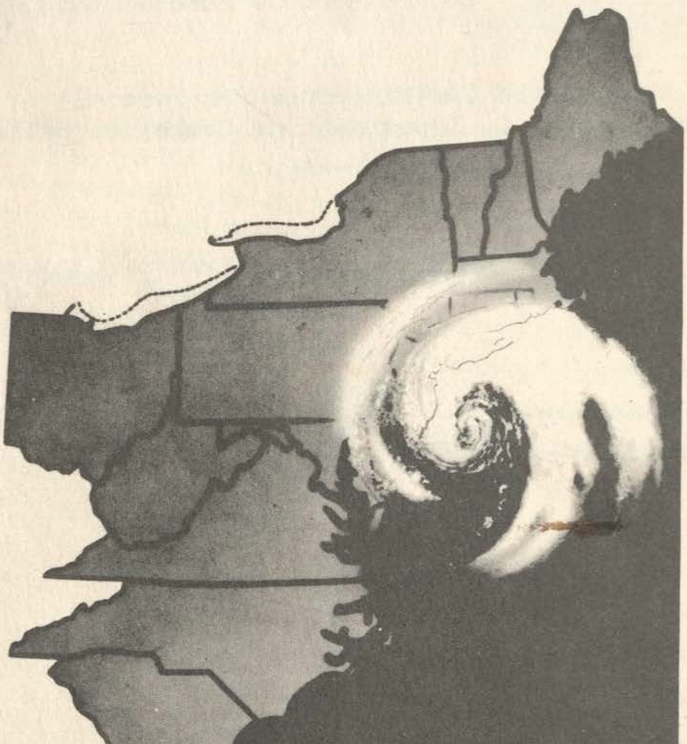
U.S.DEPARTMENT OF COMMERCE • ENVIRONMENTAL SCIENCE SERVICES ADMINISTRATION
Weather Bureau

TECHNICAL NOTE 19 -ER-1

TECHNICAL NOTE 19 - ER-1

An Experiment in Probability Precipitation Forecasting



EASTERN REGION NOTE NO.1

WEATHER BUREAU TECHNICAL NOTES

Eastern Region Notes

The Technical Notes series provides a medium for the documentation and quick dissemination of research and development results not appropriate or not yet ready for formal publication in the standard journals. The primary objective of the series is to provide continuity of documentation for the Weather Bureau's research and development effort. The Regional Notes will report on investigations devoted primarily to regional or local problems.

These notes are for limited reproduction and distribution and do not constitute formal scientific publication. Reference to a note in this series should identify it as an unpublished report.

The reports are available through the Clearinghouse for Federal Scientific and Technical Information, U. S. Department of Commerce, Sills Building, Port Royal Road, Springfield, Virginia 22151.

U.S. DEPARTMENT OF COMMERCE • John T. Connor, Secretary

ENVIRONMENTAL SCIENCE SERVICES ADMINISTRATION

Robert M. White, Administrator

U.S. Weather Bureau • George P. Cressman, Director

TECHNICAL NOTE 19-E R -1

An Experiment in Probability Precipitation Forecasting

Glenn Stallard and Staff

EASTERN REGION NOTE NO.1

WASHINGTON, D.C.

CONTENTS

AN EXPERIMENT IN PROBABILITY PRECIPITATION FORECASTING

Glenn Stallard and Staff
WBAS, Philadelphia, Pa.

1. INTRODUCTION

It is well known that perfect precipitation forecasts can be made consistently only for very short periods of time. For this reason, qualifying terms are frequently used. However, to avoid criticism, the forecaster all too often makes a categorical forecast of rain or no rain when he lacks the capacity to do so with accuracy. The use of "chance," "likely," etc., are at best unsatisfactory because of the diversity of interpretation of the terms. If the recipient of a forecast does not make the same interpretations as the forecaster, then the forecast has not been satisfactorily communicated. Actually some forecasters would use "chance" to indicate 50 percent probability while others might use the same term when the probability is believed to be less than 20 percent. As early as 1920 numerical probability forecasts were issued for the Pecos Valley by the Weather Bureau office at Roswell, N. Mex. [7]. Many years ago Brier [2] proposed the expression of forecasts in terms of probabilities for verification purposes, and his verification techniques have come into general use. Thompson [13, 14] on the other hand, proposed use of numerical probabilities as a means of making forecasts more effective and more suitable for practical application. This concept has been discussed by Case [3], George [4], Gringorton [6] and others and there has generally been a high degree of acceptance of its value. In recent years the growing interest in the use of probabilities in precipitation forecasting has been reflected in the increasing number of Weather Bureau and private meteorological organizations which express forecasts as numerical probabilities.

If forecasts are to be expressed as probabilities, then the question arises as how best to determine the probability to be assigned. Can the decision be entirely objective or must a final subjective determination be made? Numerous local studies have been made in attempts to develop objective methods of making both categorical and probability forecasts of precipitation. Usually techniques involving scatter diagrams are used, such as those described by Beebe [1] and Thompson [12] for local areas and that devised by Rothenberg [9] utilizing vorticity advection at 500 mb. However, these procedures are almost invariably valid for only a portion of the forecast period, for only certain seasons of the year; or have other limitations. Glahn and Allen [5] have indicated that a multiple discriminant analysis method developed with the aid of electronic computers may not be more effective. In any case the cost of operation of high speed computers will limit their use in local studies for some time to come. Consequently, some subjective modification or determination is necessary for most, if not all, probability forecasts. Williams [15] has reported on a subjective approach to probability forecasting of precipitation and temperatures at Salt Lake City during the winter of 1949-50. In that experiment forecasters assigned confidence factors to their forecasts with good results. After analyzing the results of forecasts made during four winter seasons at San Francisco, Root [8] concluded that experienced forecasters could assign economically useful subjective numerical probabilities of the occurrence of rain for periods up to two days in advance. Sanders [10, 11] has made extensive studies of subjective probability precipitation forecasting as a part of the synoptic laboratory program at M.I.T. He found that forecasters had skill in subjectively assigning probabilities to a wide variety of meteorological events over periods of time up to three days.

2. FORECAST PROCEDURES

During the period from Feb. 1, 1963, through Jan. 31, 1965, an experiment in subjective probability precipitation forecasting was conducted at WBAS, Philadelphia. Probabilities were assigned four times daily by the duty forecaster for each 12-hr. interval of the valid forecast periods. An exception was at 1600 GMT (1100 EST) when "THIS AFTERNOON" (1300-1900 EST) constituted the first forecast interval, and a probability forecast was not made for "TOMORROW NIGHT". Generally categories were listed at 10 percent intervals, but smaller intervals were frequently used for very low probabilites and occasionally 95 percent or 98 percent were used for very high probabilities. Zero and 100 percent were discouraged for all forecast intervals except the first. The forecasts were made for Philadelphia with a measurable amount of precipitation at International Airport required for verification.

For comparison climatological probabilities were computed, based upon the most recent 20 yr. of record. 22.5 percent was found to be the average climatological probability for the day-night intervals, 18 percent for the afternoon, and 21 percent for daytime probability.

3. VERIFICATION

If the only aim in probability forecasting were to arrive at a value near the actual frequency of occurrence of precipitation, excellent results could be obtained by forecasting at all times a probability near the climatological average. However, this would serve very little purpose, and the goal of the probability forecaster should be to assign high probabilities for days when precipitation threatens and very low probabilities for days when no precipitation is expected. This, of course, is leaning toward categorical forecasting in which all forecasts would be for 100 percent or 0 percent occurrence of precipitation. Such forecasts, if perfect, would give a skill score of .00 in the equation

$$S = \frac{N_p\ (F - 1.00)^2 + N_{np}\ (F - .00)^2}{(N)}$$

Where N_p is the number of cases with precipitation and N_{np} is the number of cases without precipitation. F is the assigned forecast probability and N is the total number of cases.

In analyzing the results of this experiment average skill score values were computed by months and by forecast categories summarized for each forecast period. In the summaries, forecasts were listed in 10 percent categories except that all forecasts with greater than 90 percent probabilities were combined and the average value used. In like manner, all forecasts with probability under 10 percent were combined and the average value used. The cases in the 25 percent category were equally divided between 20 percent and 30 percent and the cases in the 15 percent category were equally divided between 10 percent and 20 percent, resulting in some fractions in the computation tables. It was found that average skill scores ranged from .06 for the 6 to 12 hr. period (this afternoon) to .12 for the 36 to 48-hr. period (second day). In like manner climatological skill scores were computed for each of the forecast categories, but using the climatological probability for F.

Sanders [10] described a value which he called "reduction over climatology" and developed an equation in which the percent reduction over climatology is equal to 100 minus the penalties for lack of resolution. These terms, as used in this study, were defined as follows:

$$R = \text{percent reduction over climatology} = \frac{100(S_c - S)}{S_c}$$

where S is the skill score and S_c is the skill score for climatology.

$$P_1 = \text{penalty for lack of reliability} = \frac{100(F - O)^2}{S_c}$$

where F is forecast probability and O is actual percentage of occurrence.

$$P_2 = \text{penalty for lack of resolution} = \frac{\frac{100}{N} \sum (O - o_i)^2}{S_c}$$

where N is number of cases in the category and o_i is the individual observed value.

With a perfect forecast R would equal 100, and R would equal 0 for a forecast equal to the climatological probability. For forecasts poorer than climatology R has a negative value. Forecasts have perfect reliability when the actual percentage of occurrence coincides with the forecast probability. For perfect resolution it is necessary that no individual occurrence deviate from the average percentage occurrence for its group. These tests were applied to each forecast category and to the combined forecasts for each forecast interval. It was found that the reduction over climatology, as might be expected, was best for the earliest period and decreased with extension of the period. Reliability was found to be generally good, but lack of resolution was evident, particularly for periods beyond 36 hr.

Table 1 summarizes the results of all forecasts made at 0400 and 1600 GMT for the period 12 to 24 hr. after map time. Table 2 gives the results of forecasts for period 24 to 36 hr. after map time and table 3 represents the period from 36 to 48 hr. after map time. In the first two periods there were 1462 cases, but only 731 in the last since only 0400 GMT forecasts were available for this period. Results show that in the 12 to 24-hr. period each forecast category verified to within 10 percent of the assigned probability. At probabilities 80 percent and under, the error ranged from 0 to 3 percent and the average error for all cases was 1 percent. The forecast probability actually averaged 18 percent, the percentage of occurrence.

Table 1. - 0400 and 1600 GMT forecasts combined for period 12 to 24 hr. after map time. F is forecast category; N is number of forecasts made; N_p, the number of precipitation occurrences; N_{np}, the number of cases without precipitation; $O=N_p/N$, the observed frequency of occurrences; S is skill score; S_c is skill score based on climatology; R is percent reduction over climatology; and P_1 and P_2 are penalties for lack of reliability and lack of resolution, respectively.

F	N	N_p	N_{np}	O	S	S_c	R	P_1	P_2
.98	18	16	2	.89	.11	.54	80	2	18
.90	41	34	7	.83	.15	.51	71	1	28
.80	35	28	7	.80	.16	.49	67	0	33
.70	32	22	10	.69	.22	.43	49	0	51
.60	61	37	24	.61	.24	.38	37	0	63
.50	23	12	11	.52	.25	.34	26	0	74
.40	49	18	31	.37	.23	.25	8	0	92
.30	67.5	22	45.5	.33	.22	.23	4	0	96
.20	157	30.5	126.5	.19	.16	.16	0	0	100
.10	197.5	20.5	177	.10	.09	.11	18	0	82
.02	781	22	759	.02	.03	.07	57	0	43
	1462	262	1200						
Av. .18				.18	.09		41	0	59

$$S = \frac{N_p(F - 1.00)^2 + N_{np}(F - .00)^2}{N}$$

$$S_c = \frac{N_p(.225 - 1.00)^2 + N_{np}(.225 - .00)^2}{N}$$

$$R = \frac{100(S_c - S)}{S_c}$$

$$P_1 = \frac{100\,(F-O)^2}{S_c} \qquad P_2 = \frac{\frac{100}{N}\sum(O-O_i)^2}{S_c}$$

Table 2. - 0400 and 1600 GMT forecasts combined for periods 24 to 36 hr. after map time. F is forecast category; N is number of forecasts made; N_p, the number of precipitation occurrences; N_{np}, the number of cases without precipitation; $O = N_p/N$, the observed frequency of occurrence; S is skill score; S_c is skill score based on climatology; R is percent reduction over climatology; and P_1 and P_2 are penalties for lack of reliability and lack of resolution, respectively.

	F	N	N_p	N_{np}	O	S	S_c	R	P_1	P_2
	.90	15	12	3	.80	.17	.49	65	2	33
	.80	19	16	3	.84	.13	.51	75	0	25
	.70	29	20	9	.69	.21	.43	51	0	49
	.60	70	35	35	.50	.26	.33	21	3	76
	.50	24	13	11	.54	.25	.35	29	0	71
	.40	55	24	31	.44	.25	.29	14	1	85
	.30	70	23.5	46.5	.34	.22	.23	4	1	95
	.20	196.5	49.5	14.7	.25	.19	.19	0	1	99
	.10	303.5	32	271.5	.11	.09	.18	50	0	50
	.03	680	36	644	.05	.05	.08	38	1	61
		1462	261	1201						
Av.	.16				.18	.11		33	1	66

$$S = \frac{N_p\ (F - 1.00)^2 + N_{np}\ (F - .00)^2}{N}$$

$$S_c = \frac{N_p\ (.225 - 1.00)^2 + N_{np}\ (.225 - .00)^2}{N}$$

$$R = \frac{100(S_c - S)}{S_c} \qquad P_1 = \frac{100\ (F - O)^2}{S_c}$$

$$P_2 = \frac{\frac{100}{N}\sum(O - o_i)^2}{S_c}$$

Table 3. - 0400 GMT forecasts for period 36 to 48 hr. after map time. F is forecast category; N is number of forecasts made; N_p, the number of precipitation occurrences; N_{np}, the number of cases without precipitation; $O = N_p/N$, the observed frequency of occurrences; S is skill score; S_c is skill score based on climatology; R is percent reduction over climatology; and P_1 and P_2 are penalties for lack of reliability and lack of resolution, respectively.

F	N	N_p	N_{np}	O	S	S_c	R	P_1	P_2
.90	1	0	1	.00	.81	.04	-1925	202.5	0
.80	2	2	0	1.00	.04	.63	94	6	0
.70	10	7	3	.70	.21	.45	53	0	47
.60	27	14	13	.52	.26	.34	24	2	74
.50	16	5	11	.31	.25	.23	- 9	16	93
.40	30	10	20	.33	.23	.24	4	2	94
.30	46	15.5	30.5	.34	.22	.24	8	1	91
.20	123.5	25.5	98.0	.21	.17	.17	0	0	100
.10	190.5	30	160.5	.15	.14	.14	0	1	99
.04	285	20	265	.07	.07	.08	13	1	86
	731	130	601						
Av. .16				.18	.12		5	4	91

$$S = \frac{N_p\,(F - 1.00)^2 + N_{np}\,(F - .00)^2}{N}$$

$$S_c = \frac{N_p\,(.225 - 1.00)^2 + N_{np}\,(.225 - .00)^2}{N}$$

$$R = \frac{100\,(S_c - S)}{S_c} \qquad P_1 = \frac{100\,(F - O)^2}{S_c}$$

$$P_2 = \frac{\frac{100}{N}\sum (O - o_i)^2}{S_c}$$

Table 4. - Summary of forecast results for all time periods.

Time of Forecast (GMT)	Forecast Period (hr. after map time)	N	N_p	O	F	S	R	P_1	P_2
1600	06-12	731	96	.13	.16	.06	53	1	46
1000 + 2200	06-18	1462	262	.18	.18	.10	49	0	51
0400 + 1600	12-24	1462	262	.18	.18	.09	41	0	59
1000 + 2200	18-30	1462	262	.18	.17	.11	36	0	64
0400 + 1600	24-36	1462	261	.18	.16	.11	33	1	66
0400	36-48	731	130	.18	.16	.12	5	4	91

N total number of cases

N_p number of cases with precipitation

O percentage of occurrence of precipitation

F forecast probability

S forecast skill score

R percentage reduction over climatology

P_1 penalty for lack of reliability

P_2 penalty for lack of resolution

In the period 24 to 36 hr. after map time, forecast error ranged from 1 to 10 percent and was 5 percent or less for all forecast probabilities 50 percent or below. In this period, the average forecast error was 3 percent and the average forecast probability was 2 percent less than the percentage of occurrence.

Forecast error in the 36 to 48-hr. period proved to be considerably greater for individual categories, but ranged only from 1 percent to 7 percent for forecasts of 40 percent or less. The average error for all forecasts was 4 percent and the average forecast probability was again 2 percent less than the percentage of occurrence.

The 1000 and 2200 GMT forecasts were combined for the periods 6 to 18 hr. and 18 to 30 hr. after map time with results quite similar to the first two periods in the 0400 and 1600 GMT forecasts. The average error in the first 12-hr. period was found to be 1 percent and the average error in the second period 2 percent.

Table 4 summarizes the results of all forecasts. It will be noted that precipitation was over-forecast only for the afternoon period, reliably forecast for all other periods up to 24 hr. and under-forecast for the extended time periods. Resolution decreased with the extension of forecast periods and deteriorated rapidly after 36 hr.

4. POINT VERSUS AREA VERIFICATION

Although local forecasts are made for areas on the order of several hundred square miles in size, the practice is to verify precipitation forecasts by observations at one point. It is well known that showers may affect 10 percent or even less of an area and that other types of precipitation may result in a measurable amount of rainfall for only a portion of the area. While the law of averages will result in the observation point's receiving representative rainfall in the long run and making probability forecasts reliable, a high degree of resolution is impossible. This is due to the fact that there must be wide variation in observed values when 0 and 100 percent are the only ones observed. If verification were made from a network of 10 stations, intermediate values at 10 percent intervals could occur in observations and a fair measure of resolution could be applied to the forecast product.

During a trial period in the fall and winter of 1964-65 Philadelphia precipitation forecasts were verified by observations from 10 stations within the meteorologically homogeneous area extending from Baltimore, Md. to Newark, N. J. (fig. 1). Table 5A shows a comparison of skill scores with results from the standard verification program for the 5 months, September through January, based on forecasts issued at 0400 and 1600 GMT. Data from the 0400 GMT forecast for TOMORROW and from the 1600 GMT forecast for TONIGHT were combined and percentage reductions over climatology computed with results given in table 5B. Even though the observed percentage of occurrence of precipitation was near the climatological average with the area system, the use of the 10 stations in the verification resulted in an improvement in resolution and in a greater percentage reduction over climatology. The occurrence of a higher percentage of precipitation from the 10 stations is attributed primarily to the verification of a trace when the duration exceeded 1 hr.

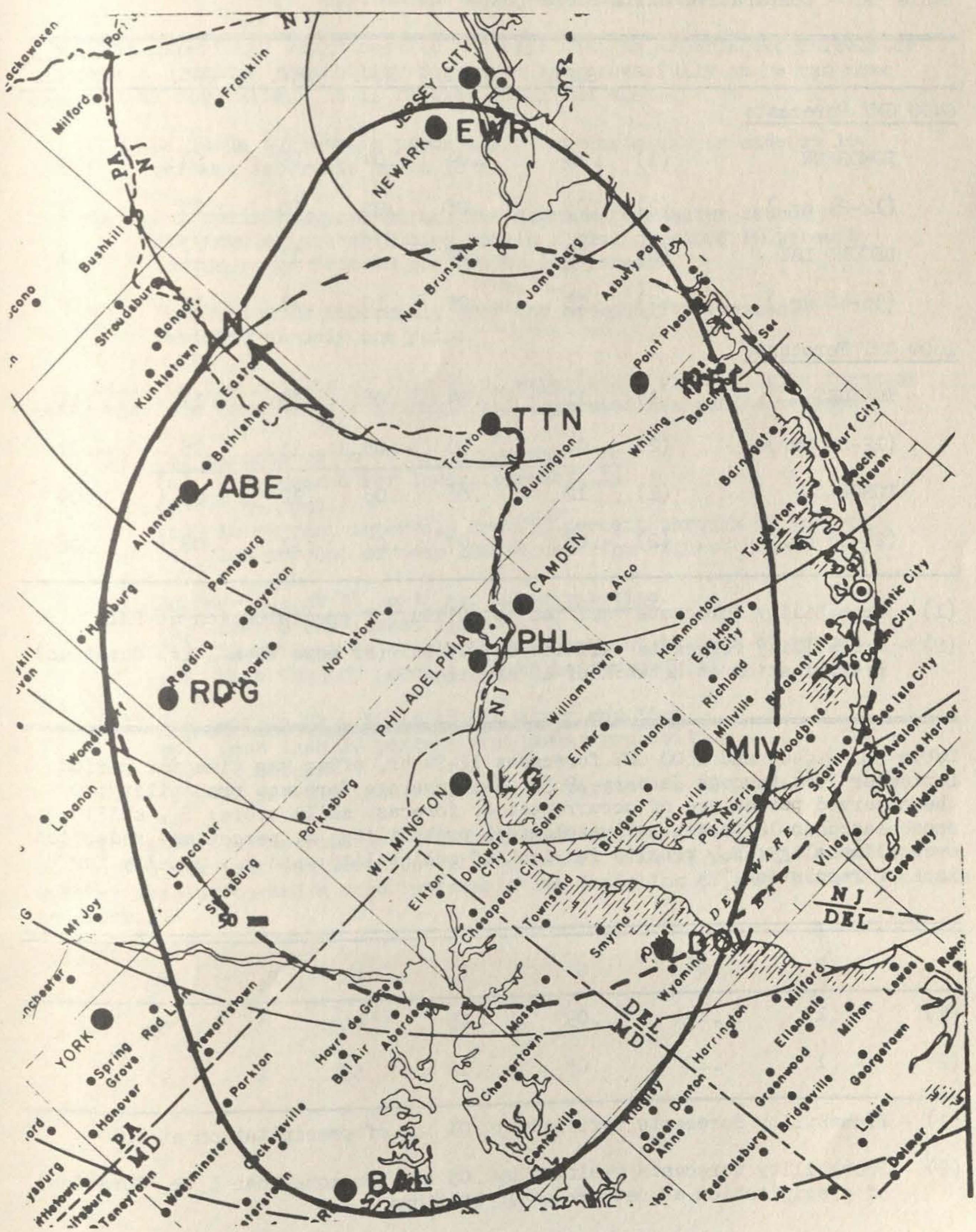


Figure 1. - Precipitation forecast verification area, Philadelphia, Pa.

Table 5A. - Comparative Skill Score (S)

		SEPT.	OCT.	NOV.	DEC.	JAN.	5-Mo. PERIOD
0400 GMT Forecasts							
TOMORROW	(1)	.10	.04	.04	.09	.04	.06
(12-24 hr.)	(2)	.06	.05	.03	.10	.03	.05
SECOND DAY	(1)	.18	.02	.11	.13	.11	.11
(36-48 hr.)	(2)	.13	.03	.10	.11	.11	.09
1600 GMT Forecasts							
TONIGHT	(1)	.11	.04	.08	.19	.14	.11
(12-24 hr.)	(2)	.04	.06	.03	.13	.08	.07
TOMORROW	(1)	.12	.06	.06	.12	.09	.09
(24-36 hr.)	(2)	.08	.07	.07	.11	.08	.08

(1) - Probability forecasts verified by .01 in. of precipitation at PHL

(2) - Probability forecasts verified by .05 in. (or more than 1 hr. duration) precipitation at network of 10 stations

Table 5B. - 0400 and 1600 GMT forecasts 12-24 hr. after map time for period September 1964 through January 1965. F is average forecast probability; O, the observed percentage of occurrence; S, forecast skill score; S_c, skill score based on 12 percent climatological probability; R, percentage reduction over climatology; P_1, penalty for lack of reliability; and P_2, penalty for lack of resolution.

	F	O	S	S_c	R	P_1	P_2
(1)	.17	.14	.09	.15	40	1	59
(2)	.17	.22	.06	.12	55	2	43

(1) - Probability forecasts verified by .01 in. of precipitation at PHL

(2) - Probability forecasts verified by .05 in. (or more than 1 hr. duration) of precipitation at network of 10 stations.

5. CONCLUSIONS AND RECOMMENDATIONS

This experiment would seem to indicate that an experienced forecaster can make subjective probability forecasts as successfully as he can make categorical forecasts. It is further concluded that

(1) Reliable subjective probability forecasts can be made at 10 percent intervals up to 36 hr.

(2) For periods beyond 36 hr. the average forecaster cannot distinguish probabilities within a plus or minus 10 percent in the range from 40 percent to 100 percent.

(3) Resolution is inherently poor for probability forecasts verified at only one point.

Based on the results of this 2-yr. experiment, the following forecast categories are considered reasonable and suggested for public release:

(1) For periods up to 24 hr. after map time.
- (a) 0 or near 0 for lowest probability
- (b) 5 percent
- (c) 10 percent intervals from 10 percent through 90 percent
- (d) 100 percent or near 100 percent for highest probability

(2) For periods of 24 to 36 hr. after map time.
- (a) Near 0 for lowest probability
- (b) 10 percent intervals from 10 percent through 80 percent
- (c) More than 80 percent for highest probability

(3) For periods from 36 to 48 hr. after map time.
- (a) Less than 10 percent for lowest probability
- (b) 10 percent
- (c) 20 percent intervals from 20 percent to 80 percent

It is further recommended that consideration be given to using a network of 5 or 10 stations for verification so that it will be possible to obtain greater resolution and a more representative depiction of what actually occurs in nature.

REFERENCES

1. R. G. Beebe, "Forecasting Winter Precipitation for Atlanta, Ga.," Monthly Weather Review, vol. 78, No. 4, Apr. 1950, pp. 59-68.

2. G. W. Brier, "Verification of Forecasts Expressed in Terms of Probability," Monthly Weather Review, vol. 78, No. 1, Jan. 1950, pp. 1-3.

3. S. Donald Case, Jr., "Some Applications of Meteorological Probabilites," Journal of Applied Meteorology, vol. 1, No. 4, Dec. 1962, pp. 489-496.

4. Joseph J. George, Weather Forecasting for Aeronautics, Academic Press, New York, 1960, pp. 1-6.

5. Harry R. Glahn and Roger A. Allen, "A Comparison of Scatter Diagram Analysis with Discriminant Analysis and a Note on Maximizing the Skill Score," Monthly Weather Review, vol. 92, No. 11, Nov. 1964, pp. 509-512.

6. Irving I. Gringorton, "Probability Estimates of the Weather in Relation to Operational Decisions," Journal of Meteorology, vol. 16, No. 6, Dec. 1959, pp. 663-671.

7. C. Hallenbeck, "Forecasting Precipitation in Percentages of Probability," Monthly Weather Review, vol. 48, No. 11, Nov. 1920, pp. 645-647.

8. Halbert E. Root, "Probability Statements in Weather Forecasting." Journal of Applied Meteorology, vol. 1, No. 2, June 1962, pp. 163-168.

9. Leon Rothenberg, A Technique for Deriving an Objective Precipitation Forecast Scheme Using Numerically Derived Predictors, Office of Forecast Development, U. S. Weather Bureau, Technical Note No. 11, 1963.

10. Frederick Sanders, The Evaluation of Subjective Probability Forecasts, Scientific Report No. 5, Dept. of Meteorology, Massachusetts of Technology, 1958.

11. Frederick Sanders, "On Subjective Probability Forecasting," Journal of Applied Meteorology, vol. 2, No. 2, Apr. 1963, pp. 191-201.

12. J. C. Thompson, "A Numerical Method for Forecasting Rainfall in the Los Angeles Area," Monthly Weather Review, vol. 78, No. 7, July 1950, pp. 113-124.

13. J. C. Thompson, "On the Operational Deficiencies in Categorical Weather Forecasts," Bulletin of the American Meteorological Society, vol. 33, No. 6, June 1952, pp. 223-226.

14. J. C. Thompson, "Economic Gains from Scientific Advances and Operational Improvements in Meteorological Prediction," Journal of Applied Meteorology, vol. 1, No. 1, Mar. 1962, pp. 13-17.

15. Philip Williams, Jr., "The Use of Confidence Factors in Forecasting," Bulletin of the American Meteorological Society, vol. 32, No. 8, Oct. 1951, pp. 279-281.