the cost-loss ratio situation. *Mon. Wea. Rev*., **113**, 362-369.

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

Neuman, C. J. 1979: On the use of deep-layer mean geopotential height fields in statistical prediction of tropical cyclone motion. Preprints, *Sixth Conference on Probability and Statistics in Atmospheric Sciences*, Banff, Alberta, Amer. Meteor. Soc., 32-38.

NRC 2003: *Fair weather*. National Research Council, National Academies Press, 220 pp.

_____ 2006: *Completing the forecast*. National Research Council, National Academies Press, 112 pp.

NWS, 1971: Operational forecasts derived from primitive equation and trajectory model output statistics (PEAT MOS)–No. 1. *Tech. Procedures Bul*. **68**, National Weather Service, NOAA, U.S. Department of Commerce, 7 pp.

_____, 2011: *NOAA's National Weather Service Strategic Plan*. National Weather Service, NOAA, U.S. Department of Commerce, 44 pp. [Available at: www.noaa.gov/com/weatherreadynation/files/strategic_plan.pdf]

Palmer, W. C., and R. A. Allen, 1949: Note on the accuracy of forecasts concerning the rain problem. Unpublished U.S. Weather Bureau manuscript, 4 pp.

Raftery, A. E., T. Gneiting, F. Balabdoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155-1174.

Roberts, C. F., J. M. Porter, and G. F. Cobb, 1969: Report on weather bureau forecast performance 1967-68 and comparison with previous years. *ESSA Tech. Memo.* **WBTM FCST 11**, Office of Meteorological Operations, U.S. Weather Bureau, ESSA, U.S. Department of Commerce, 44 pp.

Roberts, H. V., 1968: On the meaning of the probability of rain. *Proceedings of the First Statistical Meteorological Conference*, Hartford, Connecticut, Amer. Meteor. Soc., 133-141.

Ruth, D. P., B. Glahn, V. Dagostaro, and K. Gilbert, 2009: The performance of MOS in the digital age. *Wea. Forecasting*, **24**, 504-519.

Sanders, F. 1963: On subjective probability forecasting. *J. Appl. Meteor.*, **2**, 191-201.

Teweles, S. and S. B. Wobus, 1954: Verification of prognostic charts. *Bul. Amer. Meteor. Soc*., **35**, 455-463.

Thompson, J. C., 1950: A numerical method for forecasting rainfall in the Los Angeles area. *Mon. Wea. Rev*., **78**, 113-124.

_____, 1962: Economic gains from scientific advances and operational improvements in meteorological prediction. *J. Appl. Meteor*, **1**, 13-17.

Tribus, M., 1969: *Rational Descriptions, Decisions and Designs*. Pergamon Press, 478 pp.

_____, 1970: Uncertainty and the weather. *Bul. Amer. Meteor. Soc*., **51**, 4-10.

Unger, D. A., 1985: A method to estimate the continuous ranked probability score. Preprints *9th Conference on probability and Statistics in the Atmospheric Sciences.* Virginia Beach, Virginia, Amer. Meteor. Soc., 206-213.

U.S. Government, 1998: Federal civil service requirements for meteorologist positions GS 1340. U.S. Office of Personnel Management.

Vernon, E. M., 1947: An objective method of forecasting precipitation 24-48 hours in advance at San Francisco, California. *Mon. Wea. Rev*., **75**, 211-219.

WB, 1968a: Experimental computer forecasts of maximum and minimum surface temperature. *Tech. Procedures. Bul*. **16**, Weather Bureau, ESSA, U.S. Department of Commerce. 9 pp.

_____, 1968b: Operational forecasts with the sub-synoptic advection model (SAM). *Tech. Procedures. Bul.,* **14**, Weather Bureau, ESSA, U.S. Department of Commerce, 19 pp.

_____, 1969: Operational forecasts with the sub-synoptic advection model (SAM) No. 3. *Tech. Procedures Bul*. **21**, Weather Bureau, ESSA, U.S. Department of Commerce, 12 pp.

_____, 1970: Use of P. E. input in objective temperature forecasts. *Tech. Procedures Bul*. **42**, Weather Bureau, ESSA, U.S. Department of Commerce, 4 pp.

Wilks, D. S., 1997: Resampling hypothesis tests for autocorrelated fields. *J. Climate*, **10**, 65-82.

_____, 2011: *Statistical Methods in the Atmospheric sciences*, 3rd edition. Elsevier, 676 pp.

Wilson, L. J., and M. Vallee, 2003: The Canadian updateable model output statistics (UMOS) system: Validation against perfect prog. *Wea Forecasting*, **18**, 288-302.

Woodcock, F., and C. Engel, 2005: Operational consensus forecasts. *Wea Forecasting*, **20**, 101-111.

WMO, 2001: Guidelines for the education and training of personnel in meteorology and operational hydrology. WMO-No. **258**, Fourth Edition, World Meteorological Organization, Geneva, 123 pp.

_____, 2008: Guidelines on communicating forecast uncertainty. WMO *Tech Document* **4122**, World Meteorological Organization, Geneva, 25 pp.

Yacowar, N., 1968: Objective temperature forecasting. *Proceedings of the First Statistical Meteorological Conference*, Hartford, Connecticut, Amer. Meteor. Soc., 89-95.