Reply

ALLAN H. MURPHY[1]

Department of Atmospheric Sciences, Oregon State University, Corvallis 97331

11 August 1980 and 27 October 1980

Glahn (1980) makes several points with respect to the author's paper (Murphy, 1978) and the general problem of evaluating probability of precipitation (PoP) forecasts. In some cases, these points also relate to issues discussed in subsequent comments by the author (Murphy, 1979, 1980) and in recent publications by Smith (1979a,b). However, the author is solely responsible for the content of this reply. Glahn's points will be considered in the order in which they were raised.

The author's original paper (Murphy, 1978) did *not* conclude that PoP forecasts "should be evaluated in terms of areal coverage rather than point by point with the area" (Glahn). Moreover, the author's subsequent comments on this topic (Murphy, 1979, 1980) have specifically recommended performing an evaluation in terms of areal coverage *in addition to* the traditional practice of evaluating PoP forecasts at a single point (i.e., the official raingage). For example, Murphy (1980) concludes ". . . it should be emphasized that this writer views these and many other new approaches to forecast evaluation as a means of obtaining valuable information to

[1] Supported in part by the National Science Foundation under Grant ATM80-24060.

0027-0644/81/051121-02$04.50

supplement the results of existing evaluation programs" (p. 231). Perhaps the original paper did not make this point sufficiently clear.

PoP forecasts are indeed point forecasts. Moreover, these forecasts generally can be interpreted to be average point probability forecasts for the relevant forecast area(s). This fact, together with the fact that an average point probability is mathematically equivalent to an (unconditional) expected areal coverage, implies that PoP forecasts are expected areal coverage forecasts as well as point forecasts. As a result, a PoP forecast can and should be verified in terms of the occurrence of precipitation at a point *and* in terms of the observed areal coverage of precipitation. In effect, the existence of two definitions of the forecast event (precipitation at a point and expected areal coverage of precipitation) leads to two different approaches to the evaluation problem. Which approach is considered to be the "basic" approach depends primarily on the specific purpose for which (or context within which) a verification program or study is undertaken. That is, verification of PoP forecasts as point forecasts may be more appropriate in one context, whereas verification of PoP forecasts as areal-coverage forecasts may be more appropriate in another context. For example, some considerations related to the presentation and use of forecasts, such as those mentioned by Glahn, seem to favor point evaluation of PoP forecasts. On the other hand, considerations related to the state-of-the-art of precipitation forecasting and to the existence of practical "limits" on the performance of forecasters (as well as to users concerned with the occurrence of precipitation over large areas) appear to favor an areal-coverage approach to the evaluation problem. In the author's opinion, the question of which approach (or corresponding set of measures) is considered to be basic has no simple or universal answer.

Glahn mentions the problem of the variation of precipitation probabilities over the forecast area. This problem has two dimensions, one of which relates to forecasting and the other of which relates to verification. The proper solution to the forecasting problem, as Glahn indicates, is to formulate and issue separate forecasts for those portions of the forecast area for which the probability of precipitation is appreciably different. Unfortunately, this option seldom is exercised in an operational context (in the author's experience). In this regard, the possible issuance of two or more PoP forecasts for a forecast area seems to warrant much greater emphasis in practice, particularly in those locations in which topographic or land-water differences significantly affect the likelihood of occurrence of precipitation over the forecast area or in which the official raingage is situated on the outskirts of a major metropolitan area. With regard to the (point) verification problem, Glahn correctly points out that the use of multiple gages can help to solve this problem in situations in which the point probability varies over the forecast area. Moreover, even when the forecast area has been divided into subareas (with a PoP forecast for each subarea), the author believes that it may be desirable for the purposes of verification to have more than one raingage in each subarea.

Glahn advocates the use of multiple gages in a forecast area in order to increase the effective size of the sample and thereby obtain more stable statistics in *point* verification. The author is in general agreement with Glahn's comments regarding the order of magnitude of the effect (on sample size) of increasing the number of raingages and the need for a network of raingages if stable statistics are required for relatively short time periods. However, the magnitude of the effect of networks of raingages on an overall composite Brier score (Brier, 1950) might be larger than Glahn suspects. In any case, this discussion is not directly relevant to the issue of primary concern here, namely point versus areal-coverage evaluation of PoP forecasts.

Some problems related to areal-coverage verification of PoP forecasts are raised by Glahn in his concluding paragraph, including the effect of the size of the area and the number of raingages on the magnitude of the terms in the author's partition of the Brier score. These issues were briefly addressed in the author's original paper (Murphy, 1978, pp. 1684–1685). While these problems warrant careful consideration, they hardly seem to present insurmountable obstacles to the formulation of a sound, effective program of areal-coverage evaluation of PoP forecasts. For example, an appropriate set of guidelines could be established to minimize the effects of unavoidable changes in the size of a forecast area or in the number of raingages used to define the area. Glahn also mentions that he believes that the use of the squared error term in the partition (instead of the complete Brier score) might tend to discourage forecasters from subdividing forecast areas even when they believed that the probabilities varied significantly over the area(s). Some back-of-the-envelope calculations (by Glahn and the author separately) reveal that this possibility exists, but its importance and impact in practice remain unclear at present. In any case, the routine computation of Brier scores presumably would indicate the presence of such an effect and thereby enable forecasters to take this factor into account in making decisions concerning the subdivision of the relevant forecast area(s).

In summary, the author does not advocate—and has not advocated—that point verification of PoP forecasts be replaced by areal-coverage evalua-

tion of such forecasts. The use of the two terms in the author's partition is recommended as a means of obtaining valuable information concerning the quality of PoP forecasts as areal-coverage forecasts (an equally valid interpretation of such forecasts) and concerning the variability of precipitation occurrence over the forecast area, respectively. Since areal-coverage verification of PoP forecasts is not feasible on an operational basis at the present time, it seems inappropriate to speculate about the potential usefulness of information concerning trends in squared error scores (or in skill scores derived therefrom). In any case, the fact that the Brier score and its associated skill score have been used as the basic verification measures since the initiation of the PoP forecasting program in 1965 should not serve as a deterrent to the introduction of new and potentially useful evaluation measures. In this regard, it is surprising and somewhat disappointing to note that few if any of the more recently developed measures are used in current operational and experimental verification programs. Some of these procedures certainly would provide useful supplemental information concerning the quality and value of the relevant forecasts.

REFERENCES

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

Glahn, H. R., 1980: Comments "On the evaluation of point precipitation probability forecasts in terms of areal coverage." *Mon. Wea. Rev.*, **108**, 1120–1121.

Murphy, A. H., 1978: On the evaluation of point precipitation probability forecasts in terms of areal coverage. *Mon. Wea. Rev.*, **106**, 1680–1686.

——, 1979: Reply. *Mon. Wea. Rev.*, **107**, 1225–1226.

——, 1980: Limits to forecast accuracy. *Bull. Amer. Meteor. Soc.*, **61**, 230–231.

Smith, D. L., 1979a: Comments "On the evaluation of point precipitation probability forecasts in terms of areal coverage." *Mon. Wea. Rev.*, **107**, 1224–1225.

——, 1979b: Eighty-five percent and holding—a limit to forecast accuracy? *Bull. Amer. Meteor. Soc.*, **60**, 788–790.