NOAA Technical Memorandum NWS SR-89

AN EXAMINATION OF PROBABILITY OF PRECIPITATION FORECASTS IN LIGHT OF RAINFALL AREAL COVERAGE

Daniel L. Smith, WSFO, Birmingham, Alabama

Scientific Services Division
Southern Region
Fort Worth, Texas
April 1977

UNITED STATES
DEPARTMENT OF COMMERCE
Juanita M. Kreps, Secretary

NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION
Robert M. White, Administrator

National Weather
Service
George P. Cressman, Director

AN EXAMINATION OF PROBABILITY OF PRECIPITATION FORECASTS IN LIGHT OF RAINFALL AREAL COVERAGE

Daniel L. Smith
National Weather Service Forecast Office
Birmingham, Alabama

ABSTRACT

Twelve-hour composites of weather radar echoes are used to derive estimates of rainfall coverage in forecast zones of southern Alabama and northwest Florida. Coverage estimates are then compared with probability of precipitation (PoP) forecasts to assess the skill and reliability shown by the forecasts. In addition, it is shown that areal coverage and PoP forecasts are closely related and a knowledge of the former, in real time, could lead to improved forecasts.

I. INTRODUCTION

Forecasts of probability of precipitation (PoP) are a routine part of National Weather Service (NWS) public forecasts. Unlike parameters such as temperature, however, PoP forecasts are not easily verified soon after they are made. Using rainfall observations from a single station in the forecast area, the usual procedure, it is only after a considerable body of forecasts has been assembled that verification is possible. Even then questions arise as to the representativeness of the observations. This study was undertaken in part to explore an alternate verification procedure; namely, PoP forecasts are compared with areal coverage of precipitation, as deduced from hourly radar observations. The study centers on summertime (June-August) rainfall in Southern Alabama and northwest Florida. In the "scattered shower" regime which frequently prevails in this area in summer our PoP forecasts are known to be deficient. Our aim is to develop a real time verification system which will lead to improved PoP forecasts.

Before continuing, a discussion of NWS PoP forecasts and the concept of forecast zones is in order. At least twice each day each NWS Forecast Office (WSFO) issues a forecast of the basic weather elements for three or four successive 12-hour periods (e.g., "today", "tonight", "tomorrow"). The forecasts apply to zones within each state. Zones usually consist of a few counties and are so designated that the weather is homogeneous within the zone. It is intended that the zone forecast be used as a local forecast for any point (community) in the zone. PoP forecasts <u>for any given point</u> in the zone, for each 12-hour period, are a part of the zone forecast. Since the weather is considered to be homogeneous, a point probability forecast is numerically equal to the <u>average</u> point probability forecast for the zone. This equality is significant, as we shall see later.

2. POINT PROBABILITY AND AREAL COVERAGE OF PRECIPITATION

PoP forecasts are routinely verified by comparison with precipitation observations at an official raingage within a zone. If the forecasts are reliable rain will be observed on three of ten occasions when 30% is forecast, five of ten occasions when 50% is forecast, and so on. While this verification system serves a variety of useful purposes for which it is well suited, it nevertheless has several limitations:

-- Only relatively few zones are verified...

-- Regardless of our assumption of homogeneity, the raingage may not be representative of the zone...

-- A single PoP forecast (with the exception of 0 or 100%) cannot be verified "right" or "wrong"....in fact, at least a season is required to accumulate a sufficient number of forecasts for a reasonable verification of reliability.

It is a painful experience to forecast an 80% chance of precipitation and watch it rain, "everywhere but at the official raingage!" Even though in concept this is what should occur on two out of ten 80% forecasts, this is only one argument frequently used by those who suggest that it would be better to verify PoP forecasts by using observations of areal coverage of precipitation within a zone. Occasionally, and sometimes without realizing the difference, forecasters indicate they would rather forecast areal coverage than point probability!. This is an interesting prospect but it is not without problems. The probability of precipitation at a given point in the forecast area (P) is related to the expected, or conditional, areal coverage (A) by the simple expression

$$P = C \times A$$

where C is the probability that precipitation will occur somewhere in the area (call it the areal probability). A is the expected areal coverage if it rains in the forecast area. It can be easily demonstrated that forecasts of expected areal coverage are inherently less useful than forecasts of point probability. What purpose is served, for example, by telling someone that, "if it rains today, the rain will cover 50% of the area." The listener's response will probably be, "O.K., so what's the chance it will rain (meaning either somewhere in the area or, more likely, 'on me')?" In this exchange the listener is given A but responds by wanting C (or P) as well. What he really wants is the probability that he will be rained on during the period...which is just what the NWS forecasts attempt to provide.

While this study will not argue in favor of forecasts of expected areal coverage we feel there is much to be learned from a comparison of point probability forecasts and coincident observations of areal coverage. After the fact, regardless of whether or not it rained in the forecast area during a given period, the chance that a given point in the area received rain is just the same as the areal coverage of precipitation. Winkler and Murphy (1976) show that the average point probability is the same as the unconditional expected areal coverage. Unconditional areal coverage is the product of conditional areal coverage and areal probability, or A x C. Since our point probability forecasts are in fact average point probabilities (we assume homogeneity) we are in effect forecasting unconditional areal coverage.

For a given forecast period, Nature reveals to the forecaster (after-the fact) the best, or more desirable, forecast he could have made of the probability of rain at any given point in the area...namely the areal coverage. Before-the-fact, however, there are definite limits to the forecaster's ability to realize this "best" PoP forecast. Consider the flip of a coin as an analogy to areal probability. On flipping the coin, "heads" or "tails" are always equally likely outcomes. The probability of heads is 50% and this can be considered the forecaster's "limiting probability". If he forecasts heads with any other probability he relies on sheer luck. After the flip, if heads occurs, we can argue the fact that a 100% chance of heads would have been the most desirable forecast -- regardless of the fact that the forecaster had no way of reliably making such a forecast! Fortunately (or perhaps unfortunately for the forecaster) the areal probability is not always 50%. The forecaster's job is to assign a probability as close to 0 or 100% as possible in any given case. After-the-fact he knows how successful he was.

If it fails to rain anywhere in the forecast area during the forecast period clearly the best possible point probability forecast would have been 0%. But this does not mean that a forecast of, say, 20% was necessarily "bad". Since the forecaster is limited in his ability to resolve rain/no rain cases he can only approach the 0% and 100% limits with his forecasts - sometimes with more success than at other times. In the case of the 20% forecast he was quite sure (80%) that it would not rain, but not certain (100%). By the same token, an 80% forecast in the no rain case would not have been necessarily "wrong". It shows an inability, in this single case, to resolve the situation (rain or no rain) but it still allows a 20% chance of the correct outcome. The 80% forecast is certainly not unreliable -- no single PoP forecast can be judged as to its reliability. PoP forecasts allow the forecaster the luxury of stating his assessment of the likelihood of a rain event. In the latter case the forecaster has said that in eight such cases as this, out of ten, it will rain.

3. SCOPE OF THE STUDY

Consider the following situation: a 40% PoP is issued for a particular zone on a given day. The next day, after analyzing the weather, the forecaster concludes that conditions are essentially unchanged from the previous day (not an unlikely event in the study area in summer). He wants to forecast the same likelihood of rain as existed the previous day, but was his 40% reasonable? How does he evaluate his PoP forecast from the previous day? Most zones contain no stations from which rainfall observations are immediately available. Even if the zone does contain a verifying station there may well be a tendency to misinterpret the significance of the observation. Desiring to optimize their Brier Score (see Brier, 1950, or numerous more recent contributions) some forecasters may simply hope for rain at a verifying station anytime their

PoP forecast exceeds 50% and hope for no rain anytime it is less than 50%. Ignoring all but the present forecast they give no heed to their overall reliability*. If it rains at a verifying station on a 40% forecast they may feel they "underforecast" and in the example above compensate by raising the PoP the next day.

If rain is equally likely at all points in the zone (homogeneity) then the chance of rain at any given point is the same as the after-the-fact areal coverage. Unfortunately, even if the zone contains a verifying station, for a given case the occurrence of rain at that station is not indicative of areal coverage. Beebe (1952) has shown that for an area roughly the size of a zone thirty to forty observations are required to accurately estimate the areal coverage! However, the total area covered by radar echoes during a given period provides a good estimate of the areal coverage. For this study we composited hand-drawn hourly radar overlays for the 12-hour "night" and "day" periods (0035-1135 GMT, 1235-2335 GMT) for June 1 to August 8, 1976. We confined the study to that part of WSFO Birmingham's forecast area which was within range of the Pensacola, Florida, WSR-57 radar. Fig. 1 shows the Birmingham forecast zones -- four in Alabama and all four in Florida were included in the study. We examined summertime rainfall because a definite need exists for more information about scattered showers and our ability to forecast them. This is the season when an immediate post-analysis of rainfall probability forecasts can most likely result in improved forecasts for the next forecast cycle.

On the hourly radar overlays all echo areas were carefully outlined. Two significant limitations should be kept in mind when assuming that 12-hour composites of these overlays represent total areal coverage of rainfall:

-- The composite will overestimate rainfall coverage because not all echoes represent rain at the ground. Some evaporation may occur and at long ranges the echoes appear somewhat larger than the true rain area.

-- The composite will underestimate areal coverage because only hourly overlays were used. Some echo areas between hours were probably missed and moving cells will result in systematically smaller composited areas.

Fortunately, these two effects tend to counteract each other.#

----

*For a forecaster with any skill at all this may not be a bad thing -- reliability may well take care of itself. But a not-so-skilled forecaster who may not have yet developed a feel for his own limitations must pay close attention to his reliability.

#Editor's Note: Almost certainly what we have here is a net underestimation of affected area.

First period PoP forecasts were tabulated for each of the eight zones along with the coincident areal coverages as estimated from the 12-hour composites. Echo coverage was determined to the nearest 10% with an estimated accuracy of ±10%. PoP forecasts were extracted from the zone forecasts. Because precipitation is not mentioned in the zone forecast unless the PoP is 20% or greater we could only form the large category "less than 20%" to include PoP forecasts of 0, 5, and 10%.

In many of the following analyses all Florida zones are grouped together and all Alabama zones are grouped together. The grouping is somewhat arbitrary but might reveal interesting effects of the forecasters' thought processes. While summertime forecasts are seldom greatly different for southern Alabama and the Florida Panhandle, forecasters might consciously or otherwise inject differences because the two areas are separated in the forecasts released to the public. Separate forecasts are seldom written for each zone: zones may be combined in various ways as the weather dictates but Florida zones are never combined with Alabama zones.

4. PoP FORECASTS AND AREAL COVERAGES: COMPARISON OF SUMMERTIME AVERAGES

We begin our examination of the data by looking at averages without comparing specific PoP forecasts and areal coverages. Recall that forecasts grouped as "less than 20%" contain unknown numbers of 0's, 5's and 10's. For purposes of averaging we took a value of 5% for these forecasts; the true value, however, was most likely closer to zero since 0% forecasts were more numerous than 10% forecasts. Estimated coverages of less than 10% represent cases where there were echoes in a zone but coverage was less than 10%.

Table I shows averages for the four Florida zones, the four Alabama zones and all eight zones combined. Also shown in the table are estimated average rainfall frequencies from easily available raingages. Only four of the eight zones contained such gages: these zones and stations are AL15 (Dothan), AL16 (Mobile), FL02 (Pensacola) and FL04 (Apalachicola). Rainfall frequencies shown are within a few percent of the climatological frequencies for these stations indicating that the study period was a near normal season. Also contained in the table are averages for each group of zones of the frequency of occurrence of radar echoes somewhere in the zone. To the extent that the radar composites depict rainfall during the 12-hour periods, these frequencies are mean values for the areal probability (C).

Table I contains much useful information and the data are consistent, as we shall see. We might guess that it rains somewhere in the Florida Panhandle just about every day during the summer ("scattered showers"). The table shows that, averaged over all four Florida zones, echoes occurred somewhere in a zone on about 75% of the days and about 50% of the nights. Averages for the four Alabama zones are similar. These values are consistent with those of Beebe (1952) who showed that for the combined day and night period rain occurred somewhere within a zone-sized area centered on Birmingham about 80% of the time. In the summer there is little climatic difference between Birmingham and southern Alabama. Beebe's 24-hour value must be at least as large as our largest 12-hour value (~75%). Rain occurred not just somewhere but over at least 20% of a given zone on about 30% of the nighttime periods and about 55% of the daytime periods (and again Florida and Alabama show similar averages). The significance of 20% coverage will be emphasized later.

The frequency of occurrence of echoes (rain) somewhere in the zone... average areal probability...conveys no information about areal coverage. Considering all days, even days when no rain was observed anywhere in the zone, Table I shows that the average areal echo coverages for Florida and Alabama were about 30% during the day and 15% at night. Since the observed areal coverage is the same as the average point probability -- which is just what we attempt to forecast -- we can reasonably expect the average of our forecast probabilities to be near the average areal coverages just described. The table shows that they were indeed close for the study period. Averaging 0% PoP forecasts as 5% might have caused the nighttime forecast averages to be slightly too high in comparison with the areal coverages (20% vs ~15%). During the day the bias is less pronounced because there were fewer "less than 20%" forecasts. It will be seen later, however, that there was a tendency to forecast PoPs slightly too high, particularly at the lower PoP values.

Since we assume the zones to be meteorologically homogeneous, for a given instance the probability of rain at any point in the zone is the same as the average point probability (that is, the average of the probabilities at all points). It is also equal to the areal coverage of precipitation. If we average over all days in the study period (including no-rain days) rainfall frequency at any point should be the same as the average point probability (or average forecast PoP) and it should also be the same as the average areal coverage. Table I shows that the average rainfall frequencies derived from the observations of the four available stations are indeed close to the values of the other two parameters. The fact that the rainfall frequency for the Florida zones appears low in comparison might indicate that the two stations used, both on the coast, are not representative of the zones. This is in fact the kind of feature upon which we hope this study might shed some light.

Recall the earlier equation, P = C x A. A is the expected areal coverage, conditional on the fact that it rains somewhere in the area. The radar data show that for the study area it rains somewhere in each zone on about three-fourths of the daytime periods. It is not unreasonable then to imagine a given day as being one on which the forecaster is quite sure it will rain in the zone, somewhere. What information can our analysis of areal coverage yield to the forecaster for use on such a day? If only the rain days are averaged, that is the days when there were radar echoes somewhere in the zones, Table I shows that the areal coverage is about 10% higher than the average coverage on all days. In other words, on such days the chance of rain at a given point is about 10% greater than the climatological frequency. Finally, consider for the Florida or Alabama zone groups the product of the average frequency of occurrence of rain somewhere in the zone (C) and the average areal coverage for rain days only (A). The product should be; 1) the average point probability (P), and, as we have seen, 2) the average areal coverage on all days, and 3) the observed average rainfall frequency. The reader may satisfy himself that the values in Table I yield products which are precisely 2) and very close to 1) and 3).

5. PoP FORECASTS AND AREAL COVERAGES: FREQUENCIES OF USE COMPARED WITH FREQUENCIES OF OBSERVATION

Having considered summer-long averages of the data, we now examine more closely the use and occurrence of individual PoP and coverage values. Fig. 2 shows the frequency of use of each PoP forecast value along with the frequency of occurrence of each decile areal coverage value. Alabama and Florida zones are considered separately and different graphs are shown for the day and night periods. Small numbers along the curves show the actual numbers of forecasts and occurrences. Note, however, that no attempt is made to compare PoP forecasts with coincident coverage values. An outstanding feature in comparing the curves in Fig. 2 is that PoP values of 20% and 30% were forecast far more frequently than corresponding coverages were actually observed. This was true for both groups of zones and at night as well as during the day. During the daytime period, the "overuse" of certain PoP values extended to 40% and 50%.

The graphs do not reveal whether this overuse represents an over- or under-forecasting bias, but the particular abundance of 20% values is of special interest. Several factors are probable contributors to the high frequency:

-- Forecasters may overestimate the areal extent of summer showers. Observation of a few showers on radar during the period, or rainfall at an observing station, may immediately condition many forecasters to think in terms of at least 20% coverage.

-- 20% is a significant value because it is the threshold for inclusion of the word "rain" in the zone forecast. Consequently, if the forecaster estimates the PoP to be greater than 0%, but less than 20%, he may "inflate" his forecast to 20% so as to be able to include some mention of rain in the forecast, "just in case". The feeling seems to be that people do not expect to be rained on with a 20% probability forecast anyway so 20% is about the same as 0%...unless someone does get rained on! (The feeling probably also extends to the idea that if it fails to rain at a verifying station on a 20% forecast the scoring penalty is not much greater than if the forecast had been 0%.)

-- Of the PoP values the forecaster is allowed to use (0, 5, 10, 20, 30, ...90, >95) most are above the summer climatological rainfall frequency for the study area. In fact, the 20% threshold leaves him no "mentionable" values below the nighttime climatological value (15-20%) and only the 20% and perhaps 30% values below the daytime climatological frequency (~30%). Hughes (1965) has indicated that having far more values on one side of climatology than on the other may create a psychological problem for the forecaster which results in significant over- or underforecasting.

The peaks at 30% in the daytime forecast curves, particularly outstanding in the Florida zones, are especially interesting. As the curves showing observed coverage indicate, 30% rainfall coverage is not significantly more frequent than coverage of 20% or 40%. Yet the 30% value seems to be a favorite of the forecasters!. No doubt the explanation lies in the fact that 30% is about the daytime climatological rainfall frequency for the area (Table I). Fig. 2 indicates the strong, although not unusual, tendency to concentrate forecasts around the climatological frequency: a problem to which special attention should be given, particularly in the Florida zones.

For the daytime period the tendency toward overuse of certain PoPs extended to 40% and 50% but diminished in frequency so that 60% PoPs were forecast with the same frequency as observations of 60% coverage. Forecasts of PoP values above 60% were made with lower frequencies than corresponding areal coverages. Areal coverage greater than 60% occurred on about 20% of the daytime periods yet PoPs greater than 60% were forecast less than 5% of the time. Why the forecast and observed frequency curves cross at 60% is not clear. The crossover occurs at a lower value for the nighttime curves, at around 40% to 50%, suggesting that this is a reflection of climatology.

The observed frequency curves in Fig. 2 indicate that areal coverages from 40% to 100% occurred with about equal frequencies, although the total number of such occurrences in the small data sample makes this conclusion tentative, at best. Each coverage value occurred with a frequency of about 5% for the daytime periods and somewhat less frequently at night. Nighttime PoP forecast frequencies from 40% to 70% match frequencies of corresponding areal coverages quite well, but above 70% forecast frequencies fell to zero...no such PoPs were forecast! More significantly, forecasters apparently failed to recognize, at least for the study period, that daytime areal coverage of 90% or even 100% was about as likely as coverage of only 40%. They forecast a PoP of 40% about 15% of the time but never forecast PoPs of 90% or 100%!

In general, then Fig. 2 shows that frequencies of forecasts for PoPs below about 50% were higher than the frequencies of occurrence of corresponding areal coverages. In the middle range, 50% to 70%, forecasts and observations occurred with closely matching frequencies. At higher values, above about 70%, PoP forecasts were seldom made but areal coverages as high as 100% were not uncommon. Likely reasons for the tendency to overuse the lowest PoP values (around 20%) have already been given and it is not difficult to envision procedures which should correct this bias and bring the forecast PoP and observed areal coverage curves into closer agreement. But what about the underuse of PoPs above about 70%? Two interesting concepts may shed some light on this bias. First, consider the earlier equation, $P = C \times A$. A forecaster can arrive at a PoP value of 70% or more if he is virtually sure that it will rain in the zone ($C \cong 100\%$) and he expects the areal coverage to be 70% or more. Alternatively, he can be more than 70% sure that it will rain at least somewhere in the zone and expect areal coverage of nearly 100% if it does. For a PoP of 70% neither C nor A can be below 70%. In other words, a high confidence of rain in the area and an expectation of large areal coverage.

Whether or not they customarily consider the problem in just these terms, perhaps it is not surprising that the forecasters tend to underuse high probabilities. Even if they are quite sure of large areal coverage there is still the chance that the rain will not materialize in the period so they may use areal probability to "hedge" the forecast...assigning an areal probability of something less than 100%. This idea will be explored further below.

A second concept which might explain underuse of high probabilities involves reliability. Hughes (1965) has pointed out that in perfect forecasting one would have only 0% and 100% forecasts, with several times more 0's than 100's, the ratio depending on the climatological rainfall frequency. Fig. 3 is adapted from Hughes' study and shows the frequencies at which various PoP values were forecast for Chicago. The study period is not important. The broken line represents the expected frequency distribution based on the assumption of nearly perfect forecasts. The forecasts are clustered at 0% and 100% with a scattering of intermediate "imperfect" PoP forecasts. The Chicago curve is similar to those in Fig. 2. But is it reasonable to expect our PoP forecasts to resemble the upper end of the broken curve in Fig. 3? It is true that for a specific point a set of perfect forecasts would consist of PoP = 100% on each rain day and PoP = 0% on each non-rain day. Such a set is at least theoretically possible. After-the-fact, for a given period, there is no intermediate outcome at the single point: it either rained or it did not. However, our PoP forecasts are not for specific points but rather are for any given point within a zone. They are average point probabilities and for any given period there is some chance, ranging from 0% to 100%, that after-the-fact any particular point had rain. The chance depends on the areal coverage in the zone. In other words, perfect resolution, in the sense of 0% and 100% PoPs, is not possible for zone forecasts. For the daytime periods included in this study, considering only the Alabama zones, the best possible resolution would have been obtained by forecasting a 0% PoP on each of the 47% of the periods when it failed to rain anywhere in the zones and forecasting a 100% PoP on each of the 4% of the periods when it rained everywhere in the zones. A forecast of 0% or 100% on any other of the 49% of the periods would have been wrong somewhere in the zones! Thus it seems improper, for zone forecasts, to expect resolution to approach the broken curve in Fig. 3. Rather, a perfect resolution should be thought of as approaching the curve showing frequency of occurrence of areal coverage. Such a curve will always show a peak at 0% coverage (except in wet climates) but will not show the secondary peak at 100% (unless extensive rains are common).

6. PoP FORECASTS AND AREAL COVERAGES: COMPARISON OF COINCIDENT DATA

So far we have considered only the individual average characteristics of PoP forecasts and observed areal coverages. We have examined over- and underuse of forecast PoP values but we have gained little insight into possible over- or underforecasting bias because no attempt has been made to compare forecasts with coincident observations of coverage. We now turn our attention to this aspect. Fig. 4 shows the average forecast PoP for various observed values of areal coverage (broken lines). Periods were separated according to areal coverage and the forecast PoPs for each decile coverage were averaged. The figure also shows average observed areal coverage for various forecast PoP values (solid lines). In the latter case periods were separated according to PoPs and the areal coverages were averaged. Different graphs are shown for day and night periods and Alabama and Florida zones are considered separately.

Consider first the broken curves in Fig. 4. In a way, these curves graphically depict prefigurance. They show how well given extents of areal coverage were forecast. Thus, they give an indication of the forecasters' resolution -- in other words, they show how well the forecasters were able to resolve areal coverage. Probably the most outstanding feature of all broken curves is that even for large coverages the average forecast PoPs did not exceed about 50%. Murphy (1977), using relatively high density raingage networks to deduce areal coverage, has shown that PoP forecasts from St. Louis, Missouri, and Rapid City, South Dakota, reveal the same characteristics as shown by the Birmingham data in the broken lines in Fig. 4.

There is a tendency to conclude that the broken curves graphically depict a serious underforecast bias for all PoP forecasts above roughly the climatological rainfall frequency -- for both night- and daytime periods. Below climatology, overforecasting appears to be the problem. Fig. 5 is a closer look at average forecast PoPs for high and low coverage daytime periods. For the combined Alabama and Florida zones 100% coverage was observed on thirty-two of the daytime periods. Fig. 5 shows the frequency distribution of PoP forecasts for those days. The average forecast PoP was 53% and PoPs of about 50% were the mode as well as the average for the data set. While there was a secondary maximum with about a fourth of the forecasts at 70% to 80%, two-thirds of the forecasts were from 40% to 60%. A second curve in Fig. 5 shows the frequency distribution of PoP forecasts for the sixty combined Alabama and Florida daytime cases when the observed areal coverage was 10%. The average forecast PoP was 26% and only one-fifth of the forecasts were below 20% (that is, 10%, 5% or 0%). Fig. 5 even more clearly seems to indicate a forecast bias.

Do the above results indicate serious problems? Are they in fact surprising? To characterize apparent problems revealed by Fig. 4 as "over-" or "under-forecast" bias is really a misuse of terms because such bias is usually a characterization of reliability. As we shall show below, the forecasters' reliabilities were actually quite good during the study period! Overforecasting is said to exist if, say, for a set of 40% PoP forecasts the average observed areal coverage is anything less than 40%. Regardless of an underforecast bias which we might infer from the broken curves of Fig. 4, the solid curves in the same figure, discussed below, show that if any bias existed at all it was an overforecast bias! How then do we explain the slope of the broken curves? Recalling the disproportionate numbers of forecast PoPs and areal coverages -- at either extreme -- which were revealed by Fig. 2, the slopes of the curves should not be particularly surprising. We have seen that even if forecasters are highly skilled at assessing the expected areal coverage they should still forecast PoPs somewhat lower than the observed areal coverage because, practically speaking, they cannot be 100% sure that it will rain in the zone (that is, $C<100\%$). If the forecast PoP exceeds the areal coverage it can only mean that the forecaster overestimated the expected coverage (since $C\leq 100\%$).

However, if the forecast PoP is less than the observed coverage, which is what we see in the broken curves of Fig. 4 (above climatology), it signifies that the forecaster underestimated either the expected areal coverage or the areal probability, or both. The ability to resolve the areal probability as 0% or 100% and also resolve the expected areal coverage are both measures of the forecasters' skill. How closely the broken curves in the figure approximate the diagonal lines is thus a measure of skill. We prefer this interpretation of the curves over a characterization of bias. Since the broken curves deviate progressively farther from the diagonal lines as observed coverage increases, it is obvious that forecasters exhibit poor skill in the higher probability ranges. In general, the same conclusion is usually reached by other studies of forecasters' skill. The lack of skill is manifest either as underconfidence of areal probability (forecasting C too low) or underestimation of expected areal coverage (forecasting A too low). Most likely both effects frequently combine to result in a PoP forecast which is too low in "rain" cases (cases in which the probability of rain is above the climatological frequency). Below climatology PoP forecasts were generally too high. As we showed, since one cannot be more than 100% sure that it will rain in the zone ($C \nless 100\%$) this result can only come about from overestimating the expected areal coverage.

Lack of resolution of the forecasts in the middle PoP range is another interesting feature revealed by the broken curves in Fig. 4. Notice in the daytime figures that the average forecast probability remains between 30% and 40% as the areal coverage increases from 20% to 70%! This feature is not quite as pronounced at night. Since we are dealing with PoP values as forecasts for any given point and are not considering whether or not it rains at a particular point, resolution in its usual sense -- the ability to forecast only 0% and 100% (with success) -- seems to have no meaning in our context. This idea was discussed earlier in Section 5. Nevertheless, the broken curves can be thought of as depicting a kind of resolution insofar as they show the forecasters' ability to resolve areal coverage. When the coverage is high the average point probability is just as high and good "resolution" demands that the forecaster recognize such a situation with a high PoP. Like either end of the broken curves, the middle seems to indicate that forecasters lack good resolution (skill) in the summertime, particularly during the daytime period.

We now turn our attention to the other half of Fig. 4, namely the information contained in the solid curves. These curves can be thought of as depicting something like post agreement. They show how well given forecast PoP values verified against observed areal coverage. The degree to which the plotted points lie along the diagonal lines is also a measure of reliability. Even though we plot forecast PoP averages against areal coverage these graphs are actually the same as those normally used which show frequency of occurrence of precipitation at a verifying station plotted against average PoPs (for example, Cummings, 1971, 1974). This is because frequency of precipitation at a given point and average areal coverage are the same as long as the area is homogeneous in terms of rainfall distribution.

On the average, the PoP forecasts, particularly for the daytime period, were fairly reliable. For the combined zones eight 80% forecasts were made for the daytime period and the average areal coverage was 83%! However, when the zones are separated, as in Fig. 4, we find that the areal coverage was 100% on each of the four Florida forecasts of 80% PoP. On the four Alabama 80% forecasts the observed coverages were 20%, 70%, 80% and 90%. With the exception of the lowest Alabama observation this result is good, considering the small data sample. Earlier we commented on the desirability of grouping the Alabama and Florida zones separately. It would be interesting to know whether the differences just cited in areal coverage were real or the result of the small data sample.

We know from earlier discussion that forecasters made far fewer forecasts of PoP greater than 50% than were called for by the areal coverage which was subsequently observed. The low number of forecasts makes it difficult to assess the reliability above 50%. Fig. 4 suggests, however, that when forecasts of greater than 50% were made for the Alabama zones they tended to overforecast the actual coverage (the solid curve falls below the diagonal). This seems particularly true for the nighttime forecasts. For the Florida zones in the daytime there was a tendency to underforecast the coverage while at night the 70% forecasts (there were none above 70%) greatly underforecast the coverage. It should be pointed out that these results are not inconsistent with those we deduced from the broken curves in the same figure. Because of skill limitations forecasters used PoP values above 50% with far less frequency than high areal coverage was actually observed. Thus, the slope of the broken curves is inevitable. When forecasts of PoP greater than 50% (or any other value for that matter) were issued various degrees of over- and underforecast bias were apparent in the result.

7. PoP FORECASTS AND AREAL COVERAGES: VARIATION AMONG ZONES

In previous sections the four zones in each state were combined in order to increase the sample size for averaging and in hopes of revealing differences in both forecasts and areal coverages in the two areas. We now examine the zones separately to see if differences exist at the scale of zones. The analysis presented in Table 2 follows that of Murphy and Winkler (1977). They showed that forecasters at Rapid City, South Dakota, exhibited skill at distinguishing different point probabilities for points within a zone-size area which was not homogeneous with respect to rainfall coverage. We do not consider points within a zone, but rather zones within a larger portion of a state. Table 2 shows the frequency of occurrence of different areal coverages in the four Alabama and four Florida zones. Also shown are the frequencies at which forecasters used one, two, three or four different PoP values for the four zones. In distinguishing different areal coverages we required the values to differ by 20% or more to account for limitations in determining coverage from the radar composites.

It can be seen that at night areal coverage was the same (one value) about half of the time...both in Florida and Alabama. Forecasters did reasonably well in forecasting this; they overused a single PoP value about 10 to 15% of the time. Of course, the information in Table 2 does not indicate the forecasters' success at assigning the correct PoP (coverage) value! It is significant that three or four different areal coverages occurred 24% (10%) of the time in Alabama (Florida) for the night periods. Except for a single use (approximately one percent) of three values in Alabama, forecasters never used more than two PoP values for either group of zones.

For the daytime periods forecasters failed to realize with their PoP values the variation of areal coverage shown by the radar data. Note that never more than two PoP values were used for a given period. Nature, on the other hand, assigned three or four distinct areal coverages 23% of the time in Alabama and 34% of the time in Florida. The same PoP value was assigned to all four zones 60% of the time in Alabama and 84% of the time in Florida. Such uniformity was realized in areal coverage only 33% and 27% of the time in Alabama and Florida, respectively! These data would seem to indicate that forecasters should try much harder to identify features (meteorological as well as topographical) which might lead to the use of different PoP values. When more than one value was used in Alabama, zone AL16 was usually the "oddball" even though on the average its PoP was the same as the other Alabama zones (~20% nighttime, ~30% daytime). Interestingly, the radar data seem to indicate that AL16 had greater and more variable coverage than the other zones (~25% nighttime, ~40% daytime)! In Florida, in both the night- and daytime periods, forecasters frequently used a single PoP value for the inland zones (FL01 and FL03) and a second value for the coastal zones (FL02 and FL04). Such differentiation accounts for most of the two-value forecasts (24% of the night- and 16% of the daytime forecasts). For most of the daytime periods the Florida zones had two or more areal coverages but there was no apparent tendency for the inland zones to show one coverage value and the coastal zones to show another.

During the period of the study, variation of areal coverage of rainfall from zone to zone was common. This was particularly true during the daytime. It is hoped that a further examination of the radar composites will lead to a better understanding of the causes of these variations.

8. SUMMARY

This study has presented first results of an investigation of Alabama and northwest Florida rainfall and precipitation forecasting. Our investigation uses weather radar data to infer areal coverage of summertime showers. Results of this preliminary study are sufficiently encouraging for us to extend the study area to northern Alabama. The WSR-57 radar at Centreville, Alabama, near Birmingham, will be used for that area. Our aim is to improve precipitation forecasting, particularly PoP forecasts in the summer, at the Birmingham WSFO. Results, of course, should be generally applicable to forecasters elsewhere. In assessing results presented in this study we must not lose sight of the limited data sample which was used. Our summaries and statistics involving areal coverage and PoP forecasts are similar to those derived by others from different data and techniques, thus lending support to our conclusions.

In Section 3 we presented the situation of a forecaster who is faced with what he thinks is the same likelihood of rain as existed on the previous day. How can he assess his PoP forecast from the previous day? Subsequent discussion showed the utility of areal coverage observations in making this assessment. Zone to zone variation of areal coverage of rainfall was found to be common during the study period suggesting that the forecaster should make every effort to differentiate between zones. In the long run, climatology of areal coverage in the zones can help the forecaster improve his overall skill and reliability. On a real time basis, however, improvement can come from an immediate verification of his expected areal coverage. In the frequent summertime situation when the forecaster determines a high areal probability (C) he has only to examine the 12-hour radar composite to assess the accuracy of his areal coverage forecast ($PoP \cong A$ if $C \cong 100\%$). The radar composite is easily available at those forecast offices which are collocated with radar stations. Unfortunately, such is not the case at Birmingham and a way has yet to be found to make the composite available. Several approaches are under investigation.

ACKNOWLEDGEMENTS

The author extends thanks to the staff of the Pensacola Weather Service Office for providing an excellent set of radar data. Forecasters at the Birmingham WSFO engaged in many helpful discussions. Kevin Henderson, a student assistant, helped to reduce the data and discovered the hard part of meteorology! Finally, the comments of Allan Murphy, NCAR, Boulder, Colorado, were especially encouraging and helpful.

References

Beebe, Robert G., 1952: The distribution of summer showers over a small area. Mon. Wea. Rev., 80, 95-98.

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

Cummings, Allen D., 1971: Some aspects of probability of precipitation forecast performance by WSFOs in the Southern Region -- November 1969 through October 1970. NOAA Technical Memorandum NWS SR-57, 35 pp.

Cummings, Allen D., 1974: Some aspects of PoP forecast performance by Southern Region WSFOs -- revisited. NOAA Technical Memorandum NWS SR-77, 18 pp.

Hughes, Lawrence A., 1965: On the probability forecasting of the occurrence of precipitation. Washington, D.C., U. S. Dept. of Commerce, ESSA, Weather Bureau, Technical Note 20-CR-3. 36 pp.

Murphy, Allan H. and Robert L. Winkler, 1977: Experimental point and area precipitation probability forecasts for a forecast area with significant local effects. Atmosphere, 15, 54-71.

Murphy, Allan H., 1977: Personal communication.

Winkler, Robert L., and Allan H. Murphy, 1976: Point and area precipitation probability forecasts: some experimental results. Mon. Wea. Rev., 104, 86-95.

Table 1. Averages derived from forecast probabilities and observed radar echo coverage during the study period (June 1 to August 8, 1976).

	NIGHT			DAY		
	Fla	Ala	Comb	Fla	Ala	Comb
Average forecast PoP	20	20	20	32	29	30
Average areal echo covg (all days)	16	17	16	35	30	32
Average areal echo covg (rain days only*)	31	29	29	46	40	42
Average rainfall frequency	13	16	15	25	31	28
Average freq of occurrence of echo somewhere in zone	52	59	56	77	76	76

*Rain days are days (periods) when an echo occurred somewhere in a zone, regardless of coverage.

Table 2. Frequencies of use (observation) of 1-, 2-, 3- or 4 separate PoP values (areal coverages) for the four Florida and four Alabama zones used in the study.

		Frequency of Use/Observation (%)				
Night (00-12 GMT)		1 Value	2 Values	3 Values	4 Values	
AL zones	Fcst PoP	59	40	1		
	Obs Covg	47	29	18	6	68 periods
FL zones	Fcst PoP	76	24			
	Obs Covg	59	31	7	3	
Day (12-00GMT)						
AL zones	Fcst PoP	60	40			
	Obs Covg	33	33	19	4	67 periods
FL zones	Fcst PoP	84	16			
	Obs Covg	27	39	33	1	

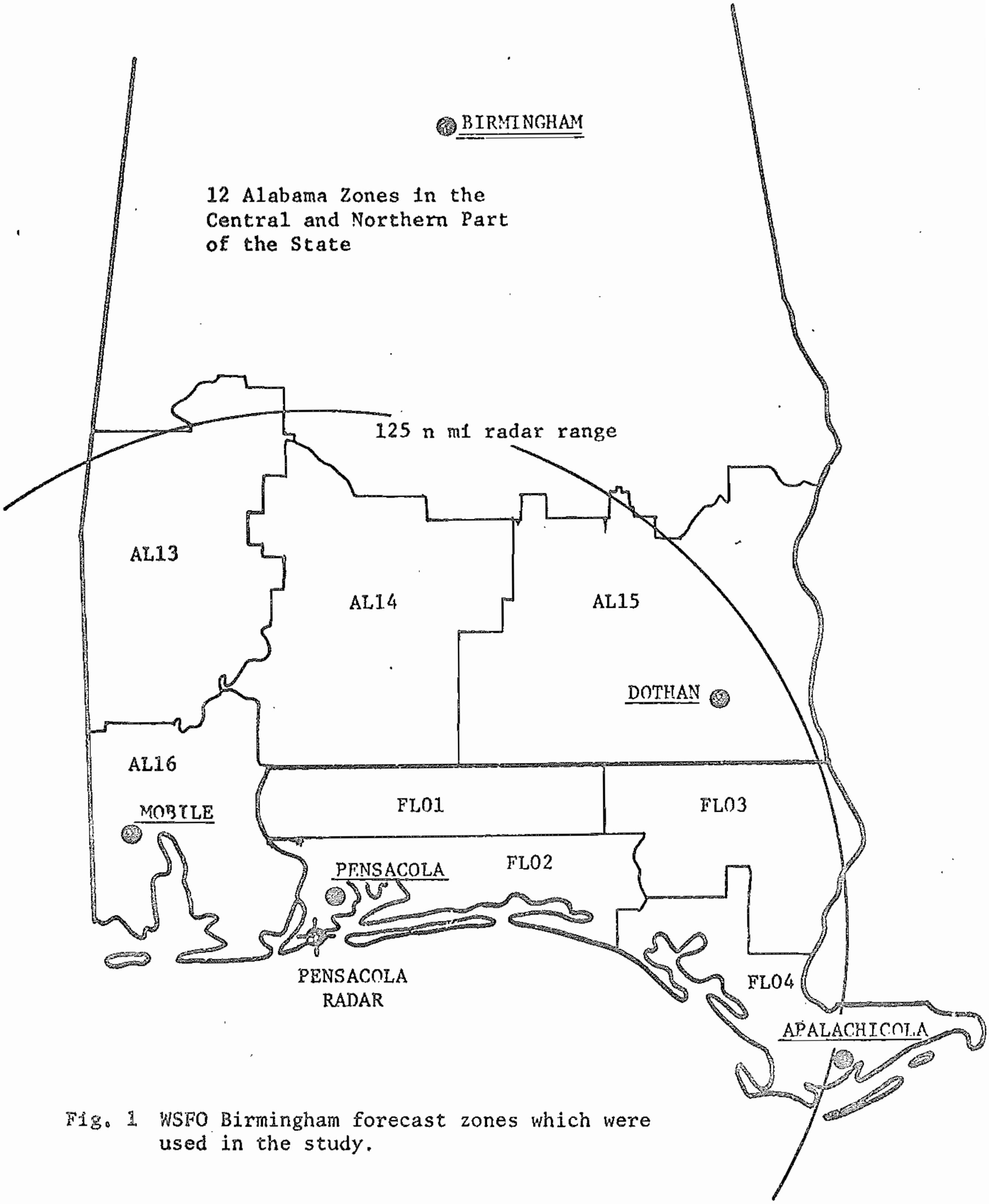

Fig. 1 WSFO Birmingham forecast zones which were used in the study.

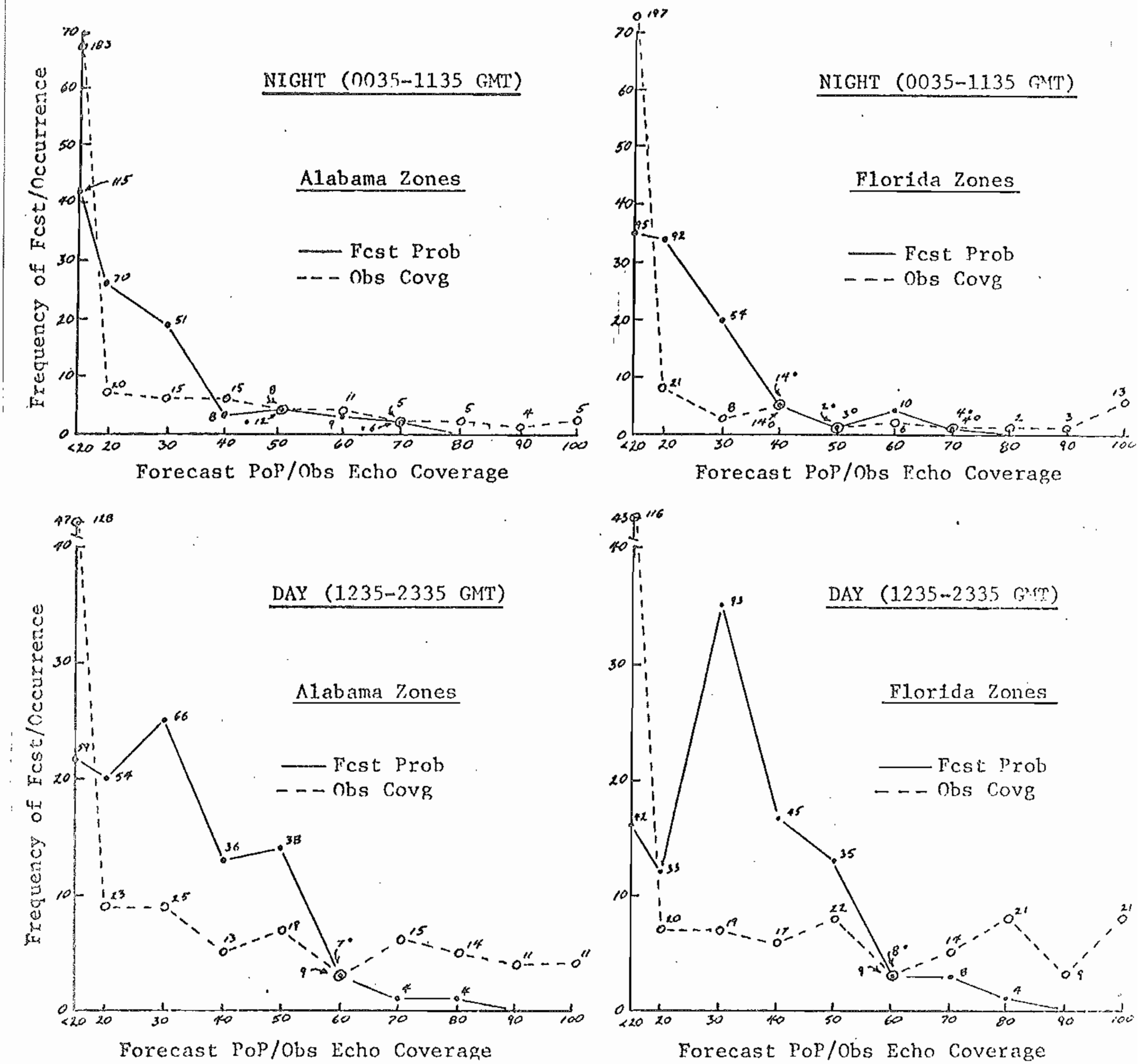

Fig. 2 Frequency of occurrence of various values of areal coverage (broken lines) and frequency of forecast of various values of the probability of precipitation (solid lines). Study period is June 1 to August 8, 1976. Small numbers indicate number of occurrences.

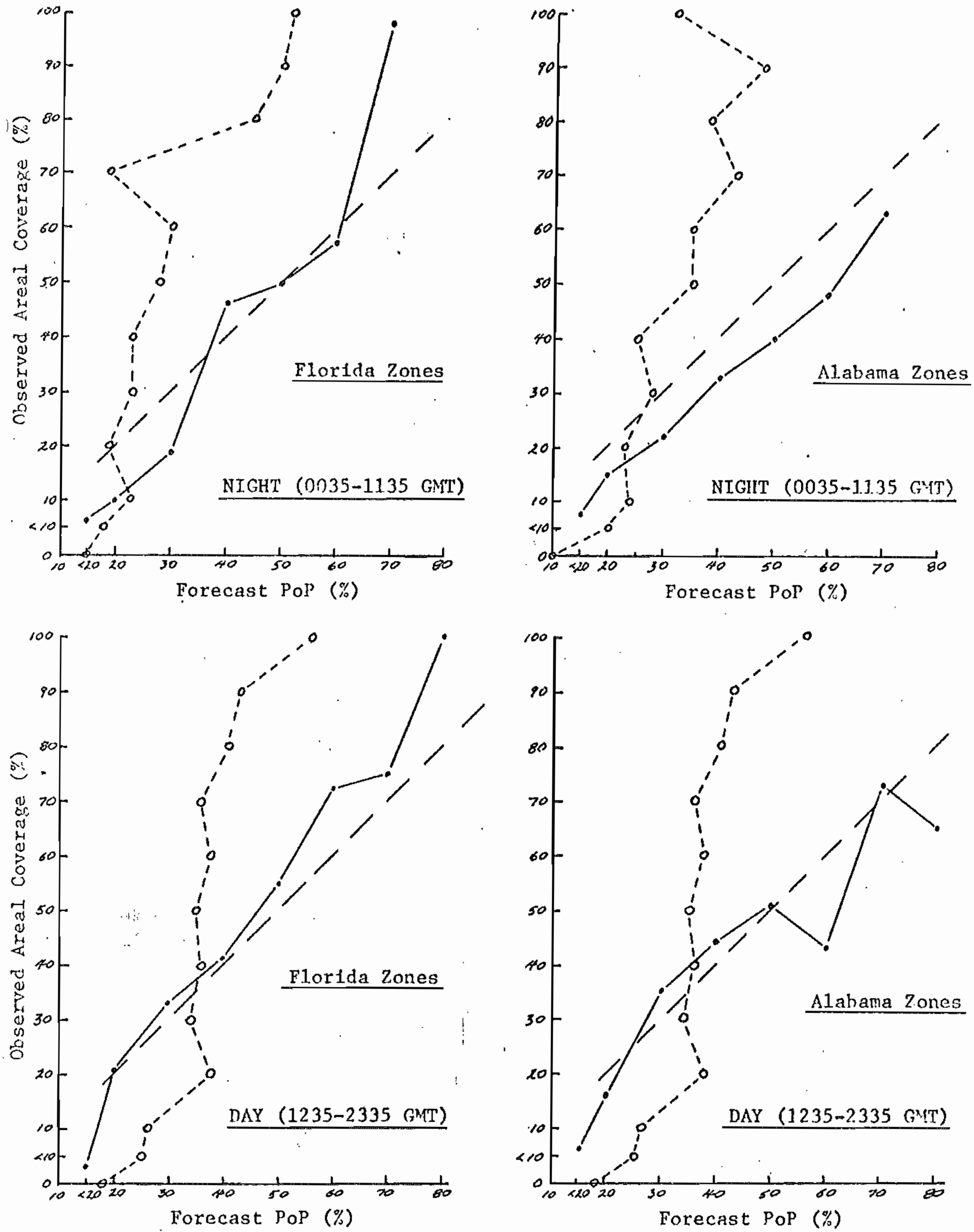

Fig. 4 Average forecast PoP for various values of observed areal coverage (broken lines) and average observed coverage for various values of of forecast PoP (solid lines). "<10" on the ordinate has meaning only for the broken curves while "<20" on the abscissa should be read only for the solid curves.

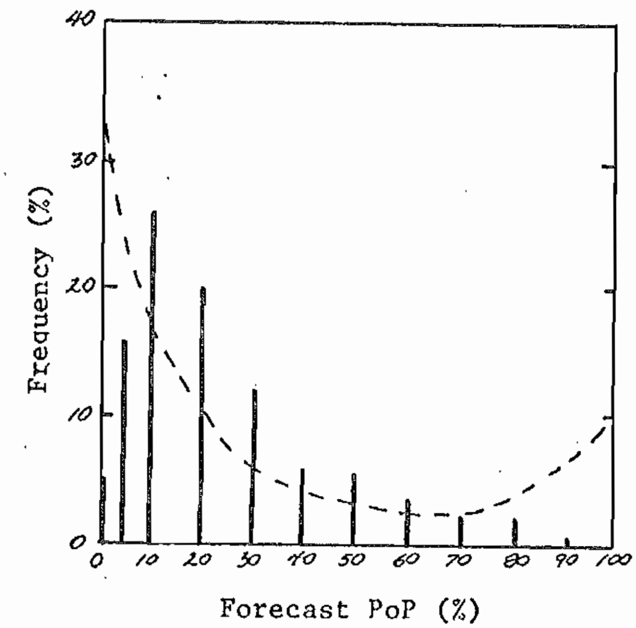

Fig. 3 Frequency of use of various values of forecast PoP for a sample of data from Chicago (after Hughes, 1965). See text for explanation of broken line.

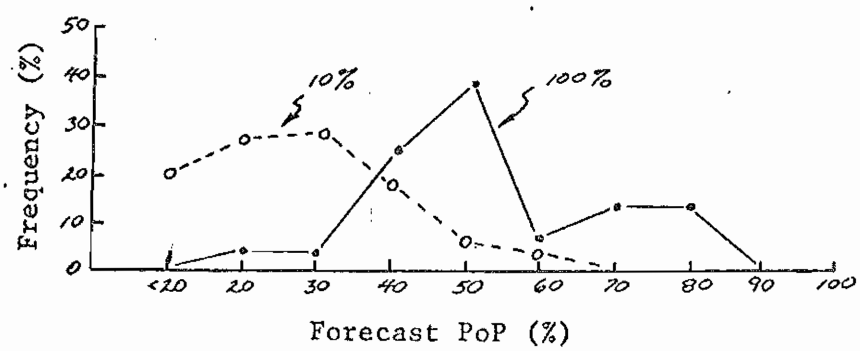

Fig. 5 Frequency of use of forecast PoPs for 32 daytime periods with observed areal coverage of 100% (solid line) and for 60 daytime periods with coverage of 10% (broken line). Alabama and Florida zones are combined.