C

No. 1

Region IV - Technical ~~Note Series~~ Memorandum

SOME NOTES ON PROBABILITY FORECASTING

by

EDWARD D. DIEMER

with

Introduction by Hazen H. Bedke
Weather Bureau Regional Office
Salt Lake City, Utah

September 1965

ENVIRONMENTAL SCIENCE SERVICES ADMINISTRATION
U. S. WEATHER BUREAU

Region IV Technical Note No. 5, September 1965

SOME NOTES ON PROBABILITY FORECASTING

Edward D. Diemer

Introduction by Hazen H. Bedke

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

This discussion on probability by Dr. Diemer sets forth some cogent arguments for the use of probability statements in our weather support to user agencies. This lucid discourse is quite timely at this stage of our operations as we explore the feasibility of making probability judgments for public dissemination as a standard practice in the Western Region, and, in fact, the Weather Bureau at large.

There will be, for some time to come, a degree of uncertainty in our weather forecasts. Making use of probability statements will permit the forecaster to describe more specifically his meteorological assessment of the future weather conditions, making use of the whole spectrum of possibilities. In our society, we not only have the rich and the poor, but we have several gradations in between, commonly known as the middle class. Similarly, we not only have "rain" or "no rain" forecasts, but forecasts in between, say 30% probability of rain. Why should be deny ourselves of this means of expressing our assessment of the meteorological situation?

Although probability forecasts have met with some opposition from various quarters, I know of no opposition from a scientific point of view. The uncertainty of the wisdom of using probability stems from an apprehension that the "public" will not understand probability forecasts. I believe the resolution of this problem rests with an education program of the public. So let us begin! The future development of the country has always leaned with its scientific accomplishments and logical rationale. We as meteorologists have a role to play in this vital developmental growth.

We in the Western Region will depend on the industry, initiative, and dynamic approach with which our MIC's and supporting forecasters pursue the objectives of our program to make full use of probability statements in our support to user agencies and, primarily, the public.

All of you who read this publication will gain a better insight into probabilities and a better appreciation of the use of probabilities in our forecasting program.

Hazen Bedke - Regional Director

II. ABILITY TO MAKE PROBABILITY FORECASTS

Let us consider some of the aspects of weather forecasting. First of all, every weather forecast which has been made has involved a probability. The degree of involvement has varied. For instance, a categorical forecast of rain is identical to a 100% probability forecast of rain. The forecasts, "chance of thunderstorms", or "chance of scattered thunderstorms", are both probability forecasts. The probability is stated verbally. The aviation forecast, "C 20 ⊕ OCNL TRW CHC 5 ⊕2 TRW G40", divides the future weather events into three categories, each with an implied different probability of occurrence.

Suppose that "no rain" forecasts are correct 95% of the time and "rain" forecasts are correct 60% of the time. In effect, the categorical forecaster has sorted all instances into two categories of rain probability, 5% and 60% respectively. Why should the forecaster be restricted to the use of these two probabilities? There is ample evidence to show that he can subdivide these into a larger set of probability categories, thus enhancing the usefulness of his advice!

In the past, each forecaster has determined a probability threshold upon which the forecast weather event is inserted into a certain category; i.e., showers, scattered showers, widely scattered showers, etc. In other words, under certain synoptic conditions, the forecaster realizes that, as far as he can determine, several possible types or degrees of weather may follow. Some forecasters may forecast "snow" at the slightest possibility; others may require more synoptic evidence or a greater possibility. Thus, each forecaster has individually decided on the significance of the forecast to the user and worded his forecast accordingly. So if a particular forecaster thinks that a remote chance of snow is important, his threshold for a snow forecast occurs at a low possibility; whereas, another forecaster may require a high possibility for a categorical snow forecast.

There are several obvious difficulties with this system of issuing categorical forecasts. First, it is <u>not</u> the job of the forecaster to make operational decisions for the user. In fact, this is impossible since there are many users with as many different needs. Secondly, although several forecasters may agree on the probability of a weather event, each may word his forecast differently (partly due to his own opinion as to what the user needs). Thirdly, all the available information is not included in the forecast.

The numerical probability forecast solves these difficulties. The meteorological advice "20% probability of snow tonight" leaves each of the users in a position to evaluate the forecast with respect to his own needs. Also, the users have a definite number which can be recorded and evaluated--rather than a vague and undefined verbal statement. Different forecasters will word the same forecast in the same manner. The probability value passes all the forecaster's information to the user.

Several frequent objections to precipitation probability forecasts are:

(1) There are no objective methods available to make the probability forecasts. (In some cases objective aids are available.)

(2) The forecaster is incapable of delineating subjectively the probability values .0, .02, .05, .10, .20, . . . , 1.00.

The first objection is valid from the purely scientific point of view. However, several empirical studies have shown that forecasters are capable of determining subjective probabilities rather accurately. Therefore, we are justifying our use of subjectively determined probabilities on the results of these studies, and concluding that objective methods are helpful when available, but not absolutely necessary.

The second objection is also answered by the results of the same probability forecast studies referred to above.

Among the many types of forecast weather elements, how many are forecast objectively, how many are forecast subjectively, how many are forecast by a combination of both? If we were to limit our forecasts to purely objective ones, we would issue very few forecasts.

The results of one of these studies which demonstrates the ability of forecasters to determine skillfully probability values are presented in Figure 1 taken from a study by Williams at Salt Lake City. The reliability of a probability forecast is indicated by how close the forecast probabilities agree with the observed occurrence. Thus, if a 20% probability is used 100 times, perfect reliability would be to have rain observed in 20 cases and no rain in 80 cases. Note that in the Salt Lake data given in Figure 1, the low probability values of 0, 02, 05, 10, and 20% show the best reliability. The amount of over-forecasting precipitation increases with higher forecast probability values.

In an article in the Journal of Applied Meteorology, April 1963, Sanders has summarized the utility of objective aids which is well worth quoting:

> "What is the relationship between subjective and objective forecasts? Too often there is a grimly isolated competition to prove which is superior, a situation which the author [Sanders] considers detrimental to the effectiveness of forecasting. If the objective technique wins, it may be regarded as a 'forecast method' and used first grudgingly and finally uncritically to supplant human judgment. If the objective technique is inferior, it may be denoted a 'forecast aid' and largely forgotten.
>
> "A healthier state of affairs, it would seem, is based on the premise that a forecast is a fallible judgment which can use all the objectively processed help it can get. The objective technique provides a probability reference point which the forecaster 'sharpens' by critical appraisal with the use of additional information.
>
> "In the author's experience [Sanders] there are no objective predictions which cannot be improved upon by the forecaster in

this way, even when the objective method produces results which are superior to subjective forecasts made before its introduction. Because of the flexibility and versatility of the human mind, this situation seems likely to continue in the foreseeable future. The important point is that the subjectively modified objective result is the best product.

"The amount of subjective improvement obtainable varies with the individual forecaster and with the objective technique. In a particular instance, the small amount of improvement which the human can provide may not be worth the expense of obtaining it, but a judgment on this matter can hardly be formed without knowledge of the sensitivity of the operational decision to the quality of the forecast."

Ability to Beat the Probability Forecast System

The first question which may enter a forecaster's mind as he contemplates precipitation probabilities is, "How can I bias my forecast to make my forecast verify as well as possible, with a minimum of effort?" The straightforward answer is that the verification system to be used in the RO4 probability forecast program cannot be "beat", "played", or "biased". For a detailed justification of this statement, refer to Sanders, J.A.M., April 1963, or to Brier and Allen in The Compendium of Meteorology, Page 481. Essentially, the content of their work is presented in a simplified form below.

Consider the scoring method, which is based on the equation

$$E = 1/N \sum_{i=1}^{N} (F_i - O_i)^2,$$

where E is the "error", N is the number of forecasts, F is the actual probability forecast, and O is the actual observed probability. The O, observed probability, is assigned the value zero if no rain occurs and one if rain occurs. (This is the E value computed on the "Precipitation Probability Data Sheet", less the 1/N factor.)

As an example of calculation, let the actual probability values forecast be: $F_1 = .20$ (20%), $F_2 = .40$, and $F_3 = .00$. Let the corresponding O values be $O_1 = .00$, $O_2 = 1.0$, and $O_3 = .00$. Then E is evaluated as:

$$E = 1/3 \sum_{i=1}^{3} (F_i - O_1)^2$$

$$= 1/3 \left\{ (F_1 - O_1)^2 + (F_2 - O_2)^2 + (F_3 - O_3)^2 \right\}$$

$$E = 1/3 \quad (.20 - 0.0)^2 + (.40 - 1.0)^2 + (0.0 - 0.0)^2$$

$$= 1/3 \quad .040 + .36 + 0.00$$

$$\simeq .13$$

Now the best score, minimum E, occurs when for all forecasts 0% is forecast when no rain occurs and 100% is forecast when rain occurs. E is then zero. In the above example, $F_1 = 0\%$, $F_2 = 100\%$, and $F_3 = 0\%$ would have been a perfect forecast and E would have been zero instead of .13.

One of the simplest probability forecasts to make is climatological expectancy. The forecast's object is to improve upon the known climatological forecast. Therefore, to get the best score, a forecaster must assign reliable probabilities which deviate from climatology, approaching 0% and 100% for the no rain and rain situation, respectively. This deviation from climatology has been called "sharpness" or "resolution". The perfect forecast, error E = 0, is one in which all forecasts are perfectly resolved into the two categories, 0% and 100%.

An example of skill in departing from a climatological forecast is illustrated in the table below:

Probability Forecast	Climat Forecast	Observed Precipitation
.10	.10	0.00
.50	.10	0.32
.20	.10	0.00
.05	.10	0.00
.05	.10	0.00

The Sanders Score for the probability forecasts is .061, and for the Climat forecast .170. In this case, the forecaster made one forecast as skillfully as Climat, three forecasts more skillfully than Climat, and one forecast less skillfully than Climat. The net result was an improvement over Climat since .061 is less than .170. The per cent improvement of the probability forecasts over the Climat forecasts is calculated to be,

$$\frac{.170 - .061}{.170} = + 64\%$$

Note: suppose that the Sanders Score for the probability forecasts had been .261 instead of .061. Then the per cent improvement of the probability forecasts over the Climat forecasts would have been,

$$\frac{.170 - .261}{.170} = - 54\%$$

The minus sign indicates a "negative" improvement, that is, the probability forecasts were less skillful than the Climat forecasts.

This departure from climatology must be made "reliably"; that is, rain should be observed with the same frequency as forecast. For instance, rain should be observed on 40 cases out of one hundred-40% probability forecasts. Here is where the forecaster may try to beat the system by systematically over or under forecasting precipitation. The following table will demonstrate that this cannot be done to achieve the best score possible.

Case	Probability Forecast	No. of Forecasts	Times Rain Occurred	Times No Rain Occurred	E
(1)	40%	100	40	60	.24
(2)	40%	100	30	70	.22
(3)	40%	100	50	50	.26
(4)	30%	100	30	70	.21
(5)	50%	100	50	50	.25

Cases (1), (4), and (5) represent perfect reliability. Case (2) illustrates an example of "over" forecasting; that is, rain was forecast to occur 40 times out of 100 cases, and it actually occurred 30 times. Case (3) indicates "under" forecasting. That is, rain was forecast to occur 40 times out of 100 and it actually occurred 50 times.

A forecaster who wishes to deliberately bias his forecast may do so by either under forecasting or over forecasting precipitation. The above table shows the effects of deliberately under forecasting or over forecasting precipitation and how the forecaster's score suffers in both instances.

In comparing Case (5), perfect reliability, with Case (3), under forecast, we find that the systematic under forecast raises the E score from .25 to .26. Similarly, in comparing Case (4), perfect reliability, with Case (2), over forecast, we find that the over forecasting raises the E score from .21 to .22.

Upon further reflection and comparisons of E scores, it might appear that there is an advantage to over forecasting since the E score does decrease from .24 (Case (1) perfect reliability) to .22 (Case (2) over forecasting). However, to over forecast, the forecaster had to first arrive at 30% as being the true probability, that is, the probability which would be observed. This is necessary since a systematic forecast deviation must be deviated from something--in this case the true probability value. So, to over forecast knowingly (i.e., try to "beat" the verification scheme) the forecaster would have to determine 30% as near the true probability but issue 40% as his forecast. His E score for the

40% (Case (2) over forecast) is .22. Had he not over forecast, his forecast probability would have been 30% (Case (4) perfect reliability) with an E score of .21, which is better than the over-forecast score. Consequently, the forecaster cannot improve his score by over forecasting or by under forecasting precipitation.

In summary, to get the best score, the forecaster must deviate as far from climatology as possible, approaching 0% and 100% probability values for the no-rain and rain cases respectively. This is called resolution or sharpness. The forecaster must recognize nearly certain events as often as possible, using the appropriate high or low probability values. Also, to get the best score, the forecaster must show reliability or validity in his forecast. That is, the forecast probability must correspond as closely as possible to the observed frequency of precipitation occurrence. A systematic over or under forecast does not have reliability and so does not produce the best score.

III. USE AND ACCEPTANCE OF PROBABILITIES

Do probabilities have utility in Weather Bureau guidance material? There is a general affirmative agreement among forecasters on this question. Undoubtedly probability values communicate much valuable information in a direct and simple manner.

Do probabilities have utility in public forecasts? This question provokes some disagreement among forecasters and the lay public. Let us consider some of these disagreements.

The question has been raised, "Will the public understand probability forecasts?" At this point the "public" is usually divided into the "General Public" and "Industry." It is usually agreed that "Industry" can understand probability forecasts and that such forecasts are beneficial to "Industrial" users. As for the "General Public" there is considerable doubt that all "the men on the street" will understand initially and appreciate the added information given in a probability forecast as compared to a categorical forecast. Indeed, there is reason to believe that a small minority of the "General Public" in essence will always want the forecaster to make their operational decisions for them. They don't want and won't appreciate the added information a probability forecast gives them. Although this fact is disturbing (more to some forecasters than others), there is reason to be optimistic that the greater majority of the "General Public" can be taught to understand, to appreciate, and to use probability forecasts. Where probability forecasts have been properly introduced to the public, they have met with favorable acceptance by the vast majority of people.

For example, Seattle has used probabilities successfully in all their local forecasts. San Francisco has used probabilities in their local forecasts issued during the winter season. St. Louis has used probabilities in all of their local and zone forecasts and occasionally in the state forecasts. Some private corporations, such as Travelers Research at Hartford, Connecticut, promote their companies by making public weather forecasts in terms of probabilities.

The Central Office of the Weather Bureau recognizes the need for some public education regarding the meaning and use of probability forecasts and are presently preparing such a program. This educational program is expected to be completed or well under way before any decision to make general public dissemination of probability forecasts is made.

IV. DETERMINING PROBABILITIES

The current state of the science of meteorology suggests that a judicious blending of subjective and objective estimates of probability is the best method to use in making probability forecasts. Determination of probabilities by purely objective techniques for most forecast elements is and will continue to be rare. On the other hand, purely subjective techniques of forecasting and at some stations of assigning forecast probabilities are in general use.

Examples of probability forecasts made by mostly subjective estimates based on facsimile guidance are given in Sanders article in J.A.M., April 1963. Of particular interest in this article is the significant reliability shown in the forecasts made by M.I.T. students with no forecasting experience and only a few weeks' exposure to synoptic data and analyses. The following two examples are taken from Sanders' paper.

In Figure 2 is plotted the forecast probability versus the observed frequency of occurrences for the instructors' forecasts in the 1955-1956 seasons. These forecasts were generally for a 24-hour period, and the probabilities were in increments of 10%. The elements forecast covered a wide spectrum: ceiling, visibility, wind, precipitation, temperature, etc. Most forecasts were for the United States, but some referred to locations throughout the Northern Hemisphere. As indicated by the near coincidence of the curves, these forecasts were highly reliable.

In Figure 3 is plotted the forecast departure from the climatological probability versus observed departure from climatological probability for the forecasts of a student with <u>no</u> forecasting experience but using facsimile charts as guidance. These departures from climatology are another way of plotting a reliability curve. These forecasts were for precipitation and temperature over 24-hour periods extending 72 hours after initial data time. Note the significant reliability shown in these forecasts. Similar curves for eight other students are presented in the paper.

A study conducted by Root in San Francisco clearly demonstrates that the San Francisco forecasters have skill in assigning probability values to precipitation events (J.A.M., June 1962). In this case, experienced forecasters had facsimile, teletype, and objective aids available to use in constructing their forecasts. Figure 4 is the reliability curve for the local forecasts of rain made at San Francisco during four winter seasons.

As mentioned above, the best approach to probability forecasting is through subjective improvement of the forecast given by objective techniques. Objective forecast studies can easily be designed to yield results in terms of probabilities. Several examples from studies made by Williams at Salt Lake City are illustrated in Figures 5 and 6.

Studies of this type will usually not divide probabilities into all the categories used in the actual forecast. Many times the results will be highly skewed showing probabilities only at the high or low end of the scale. Nonetheless, these results are very useful in practical forecasting. In developing forecast studies, forecast offices should incorporate NMC prognostic parameters as well as observed parameters and when appropriate have the final result given in terms of probability. Studies of this type normally will be for those weather phenomena which are most important; i.e., snow, heavy rain, critical temperatures, damaging wind, etc.

There are numerous published examples of objective forecasting techniques. Many of these can be adapted so as to apply to different stations. Also, there have been publications dealing with the procedures of developing objective aids.

Several pertinent references are:

(1) Office of Forecast Development, Technical Note No. 11.
(2) "Some Techniques for Deriving Objective Forecasting Aids and Methods," AWSM 105-40.
(3) "Selected Bibliography on Forecast Development," compiled by J. O. Ellis, Technical Development Laboratory, Systems Development Office.

An important factor in determining forecast probabilities is the definition of the occurrence of the event. For example, during an "academic" test of forecasting precipitation probabilities, it is useful to define the occurrence of a precipitation event as measurable rain in the airport rain gauge. Such a definition may not be acceptable, however, when verifying probability forecasts issued to the public for such metropolitan areas as Salt Lake City. Here the frequency of occurrence of measurable precipitation varies significantly over the area and is largely a function of location with reference to the mountains and the lake. A more general definition of a precipitation event is needed; It may be relatively simple to decide on this more general definition; but, unfortunately, the climatological expectancy of such a definite event is not usually available. Consequently, in such cases some changes may be necessary in the definition and verification of a precipitation event when a test program ends and operational probability forecasts are begun and issued to the General Public.

V. ODDS AND PROBABILITY

Odds and probability are frequently confused. While there is a relationship between these two concepts, the statement of odds in terms of probabilities or vice versa is frequently given incorrectly. For example, if the odds are 2 to 1 that it will rain, what is the probability of rain? The correct answer is 66.67%. Because the term "odds" may be misinterpreted and because it is closely associated with betting procedures, <u>it should never be used in a weather forecast</u>..

One or two betting examples here will help point out the difference between odds and probability. Suppose two people make a bet at even odds, 1 to 1, on a horse running in a certain race. The fact that the bet was made at even odds, 1 to 1, means that each person could see no more reason for the horse to win than he can find reason for the horse to lose. Their estimate is that in two identical races the horse should win once. In other words, there is a 50% probability of the horse winning the race upon which the bet was placed. By way of comment, after the race neither person involved in the bet could say for certain that he had placed the correct odds on the race. However, after two similar races if the horse won one and lost one, there would be some indications that the odds were correct. But after 100 similar races in which the horse won 50 and lost 50, most people would be convinced that the original 1 to 1 odds were perfectly correct. (Relating this to weather forecasting, a statement of a 50% probability of rain for tomorrow cannot be verified by the occurrence or nonoccurrence of rain tomorrow. Rather, it will take the tabulation of many forecasts of 50% probability of rain and actual occurrences before the statement can be verified. Thus, the forecast verification by the Sanders Score and discussion in Section II is really an administrative procedure to evaluate forecasts on a daily basis, rather than a true verification of a given forecast.)

Now let us change horses. Suppose you figure there is only a 33.33% chance of a horse to win. Would you bet at even odds? If the 33.33%

probability was correct, the corresponding odds would be 2 to 1, meaning the horse would lose two races for each race he won--under similar conditions. Note, these odds could easily be mistaken to mean two wins in three tries, or 66.67% probability.

A short table of percentages and corresponding odds is given below. The odds, 3 to 2, or 40% probability (percentage) means that out of 5 cases the event will occur 2 times and not occur 3 times.

Percentage	Odds	Percentage	Odds
0.10%	1000-1	50.00%	1-1
0.20%	500-1	60.00%	2-3
0.99%	100-1	71.42%	2-5
1.96%	50-1	80.00%	1-4
5.00%	19-1	90.00%	1-9
10.00%	9-1	99.01%	1-100
20.00%	4-1	99.80%	1-500
29.41%	12-5	99.90%	1-1000
40.00%	3-2		

APPENDIX

Error (E) Table

The following table is inserted here as a ready reference and solution of the error equation:

$$E = (F - O)^2$$

for given values of forecast probabilities versus the occurrence or nonoccurrence of the forecast event.

Probability %	Event Not Observed 0	Event Observed 1.0
0	.00	1.00
02	.0004	.9604
05	.0025	.9025
10	.01	.81
20	.04	.64
30	.09	.49
40	.16	.36
50	.25	.25
60	.36	.16
70	.49	.09
80	.64	.04
90	.81	.01
100	1.00	.00

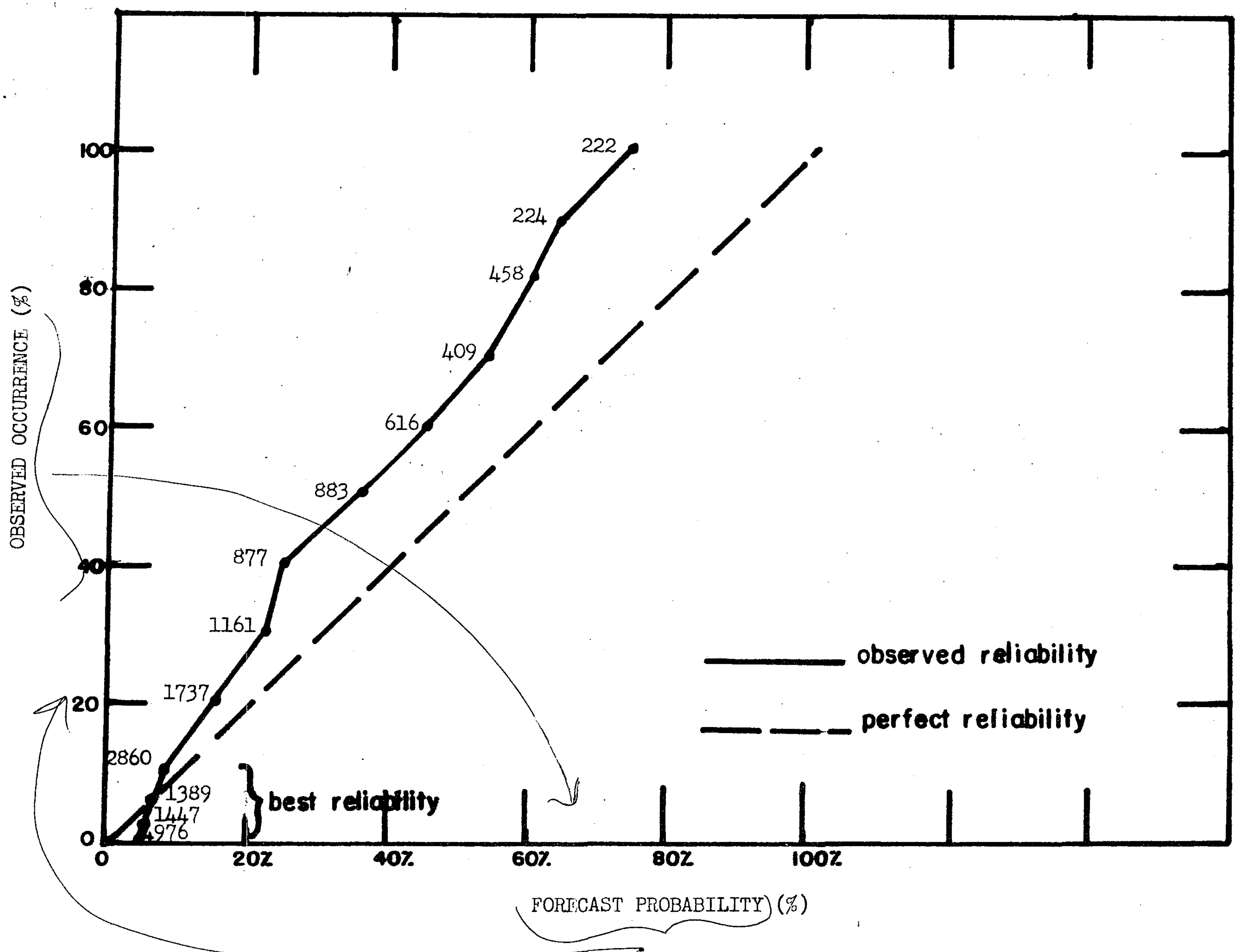

Figure 1. Forecast probability vs. observed relative frequency of occurrence by USWB, Salt Lake City forecasters, October 1960 through September 1964. Number of forecasts is given next to each data point. (After Williams)

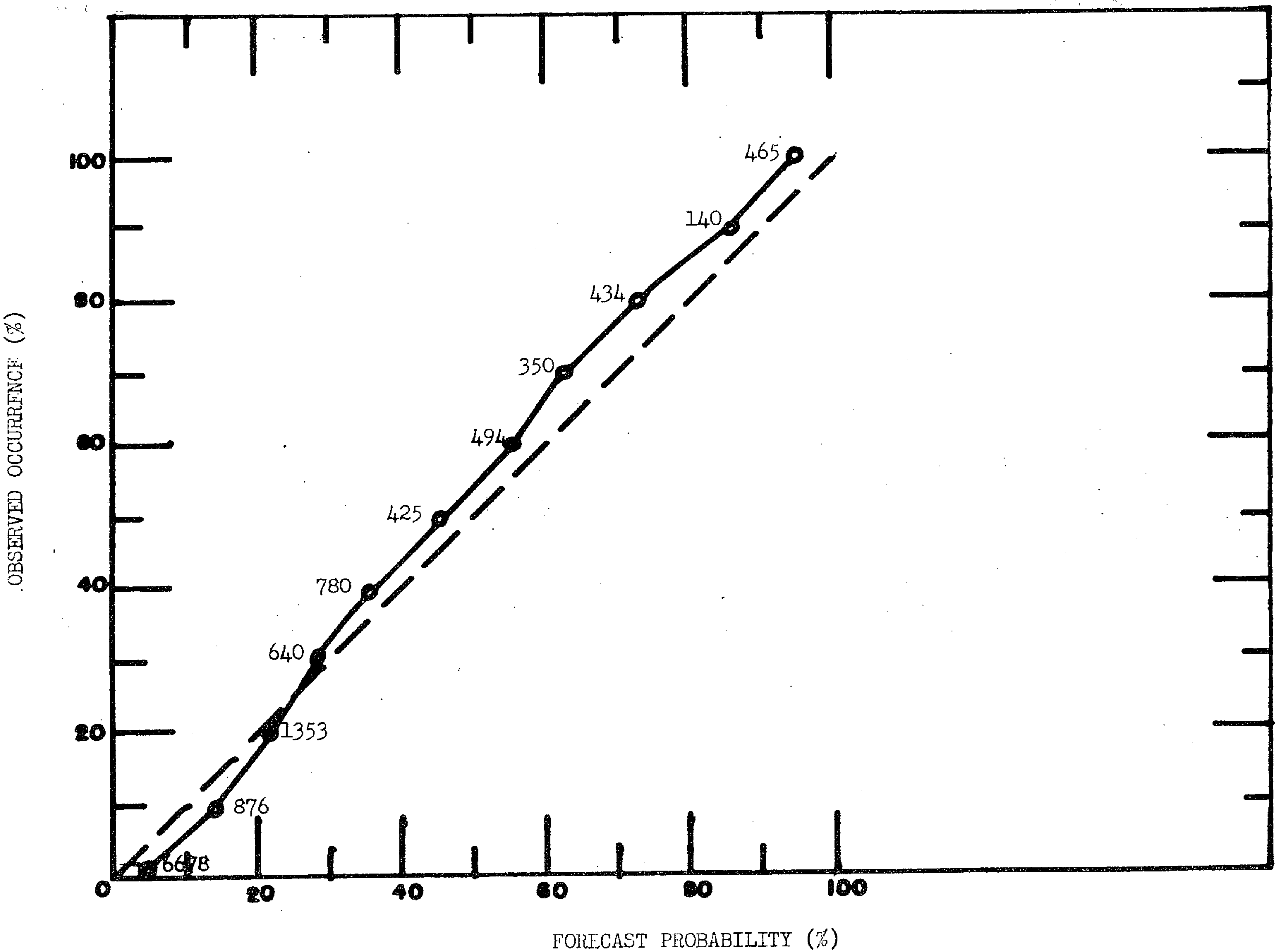

Figure 2. Forecast probability vs. observed relative frequency of occurrence for MIT instructors' forecasts in 1955-1956 seasons. Number of forecasts is given next to each data point. (After Sanders)

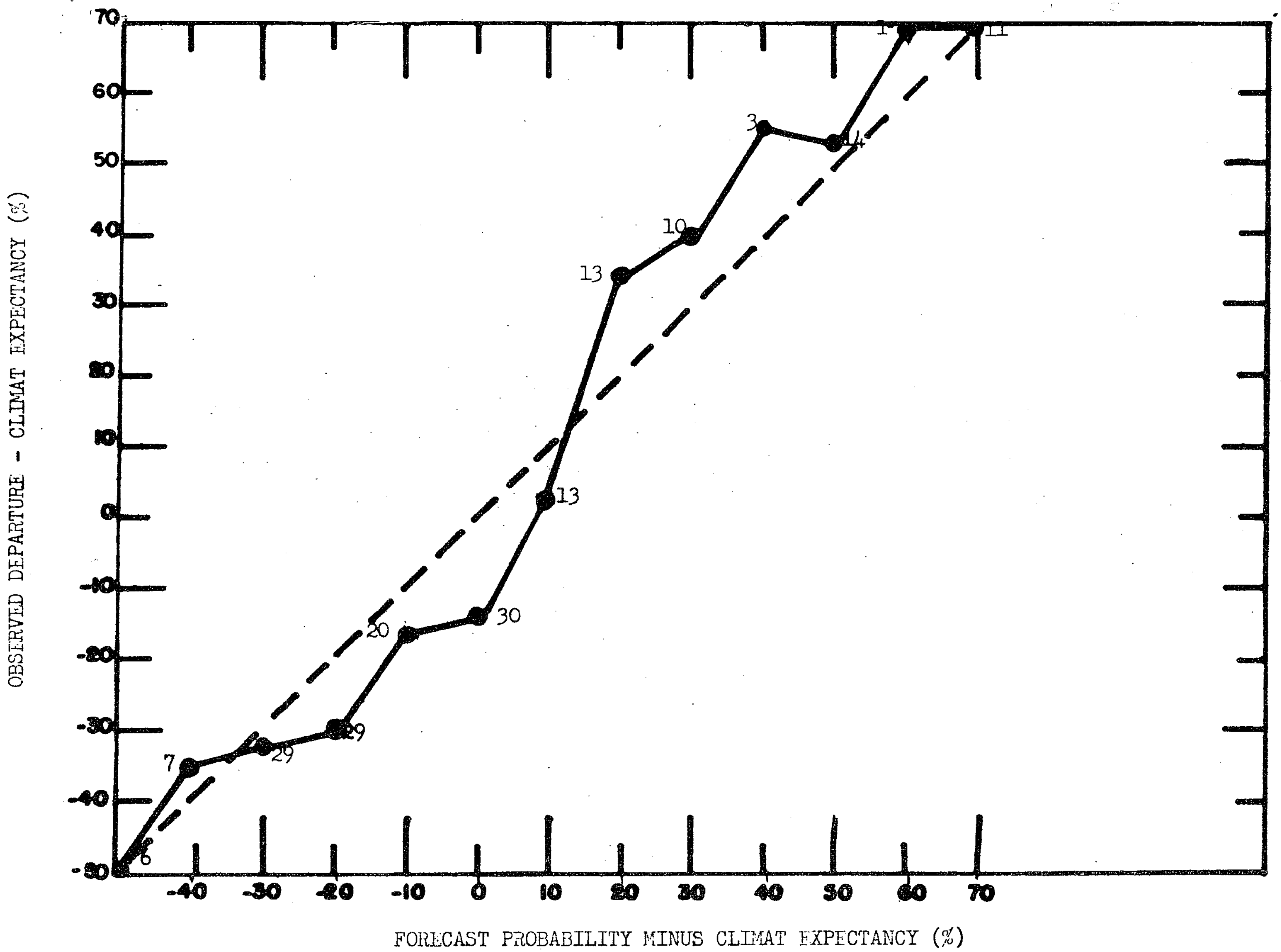

Figure 3. Departure of forecast probability from climatological expectancy vs. observed departure of relative frequency of occurrence for student forecasts in 1955-1956 seasons. Number of forecasts is given next to each data point. (After Sanders)

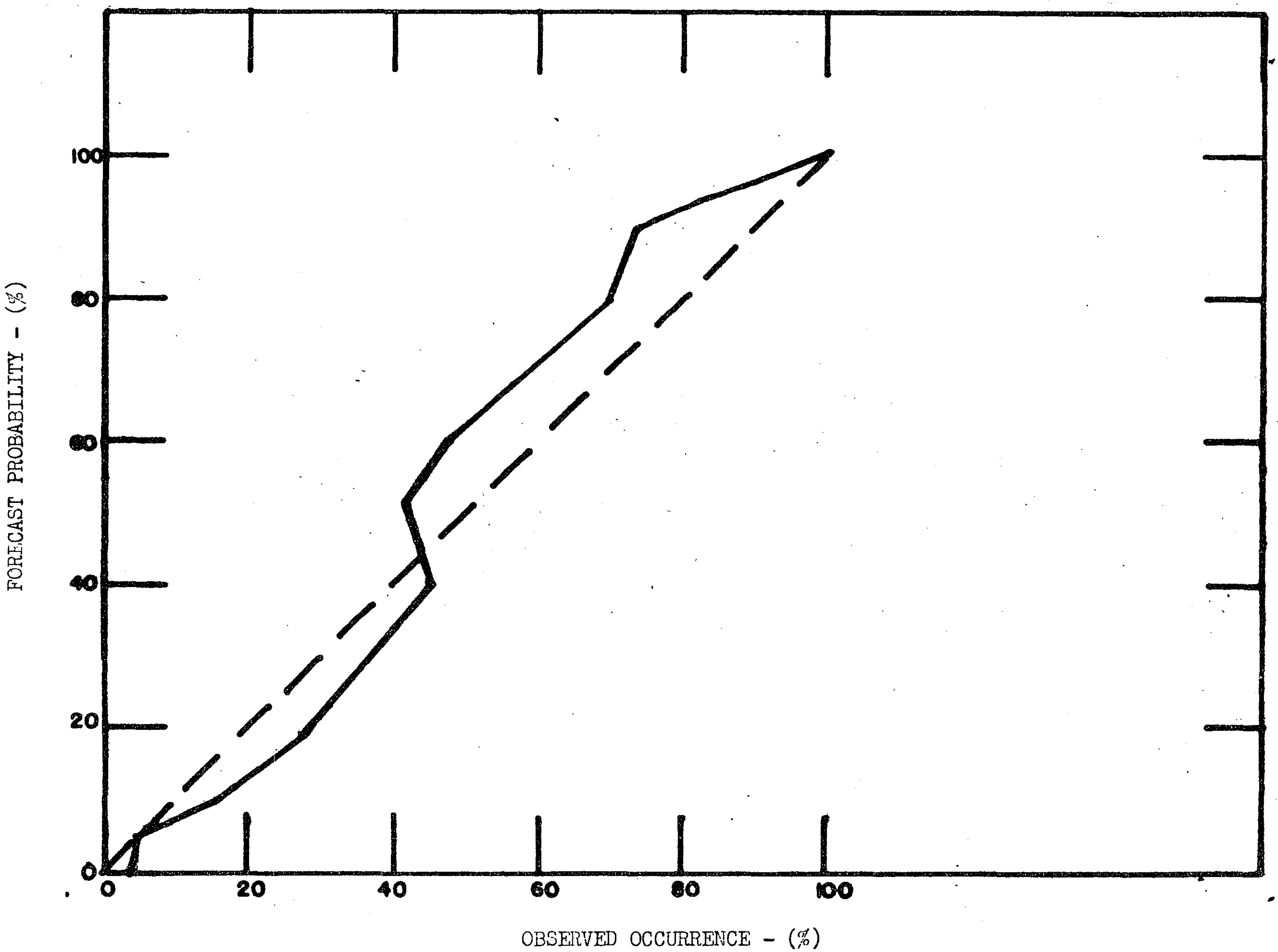

Figure 4. Comparison of the forecast probability with the observed per cent occurrence for the rain probability forecasts made at San Francisco, California during the months December through February of the winters 1956-57 through 1959-60. (After Root)

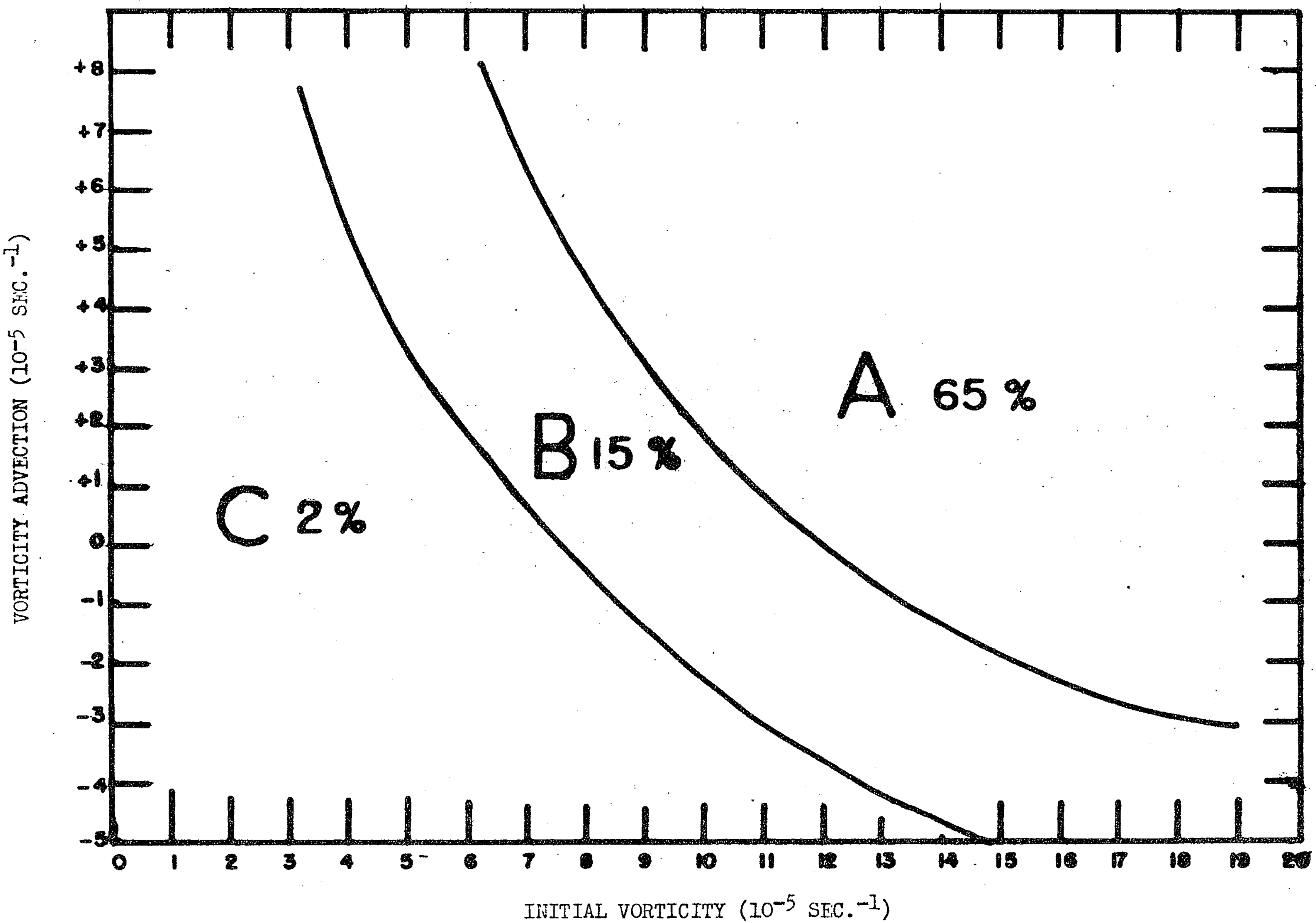

Figure 5. Relationship between initial vorticity, vorticity advection, and the occurrence of measurable precipitation at Salt Lake City October 1961 through April 1962. (After Williams)

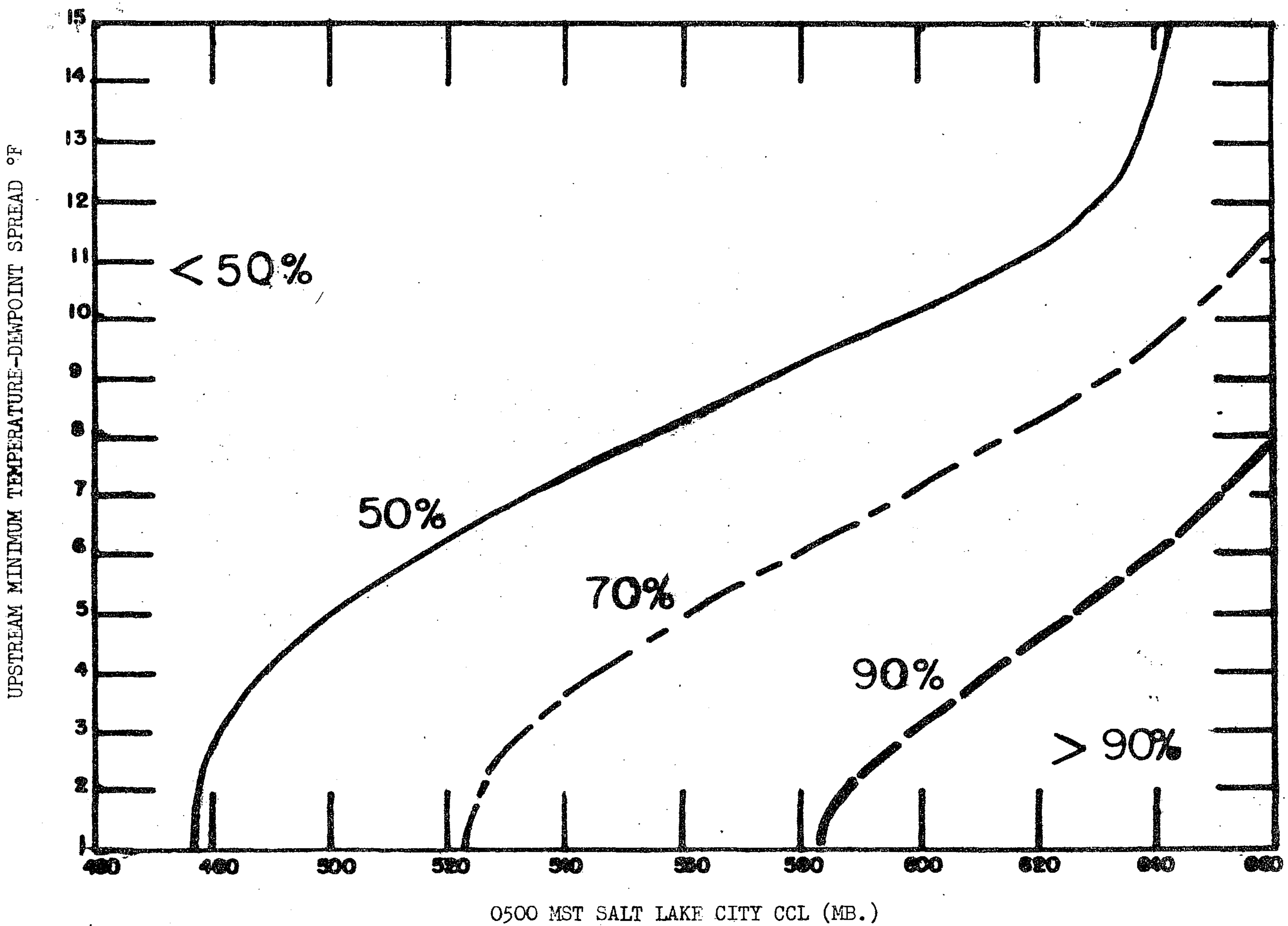

Figure 6. Relationship between minimum temperature-dewpoint spread in the 700-500 mb. layer and the CCL to the frequency of occurrence of showers at Salt Lake City. June 15-Sept. 15, 1957, 1958, 1959. (After Williams)