

On the Evaluation of Point Precipitation Probability Forecasts in Terms of Areal Coverage

ALLAN H. MURPHY

Environmental and Societal Impacts Group, National Center for Atmospheric Research,¹ Boulder, CO 80307

(Manuscript received 10 April 1978, in final form 11 September 1978)

ABSTRACT

Probability of precipitation (PoP) forecasts can often be interpreted as average point probability forecasts. Since the latter are equivalent to (unconditional) expected areal coverage forecasts, PoP forecasts can be evaluated in terms of observed areal coverages in those situations in which observations of precipitation occurrence are available from a network of points in the forecast area. The purpose of this paper is to describe a partition of the average Brier, or probability, score—a measure of the average accuracy of average point probability forecasts over the network of points of concern—that facilitates such an evaluation. The partition consists of two terms: 1) a term that represents the average squared error of the average point probability forecasts interpreted as areal coverage forecasts and 2) a term that represents the average variance of the observations of precipitation occurrence in the forecast area. The relative magnitudes of the terms in this partition are examined, and it is concluded (partly on the basis of experimental data) that the variance term generally makes a significant contribution to the overall probability score. This result, together with the fact that the variance term does not depend on the forecasts, suggests that the squared error term (rather than the overall score) should be used to evaluate PoP forecasts in many situations. The basis for the interpretation of PoP forecasts as average point probability forecasts and some implications of the results presented in this paper for the evaluation of PoP forecasts are briefly discussed.

1. Introduction

A probability of precipitation (PoP) forecast issued by a National Weather Service (NWS) forecaster is a forecast of the probability of measurable precipitation (i.e., ≥ 0.01 inches) during a specified period (generally 6 or 12 h) at the official raingage. Frequently, however, the probability of precipitation is considered—by forecasters and users of the forecasts—to be the same at every point in the forecast area, in which case a PoP forecast represents an *average* point probability forecast.² Whether or not the point probability is uniform over the forecast area, the accuracy of a PoP forecast is currently determined by comparing the point probability with the precipitation observation at the official raingage, using one or more evaluation measures such as the Brier, or probability, score (Brier, 1950). From time to time NWS forecasters have expressed some dissatisfaction with this procedure. Specifically, the forecasters have indicated that scores based upon the evaluation of PoP forecasts at a single point do not provide a satisfactory measure of the accuracy

of these forecasts, and they have expressed an interest in having such forecasts evaluated at more than one point in the area or over the entire area. However, problems related to the areal interpretation and evaluation of PoP forecasts, as well as the general lack of suitable networks of raingages in the areas of concern, have tended to discourage investigations of alternative evaluation procedures.

Curtiss (1968), and more recently Winkler and Murphy (1976), have shown that an average point probability forecast is equivalent to an expected areal coverage forecast, where the latter represents the expected fraction of the forecast area covered by measurable precipitation.³ The existence of this relationship, together with the fact that many PoP forecasts can be considered to be average point probability forecasts, indicates that these forecasts could be evaluated in terms of observed areal coverages. Any such evaluation procedure would necessarily require precipitation observations from

¹ The National Center for Atmospheric Research is sponsored by the National Science Foundation.

² The basis for the interpretation of PoP forecasts as average point probability forecasts is discussed in greater detail in Section 4.

³ This relationship should not be confused with the relationship between the average point probability \bar{p} , the area probability a and the conditional expected areal coverage e_c , first described by Hughes (1965). According to this relationship, $\bar{p} = ae_c$, where a is the probability of precipitation at one or more points in the forecast area and e_c is the expected areal coverage of precipitation in the forecast area given that precipitation occurs at at least one point in the area.

a network of points in the forecast area. While such data are generally difficult to obtain, Smith (1977) has recently demonstrated that radar data can be used to derive estimates of the observed areal coverage of measurable precipitation. Moreover, Winkler and Murphy (1976) and Murphy and Winkler (1977) have conducted experiments in which NWS forecasters have formulated point and area (including areal coverage) precipitation probability forecasts, and these forecasts were then evaluated using precipitation data from networks of raingages in the areas of concern. These studies, and the existence of a relationship between average point forecasts and areal coverage forecasts, suggest that further investigation of the problem of evaluating PoP forecasts over an area may be warranted at this time.⁴

The purposes of this paper are 1) to describe a relationship between the average probability score—a measure of the average accuracy of average point precipitation probability forecasts over a forecast area—and a measure of the average accuracy of such forecasts when they are interpreted as expected areal coverage forecasts; and 2) to discuss some implications of the existence of this relationship for the evaluation of PoP forecasts. In Section 2 we demonstrate that an average point probability forecast is equivalent to an expected areal coverage forecast. A partition of the average probability score is derived in Section 3a. The partition consists of two terms, a term that represents a measure of the accuracy of areal coverage forecasts and a term that represents the variance of the observations of precipitation occurrence over the area of concern. The relative magnitudes of the terms in this partition are investigated in Section 3b. In Section 4 the basis for the interpretation of PoP forecasts as average point probability forecasts is briefly examined, and some implications of the results presented in this paper for the evaluation of PoP forecasts are discussed.

2. Point probability forecasts and expected areal coverage forecasts

Consider a forecast area defined by a set of n points ($i = 1, \dots, n$). This set could represent a network of raingages or a grid of points for which radar data are available. Let p_i denote the forecast probability of precipitation at point i . Further, let d_i denote an indicator variable, where $d_i = 1$ if measurable precipitation occurs at point i and $d_i = 0$ otherwise. Note that, if $E(d_i)$ denotes the expected value of d_i , then

$$E(d_i) = 1 \times \Pr\{d_i = 1\} + 0 \times \Pr\{d_i = 0\},$$

⁴ Some aspects of this problem were considered in a study by Curran and Hughes (1968).

or

$$E(d_i) = p_i \tag{1}$$

[$p_i = \Pr\{d_i = 1\}$]. That is, the probability p_i is the expected value of the indicator variable d_i .

The areal coverage of precipitation in the forecast area is simply \bar{d} , where

$$\bar{d} = (1/n) \sum_{i=1}^n d_i. \tag{2}$$

Now, if we let e denote an expected areal coverage forecast, then

$$e = E(\bar{d}),$$

or from (2),

$$e = E[(1/n) \sum_{i=1}^n d_i],$$

or

$$e = (1/n) \sum_{i=1}^n E(d_i),$$

or from (1),

$$e = (1/n) \sum_{i=1}^n p_i,$$

or

$$e = \bar{p}, \tag{3}$$

where \bar{p} is the average point probability. Thus, an average point probability forecast is equivalent to an expected areal coverage forecast. This relationship was first discovered by Curtiss (1968) (see also Winkler and Murphy, 1976).

3. Measures of the accuracy of point forecasts and areal coverage forecasts

a. A partition of the average probability score

The probability score is a measure of the accuracy of probability forecasts (Murphy, 1977), and this measure is used on a routine, operational basis to evaluate PoP forecasts formulated by NWS forecasters. If we denote the probability score for the point forecast p_i by $PS(p_i, d_i)$, then

$$PS(p_i, d_i) = (p_i - d_i)^2 \tag{4}$$

(Brier, 1950).⁵ Therefore, the average probability score for an average point probability forecast \bar{p} , when this forecast is evaluated at all n points in the area, is $\overline{PS}(\bar{p}, d_i)$, where

$$\overline{PS}(\bar{p}, d_i) = (1/n) \sum_{i=1}^n (\bar{p} - d_i)^2,$$

or

$$\overline{PS}(\bar{p}, d_i) = (1/n) \sum_{i=1}^n (\bar{p}^2 - 2\bar{p}d_i + d_i)$$

($d_i^2 = d_i$), or

$$\overline{PS}(\bar{p}, d_i) = \bar{p}^2 - 2\bar{p}\bar{d} + \bar{d} \tag{5}$$

⁵ The Brier score employed in this paper is one-half the original Brier score.

[see Eq. (2)]. Now, adding and subtracting \bar{d}^2 from the right-hand side (RHS) of (5), we obtain

$$\overline{\text{PS}}(\bar{p}, \bar{d}_i) = (\bar{p} - \bar{d})^2 + \bar{d}(1 - \bar{d}). \quad (6)$$

The first term on the RHS of (6) is the squared difference between the average point probability \bar{p} and the observed areal coverage \bar{d} . Since $\bar{p} = e$, the expected areal coverage, this term represents the squared error of the forecast e (or \bar{p}). If we denote this measure of accuracy by $\text{SE}(\bar{p}, \bar{d})$, then

$$\text{SE}(\bar{p}, \bar{d}) = (\bar{p} - \bar{d})^2. \quad (7)$$

The range of values of $\text{SE}(\bar{p}, \bar{d})$ is the closed unit interval [0,1].

The second term on the RHS of (6) is simply the variance of the observations of precipitation at the n points defining the forecast area (i.e., the d_i , $i = 1, \dots, n$). If we denote this term by $\text{VAR}(d_i)$, then

$$\text{VAR}(d_i) = \bar{d}(1 - \bar{d}). \quad (8)$$

$\text{VAR}(d_i)$ is of course non-negative, with a maximum value of $1/4$ (when $\bar{d} = 1/2$) and a minimum value of 0 (when $\bar{d} = 0$ or 1).

From (7) and (8), Eq. (6) can be rewritten as

$$\overline{\text{PS}}(\bar{p}, \bar{d}_i) = \text{SE}(\bar{p}, \bar{d}) + \text{VAR}(d_i). \quad (9)$$

Therefore, the average probability score for an average point probability forecast \bar{p} , evaluated at the n points defining the forecast area, is the sum of 1) the squared error of the forecast \bar{p} interpreted as an expected areal coverage forecast (and compared with the observed areal coverage \bar{d}) and 2) the variance of the observations of precipitation at these n points. Thus, Eq. (9) represents a partition of the average probability score $\overline{\text{PS}}(\bar{p}, \bar{d}_i)$ into two terms, one of which is a function of the forecast and the observations and the other of which is a function only of the observations.

Heretofore we have been concerned with a single forecast (\bar{p}). However, we can readily obtain similar expressions for a sample of k average point probability forecasts \bar{p}_j ($j = 1, \dots, k$). Then, for example, (9) would become

$$\overline{\text{PS}}(\bar{p}_j, d_{ji}) = \overline{\text{SE}}(\bar{p}_j, \bar{d}_j) + \overline{\text{VAR}}(d_{ji}), \quad (10)$$

where

$$\overline{\text{PS}}(\bar{p}_j, d_{ji}) = (1/nk) \sum_{j=1}^k \sum_{i=1}^n (\bar{p}_j - d_{ji})^2, \quad (11)$$

$$\overline{\text{SE}}(\bar{p}_j, \bar{d}_j) = (1/k) \sum_{j=1}^k (\bar{p}_j - \bar{d}_j)^2, \quad (12)$$

$$\overline{\text{VAR}}(d_{ji}) = (1/k) \sum_{j=1}^k \bar{d}_j(1 - \bar{d}_j). \quad (13)$$

Here $\overline{\text{PS}}(\bar{p}_j, d_{ji})$ represents the average probability score for the forecasts \bar{p}_j (where the average is taken over all n points in the area and over all k forecasts in the sample), $\overline{\text{SE}}(\bar{p}_j, \bar{d}_j)$ is the average

squared error of the forecasts \bar{p}_j (interpreted as expected areal coverage forecasts e_j and compared with the observed areal coverages d_{ji}), and $\overline{\text{VAR}}(d_{ji})$ is the average variance of the observations of precipitation at the n points in the area over the k forecasting occasions.

b. Relative magnitudes of the terms in the partition

The partition of the average probability score derived in Section 3a has some interesting and potentially important implications for the evaluation of PoP forecasts. However, the significance of these implications depends, to some extent, on the respective contributions of the squared error and variance terms to the (average) probability score—or, equivalently, on the relative magnitudes of $\overline{\text{PS}}(\bar{p}, \bar{d}_i)$ and $\text{SE}(\bar{p}, \bar{d})$. In this regard, it should be noted that since $\text{VAR}(d_i) \geq 0$,

$$\overline{\text{PS}}(\bar{p}, \bar{d}_i) \geq \text{SE}(\bar{p}, \bar{d}), \quad (14)$$

with equality only if $\bar{d} = 0$ or 1. Thus, unless measurable precipitation occurs at none or all of the n points defining the forecast area, the average probability score for an average point probability forecast evaluated at these n points will be greater than the squared error of this forecast interpreted as an areal coverage forecast and evaluated in terms of the observed areal coverage. Of course, an inequality similar to (14) also holds on $\overline{\text{PS}}(\bar{p}_j, d_{ji})$ and $\overline{\text{SE}}(\bar{p}_j, \bar{d}_j)$ for a sample of k average point probability forecasts \bar{p}_j (in this case, equality occurs only if $\bar{d}_j = 0$ or 1 for all j ; $j = 1, \dots, k$).

We now consider (7) and (8) and note that the squared error term is less than or equal to the variance term if and only if $(\bar{p} - \bar{d})^2 \leq \bar{d}(1 - \bar{d})$. Specifically,

$$\text{SE}(\bar{p}, \bar{d}) \leq \text{VAR}(d_i), \quad (15)$$

if and only if

$$p \leq \bar{d} + [\bar{d}(1 - \bar{d})]^{1/2} \quad \text{when } \bar{d} \leq 0.5$$

and

$$p \geq \bar{d} - [\bar{d}(1 - \bar{d})]^{1/2} \quad \text{when } \bar{d} > 0.5.$$

The values of $\text{SE}(\bar{p}, \bar{d})$ and $\text{VAR}(d_i)$ are presented in Table 1 for selected values of \bar{p} and \bar{d} . The pairs of values of these two terms for which $\text{SE}(\bar{p}, \bar{d}) \leq \text{VAR}(d_i)$ are enclosed within solid lines in the table. Note that, for $|\bar{p} - \bar{d}| \leq 0.2$, $\text{SE}(\bar{p}, \bar{d}) \leq \text{VAR}(d_i)$ for most values of \bar{p} and \bar{d} except when \bar{d} is very small or very large (in this case, when $\bar{d} \leq 0.04$ or $\bar{d} \geq 0.96$). That is, if the difference between the average point probability forecast (\bar{p}) and the observed areal coverage (\bar{d}) is relatively small, then the variance term generally will make a larger contribution to the probability score than the squared error term, except possibly for very small or very large values of the areal coverage.

It is also of interest to examine the relative

TABLE 1. Squared error of the average point probability forecast $SE(\bar{p}, \bar{d})$ and the variance of the observations $VAR(d_i)$ for selected values of \bar{p} and \bar{d} . $SE(\bar{p}, \bar{d})$ and $VAR(d_i)$ are the upper and lower values, respectively, in each pair. See text for additional explanation.

		Observed areal coverage \bar{d}										
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Average point probability (expected areal coverage) $\bar{p}(e)$	0.0	0.00 0.00	0.01 0.09	0.04 0.16	0.09 0.21	0.16 0.24	0.25 0.25	0.36 0.24	0.49 0.21	0.64 0.16	0.81 0.09	1.00 0.00
	0.1	0.01 0.00	0.00 0.09	0.01 0.16	0.04 0.21	0.09 0.24	0.16 0.25	0.25 0.24	0.36 0.21	0.49 0.16	0.64 0.09	0.81 0.00
	0.2	0.04 0.00	0.01 0.09	0.00 0.16	0.01 0.21	0.04 0.24	0.09 0.25	0.16 0.24	0.25 0.21	0.36 0.16	0.49 0.09	0.64 0.00
	0.3	0.09 0.00	0.04 0.09	0.01 0.16	0.00 0.21	0.01 0.24	0.04 0.25	0.09 0.24	0.16 0.21	0.25 0.16	0.36 0.09	0.49 0.00
	0.4	0.16 0.00	0.09 0.09	0.04 0.16	0.01 0.21	0.00 0.24	0.01 0.25	0.04 0.24	0.09 0.21	0.16 0.16	0.25 0.09	0.36 0.00
	0.5	0.25 0.00	0.16 0.09	0.09 0.16	0.04 0.21	0.01 0.24	0.00 0.25	0.01 0.24	0.04 0.21	0.09 0.16	0.16 0.09	0.25 0.00
	0.6	0.36 0.00	0.25 0.09	0.16 0.16	0.09 0.21	0.04 0.24	0.01 0.25	0.00 0.24	0.01 0.21	0.04 0.16	0.09 0.09	0.16 0.00
	0.7	0.49 0.00	0.36 0.09	0.25 0.16	0.16 0.21	0.09 0.24	0.04 0.25	0.01 0.24	0.00 0.21	0.01 0.16	0.04 0.09	0.09 0.00
	0.8	0.64 0.00	0.49 0.09	0.36 0.16	0.25 0.21	0.16 0.24	0.09 0.25	0.04 0.24	0.01 0.21	0.00 0.16	0.01 0.09	0.04 0.00
	0.9	0.81 0.00	0.64 0.09	0.49 0.16	0.36 0.21	0.25 0.24	0.16 0.25	0.09 0.24	0.04 0.21	0.01 0.16	0.00 0.09	0.01 0.00
	1.0	1.00 0.00	0.81 0.09	0.64 0.16	0.49 0.21	0.36 0.24	0.25 0.25	0.16 0.24	0.09 0.21	0.04 0.16	0.01 0.09	0.00 0.00

contributions of these two terms to the overall score for actual sets of probability forecasts. For this purpose, we shall consider the results of three experiments in which point and area precipitation probability forecasts were prepared by NWS forecasters. The experiments were conducted in St. Louis, Missouri (November 1972–March 1973), Rapid City, South Dakota (June–September 1974) and Tucson, Arizona (July–September 1977).⁶ The forecasters who participated in these experiments formulated average point probability forecasts (as well as other types of precipitation probability forecasts) for 12 h periods twice each day during the respective experimental periods.⁷ Precipitation

observations were available from networks of 20, 10 and 17 raingages, respectively, to evaluate the results of these experiments. Each network was chosen to provide reasonably complete coverage of the relevant forecast area, and these areas ranged in size from approximately 1300 to 2000 mi².

The average probability score for these forecasts, $\overline{PS}(\bar{p}_j, \bar{d}_{ji})$, the average squared error of the forecasts, $\overline{SE}(\bar{p}_j, \bar{d}_j)$, and the average variance of the observations, $\overline{VAR}(d_{ji})$, are presented in Table 2. Note that the squared error (variance) term accounts for 71.4% (28.6%), 26.8% (73.2%) and 31.3% (68.7%) of the probability score in the St. Louis, Rapid City and Tucson experiments, respectively. The seasonal dependence of the relative contributions of these two terms is quite apparent. In the summer experiments (Rapid City and Tucson) involving weather situations characterized for the most part by convective precipitation, the variance term ac-

⁶ These experiments were designed and conducted by the author in collaboration with R. L. Winkler, Indiana University. The results of the St. Louis and Rapid City experiments were described in detail in Winkler and Murphy (1976) and Murphy and Winkler (1977), respectively.

⁷ In the St. Louis experiment forecasts were formulated for three consecutive 12-hour periods (0–12, 12–24 and 24–36 h), while in the Rapid City and Tucson experiments forecasts were made for the first period only (0–12 h). Thus, while 771 fore-

casts were prepared during the St. Louis experiment, these forecasts involved only 265 different 12 h periods (see Winkler and Murphy, 1976).

TABLE 2. The average probability score $\overline{PS}(\bar{p}_j, d_{ji})$, the average squared error $\overline{SE}(\bar{p}_j, d_{ji})$ and the average variance $\overline{VAR}(d_{ji})$ for the average point probability forecasts formulated during the St. Louis, Rapid City and Tucson experiments. The contributions (percent) of the squared error and variance terms to the overall score are indicated in parentheses. See text for additional explanation.

Experiment	Number of forecasts k	Average probability score $\overline{PS}(\bar{p}_j, d_{ji})$	Average squared error $\overline{SE}(\bar{p}_j, d_{ji})$	Average variance $\overline{VAR}(d_{ji})$
St. Louis	771	0.105	0.075(71.4)	0.030(28.6)
Rapid City	222	0.082	0.022(26.8)	0.060(73.2)
Tucson	184	0.115	0.036(31.3)	0.079(68.7)

counted for approximately 70% of the overall score. On the other hand, in the winter experiment (St. Louis) involving situations characterized primarily by large-scale precipitation events, the variance term accounted for less than 30% of the score. To the extent that the areal coverages observed in these experiments are typical of areal coverages in the summer and winter for forecast areas of similar size⁸ in other locations, we can conclude that the variance (squared error) term will make a larger (smaller) contribution to the overall score than the squared error (variance) term in the summer (winter) season.⁹ Presumably, the relative contributions of these two terms during the fall and spring seasons will generally be bounded by these summer and winter extremes. In any case, it can be seen that the variance term (which does not depend on the average point probability forecasts) will account for a significant fraction of the probability score in all seasons of the year and undoubtedly dominates the squared error term in many locations during the summer.

4. Discussion

The results presented in this paper raise a number of questions related to the evaluation of PoP forecasts. For example, should such forecasts be evaluated only at the official raingage or should they be evaluated (whenever possible) at a network of points in the forecast area? Moreover, in the latter case, should an evaluator compute the average probability score for the forecasts over the network of points or should he calculate the average squared error of the forecasts by interpreting them as expected areal coverage forecasts? Clearly, the answers to these questions depend in large measure upon the nature of the PoP forecasts themselves.

⁸ The impact of the size of the forecast area on these considerations is briefly discussed in Section 4.

⁹ In any case, it seems reasonable to believe that these experimental results do provide a good indication of the likely range of areal coverages for such forecast areas.

In this regard, is a PoP forecast an individual point probability forecast for a specific point in the forecast area (i.e., the official raingage) or is it an average point probability forecast valid effectively at any given point in the forecast area (assuming that any variations in probability among the points are relatively small)? While a full treatment of the issues raised by these questions is beyond the scope of this paper, we will briefly consider some of these issues and their implications for the evaluation of PoP forecasts in this section.

With regard to the nature of PoP forecasts, an issue of fundamental importance relates to the ability of forecasters (or, more precisely, the forecast system) to make forecasts for a particular point in a local forecast area. Specifically, in view of the state of the art of precipitation forecasting and the synoptic-scale observational network on which such forecasts are necessarily based, it seems reasonable to conclude that PoP forecasts are generally applicable to an area considerably larger than that represented by a standard raingage.¹⁰ Moreover, from a practical point of view, it has never been completely clear (at least to the author) how NWS forecasters, in an operational setting, achieve a suitable compromise between their understandable desire to report a probability value that is valid at the official raingage (where their forecasts are evaluated) and the need to provide a forecast that is applicable to individuals at different points in the forecast area. This problem, which could lead to the use of average point probabilities in PoP forecasts, would appear to be particularly acute in locations in which topographic or land-water differences significantly affect the frequency of occurrence of precipitation over the forecast area or in which the official raingage is situated on the outskirts of a major metropolitan area. In a related vein, the results of a survey of NWS forecasters (Murphy and Winkler, 1974) indicate that "different forecasters prefer different definitions of a precipitation event . . . and, as a result, they often use different definitions in connection with their PoP forecasts" (p. 1451). Thus, while Chapter C-91 of the NWS Operations Manual (National Weather Service, 1977) states that a PoP forecast is the probability of measurable precipitation "at a particular point during a given time period" (p. 2), the issues discussed in this paragraph indicate that some reservations are in order concerning the *universal* interpretation of such forecasts as individual point probability forecasts valid only at the official raingage. In this regard, it seems quite likely that many PoP forecasts are, in reality, average point

¹⁰ These considerations may be the source, at least in part, of the forecasters' dissatisfaction with the "standard" evaluation procedure for PoP forecasts (see Section 1).

probability forecasts representative of most if not all of the forecast area of concern.

If a PoP forecast is (or can be considered to be) an average point probability forecast, then how should such a forecast be evaluated? At first, it might seem reasonable to use the average probability score, computed over the network of points in the forecast area, as the evaluation measure. However, we have demonstrated (in Section 3a) that this average score can be partitioned into two terms, one of which represents the squared error of the forecast (interpreted as an expected areal coverage forecast) and the other of which represents the variance of the observations of precipitation occurrence in the area of concern. The magnitude of the first term in this partition can be influenced by the forecaster; in fact, he can reduce its contribution to the overall probability score to zero by correctly forecasting the areal coverage. On the other hand, the second term in the partition does not depend on the forecast. Still, this latter term (generally) varies from forecasting occasion to forecasting occasion and contributes significantly to the overall score during all seasons of the year (see Section 3b). Thus, it seems more appropriate in such a situation to use the squared error term (alone) as the basic evaluation measure for average point probability forecasts. A standard skill score can then be formulated in terms of this measure by determining the relative improvement, in percent, of the squared error for the forecasts of concern over the squared error for (say) climatological forecasts. The use of the squared error term instead of the average probability score in such a skill score effectively redefines the range of the (skill) score to account for the fact that the forecaster cannot influence the variance term in the original partition. It should perhaps be mentioned here that the magnitude of the variance term may also be of interest, and, as a result, it may prove to be desirable to compute this quantity for each set of forecasts as well as the basic measures of accuracy and skill.

The evaluation of PoP forecasts in terms of areal coverage presents some difficulties, particularly with respect to the possible operational implementation of such an evaluation procedure. The foremost problem is, of course, the need for precipitation observations from a network of points in the forecast area. As indicated in the Introduction, Smith (1977) has shown that radar data can be used to obtain estimates of the areal coverage of precipitation, but this approach will require further testing and refinement before it can be used on an operational basis. Moreover, very few forecast areas currently have networks of raingages that can provide the relevant observations, and the number of such networks is not expected to increase significantly during the next few years. Thus, an evalu-

ation procedure for PoP forecasts based on areal coverage considerations is unlikely to be implemented on an operational basis in the near future. This situation, however, need not (and should not) preclude further studies of such procedures or their use on a trial basis in selected locations for which the relevant data are available.

When the evaluation of PoP forecasts in terms of areal coverages becomes feasible on an operational basis (from the point of view of data availability), questions related to the size of the forecast area and the number of points used to define the area will have to be resolved. In this regard, it should be noted that the distribution of areal coverage of precipitation in an area depends upon both of these factors, as well as on the character of the precipitation (i.e., large-scale or convective). For example, small areas with only a few points will tend to have a bimodal distribution of areal coverage (precipitation at all or no points), while large areas or areas with many points will tend to have a unimodal distribution (precipitation at some points). Thus, differences in these factors could lead to differences in scores, as estimated by the evaluation measures, and these latter differences would not be related to the ability of the forecasters or forecast systems of concern. Therefore, it will be necessary to establish some guidelines relative to the size of the forecast area and the number of points in the area prior to the implementation of an operational program of this type.

Finally, the evaluation of PoP forecasts using an areal coverage measure will require some additional resources (e.g., for data collection and processing) and will be somewhat more time consuming. On the other hand, this approach will yield additional information relative to the performance of NWS forecasters and the spatial variability of precipitation. Moreover, it will provide more appropriate measures of the accuracy and skill of PoP forecasts in many situations, an important consideration from the forecasters' point of view. Whether the value of this additional and more realistic information is sufficient to offset the additional expense and time involved in obtaining it would appear to be an open question at this time. Hopefully, further studies of this topic will shed some additional light on this question, as well as on the scientific issues related to the evaluation of PoP forecasts in terms of the areal coverage of precipitation.

Acknowledgments. The author would like to acknowledge the helpful comments of H. R. Glahn, L. A. Hughes and R. W. Katz on an earlier version of this paper.

REFERENCES

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

- Curran, J. T., and L. A. Hughes, 1968: The importance of areal coverage in precipitation probability forecasting. Kansas City, Mo., ESSA, Weather Bureau, Central Region, Tech. Memo. WBTM CR-24, 10 pp.
- Curtiss, J. H., 1968: An elementary mathematical model for the interpretation of precipitation probability forecasts. *J. Appl. Meteor.*, **7**, 3-17.
- Hughes, L. A., 1965: On the probability forecasting of the occurrence of precipitation. Washington, D.C., ESSA, Weather Bureau, Tech. Note 20-CR-3, 36 pp.
- Murphy, A. H., 1977: Evaluation of probabilistic forecasts: Some procedures and practices. *Weather Forecasting and Weather Forecasts: Models, Systems, and Users*, A. H. Murphy and D. L. Williamson, Ed., National Center for Atmospheric Research, 807-830.
- , and R. L. Winkler, 1974: Probability forecasts: A survey of National Weather Service forecasters. *Bull. Amer. Meteor. Soc.*, **55**, 1449-1453.
- , and —, 1977: Experimental point and area precipitation probability forecasts for a forecast area with significant local effects. *Atmosphere*, **15**, 61-78.
- National Weather Service, 1977: Probability of precipitation forecasts. *National Weather Service Operations Manual*, Washington, D.C., NOAA, National Weather Service, Chap. C-91.
- Smith, D. L., 1977: An examination of probability of precipitation forecasts in light of rainfall areal coverage. *Nat. Wea. Digest*, **2**, No. 2, 15-26.
- Winkler, R. L., and A. H. Murphy, 1976: Point and area precipitation probability forecasts: Some experimental results. *Mon. Wea. Rev.*, **104**, 86-95.