POINT AND AREA PRECIPITATION PROBABILITY FORECASTS:

SOME EXPERIMENTAL RESULTS

Robert L. Winkler

Allan H. Murphy

October 1975

Point and Area Precipitation Probability Forecasts:
Some Experimental Results[1]

Robert L. Winkler[2] and Allan H. Murphy[3]

Abstract

An experiment was conducted at the National Weather
Service Forecast Office in St. Louis, Missouri, to inves-
tigate the ability of forecasters to differentiate among
different points in a forecast area with regard to the
likelihood of the occurrence of measurable precipitation
and the relative ability of forecasters to make point and
area precipitation probability forecasts. On each fore-
casting occasion in the experimental period (November 1972
to March 1973), the forecasters made an average point pro-
bability forecast for the St. Louis metropolitan area,
point probability forecasts for five specific points in
the area, an area probability forecast, and an expected
areal coverage forecast.

The results indicate that the forecasters did not
differentiate among the five points very often, but that
this absence of differences among the point probabilities
was justified by the lack of variability exhibited by
the observations of precipitation occurence at these
points during the experimental period. Evaluations of
the average point probability forecasts, individual
point probability forecasts, and expected areal cover-
age forecasts reveal that these forecasts were quite
reliable and accurate and that they were also internally
consistent. The area probability forecasts, however,
tended not to be consistent with the other forecasts,
and the average area probability forecast was consider-
ably lower than the relative frequency of occurrence of
precipitation "somewhere in the area."

The implications of these results for precipitation probability forecasting in meteorology are briefly discussed.

1. Introduction

Probability of precipitation (POP) forecasts have been included in public weather forecasts in the United States on a regular basis for more than twenty years. Prior to 1965, POP forecasts were formulated and issued in Hartford, Connecticut, by the Travelers Weather Service and in several locations by the National Weather Service (NWS).[4] In 1965, a nationwide program involving POP forecasts was initiated by the NWS. This program has now been in existence for a decade, and the evidence presently available suggests that both forecasters and the general public consider the POP forecasts to be an important and integral part of the NWS's public weather forecasts (e.g., American Telephone and Telegraph Company [1]; Bickert [2]; Murphy and Winkler [9]).

Probability forecasts enable forecasters to express their uncertainty about future weather events in a formal, quantitative fashion. In this sense, POP forecasts represent a signif-icant improvement over categorical, or deterministic, forecasts of precipitation (e.g., "rain today") and over forecasts utili-zing vague verbal descriptions of the forecaster's uncertainty concerning precipitation (e.g., "a slight chance of rain today," "rain is likely today"). POP forecasts provide more information than precipitation forecasts not involving probability, and such additional information is valuable to potential users of the forecasts when decisions must be made in the face of uncertainty about the occurrence of precipitation.

Despite the advantages of probability forecasts over other types of forecasts, many aspects of probability forecasting in meteorology are in need of further detailed investigation, both from the standpoint of theoretical studies and from the stand-point of practical and/or experimental studies.[5] For example, since a POP forecast relates to the probability of occurrence of measurable precipitation (i.e., at least 0.01 inches) at a point in the forecast area (in general, at the official raingage),

---

[4]The Travelers Weather Service began issuing POP forecasts at Hartford in 1955, and the National Weather Service issued POP forecasts at Los Angeles and San Francisco as early as 1956 and 1957, respectively.

[5]For a recent review of probability forecasting in meteo-rology, see Julian and Murphy [8].

a POP forecast, as such, is not a particularly "rich" forecast. Specifically, many users are interested in the occurrence of (measurable) precipitation at points in the forecast area for which the probability of precipitation may be quite different than that at the official raingage. The present practice of issuing an average or uniform point probability for the entire area clearly does not satisfy the requirements of such users (unless the point probability is indeed the same at every point in the area).

In this paper, the results of an experiment involving point and area precipitation probability forecasts are presented. The experiment was designed to investigate (1) the ability of forecasters to differentiate among different points in a forecast area with regard to the likelihood of the occurrence of measurable precipitation and (2) the relative ability of forecasters to make point and area (including areal coverage) precipitation probability forecasts. In Section 2, point and area precipitation probabilities are defined, the relationships among such probabilities are discussed, and current forecasting practices are briefly discussed in the light of these different types of probabilities. The experiment is described in Section 3, and the results of the experiment are presented in Section 4. Section 5 contains a brief summary and a discussion of potential future work in the area.

2. Point and Area Precipitation Probabilities

a. Some Definitions and Relationships

When a precipitation forecast is prepared for a given forecast area for a specific time period, various types of probabilities can be considered for inclusion in the forecast. Perhaps the most basic type of probability involving precipitation is a point probability of precipitation, which is the probability of measurable precipitation at a specific point in the forecast area. A point probability can be formulated for any specific point, and in many situations, point probabilities will differ for different points in the forecast area.[6] The most extreme

---

[6] If the forecast area of concern is a very large area, such as an entire state or a region of the U.S., then the probability of precipitation obviously will vary from location to location (e.g., from city to city) within the area. In this paper, however, we are concerned with variations within a smaller area, such as a single metropolitan area. Of course, the smaller the area, the more difficult it is to differentiate among specific points in the area when formulating a forecast. However, the general idea of differentiating among points is still the same as in the case of a very large area.

situation involving differing probabilities is one in which the
forecaster feels absolutely certain that precipitation will
occur at certain points (i.e., the probability of precipitation
equals one at those points) and equally certain that precipita-
tion will not occur at certain other points in the forecast
area (i.e., the probability of precipitation equals zero at the
other points).[7] At the other extreme is the case in which the
point probability is uniform over the forecast area (i.e., the
probability is the same for all of the points in the area).
Of course, many situations fall in between these two extremes.

If a single precipitation probability is desired to sum-
marize the point probabilities for a forecast area, an appealing
summary measure is the average of the point probabilities of
precipitation for all of the points in the area. This measure,
which is called an average point probability, can be formally
defined as

$$\bar{p} = \int_0^1 f(p)\,dp \quad , \tag{1}$$

where p represents a point probability, $\bar{p}$ represents the average
point probability, and f(p) represents the distribution of
point probabilities across the forecast area. If the forecast
area is thought of in terms of a finite grid of K points rather
than in terms of a continuum, then

$$\bar{p} = \sum_{i=1}^{K} p_i/K \quad , \tag{2}$$

where p_i represents the point probability of precipitation at
point i.[8] In the special case in which the point probability
is uniform over the forecast area, $\bar{p}$ is equal to the common
value of the point probabilites. Note that since $\bar{p}$ is a summary
measure, it does not, in general, contain all of the information
provided by the set of individual point probabilities. For
example, $\bar{p} = 0.5$ could correspond to a uniform point probability

_____

[7]A difference of one between probabilities for two points
in the same forecast area would be unusual, especially if the
forecast area is not very large. However, reasonably large
differences may be caused by factors such as variations in top-
ography (e.g., a city adjacent to a mountain range, a portion
of which is included in the metropolitan area), small storms,
or isolated precipitation cells.

[8]In the remainder of this paper, a forecast area will be
defined in this manner in terms of a finite grid of points.

of 0.5, to point probabilities of one for half of the area and zero for the other half of the area, or to numerous other situations.

The primary concern in a precipitation probability forecast could be with the forecast area as a whole rather than with the individual points making up the area. Then an area probability, which is defined as the probability that precipitation will occur somewhere in the forecast area, may be of interest. If a represents an area probability and δ_i is an indicator variable that equals one if precipitation occurs at point i and zero otherwise, then

$$a = P\left(\sum_{i=1}^{K} \delta_i > 0\right) . \tag{3}$$

The sum of the indicator variables in (3) is simply the number of points in the area at which measurable precipitation occurs. For any point, the occurrence of precipitation at that point implies the occurrence of precipitation in the area. Therefore, an area probability must be at least as large as each point probability, which implies that

$$a \geq \max_{i} p_i . \tag{4}$$

Another measure relating to the entire area is the expected areal coverage, which is defined as the expected proportion of the forecast area over which precipitation will occur. If e denotes the expected areal coverage, then

$$e = E\left[\sum_{i=1}^{K} \delta_i / K\right] = \sum_{i=1}^{K} E(\delta_i)/K = \sum_{i=1}^{K} p_i/K . \tag{5}$$

Each point probability, p_i, is equal to the expected value of the corresponding indicator variable, δ_i, and e is simply the average of these expected values. Thus, from (2) and (5),

$$e = \bar{p} . \tag{6}$$

Then, from (4) and (6),

$$e \leq a , \tag{7}$$

since

$$\bar{p} \leq \max_{i} p_i \quad . \tag{8}$$

Definitions of and some relationships among point probabilities of precipitation, the average point probability, the area probability, and the expected areal coverage are provided by (1) through (8). For discussions of some of these relationships, see Curtiss [4]. For a theoretical and numerical investigation of the relationship between point and area probabilities under a simple model with circular precipitation cells of uniform size that are distributed at random over an area that is large compared to the forecast area, see Epstein [5]. Next, current forecasting practices and some problems associated with these practices are examined in light of the different types of probabilities that can be considered in forecasts of precipitation occurrence.

b. Current Forecasting Practices

As indicated in Section 1, POP forecasts are now issued on a regular basis by the NWS, and NWS forecasters have a considerable amount of experience at preparing such forecasts. The official definition of the probability which constitutes a POP forecast is an average point probability of measurable precipitation for a forecast area. In the forecasts formulated by NWS forecasters, the point probability is, in general, implicitly assumed to be uniform over the forecast area. Under this assumption, of course, the POP forecast issued to the public applies to each point in the forecast area. On the other hand, the observation of precipitation is taken at only one point, the official raingage.

When forecasters feel that the point probabilities vary over the forecast area, they occasionally issue two (or more) forecasts, each of which is applicable to a different part of the area. The issuance of more than one probability forecast does not appear to be a common practice, however, despite the fact that forecasters in several locations have expressed considerable interest in making such distinctions more frequently (Hughes, personal communication). The fact that precipitation is generally observed at only one point in a forecast area, the lack of suitable evidence regarding the forecasters' ability to differentiate among different points, and the extra effort required to make more than one forecast are factors which probably contribute to the reluctance of forecasters to issue different forecasts for different parts of the forecast area.

Although the NWS's official definition of a POP forecast is relatively straightforward, such forecasts may not always be

interpreted by the public and by forecasters according to the definition. Some members of the public may interpret a precipitation probability in terms of an area probability, an expected areal coverage, or yet some other definition. Moreover, some forecasters may have a definition other than the official definition in mind when making a precipitation probability forecast. In a recent questionnaire administered to almost 700 NWS forecasters (Murphy and Winkler [9]), the responses indicated that different forecasters prefer different definitions of the event "precipitation" and of a precipitation probability; only 40.1% of the respondents preferred the current NWS definition of a POP forecast, while 35.6% preferred an area probability, 14.5% preferred an expected areal coverage, and the remaining 9.8% gave various other answers. The responses to the questionnaire also indicated that the forecasters sometimes use definitions other than the official definition in preparing their precipitation probability forecasts. Clearly, such practices present users with difficult problems related to the interpretation and use of these forecasts.

3. Design of the Experiment

The subjects in the experiment were fourteen weather forecasters from the NWS's Weather Service Forecast Office (WSFO) at St. Louis, Missouri. Each time the forecasters were on public weather forecasting duty, they made point and area precipitation probability forecasts for the St. Louis metropolitan area. In particular, the forecasters were asked for (1) an average point probability of measurable precipitation for the entire forecast area; (2) point probabilities of measurable precipitation at five specific points (raingages) in the forecast area; (3) an area probability of measurable precipitation for the forecast area; and (4) the expected areal coverage of the forecast area by measurable precipitation. The forecasters were instructed to make the forecasts at approximately 0400 on the midnight shift and 1600 on the day shift. On each forecasting occasion, the forecasts were made for three different twelve-hour periods (on the day shift, for "tonight," "tomorrow," and "tomorrow night"; on the midnight shift, for "today," "tonight," and "tomorrow").

The forecast area was defined for the purposes of the experiment by twenty points corresponding to raingages from the Illinois State Water Survey network of raingages in the St. Louis area. Within the constraints imposed by the location of available raingages, the twenty points were chosen in such a way as to obtain a "representative" coverage of the St. Louis metropolitan area. The set of twenty raingages defined a circular area with a radius of approximately thirty nautical miles centered at a point near the Arch in St. Louis; see Figure 1 for a map of the area showing the twenty raingages. The area probability was specifically defined as the probability of measurable precipitation at one or more of the twenty points constituting the forecast area, and the expected areal coverage was defined as

RADIUS OF CIRCLE
20 Statute Miles

▲ Raingages Defining the Forecast Area

ⓐ Raingages for Which Point Probability
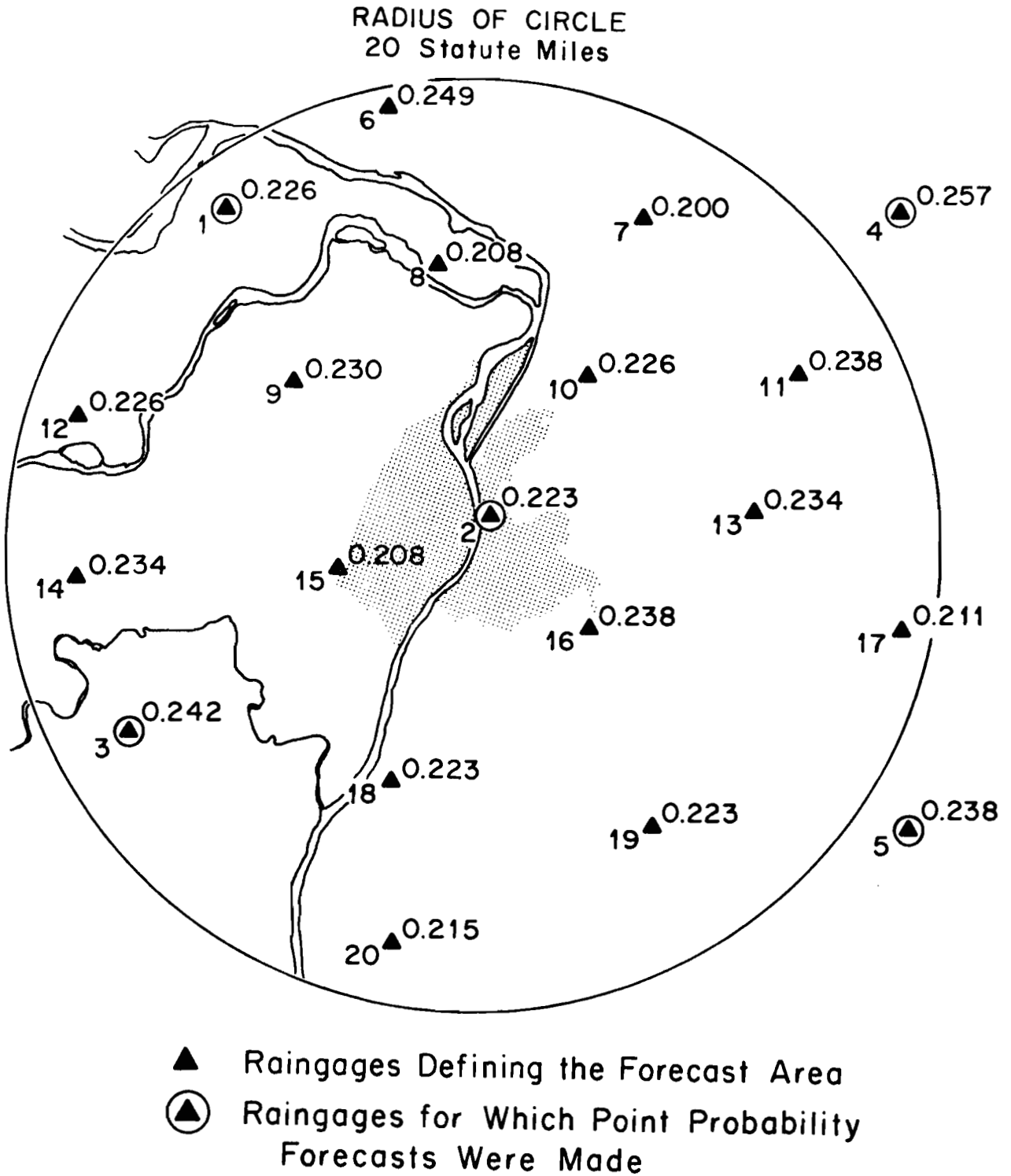   Forecasts Were Made

Figure 1. Network of twenty raingages in the St. Louis metropolitan
           area used in the experiment. Gages 1 through 5 correspond
           to the locations for which the forecasters made individual
           point probability forecasts. The relative frequency of
           occurrence of measurable precipitation during the experi-
           mental period is indicated for each gage in the network.

the expected fraction of the twenty points at which measurable precipitation would be observed. The numbers that could be used for the point and area probability forecasts were 0.00, 0.002, 0.05, 0.10, 0.20, 0.30, ..., 0.90, and 1.00, while the numbers that could be used for the expected areal coverage forecasts were limited to 0.00, 0.05, 0.10, 0.15, 0.20, ..., 0.95, and 1.00.

The five points for which point probabilities were obtained were chosen from the set of twenty points defining the forecast area. The five points consisted of a point near the center of the area and points near the northeast, southeast, southwest, and northwest corners of the area (see Figure 1). The average point probability provided by the forecasters was an average point probability for the entire area, not just an average point probability for the five points for which indvidual point probabilities were obtained.

Prior to the start of the experiment, the forecasters were given instruction sheets that included careful definitions of the different probabilities and of the forecast area (the map in Figure 1 was included) and an explanation of the procedural details of the experiment. The definitions of the probabilities were also included on the response sheets in a further attempt to avoid any confusion on the forecasters' part. The forecasts analyzed in this paper were collected over a period from November 1972 to March 1973. During that period, forecasts were made on 257 different occasions. Since three sets of forecasts (for three different twelve-hour periods) were made on each occasion, 771 sets of forecasts were prepared during the experiment.

4. Results of the Experiment

The experiment was conducted for a period of 131.5 days, or 263 shifts. During this period, forecasts were made for all but six shifts. On each occasion, forecasts were made for three different twelve-hour periods, as noted in Section 3. Thus, a total of 771 sets of forecasts was collected, with each set consisting of an average point probability ($\bar{p}_f$, where the subscript indicates that the average point probability was formulated by the forecaster, not calculated from individual point probabilities), five individual point probabilities (p_1, p_2, p_3, p_4, p_5), an area probability (a), and an expected areal coverage (e). The numbers of sets of forecasts prepared by specific forecasters ranged from three (one occasion) to 159 (53 occasions) for the fourteen forecasters, and five forecasters prepared more than 100 sets of forecasts.[9] Unless

---

[9]The five forecasters preparing over 100 sets of forecasts accounted for 642, or 83.3%, of the 771 sets of forecasts formulated during the experiment.

otherwise stated, all results in this section pertaining to the forecasts refer to the entire sample of 771 sets of forecasts.

This section is divided into three subsections. First, the forecasts are examined, and the relationships among the different types of forecasts are compared with the relationships discussed in Section 2. Next, the observations are examined, and considerations such as the variability among different points in the forecast area are investigated. Finally, the forecasts are evaluated in light of the observations.

a. The Probability Forecasts: An Ex Ante Examination

The averages of the probability forecasts and certain functions of the forecasts that are of interest are presented in Table 1. With regard to the individual point probabilities, the average values of the five probabilities are very close, ranging from 0.222 to 0.233. Thus, any differences among the five point probabilities on particular occasions "canceled out" for all practical purposes over the entire experiment. The forecasters did not consistently assign higher probabilities to any single point than to the other four points.

Differences among the five point probabilities were infrequent in the individual sets of forecasts as well as on the average. For each set of five point probabilities, a sample variance, $s_p^2 = \sum_i (p_i - \bar{p}_c)^2/4$, was computed, where $\bar{p}_c = \sum_i p_i/5$ was the calculated average of the five probabilities.[10] The average value of s_p^2 was 0.001, as shown in Table 1. This value is especially small considering that, except when probabilities less than or equal to 0.10 are involved, any difference in probabilities must be of a magnitude of at least 0.10 because of the limitations on the values that the forecasters could use for their probabilities.

The lack of much variability among the point probabilities is also indicated by the small difference, 0.244 - 0.228 = 0.016, between the average values of the largest point probability, $m = \max_i\{p_i\}$, and the average of the point probabilities, $\bar{p}_c$. Moreover, a closer examination of the forecasts reveals that for 619 (80.3%) of the 771 sets of forecasts, $p_1 = p_2 = p_3 = p_4 = p_5$. It is clear that the forecasters did not differentiate among the

---

[10]The subscript on $\bar{p}_c$, which indicates that it was calculated from the point probabilities, differentiates $\bar{p}_c$ from $\bar{p}_f$, the average point probability formulated by the forecaster.

Table 1. Averages (with standard errors) of the probability forecasts and selected functions of the probability forecasts (n = 771).*

	Average	Standard Error
p_1	0.226	0.0084
p_2	0.227	0.0085
p_3	0.233	0.0086
p_4	0.222	0.0084
p_5	0.231	0.0086
$\bar{p}_f$	0.229	0.0085
$\bar{p}_c$	0.228	0.0084
m	0.244	0.0088
s_p^2	0.001	0.0002
a	0.240	0.0090
e	0.229	0.0086
$\bar{p}_f - \bar{p}_c$	0.001	0.0007
$\|\bar{p}_f - \bar{p}_c\|$	0.005	0.0006
$\bar{p}_f - e$	0.000	0.0011
$\|\bar{p}_f - e\|$	0.007	0.0010
a - m	-0.004	0.0017
a - $\bar{p}_f$	0.014	0.0016

*$p_1, \ldots, p_5$ are the five point probabilities; $\bar{p}_f$ is the forecast average point probability; $\bar{p}_c$ is the calculated average point probability, $\sum_i p_i/5$; m is the largest point probability, $\max_i \{p_i\}$; s_p^2 is the sample variance of the point probabilities, $\sum_i (p_i - \bar{p}_c)^2/4$; a is the area probability; and e is the expected areal coverage.

five points very often.[11] The 152 cases with $p_i \neq p_j$ for at least one (i,j) pair are broken down by forecaster and by lead time in Table 2. The proportion of cases with differing point probabilities varied from 0.00 (Forecaster 4) to 0.45 (Forecaster 2) among the five forecasters who made over 100 sets of forecasts. With respect to lead time, one would expect the number of such cases to decrease with increasing lead time, but the results exhibit only a slight tendency in that direction.

The experiment yielded two average point probabilities on each occasion, one of which $(\bar{p}_f)$ was forecast directly and one of which $(\bar{p}_c)$ was calculated from the individual point probabilities. If these average point probabilities referred to the same set of points, then they should be equal, as required by (2). Since $\bar{p}_f$ referred to a network of twenty raingages while $\bar{p}_c$ referred to only five of these raingages, $\bar{p}_f$ and $\bar{p}_c$ could differ, although the difference would not be expected to be large. In fact, the average value of $\bar{p}_f - \bar{p}_c$ was only 0.001, and the average value of $|\bar{p}_f - \bar{p}_c|$ was 0.005. Thus, the two average point probabilities were very consistent. To make sure that this consistency was not an artifact related to the low variability among the point probabilities, a plot of $\bar{p}_f - \bar{p}_c$ versus s_p^2 was examined (to conserve space, this figure is not presented here), and $\bar{p}_f - \bar{p}_c$ does not appear to be related to s_p^2.

Another comparison of interest is that of $\bar{p}_f$ and the expected areal coverage, e. From (6), $\bar{p}_f$ should equal e. As shown in Table 1, the average value of $\bar{p}_f$ - e for the experiment was 0.000, and the average absolute difference $|\bar{p}_f - e|$ was 0.007. Furthermore, $\bar{p}_f$ = e on 716 (92.9%) of the 771 forecasting occasions, with the percentage of cases with $\bar{p}_f$ = e ranging from 88.2% to 96.5% among the forecasters who made over 100 sets of forecasts; and $|\bar{p}_f - e|$ was larger than 0.05 on only 32 occasions (4.2%). Therefore, the average point probability and the expected areal coverage were very consistent, as they should be.

---

[11]When the forecasters did differentiate among the five points, they usually used either two or three probability values (74 and 67 times, respectively). Four probability values were used ten times, and only once in the entire experiment did complete differentiation (five different values for the five points) occur.

Table 2. Number of cases, by lead time
and forecaster, for which the
five point probabilities were
not all equal.

Forecaster (Number of Sets of Forecasts	Period (Lead Time in Hours)			
	1(0-12)	2(12-24)	3(24-36)	All
1 (114)	11	11	8	30
2 (102)	15	15	16	46
3 (159)	10	6	10	26
4 (156)	0	0	0	0
5 (111)	10	7	7	24
Other (129)	10	10	6	26
All (771)	56	49	47	152

The consistency of point and area probabilities is also of concern. As indicated in (4), the area probability should be at least as large as the largest point probability, since precipitation at any point implies precipitation in the area. From Table 1, the average difference between a and m, the largest point probability, was -0.004. Not only was the difference small; on the average, it violated (4). A closer look at the data reveals that a < m on 59 occasions (7.6%), a = m on 689 occasions (89.4%), and a > m on 23 occasions (3.0%); differences among forecasters in this regard can be seen in Table 3. Theoretically, a can equal m only under the condition that if precipitation occurs anywhere in the area, it must occur at the point corresponding to the largest point probability. If the forecasts are to be considered internally consistent, this condition must be considered to have been satisfied for almost 90% of the situations encountered during the experiment.[12] Of course, the 59 sets of forecasts for which a < m clearly are not internally consistent. Moreover, on nine of these 59 occasions, the weaker condition that a > $\bar{p}_f$ was violated, and the average value of a - $\bar{p}_f$ was only 0.014.

The ex ante examination of the forecasts yields two key results. First, the forecasters seldom differentiated among the five individual points for which point probability forecasts were made. Second, the consistency among the point probabilities, the average point probability, and the expected areal coverage was remarkably high;[13] but these values were not always consistent with the area probability. The analyses in the following two subsections shed further light on these two results.

b. The Observations

During the experiment, forecasts were made for a total of 265 twelve-hour periods. The corresponding observations for these periods were obtained for the twenty raingages defining the forecast area, and the relative frequencies of precipitation at these gages are presented in Figure 1. The relative frequencies ranged from 0.200 to 0.257, with an average relative frequency of 0.227 and a standard deviation of 0.015.[14] Thus,

---

[12]In fact, on 580 (75.2%) of the occasions, all eight forecasts ($\bar{p}_f$, p_1, p_2, p_3, p_4, p_5, a, and e) were equal.

[13]It should be noted that the lack of variability in the forecasts might have facilitated this consistency to some extent.

[14]By way of comparison, the fifteen-year climatological relative frequency at the official raingage for the St. Louis area computed for 1950-1964 data for the months covered by the experiment is 0.184.

Table 3. Number of cases, by forecaster, for
which the area probability (a) was
less than, equal to, and greater than
the largest point probability (m).

Forecaster (Number of Sets of Forecasts)	a < m	a = m	a > m
1 (114)	21	93	0
2 (102)	31	69	2
3 (159)	1	157	1
4 (156)	0	156	0
5 (111)	2	106	3
Other (129)	4	108	17[*]
All (771)	59	689	23

[*]Forecaster 10 accounted for eleven cases with a > m, and
a = m for his remaining 25 cases.

some variability among points in relative frequency of precip-
itation existed, but the amount of variability was not great.

The relative frequencies reflect tendencies for the entire
experimental period and may or may not reflect variability among
points for individual forecast periods.[15] The frequency distri-
butions of areal coverage presented in Table 4 provide some
information about within-period variability. For all twenty
gages, the areal coverage equaled zero on 177 (66.8%) of the
occasions (implying a relative frequency of 0.332 for "precip-
itation somewhere in the area") and equaled one of 29 (10.9%)
of the occasions. For 77.7% of the periods, then, there was
absolutely no variability among the twenty gages. The average
areal coverage was 0.227 (of necessity, equal to the average
relative frequency obtained from the values in Figure 1).

For the five gages corresponding to the individual point
forecasts, the areal coverage equaled zero and one, respectively,
on 184 (69.4%) and 48 (18.1%) of the occasions, and the average
areal coverage was 0.237. Thus, there was no variability among
the five gages for 87.5% of the periods. In light of this
result, the fact that the five point probabilities were equal
for 80.3% of the sets of forecasts does not seem at all sur-
prising.

c.   The Probability Forecasts: An Ex Post Evaluation

A comparison of the average values of the point probability
forecasts (from Table 1) and the corresponding relative fre-
quencies (from Figure 1) indicates that on the average, the
forecasts and relative frequencies were very close. For the five
gages for which point probability forecasts were made, the
differences between the average probabilities and the relative
frequencies were 0.000, 0.004, -0.009, -0.035, and -0.007. The
average difference was -0.009, and the average absolute dif-
ference was 0.011. Furthermore, the difference between the
average value of $\bar{p}_f$ and the average relative frequency at the

twenty gages defining the forecast area was only 0.002, and the
difference between the average value of e and the average areal
coverage was also 0.002.

---

[15]For example, equal relative frequencies of 0.25 could
result from precipitation at all gages 25% of the time and at
no gages the remaining 75% of the time, indicating absolutely
no variability on any specific occasion. Alternatively, such
relative frequencies could result from precipitation at exactly
25% of the gages each period, with the gages receiving precip-
itation changing from period to period, indicating a great
deal of variability on each specific occasion.

Table 4. Frequency distributions of areal coverage for twenty gages defining forecast area and for five gages corresponding to the point probability forecasts.

	Twenty Gages						Five Gages	
Areal Coverage	Relative Frequency	Cumulative Relative Frequency	Areal Coverage	Relative Frequency	Cumulative Relative Frequency	Areal Coverage	Relative Frequency	Cumulative Relative Frequency
0.00	0.668	0.668	0.55	0.004	0.777	0.00	0.694	0.694
0.05	0.019	0.687	0.60	0.019	0.796	0.20	0.045	0.740
0.10	0.023	0.709	0.65	0.000	0.796	0.40	0.030	0.770
0.15	0.015	0.725	0.70	0.004	0.800	0.60	0.023	0.792
0.20	0.011	0.736	0.75	0.008	0.808	0.80	0.026	0.819
0.25	0.004	0.740	0.80	0.011	0.819	1.00	0.181	1.000
0.30	0.008	0.747	0.85	0.000	0.819			
0.35	0.015	0.762	0.90	0.038	0.857			
0.40	0.011	0.774	0.95	0.034	0.891			
0.45	0.000	0.774	1.00	0.109	1.000			
0.50	0.000	0.774						

The correspondence between the average forecasts and the observations, then, was excellent for the point probabilities, the average point probability, and the expected areal coverage. For the area probability, however, the average value of a was 0.240 and the relative frequency of "precipitation somewhere in the area" was 0.332. Thus, on the average, the area probabilities were too low by 0.092.

A more detailed look at the reliability of the forecasts is provided by a graph of probabilities versus relative frequencies. In Figure 2, such a graph is presented for the point probabilities, the average point probability, and the area probability. For the point probabilities, five curves (one for each gage) could be drawn, but since these curves were all very similar, they are summarized in Figure 2 by dashed "upper and lower curves." For each possible probability value, the largest and smallest of the five relative frequencies (one for each gage) corresponding to that probability value were found. The "upper curve" joins the largest values, and the "lower curve" joins the smallest values. Thus, the individual curves for the five gages all lie between the upper and lower curves. It appears from Figure 2 that the point probabilities and the average point probabilities tended to be slightly high (i.e., the curves are below the 45^0 line) for low probabilities and slightly low (i.e., the curves are above the 45^0 line) for high probabilities. This tendency has also been observed in previous studies of probability forecasts in meteorology (e.g., see Hughes [7]; Sanders [12]) and in psychological experiments involving probability assessment (e.g., see Peterson and Beach [11]). Also, note from Figure 2 that the area probabilities are consistently too low in the sense that the curve is above the 45^0 line and is, in general, much further from the 45^0 line than are the curves involving the point probabilities.

Another type of evaluation of probability forecasts involves scoring rules such as the Brier score (Brier [3]), which is widely used in the evaluation of probability forecasts in meteorology. For the St. Louis experiment, quadratic and logarithmic scores, denoted by Q and L, respectively, were computed as follows for each probability forecast (denoted by p):

$$Q(p) = \begin{cases} 1 - (1 - p)^2 & \text{if precipitation,} \\ 1 - p^2 & \text{if no precipitation,} \end{cases} \tag{9}$$

and

$$L(p) = \begin{cases} 1 + \ln p & \text{if precipitation,} \\ 1 + \ln(1 - p) & \text{if no precipitation.} \end{cases} \tag{10}$$
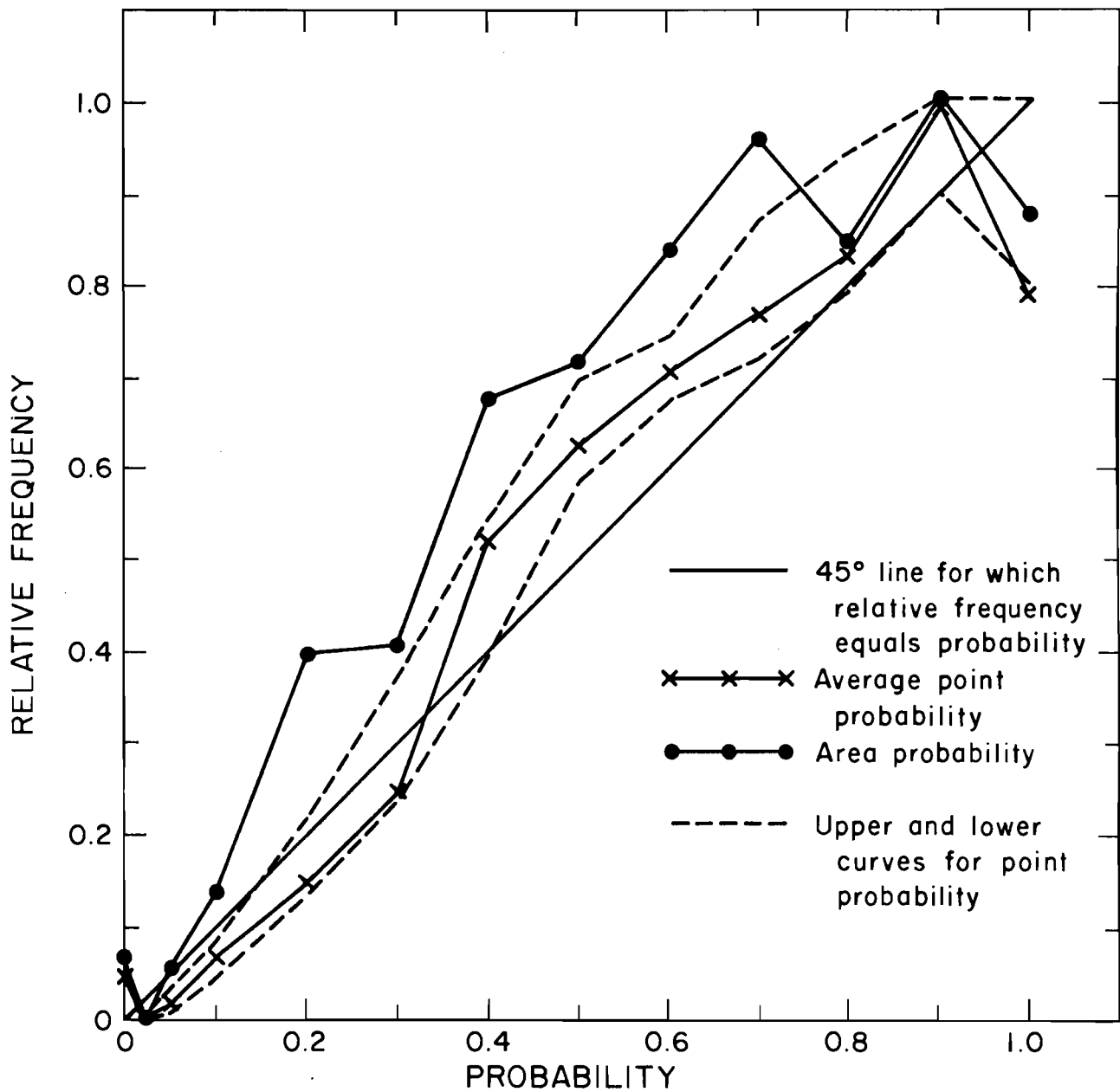
Figure 2. Probabilities versus relative frequencies for the point
and area probability forecasts. The "upper curve" connects
points of the form $(p, \max r_{i,p})$ and the "lower curve"
connects points of the form $(p, \min r_{i,p})$, where p is a
probability value and $r_{i,p}$ is the relative frequency of
precipitation on the occasions for which $p_i = p(i = 1, \ldots, 5)$.

The quadratic score Q is a linear transformation of the Brier score, and for both Q and L, a higher score indicates a "better" forecast (a perfect forecast attains a score of one for each rule). For a general discussion of scoring rules in probability assessment and evaluation, see Winkler and Murphy [13] or Murphy and Winkler [10].

The average quadratic and logarithmic scores are presented in Table 5. Overall, the average scores corresponding to the five point probabilities were quite similar, although the average scores for p_2 were somewhat higher than those for the other point probabilities. In this regard, gage 2 had the lowest relative frequency of precipitation of the gages corresponding to $p_1, \ldots, p_5$ (see Figure 1). This result is consistent with the results of empirical studies (e.g., Hughes [7]; Glahn and Jorgenson [6]) which indicate that the average Brier scores associated with precipitation probability forecasts generally decrease as the relative frequency of occurrence of precipitation decreases. The average scores for $\bar{p}_f$ were just slightly higher (by 0.007 for both Q and L) than the average of all scores for $p_1, \ldots, p_5$. On the other hand, although the ex ante examination of the forecasts in subsection 4.a indicated that the area probability was very close to the point probabilities, the average score for a was much lower than the average scores for $p_1, \ldots, p_5$ and $\bar{p}_f$. It appears that the area probabilities were clearly inferior to the point probabilities.

The effect of lead time is as expected--the average scores consistently decreased as the lead time increased. With respect to individual forecasters, forecasters 1, 3, and 5 had much higher average scores for the point probabilities and the average point probability than did forecasters 2 and 4. The average relative frequencies for the five-gage network on the occasions when forecasters 1, 2, 3, 4, and 5 made forecasts were 0.221, 0.263, 0.215, 0.231, and 0.283, respectively. To the extent that lower relative frequencies of occurrence of precipitation tend to be associated with "better" average scores, as indicated in the previous paragraph, these results would seem to suggest that forecaster 5's forecasts were more "skillful" than those prepared by forecasters 1 through 4. Also, note from Table 2 that of the five forecasters with over 100 sets of forecasts, forecaster 2 had by far the largest proportion of cases (0.45) with $p_i \neq p_j$ for some i and j, while forecaster 4 had by far the smallest proportion of such cases (0.00). For the area probability, on the other hand, forecasters 1 and 2, who set a < m much more frequently than the other forecasters (see Table 3), had the lowest average scores.

Table 5. Average quadratic and logarithmic scores.

			Probability							
Scores	Forecasts	(Number of Sets of Forecasts)	p_1	p_2	p_3	p_4	p_5	$(p_1,...,p_5)^a$	$\bar{p}_f{}^b$	a
	All	(771)	0.895	0.903	0.889	0.885	0.893	0.893	0.900	0.856
	Period 1[c]	(257)	0.909	0.914	0.898	0.905	0.907	0.907	0.907	0.873
	Period 2[c]	(257)	0.894	0.902	0.891	0.884	0.893	0.893	0.899	0.854
	Period 3[c]	(257)	0.883	0.891	0.876	0.867	0.879	0.879	0.894	0.840
Quadratic	Forecaster 1	(114)	0.899	0.902	0.868	0.861	0.882	0.882	0.923	0.820
	Forecaster 2	(102)	0.858	0.875	0.844	0.874	0.839	0.858	0.877	0.811
	Forecaster 3	(159)	0.911	0.919	0.917	0.906	0.915	0.914	0.908	0.892
	Forecaster 4	(156)	0.887	0.887	0.883	0.863	0.876	0.879	0.878	0.847
	Forecaster 5	(111)	0.906	0.914	0.905	0.908	0.916	0.910	0.906	0.874
	Other Forecasters	(129)	0.901	0.916	0.902	0.896	0.919	0.907	0.909	0.874
	All	(771)	0.652	0.676	0.632	0.628	0.649	0.647	0.654	0.549
	Period 1[c]	(257)	0.670	0.699	0.644	0.669	0.676	0.672	0.670	0.577
	Period 2[c]	(257)	0.661	0.681	0.648	0.632	0.655	0.655	0.661	0.557
	Period 3[c]	(257)	0.624	0.648	0.606	0.581	0.616	0.615	0.633	0.514
Logarithmic	Forecaster 1	(114)	0.672	0.685	0.584	0.579	0.630	0.630	0.680	0.468
	Forecaster 2	(102)	0.562	0.607	0.521	0.598	0.514	0.560	0.588	0.437
	Forecaster 3	(159)	0.699	0.733	0.714	0.694	0.714	0.711	0.710	0.653
	Forecaster 4	(156)	0.624	0.619	0.610	0.561	0.590	0.601	0.600	0.524
	Forecaster 5	(111)	0.691	0.711	0.682	0.687	0.714	0.697	0.660	0.601
	Other Forecasters	(129)	0.648	0.708	0.645	0.644	0.708	0.671	0.674	0.566

[a] This column gives the scores averaged over all point probabilities (i.e., the averages of the figures in the columns for p_1, p_2, p_3, p_4, and p_5).

[b] An average score for $\bar{p}_f$ was computed at each of the 20 raingages defining the area, and these values were then averaged over the 20 gages.

[c] Periods 1, 2, and 3 correspond to lead times of 0-12, 12-24, and 24-36 hours, respectively.

5. Summary and Discussion

One objective of the experiment reported here was to investigate the ability of forecasters to differentiate among different points in a forecast area with regard to the likelihood of the occurrence of measurable precipitation. The results indicate that the forecasters did not differentiate among points very often; all five point probabilities were equal for 80.3% of the sets of forecasts. However, the observations also showed very little variability among points; for 77.7% (87.5%) of the forecast periods, the twenty gages defining the forecast area (five gages for which point probabilities were formulated) either all received precipitation or all received no precipitation. Thus, it appears that the lack of variability among the point probabilities was largely justified by the particular weather situations that occurred during the period of the experiment. When evaluated by comparisons with relative frequencies and by scoring rules, the point probabilities were found to be quite reliable and accurate. In any event, with so little differentiation among points, it is difficult to evaluate the degree of "skill" exhibited by the forecasters in making such distinctions, especially in view of the fact that the poorest performances in terms of point probabilities (as measured by deviations between average probabilities and relative frequencies and by average scores) were exhibited by the two forecasters who differentiated among points most frequently and least frequently.

The lack of variability among points during this experiment may be due to the fact that the St. Louis area is not subject to any pronounced local effects (such as a topographical effect) or to the fact that the experiment was conducted in the winter, which tended to minimize the effects of mesoscale weather systems. An important consideration in the selection of the St. Louis WSFO for this experiment was the existence of a reasonably dense network of recording raingages in the St. Louis metropolitan area. Future experiments in locations in which prominent local effects exist and during periods of the year in which mesoscale systems are of greater importance should provide more conclusive evidence concerning the ability of forecasters to differentiate among points in a forecast area in terms of point probabilities of precipitation. The availability of networks of recording raingages will, of necessity, remain a consideration in the selection of suitable locations for such experiments.

Another objective of the experiment was to investigate the relative ability of forecasters to make point and area precipitation probability forecasts. The analysis in Section 4 revealed that the point probabilities, the average point probability, and the expected areal coverage were internally consistent in the sense that the relationships among these forecasts satisfied the conditions discussed in Section 2. The area probability, however, tended not to be consistent with the other forecasts. An evaluation of the forecasts in light of the observations revealed that the area probability tended to be too low while the other

forecasts were quite reliable and accurate. The average area probability was considerably lower than the relative frequency of "precipitation in the area," and the average scores for the area probability were much lower than the average scores for the point probabilities and the average point probability.

Since the POP forecasts issued by the NWS are point probabilities and NWS forecasters at St. Louis and elsewhere regularly make such forecasts, the fact that the point probability forecasts in the experiment were quite reliable and accurate is not surprising. Of course, these results, including the consistency of point probabilities with average point probabilities and with expected areal coverage forecasts, may not generalize to more complex situations in which greater variability exists among the individual point probabilities. Further experimentation could provide some information about the generalizability of the results. The relatively poor performance of the forecasters in terms of area probabilities could also be studied further experimentally. Perhaps some feedback and experience would enable forecasters to learn to avoid making their area probabilities too low.

Theoretical investigations of point and area probabilities could also provide valuable information. Under various assumptions concerning the stochastic nature of the process generating precipitation in an area, the amount of variation among point probabilities and the relative magnitudes of point probabilities and area probabilities could be determined. The degree of dependence among points with respect to precipitation, as reflected by probabilities such as the probability of precipitation at one point conditional upon precipitation at a second point, is a particularly crucial factor in characterizing the relationship between point probabilities and area probabilities.

The questions studied in this paper have important practical implications. For example, if a forecaster feels that the probability of precipitation varies across the forecast area, then the use of a single POP forecast such as an average point probability could be quite misleading. In such situations, the issuance of different POP forecasts for different portions of the forecast area seems advisable. Also, if a forecaster uses a definition of a POP forecast that differs from the official NWS definition, that forecaster's POP forecasts may be affected. In a similar vein, erroneous assumptions about what a POP forecast really means may cause users of POP forecasts to misinterpret such forecasts. Moreover, regardless of the official defintion of a POP forecast, a better understanding of the relationships among the various types of probabilities (e.g., area probabilities vis-a-vis point probabilities) may improve the ability of forecasters to formulate precipitation probability forecasts.

Acknowledgments

References

[1]   American Telephone and Telegraph Company. "Weather Announce-
       ment Study." New York, N.Y., American Telephone and
       Telegraph Company, Market and Service Plans Department,
       unpublished report, 1971.

[2]   Bickert, C. von E. "A Study of the Understanding and Use of
       Probability of Precipitation Forecasts in Two Major
       Cities." Denver, Colorado, University of Denver,
       Denver Research Institute, unpublished report, 1967.

[3]   Brier, G.W. "Verification of Forecasts Expressed in Terms
       of Probability." Monthly Weather Review, 78 (1950),
       1-3.

[4]   Curtiss, J.H. "An Elementary Mathematical Model for the
       Interpretation of Precipitation Probability Forecasts."
       Journal of Applied Meteorology, 7 (1968), 3-17.

[5]   Epstein, E.S. "Point and Area Precipitation Probabilities."
       Monthly Weather Review, 94 (1966), 595-598.

[6]   Glahn, H.R., and Jorgenson, D.L. "Climatological Aspects
       of the Brier P-Score." Monthly Weather Review, 98
       (1970), 136-141.

[7]   Hughes, L.A. "On the Probability Forecasting of the
       Occurrence of Precipitation." Technical Note 20-CR-3.
       Washington, D.C., U.S. Department of Commerce, Environ-
       mental Science Service Administration, Weather Bureau,
       1965.

[8]   Julian, P.R., and Murphy, A.H. "Probability and Statistics
       in Meteorology: A Review of Some Recent Developments."
       Bulletin of the American Meteorological Society, 52
       (1972), 158-165.

[9]   Murphy, A.H. and Winkler, R.L. "Probability Forecasts: A
       Survey of National Weather Service Forecasters."
       Bulletin of the American Meteorological Society, 55
       (1974), 1449-1453.

[10]   Murphy, A.H., and Winkler, R.L. "Scoring Rules in Probabil-
       ity Assessment and Evaluation." Acta Psychologica, 34
       (1970), 273-286.

[11]   Peterson, C.R., and Beach, L.R. "Man as an Intuitive
       Statistician." Psychological Bulletin, 68 (1967), 29-46.

[12]   Sanders, F.   "Skill in Forecasting Daily Temperature and
           Precipitation: Some Experimental Results."   Bulletin
           of the American Meteorological Society, 54 (1973),
           1171-1179.

[13]   Winkler, R.L., and Murphy, A.H.   "'Good' Probability
           Assessors."   Journal of Applied Meteorology, 7 (1968),
           751-758.