U. S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

NOAA Technical Memorandum NWS TM SR-57

SOME ASPECTS OF PROBABILITY OF PRECIPITATION FORECAST PERFORMANCE BY WSFOS IN THE SOUTHERN REGION -- NOVEMBER 1969 THROUGH OCTOBER 1970

SOUTHERN REGION HEADQUARTERS
SCIENTIFIC SERVICES DIVISION
FORT WORTH, TEXAS
May 1971

SOME ASPECTS OF PROBABILITY OF PRECIPITATION FORECAST PERFORMANCE BY WSFOS IN THE SOUTHERN REGION -- NOVEMBER 1969 THROUGH OCTOBER 1970

Allen D. Cummings, SSD

INTRODUCTION

Forecast verification printouts for the period November 1969 through October 1970 have recently been made available to WSFOs in the Southern Region. These printouts list verification data on temperature and probability of precipitation (PoP) for individual forecasters and for the WSFO forecast product itself. They treat the year as a whole and also provide a broad seasonal breakdown, i.e., November through April (cold season) and ~~April~~ May through October (warm season). Included are forecasts made by the WSFO for its own location and for stations receiving direct guidance from it -- 4 additional stations in most cases. Forecasts for 0600Z and 1800Z have been combined in the verification procedure. The information contained in the printouts offers an opportunity to improve the quality of forecasts through identification of individual and station biases. A known bias is one that can be reduced.

Scientific Services Division, Southern Region Headquarters, has examined in some depth the PoP verification data for the Region. Since each forecast office has access only to its own record, SSD has prepared certain of the data for all WSFOs, jointly and separately, in a convenient format so that each may compare its product with that of others and with something of a regional "average". Such comparisons have limited utility but may serve to stimulate thought and discussion which will lead to tangible improvement in PoP forecasts.

Estimates of the probability of precipitation that reach the using public represent the most important routine forecast product of the Weather Service. The credibility even of the high-impact services we provide has roots in our ability to predict with demonstrable success whether or not it will rain at some time in the future. Since we are obliged to approach this from a probabilistic standpoint, we owe to the user the greatest reliability in *our* statements that we can bring to bear on *his* problems. Certainly we *would* like to exhibit this reliability in *the* very high and very low probabilities, but the state of the science does not usually permit us to resolve the situation so neatly. The forecaster must be aware of his limitations and the MIC should be aware that these limitations vary among forecasters. It is the responsibility of the MIC to take corrective action to eliminate individual (and, consequently, station) biases that are uncovered in the verification program. SSD has directed attention only to the total forecast product; evaluation of individual forecaster verification data should be undertaken locally.

This Tech Memo is not a treatise on probability forecasting or verification. It is purposely elliptic, presupposing more than a nodding acquaintance with the subject. This is reasonable as it is intended expressly for the information of those SR forecasters directly associated with the origination, adaptation or quality control of PoP forecasts.

THE RELIABILITY DATA

Figures 1-28 show graphs of forecast probability vs. observed frequency of measurable precipitation for the WSFOs (except San Juan) in the Southern Region. Such presentations may be called "reliability graphs", since the nearness of the plotted points to the diagonal line is a measure of reliability. The number of forecasts associated with each point is entered so that an impression of "resolution" is gained.

A Regional View. Figure 1 is a composite for all forecasts in the Region during the period November 1969 through October 1970, except that the relatively few forecasts of 2% and 5% made by ABQ and FTW have been omitted. All points except that in the >95% category are within 5% of perfect reliability in Period 1 with a slight seeming deterioration evidenced in Period 2. But as a matter of fact the root-mean-square reliability error (ER) is 3.2% in Period 1 and only 2.8% in Period 2. This is not surprising since reliability should not depend on lead time. The better (lower) Brier score in the earlier period is due to a smaller resolution penalty. The Period 3 graph shows a general inability to successfully use the very high probabilities, and to some extent the "near zero" value, this far in advance; the ER is 3.6%. As expected, the figure shows a progressive concentration of forecasts near the mean climatological value (which lies in the neighborhood of 15-20%) with increasing lead time while the use of extreme probability estimates at both ends of the scale declines. Most notable perhaps is the general overforecasting of precipitation on a regionwide basis.

The record of the WSFOs. An examination of the Figures 2-28 will reveal the biases and limitations each WSFO has shown. Figure-by-figure comments will not be offered since the more interesting features are obvious; e.g., the remarkably consistent overestimation of PoP seen in Figure 17, the classic "underconfidence" exhibited in Period 1 of Figure 20, the almost complete lack of ability to depart successfully from near-climatological values exhibited in Period 3 of Figures 4 and 19 (probably due to peculiar precipitation regimes), the curious avoidance of 40% forecasts evidenced in Figure 26 (this value serves the user as well as any other and should not be arbitrarily neglected), etc. Figures 2-28 do not, of course, display the consistency of the regional composite graphs. This is due largely to the smaller sample size for a given station, since except for the lower categories not enough forecasts have been made to achieve a stable degree of reliability. The PoP values that have found their way to the user in a given locality may not exhibit characteristics greatly different than these sets, since the number of forecasts is comparable.

Generalized guidelines for greater reliability. The scatter of points in the range of probabilities beyond which reasonable reliability is not retained suggests a potential source of dramatic improvement. The scatter will be even more pronounced for some individual forecasters and it is at this formulating level that a revision of methodology must take place. A forecaster should adjust his approach to the problem and re-evaluate his limitations if he finds himself "way out in left field" in some range of categories. Such adjustments ought to be made continually to ensure that the user of the forecast can, in the long run, rely on PoP statements. This is not to encourage a false placing of forecasts in categories where the forecaster does not believe they belong in order to achieve greater reliability. Rather, he must recognize that previous statements of PoP in problem categories may have been over- or under-estimates and subsequently modify his appraisal of the types of weather situations associated with them.

There appears to be a general overuse of near-zero probability, particularly in Period 3. A greater effort should be made to identify those low-potential situations where the uncertainties are sufficient to warrant moving some of these forecasts into higher categories. Immediate results may not be so spectacular at this extremity in the long range view due to the vastly diminished returns to be expected, but a quantum step in the short run is essential to greater utility of PoP forecasts and will have an ultimate impact on performance measures as great as improvement in higher categories.

Boldness is an asset to a forecaster, but those who are most inclined that way must be wary of claiming a degree of certainty in their statements which will not suffer the test of verification gladly. Reliability is a necessary (albeit, not sufficient) characteristic of a satisfactory set of PoP forecasts. Such a set should exhibit an overall ER of certainly less than 5%, assuming a substantial body of forecasts. Goals ought to be set locally for individuals and for the WSFO product and revised as they are achieved. The ER for all periods combined, region-wide, was 3.2% from November 1969 through October 1970. Since this is not a function of lead time it should be distressing if a WSFO finds a significant increase in its ER in Period 3. This may be the case for some, since the composite ERs reflect it to a modified degree. A possible explanation may lie in more relaxed attitudes toward Period 3 when requirements of the nearer future are pressing. There is a manifest need to guard against this.

SSD has not computed ERs except on a regional basis. Each WSFO should do this for its own product and establish preliminary goals commensurate with foreseeable accomplishment. The ER is determined from reliability data in the printouts according to

$$ER = 100 \left[\frac{1}{N} \sum_{i=1}^{n} l_i (F_i - f_i)^2 \right]^{\frac{1}{2}}$$

where N is the total number of forecasts in all categories; n the number of categories; l_i the number of forecasts in the *i*th category; F_i the *i*th forecast probability category, i.e., 0, .1, .2, ...; and f_i the observed frequency of precipitation for the *i*th category. The computation is a simple one to perform.

CERTAIN ASPECTS OF THE S-SCORES

The regression lines in Figure 29 afford a graphic view of the inverse relationship of forecast lead time and ability to improve on constant forecasts of the climatological frequencies. These lines were computed by the method of least squares. Correlation coefficients for Periods 1, 2 and 3 are .535, .623 and .601, respectively. Figure 30 is the same presentation but for Period 1 only. The dashed line is the regression line for all stations except Miami. The correlation coefficient for this line is .747, which suggests that the relationship between S-score and precipitation frequency is a valid one but that Miami is not properly included, due probably to the unique environment of the forecast area. We may, then, cautiously surmise that WSFOs above the dashed line have a better-than-average performance record.

Figure 31 is included to illustrate the extent of improvement of forecast scores in the cold season as opposed to the warm season. The great difference in relative performance by Albuquerque and Miami is noteworthy. This difference may reflect a dependence of performance potential on precipitation regime as well as frequency. Warm season precipitation forecasting at these WSFOs may be a particularly fruitful area for objective forecast studies.

CONCLUDING REMARKS

The Southern Region TM series serves a multiplicity of purposes, in this case primarily that of a distribution vehicle to afford WSFOs a chance to see how their PoP product stacks up against that of others. The data presented here may also benefit WSOs receiving direct PoP guidance from a parent WSFO. No precept to guarantee improvement has been offered, but hopefully the reiteration of some basic ideas will prove helpful in bettering forecast performance.

The body of literature (much of it informal) on PoP forecasting and verification has become fairly extensive. While no specific citations have been made here, a brief list of especially cogent references is appended. The ideas above, express and implied, come largely from these sources.

ACKNOWLEDGMENT

Special thanks to Dan Smith, SSD, SRH, who prepared the figures.

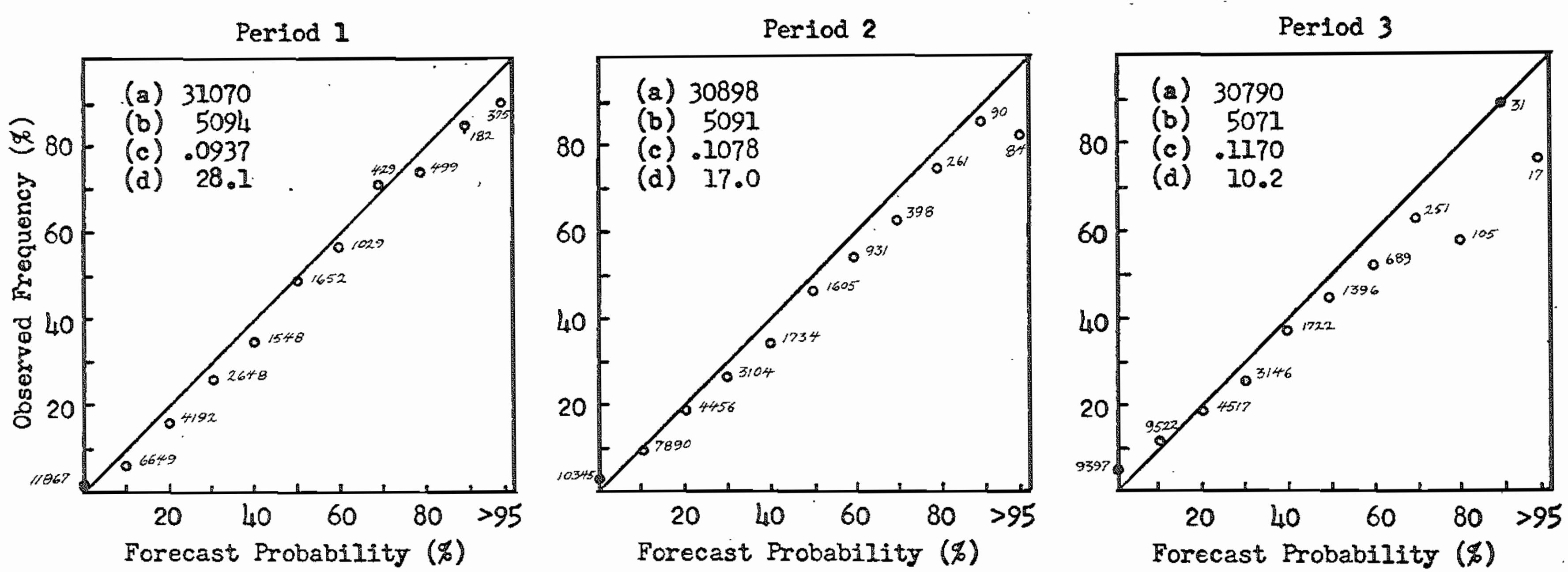
SELECTED REFERENCES

1. Epstein, E. S.,1966: Quality Control for Probability Forecasts. Mon. Wea. Rev., 94, 487-494.

2. Hughes, L. A., 1965: On the Probability Forecasting of the Occurrence of Precipitation. ESSA Tech. Note 20-CR-3, 36 pp.

3. __________, 1967: Improving Precipitation Probability Forecasts Using the Central Region Verification Printout. ESSA CRTM-15, 14 pp.

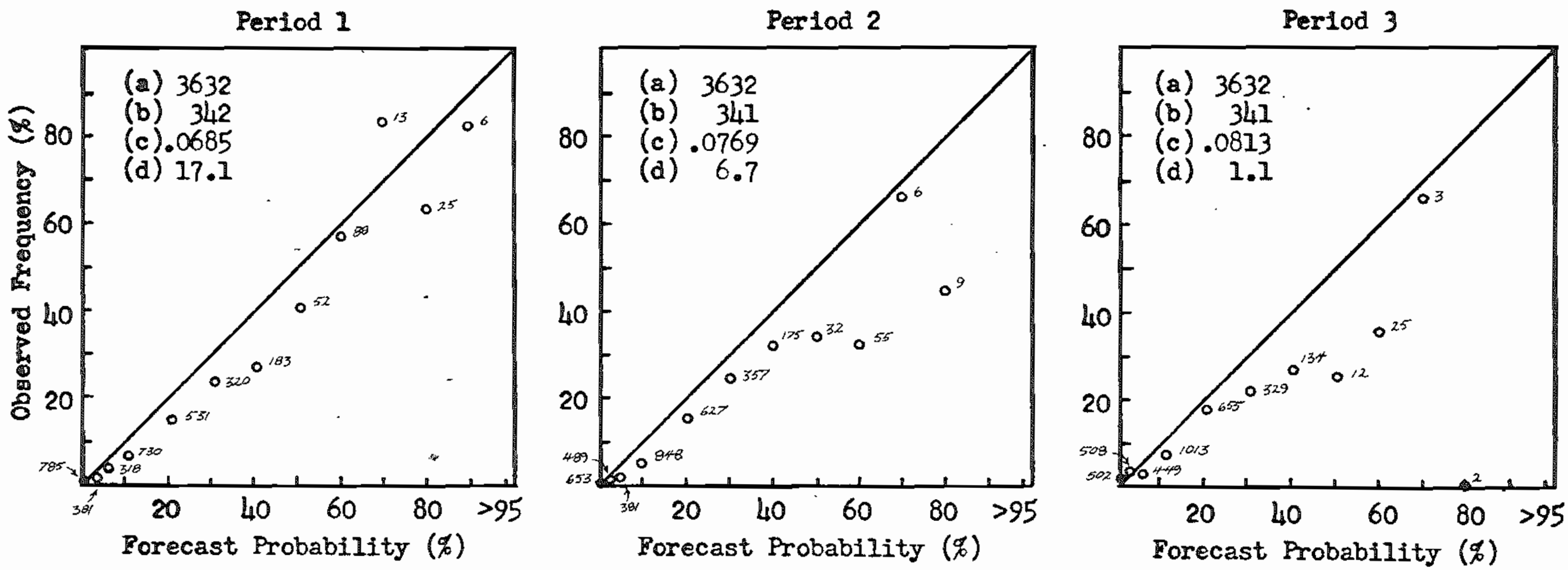
4. Sanders, F., 1963: On Subjective Probability Forecasting. J. App. Meteor., 2, 191-201.



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score (average for all WSFOs)
(d) Improvement over climatology (%) (average)

Figure 1. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

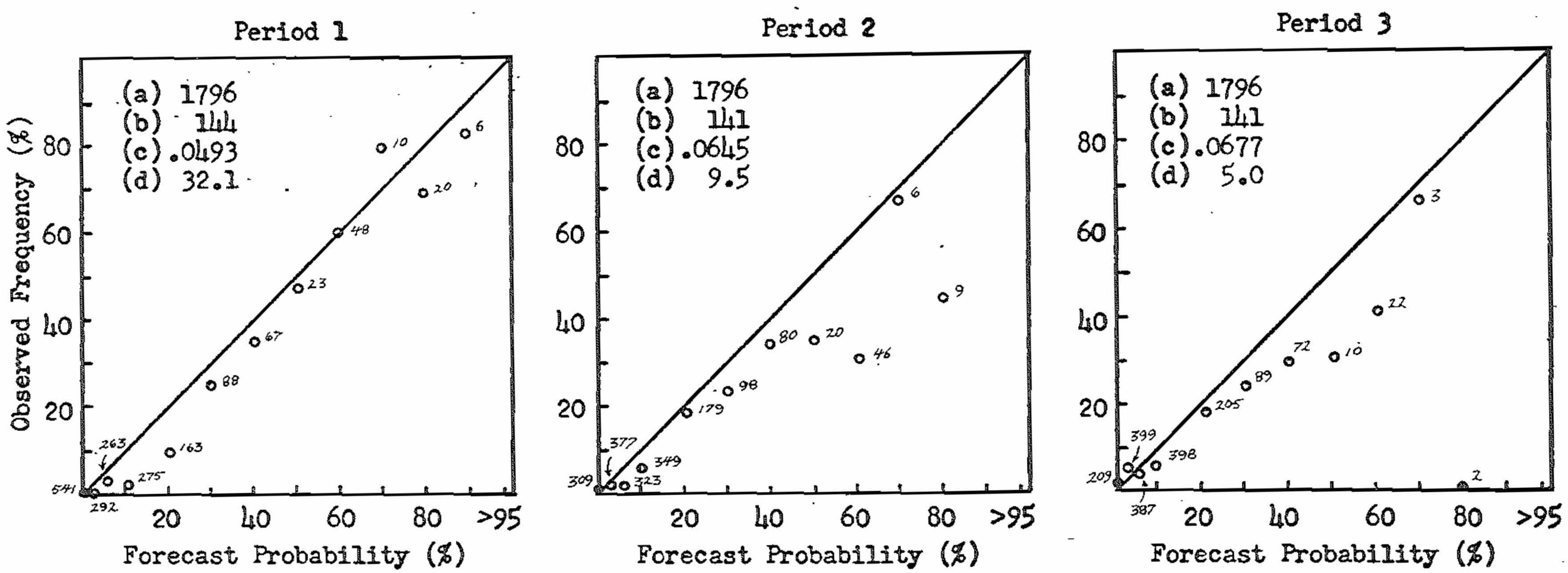
ALL STATIONS -- Whole Year



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 2. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

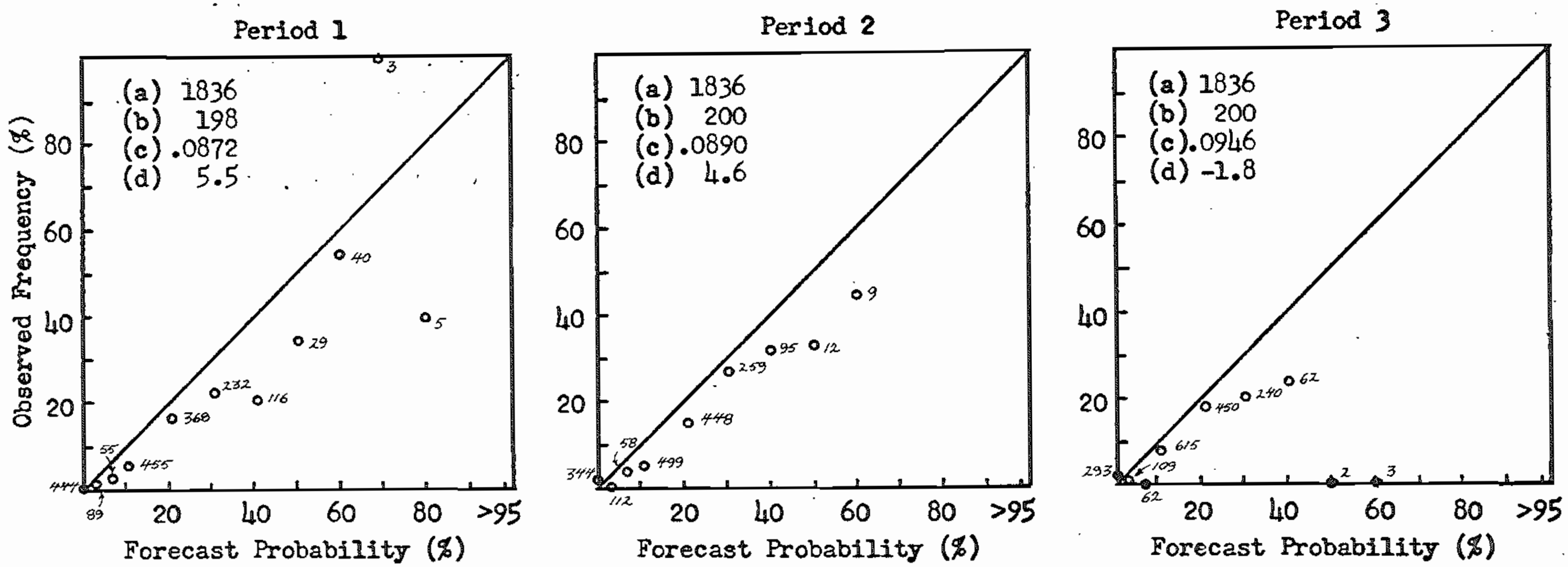
ALBUQUERQUE -- Whole Year



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 3. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

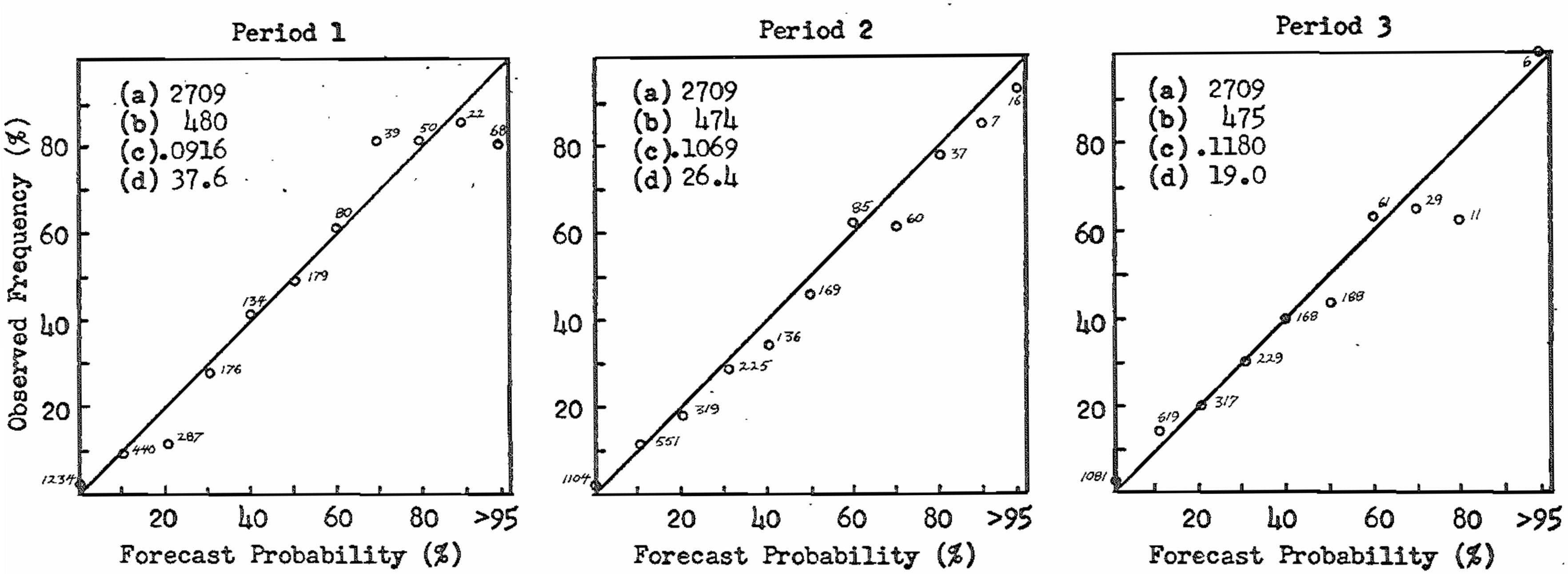
ALBUQUERQUE -- Cold Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 4. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

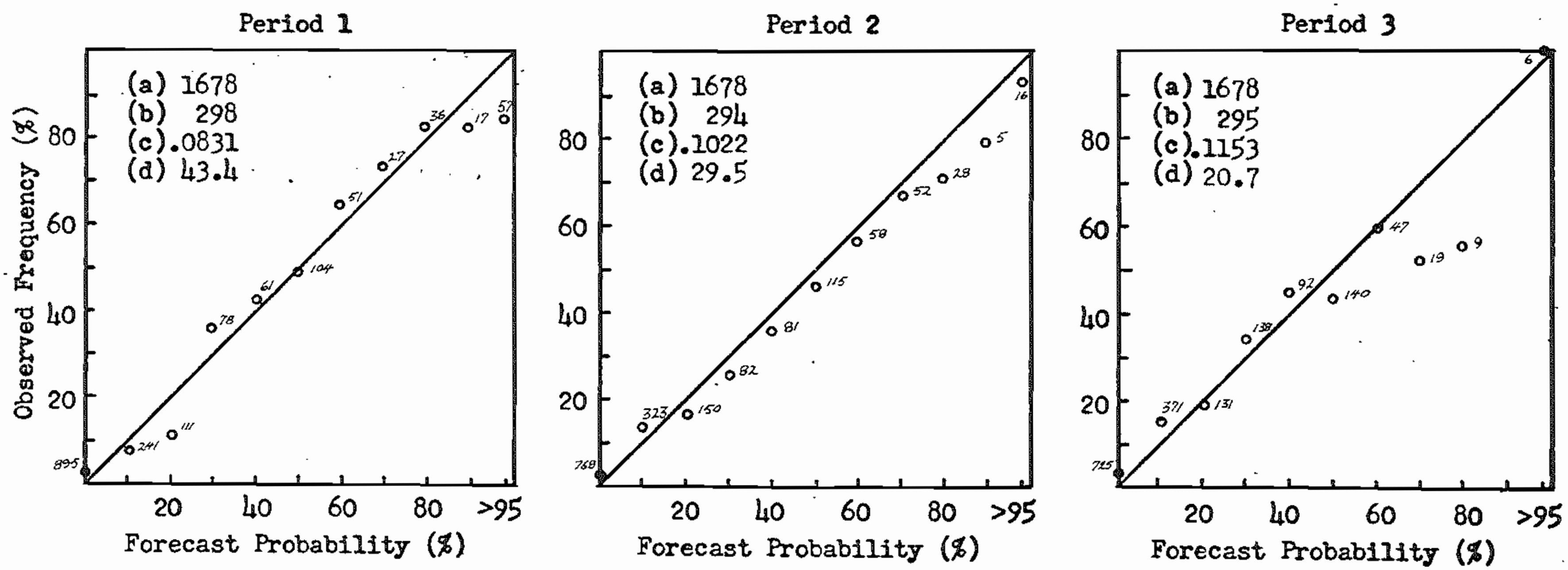
ALBUQUERQUE -- Warm Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 5. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

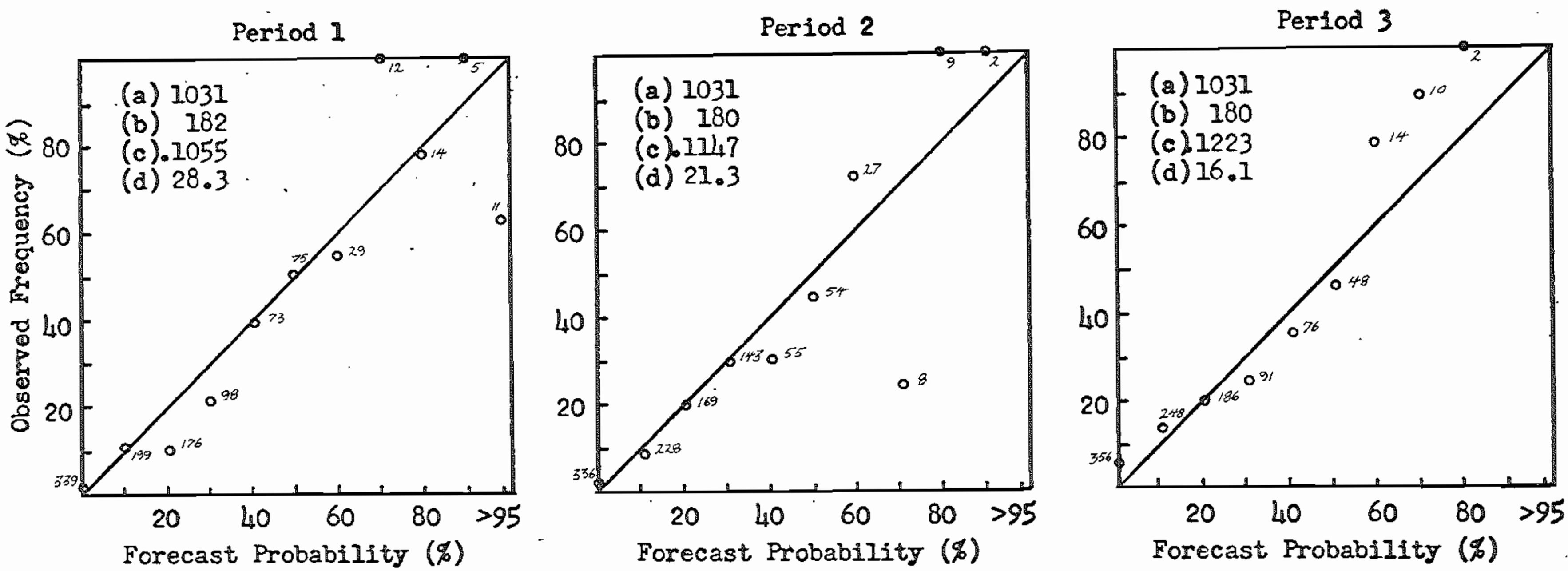
ATLANTA -- Whole Year



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 6. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

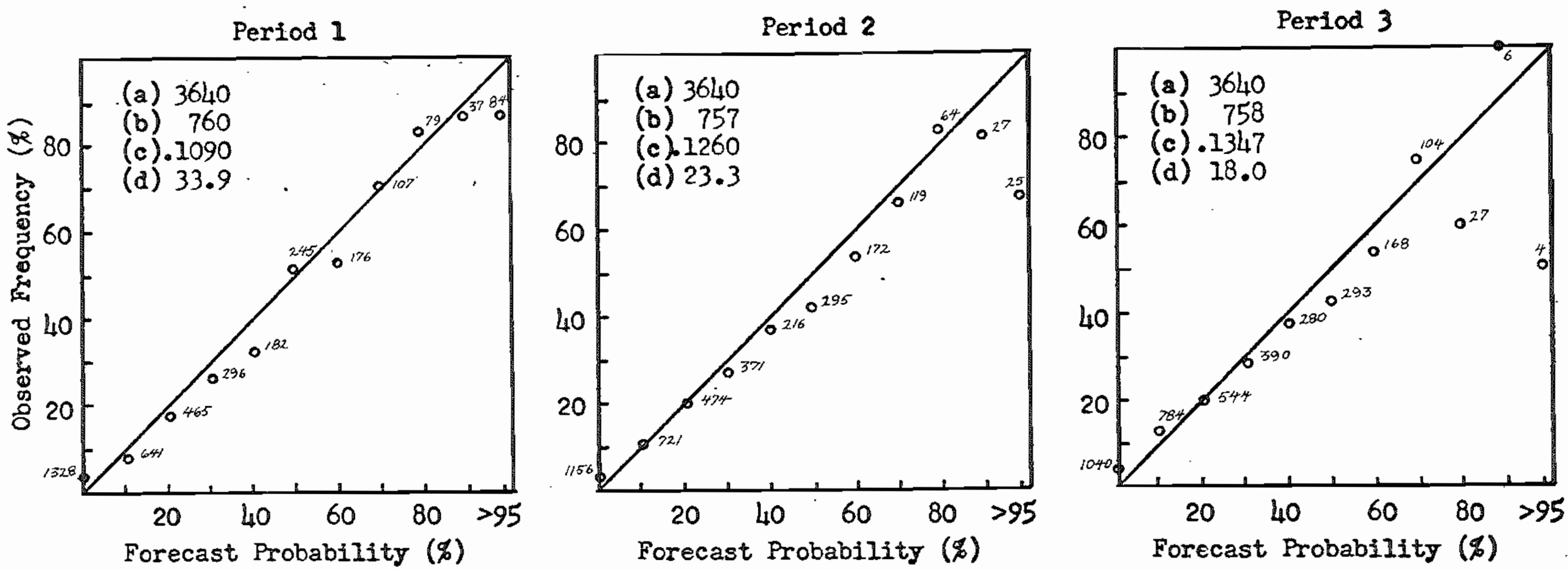
ATLANTA -- Cold Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 7. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

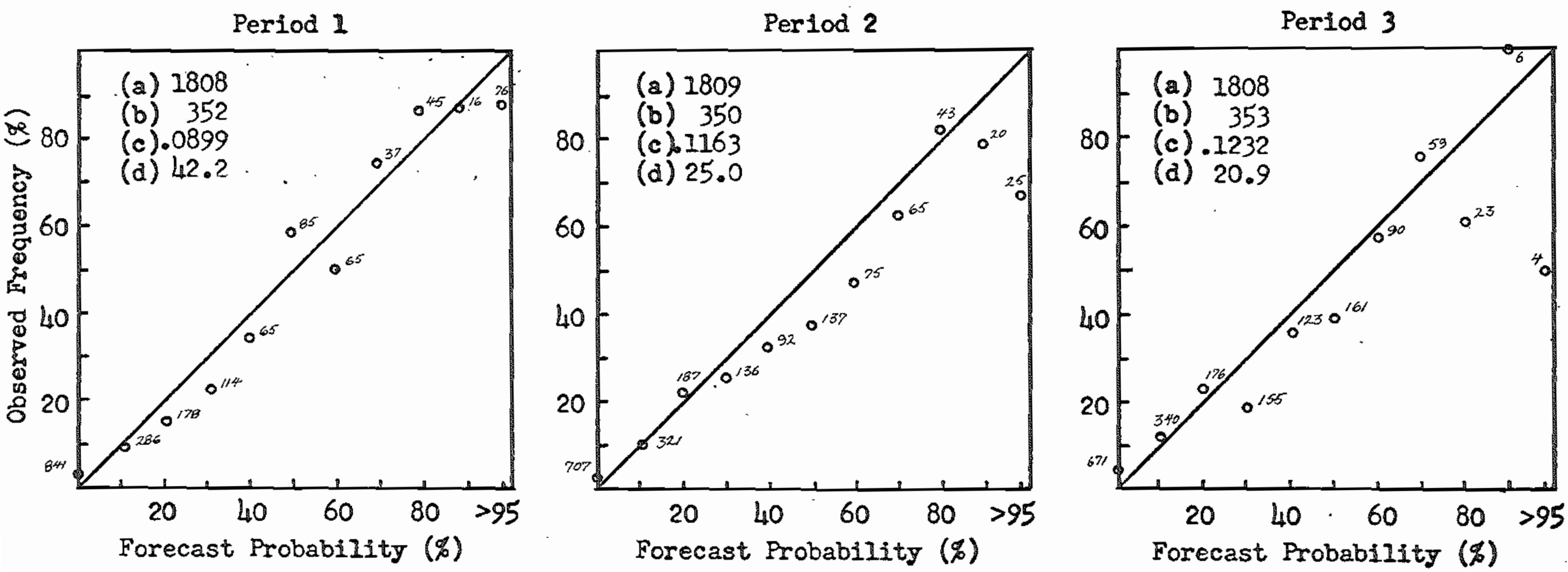
ATLANTA -- Warm Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 8. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

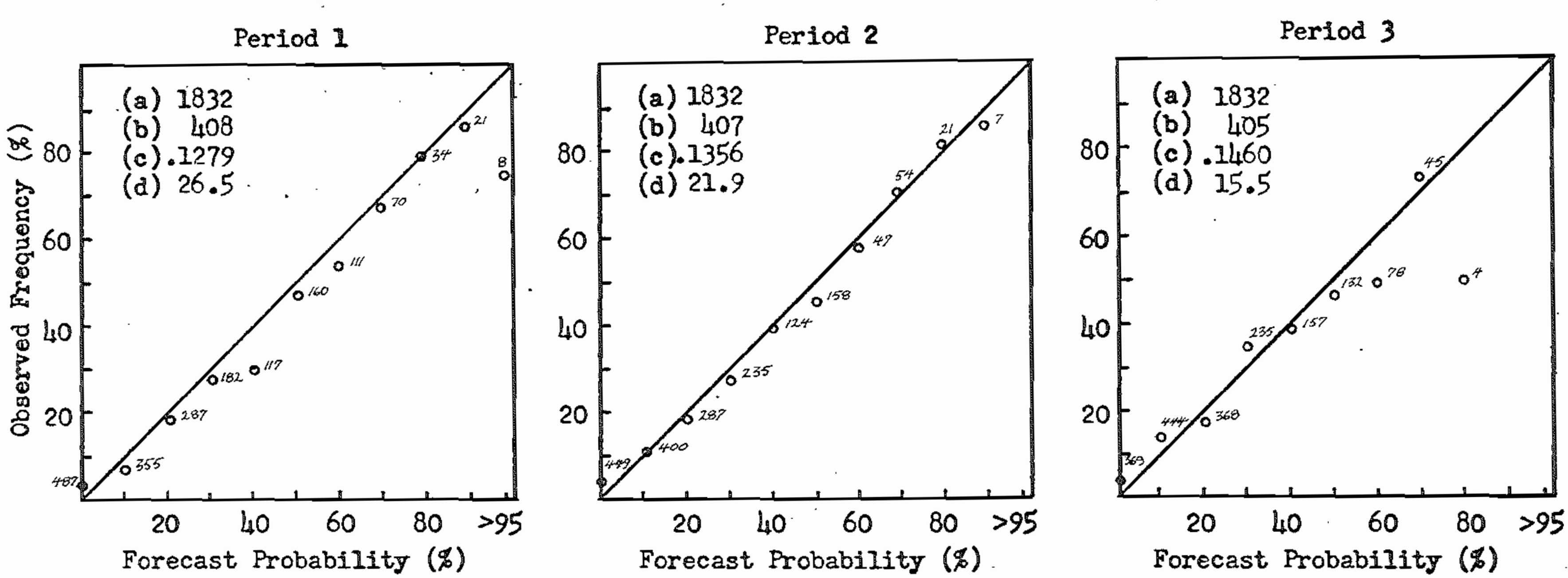
BIRMINGHAM -- Whole Year



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 9. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

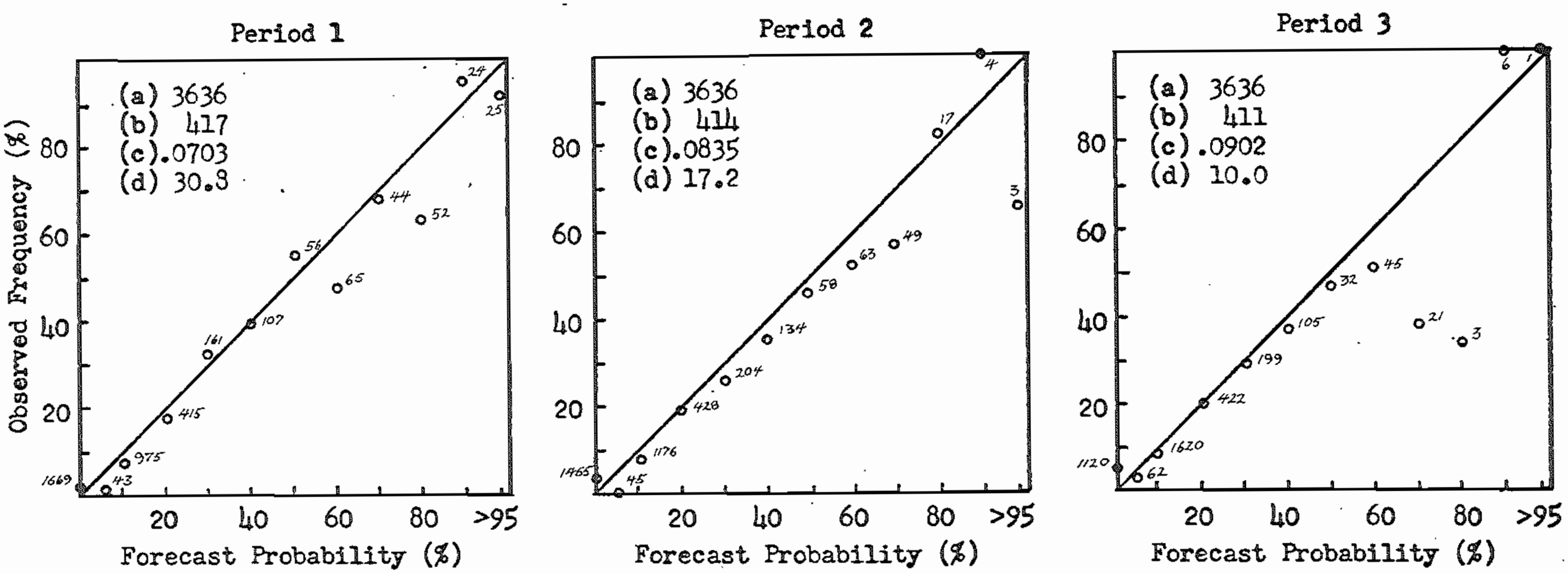
BIRMINGHAM -- Cold Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 10. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

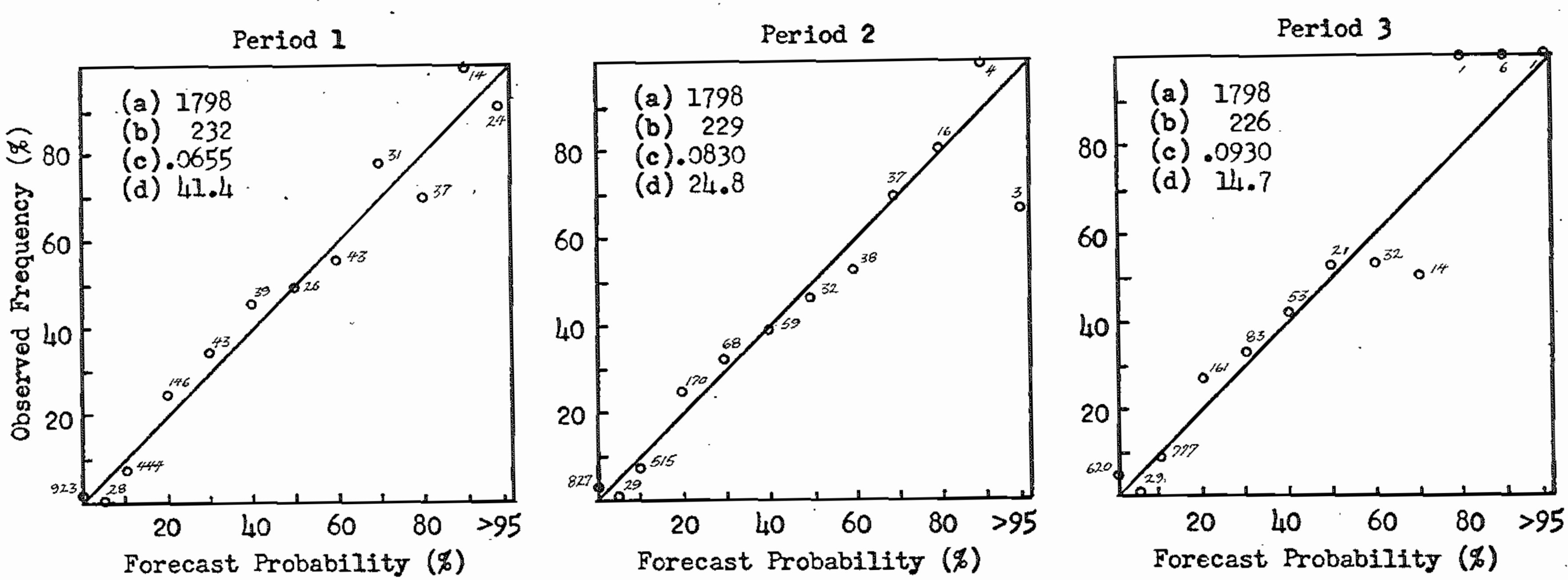
BIRMINGHAM -- Warm Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 11. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

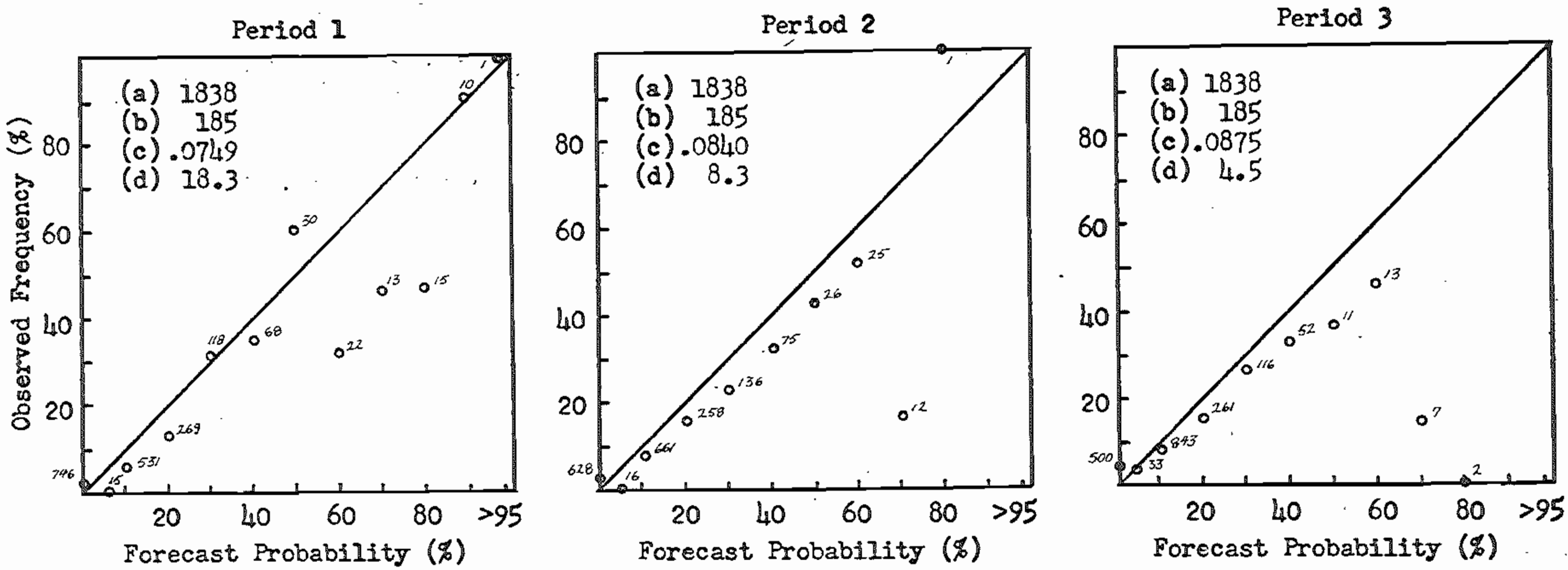
FORT WORTH -- Whole Year



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 12. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

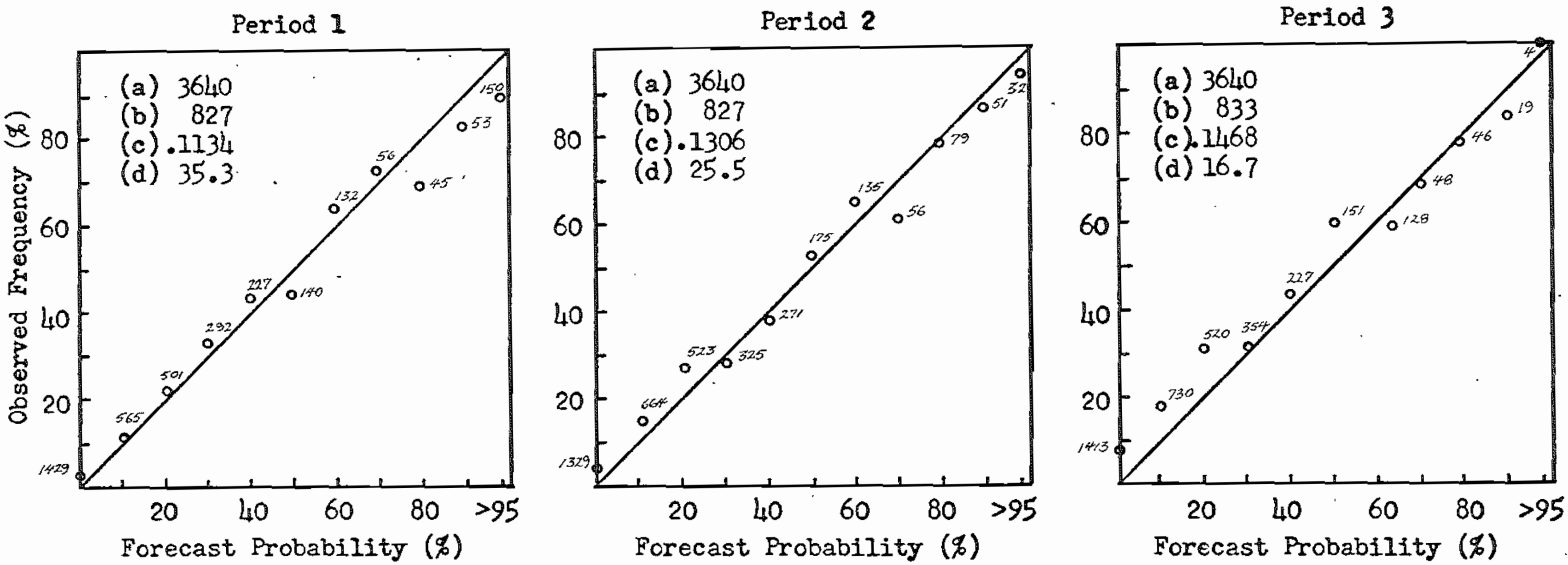
FORT WORTH -- Cold Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 13. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

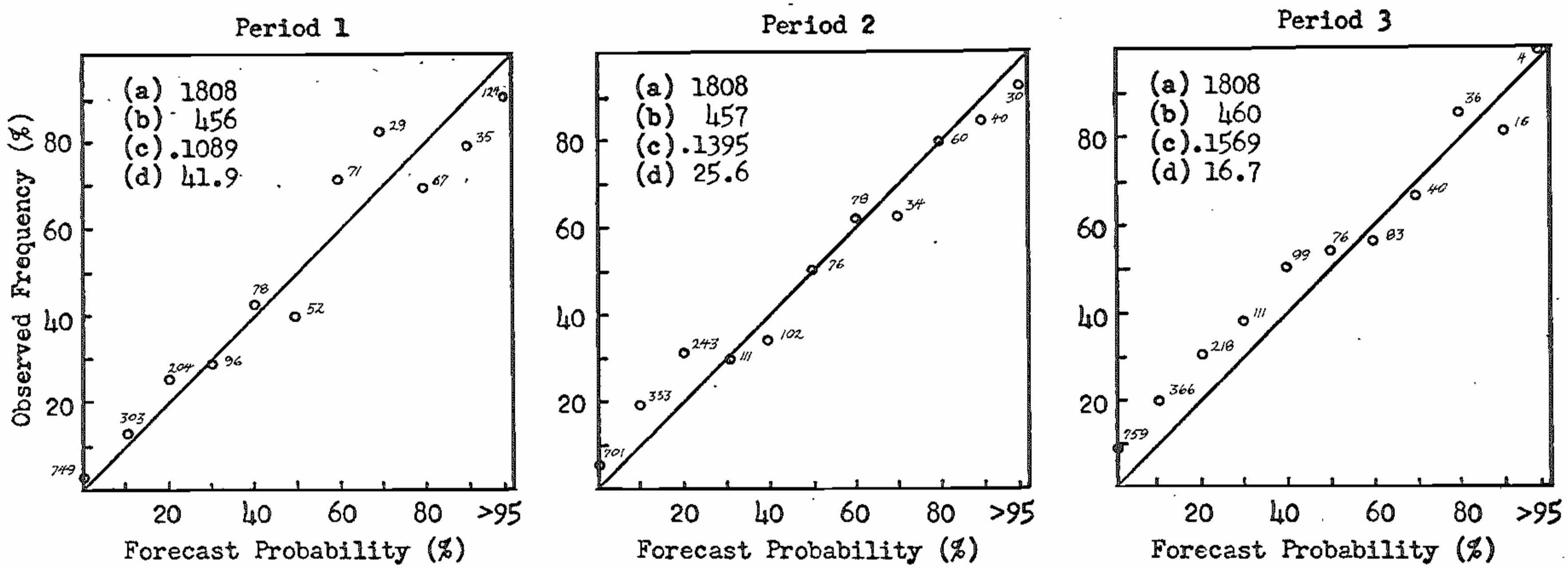
FORT WORTH -- Warm Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 14. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

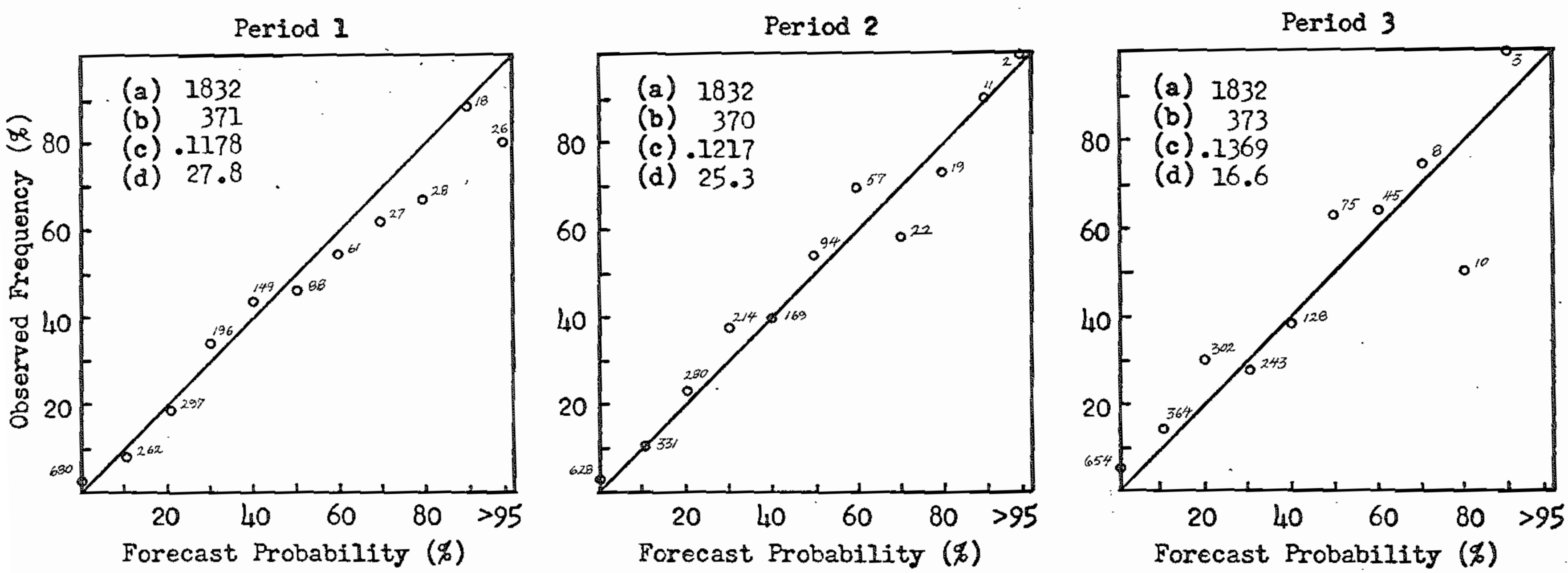
MEMPHIS -- Whole Year



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 15. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

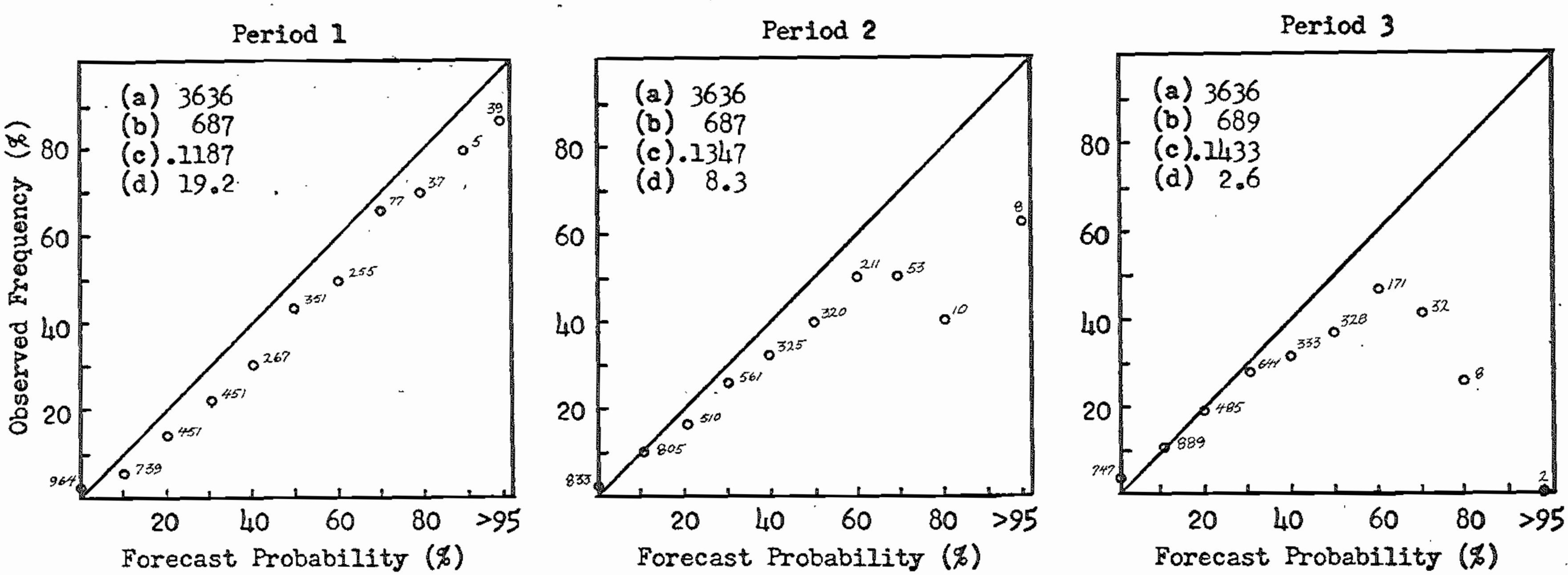
MEMPHIS -- Cold Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 16. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

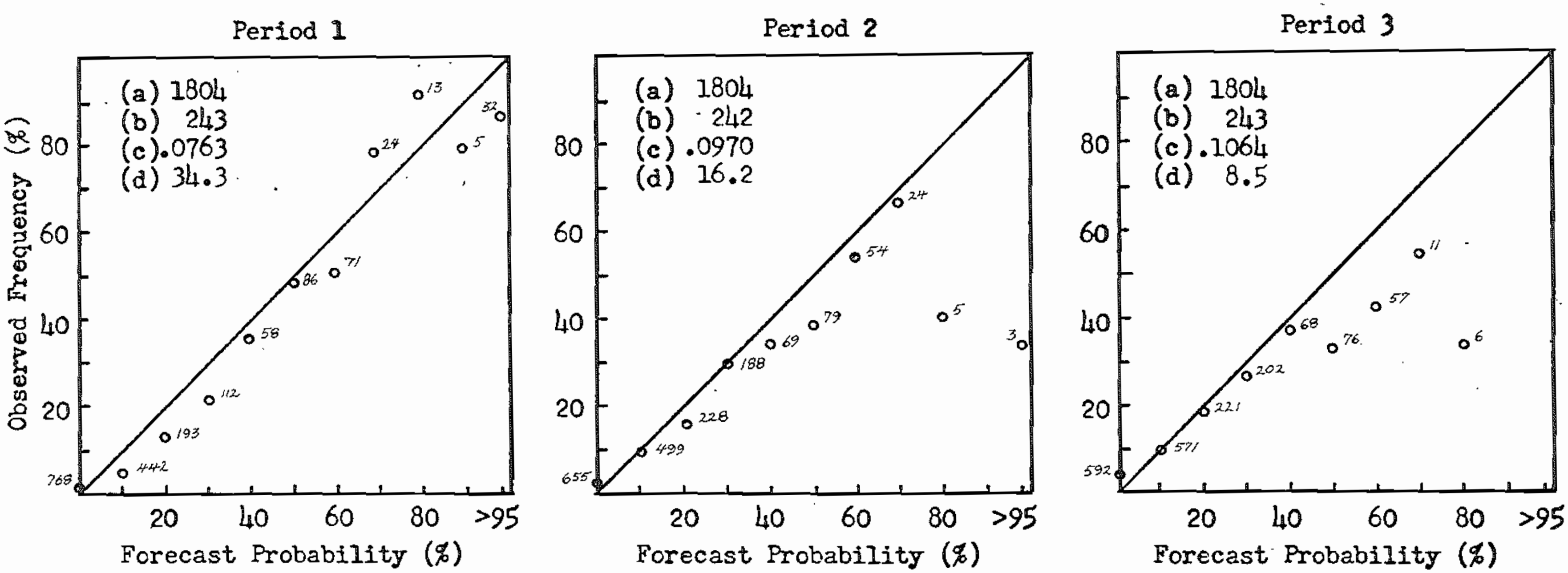
MEMPHIS -- Warm Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 17. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

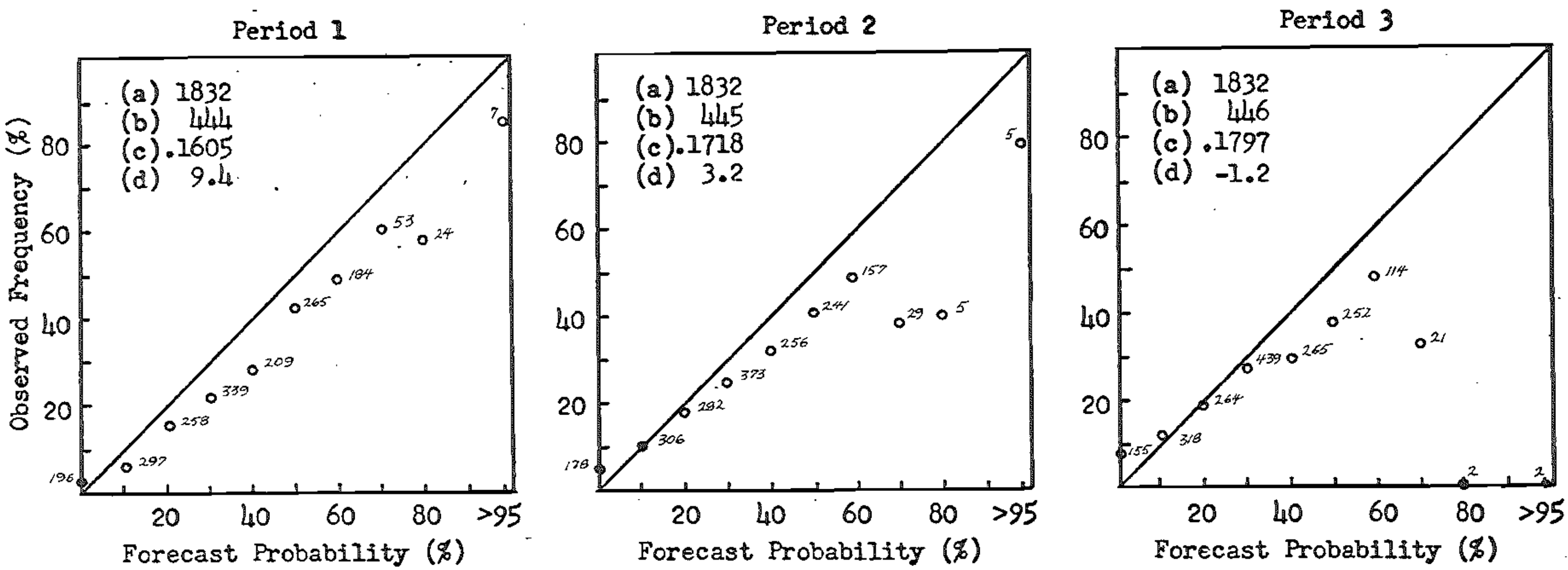
MIAMI -- Whole Year



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 18. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

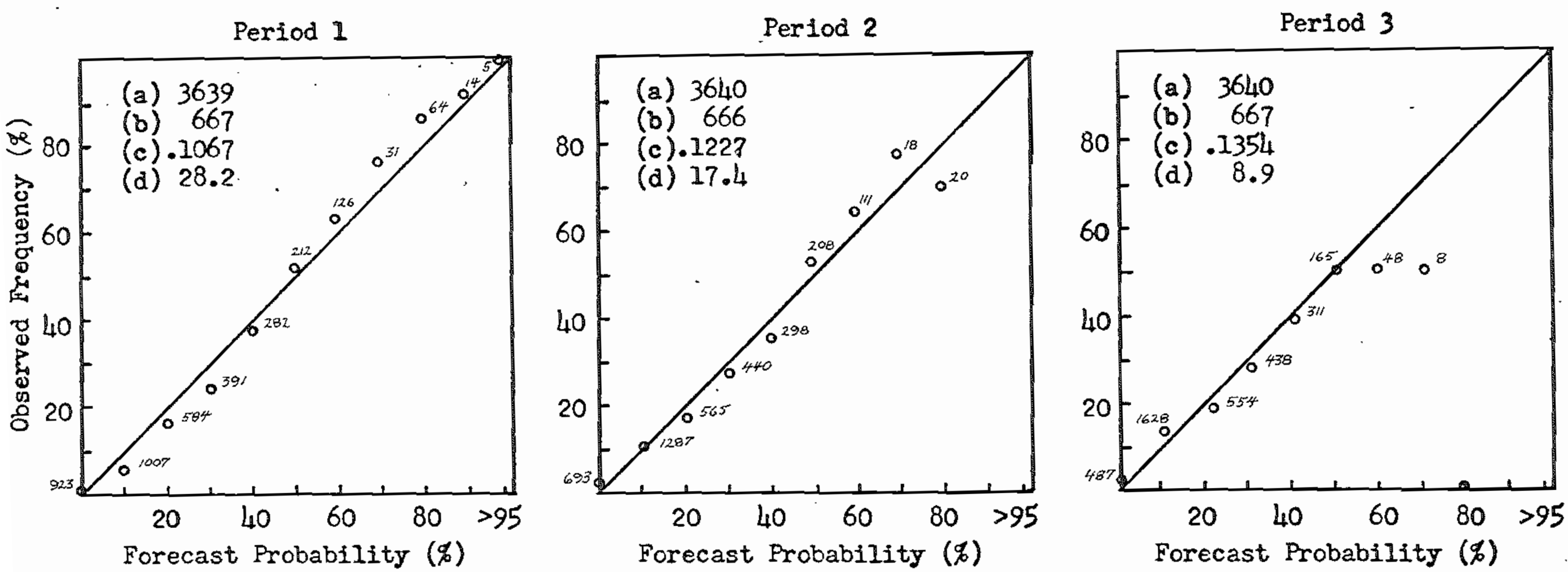
MIAMI -- Cold Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 19. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

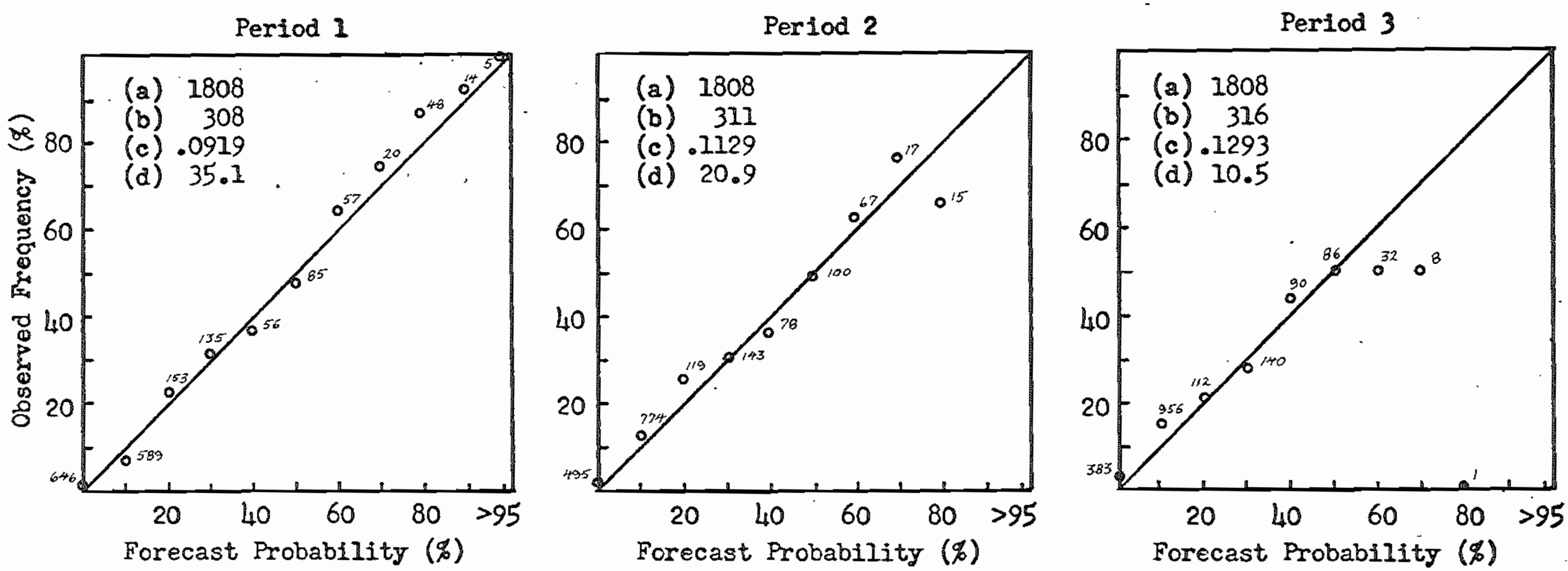
MIAMI -- Warm Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 20. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

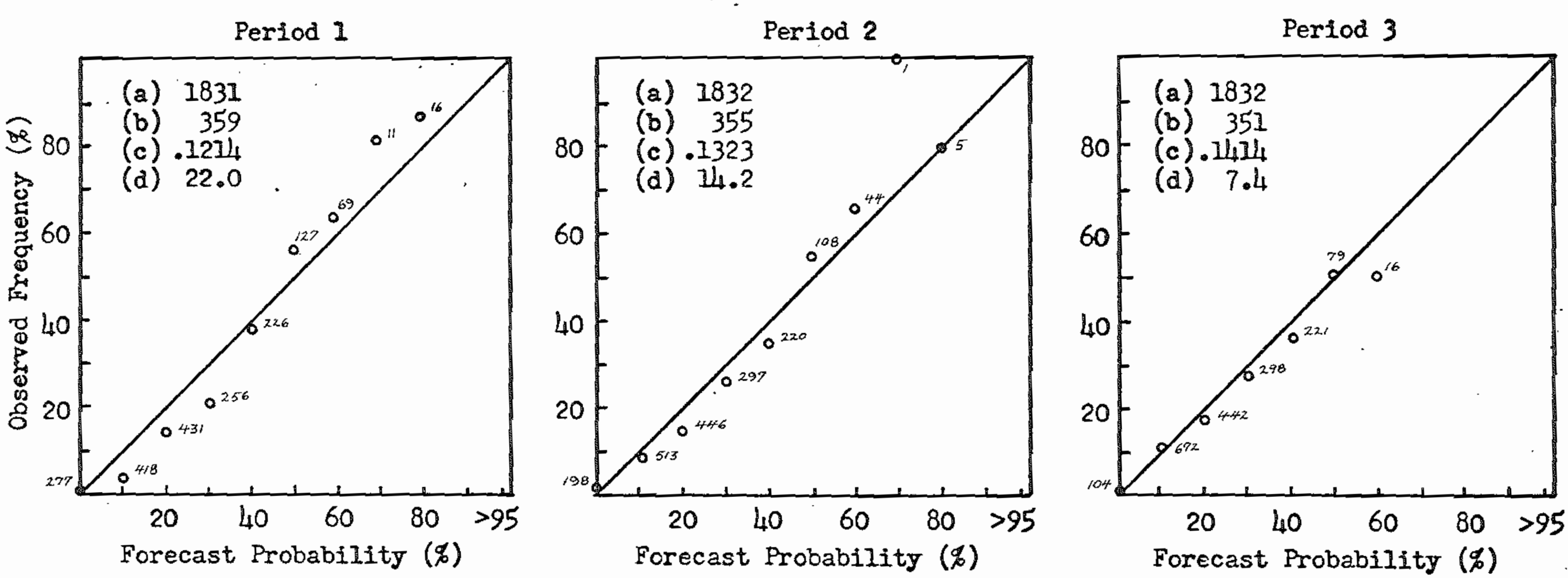
NEW ORLEANS -- Whole Year



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 21. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

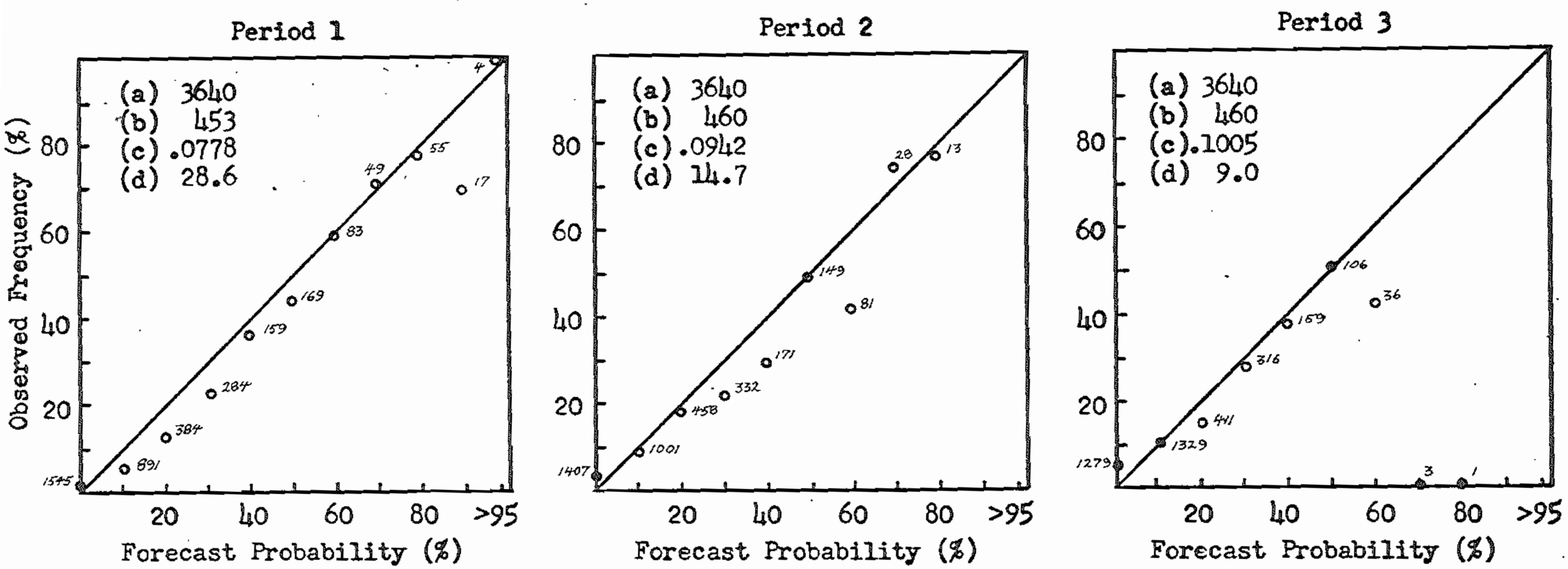
NEW ORLEANS -- Cold Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 22. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

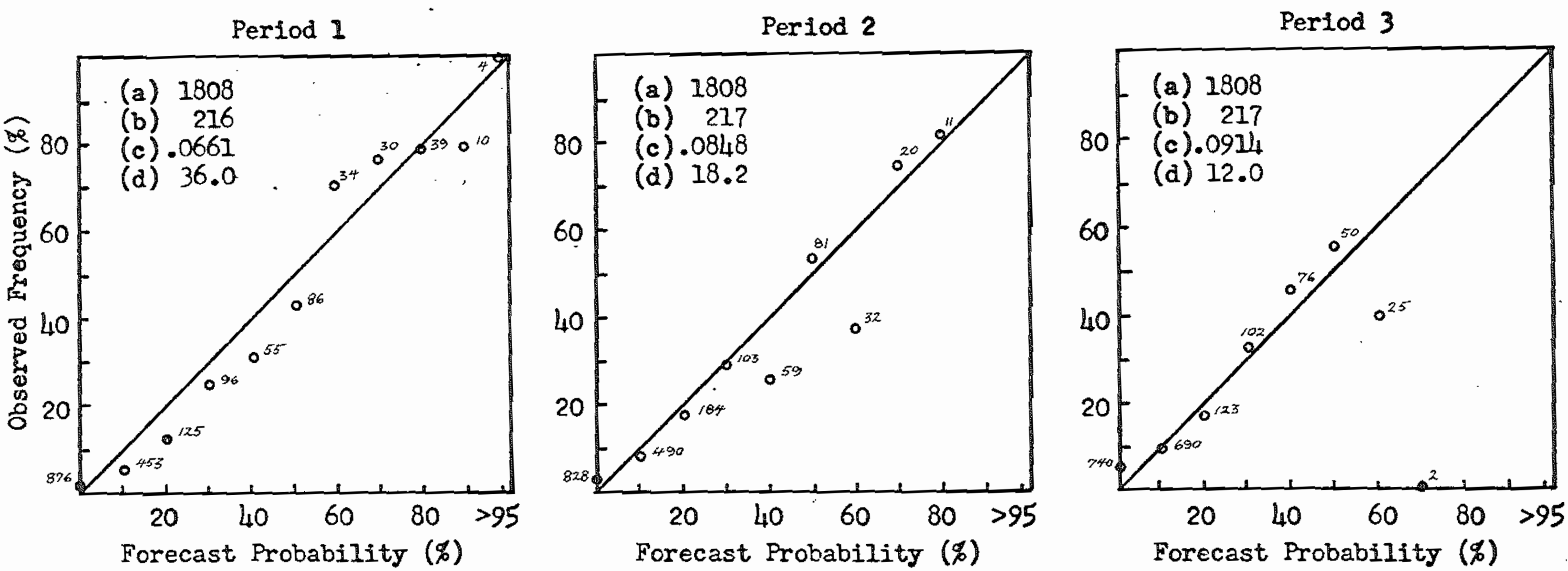
NEW ORLEANS -- Warm Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 23. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

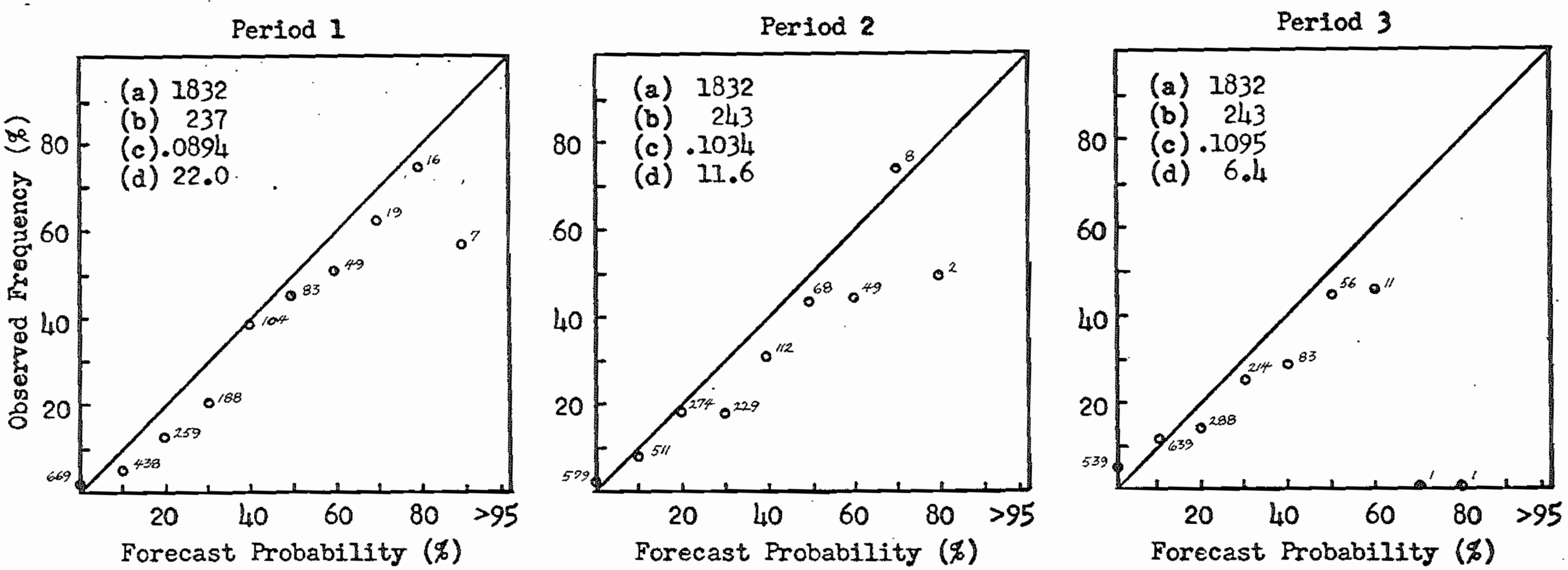
OKLAHOMA CITY -- Whole Year



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 24. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

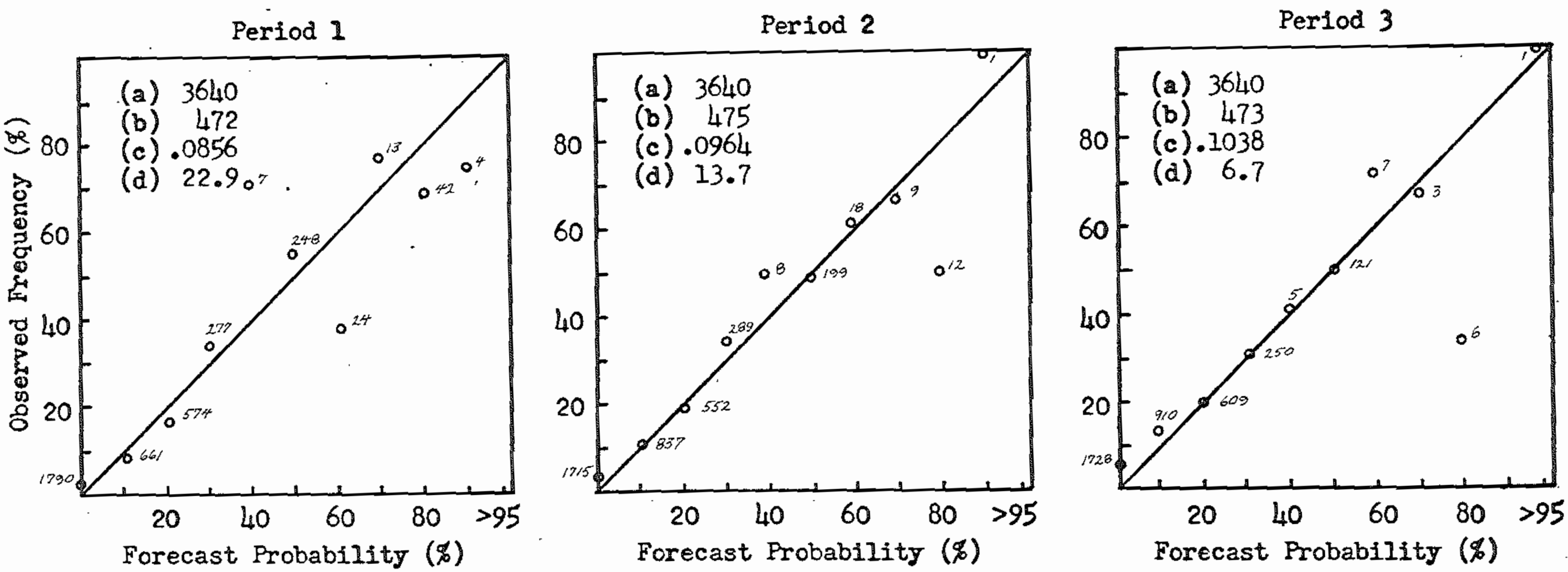
OKLAHOMA CITY -- Cold Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 25. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

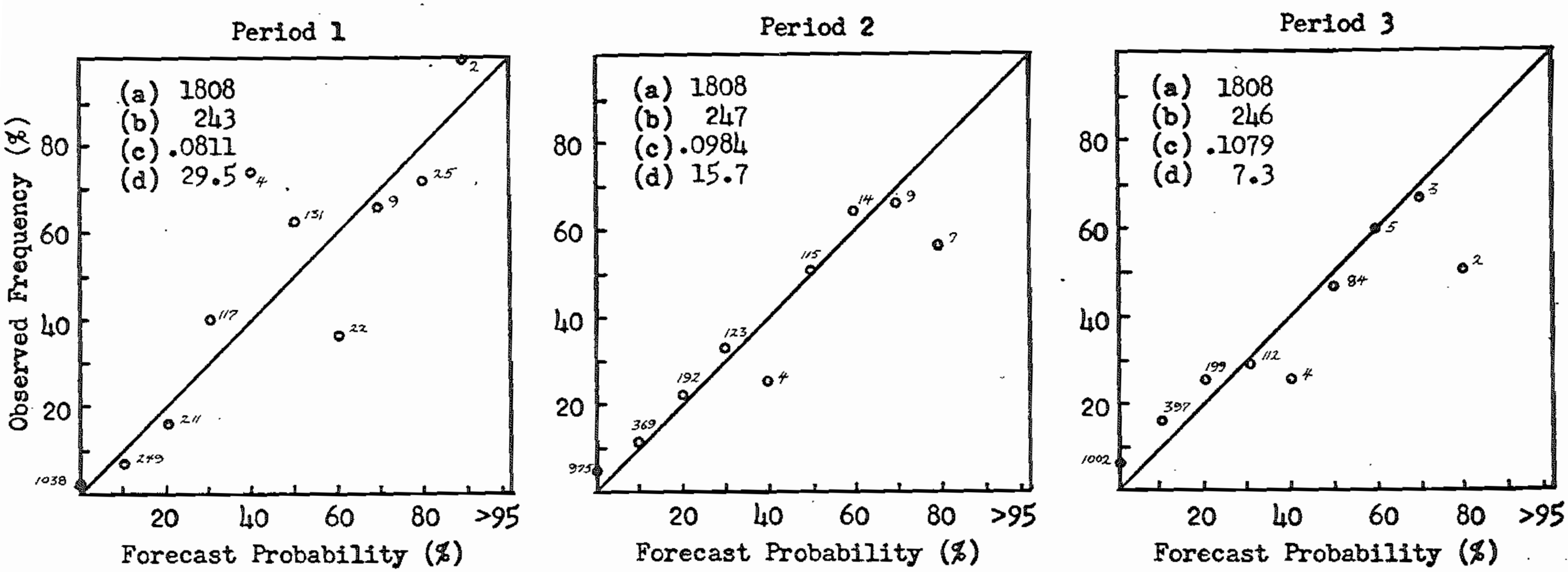
OKLAHOMA CITY -- Warm Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 26. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

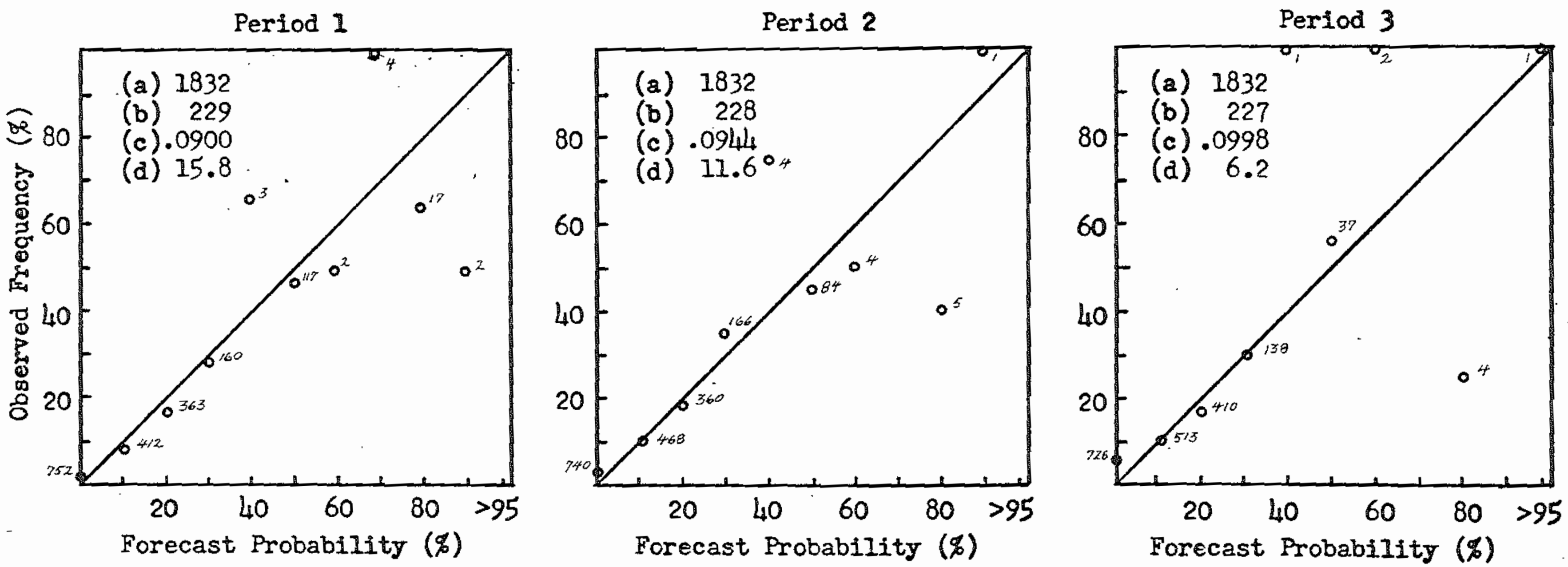
SAN ANTONIO -- Whole Year



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 27. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

SAN ANTONIO -- Cold Season



(a) Total number of forecasts
(b) Number of precipitation cases
(c) Brier score
(d) Improvement over climatology (%)

Figure 28. Forecast probability vs. observed frequency of occurrence of measurable precipitation in the indicated 12-hour periods for all forecasts (06Z and 18Z combined) issued by

SAN ANTONIO -- Warm Season

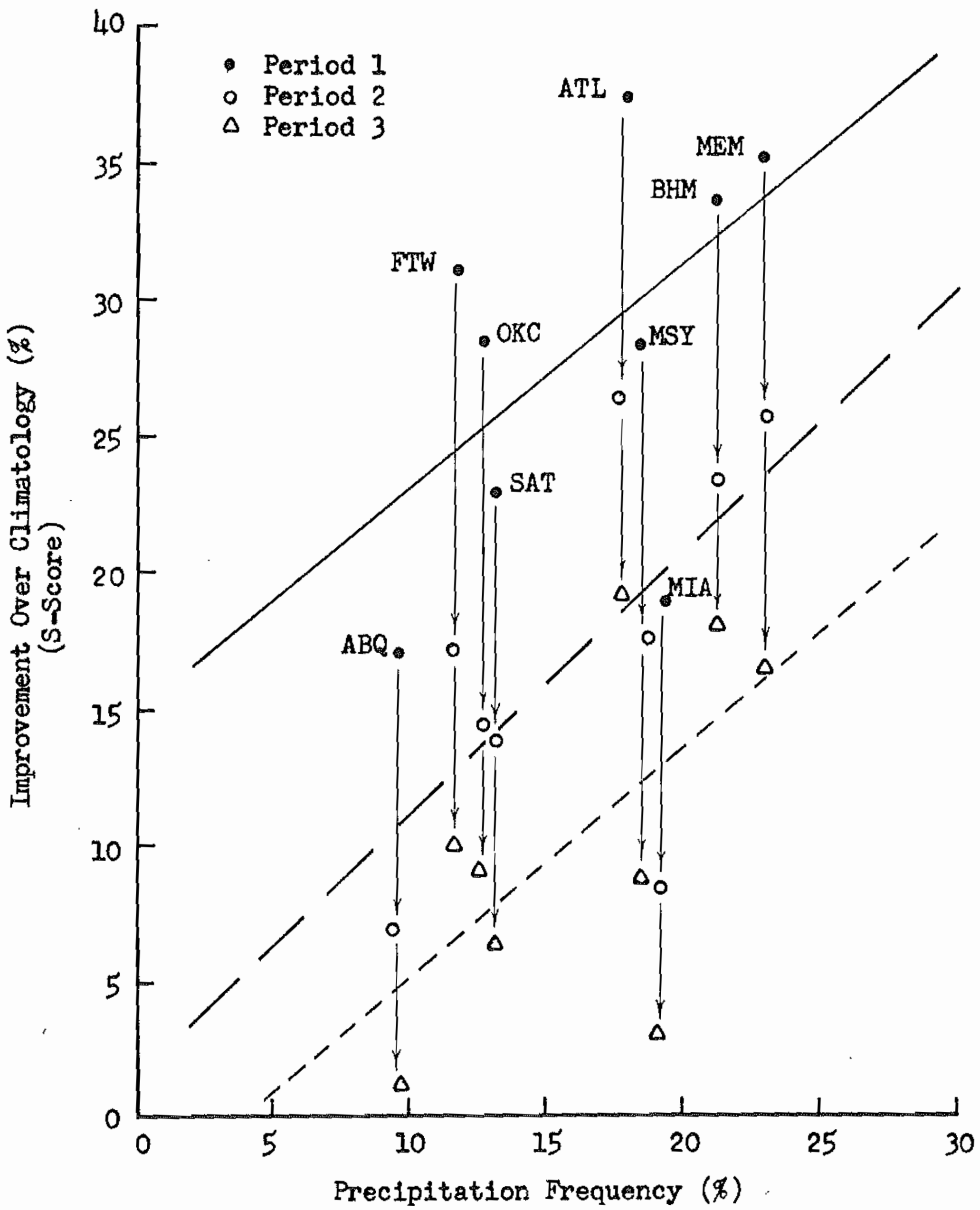


Figure 29. S-Score vs. precipitation frequency. Solid line is least-squares fit for Period 1, long dashes for Period 2, short dashes for Period 3. Plotted points are for all forecasts, whole year.

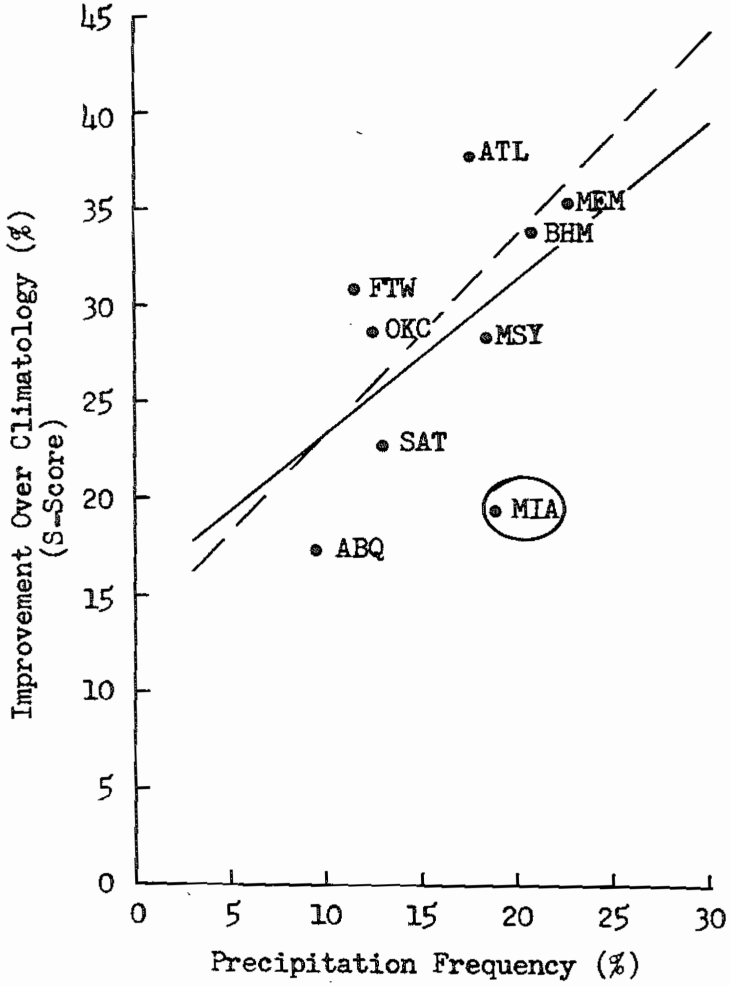


Figure 30. S-Score vs. precipitation frequency for Period 1. Solid line same as in Fig. 29; dashed line is least squares fit omitting MIA.

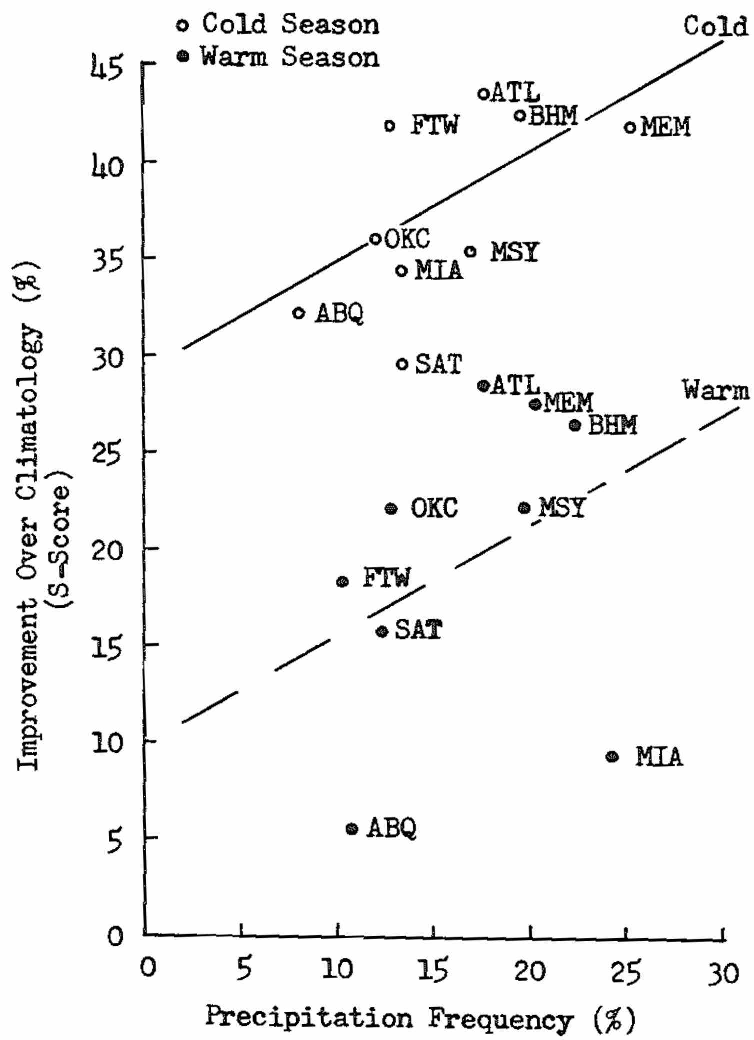


Figure 31. S-Score vs. precipitation frequency for Period 1 by seasons. Solid line is least squares fit for cold season, dashed line for warm season.