

## On the Combination of Forecast Probabilities for Consecutive Precipitation Periods

DANIEL S. WILKS

*Department of Soil, Crop, and Atmospheric Sciences, Cornell University, Ithaca, New York*

(Manuscript received 18 February 1990, in final form 10 August 1990)

### ABSTRACT

The performance of the method of Hughes and Sangster (1979) for combining precipitation probabilities pertaining to standard 12-h forecast periods is examined for 100 stations in the conterminous United States. Precipitation probabilities for both 24- and 36-h periods are investigated. Although originally derived for a small number of stations by using data from a limited time period, the original formulation is found to be remarkably robust. There is a tendency for overforecasting in the lower half of the probability range, which is most pronounced for the 36-h forecasts. A modification of the original procedure is suggested which largely corrects this problem.

### 1. Introduction

Probability of Precipitation (PoP) forecasts are routinely issued twice daily by the National Weather Service for a large number of locations within the United States. These forecasts are formulated both objectively (Carter et al. 1989) using the model output statistics (MOS) technique (Glahn and Lowry 1972), and subjectively, by human forecasters. The most familiar PoP forecasts are issued for three consecutive 12-h periods: 12–24 h, 24–36 h, and 36–48 h following either 0000 or 1200 UTC. In some cases PoP forecasts for 6-h periods are also formulated and issued, but forecasts pertaining to periods longer than 12 h are not commonly made.

For many decision problems, particularly agricultural ones, precipitation probabilities for 24- or 36-h periods could be of substantial value. An obvious example is that of irrigation scheduling (Allen and Lambert 1971). Other potential applications include scheduling of tillage, pest and plant disease management, and harvest operations. For such problems, it is often the case that wet conditions impede progress or render previous effort useless. Nonagricultural uses for longer-period PoP forecasts undoubtedly exist as well.

A method for combining routinely issued 12-h PoP forecasts to yield 24- and 36-h PoPs has been presented by Hughes and Sangster (1979, hereafter H-S). The method is comprised of elementary results from probability theory together with a statistical model reflecting the persistence of precipitation occurrence in consecutive forecast periods. The H-S method is simple and intuitively reasonable, but was derived by using only

subjective PoP forecast data from seven warm seasons (April–September) and seven cool seasons (October–March) at eight stations.

The characteristics of forecasts resulting from any new procedure are of interest, but no documentation of the performance of the H-S method has been presented to date. This paper examines the performance of both subjective and MOS PoP forecasts combined according to the H-S method for a large number of locations throughout the conterminous United States. The characteristics of these combined forecasts are compared to results obtained when independence of precipitation occurrence in consecutive 12-h periods is assumed. A simple modification of the original H-S formulation is also constructed and tested.

### 2. Forecast combination methods

A general rule for the combination of probabilities is given by the additive law of probability (e.g., Lindgren 1976). In the case of probabilities for two events, this can be written

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}. \quad (1)$$

Here, quantities in braces refer to the events of interest. In the present context, the events pertain to precipitation occurrences:  $\{A\}$  and  $\{B\}$  represent precipitation occurrence in each of two 12-h forecast periods, the intersection  $\{A \cap B\}$  represents the occurrence of precipitation in *both* of the forecast periods, and the union  $\{A \cup B\}$  represents precipitation occurrence in *either* of the two periods, *or both*. The probability on the left-hand side of (1) is thus the 24-h PoP desired.

Extension of the additive law to the combination of three 12-h PoPs into a single 36-h PoP is straightforward, and given by

*Corresponding author address:* Daniel S. Wilks, Dept. of Soil, Crop, and Atmospheric Sciences, Bradford and Emerson Halls, Cornell University, Ithaca, NY 14853.

$$\Pr\{A \cup B \cup C\} = \Pr\{A\} + \Pr\{B\} + \Pr\{C\} - \Pr\{A \cap B\} - \Pr\{B \cap C\} - \Pr\{A \cap C\} + \Pr\{A \cap B \cap C\}. \quad (2)$$

Here the event  $\{C\}$  represents precipitation occurrence in the third 12-h forecast period, and the intersection  $\{A \cap B \cap C\}$  represents the event that precipitation occurs in *all three* periods. The quantity on the left-hand side is the probability of precipitation in at least one of the three periods, i.e., the 36-h PoP.

The logical basis of (1) and (2) may be most easily understood with reference to the Venn diagram in Fig. 1. Each of the three circles represents one of the three events  $\{A\}$ ,  $\{B\}$ , and  $\{C\}$ . Regions in the figure contained in at least two circles represent the two-fold intersections  $\{A \cap B\}$ ,  $\{B \cap C\}$ , and  $\{A \cap C\}$ , i.e., precipitation occurring in both periods of each pair. The central region contained in all three circles represents precipitation occurring in all three periods, or the event  $\{A \cap B \cap C\}$ . Areas are in all cases proportional to the probabilities, so that for the case of (2), the desired 36-h PoP is proportional to the total shaded area. Geometrically, this area is given by the sum of the areas of the individual circles (each proportional to the individual 12-h PoPs), with appropriate corrections for double counting of regions contained in more than one of the three circles. Figure 1 would represent (1) if circle "C" were absent.

It should be clear that, given the individual probabilities  $\Pr\{A\}$ ,  $\Pr\{B\}$ , and  $\Pr\{C\}$ , formulation of 24- and 36-h PoPs follows directly from evaluation of the appropriate joint (pertaining to the intersections) probability or probabilities. For the case of independent events (i.e., occurrence or nonoccurrence of one event does not affect the probability of the other), this task is straightforward and direct: the joint probabilities are given by the product of the individual probabilities (e.g.,  $\Pr\{A \cap B\} = \Pr\{A\} \Pr\{B\}$ , and  $\Pr\{A \cap B \cap C\} = \Pr\{A\} \Pr\{B\} \Pr\{C\}$ ). However, occurrence of precipitation in consecutive 12-h periods is not, in

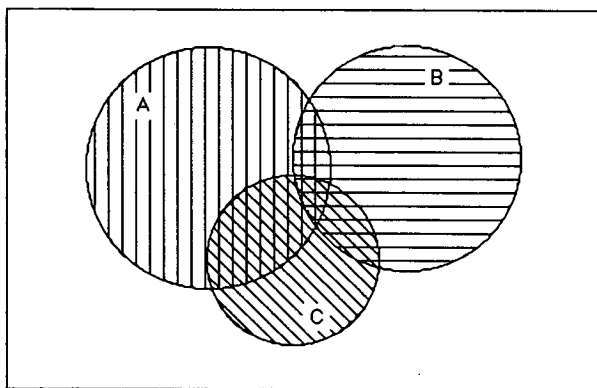


FIG. 1. Venn diagram illustrating, geometrically, the combination of three 12-h precipitation probabilities into a 36-h probability.

general, independent, and the observed positive dependence (persistence) must be accounted for in a good PoP combination procedure.

The dependence of the events can be approached through the multiplicative law of probability,

$$\Pr\{A \cap B\} = \Pr\{A | B\} \Pr\{B\}, \quad (3)$$

where  $\Pr\{A | B\}$  is the conditional probability of  $\{A\}$  given  $\{B\}$ , i.e., the probability of  $\{A\}$  considering only those occasions when  $\{B\}$  has occurred or will occur. Clearly, if the two events in (3) are independent, occurrence or non-occurrence of  $\{B\}$  does not affect the probability of  $\{A\}$ , in which case  $\Pr\{A | B\} = \Pr\{A\}$ . For events exhibiting persistence,  $\Pr\{A | B\} > \Pr\{A\}$ . Thus, using (1) with the assumption of independence of precipitation events will in general result in 24-h PoPs which are too large.

The H-S method uses the additive law of probability together with the following statistical model for the conditional probabilities:

$$\Pr\{A | B\} = [\Pr\{A\}]^k, \quad 0 \leq k \leq 1. \quad (4)$$

Using (3), this leads to

$$\Pr\{A \cap B\} = [\Pr\{A\}]^k \Pr\{B\}. \quad (5)$$

Here  $\Pr\{A\}$  is the larger of the two forecast probabilities, and the dependence parameter  $k$  is an empirical constant. Clearly for  $k = 1$ ,  $\Pr\{A | B\} = \Pr\{A\}$ , and (4) and (5) reduce to the case of independent events. As  $k$  becomes smaller,  $\Pr\{A | B\}$  becomes larger, reflecting stronger dependence between  $\{A\}$  and  $\{B\}$ . Completely dependent events would be characterized by  $k = 0$ . On the basis of data from eight stations in the north-central United States for the period October 1966 to September 1973, H-S take  $k$  to be 0.55 for the October–March cool season, and 0.70 for the April–September warm season. The final H-S model for combined precipitation probabilities is thus

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A\}^k \Pr\{B\}. \quad (6)$$

The H-S method exhibits several desirable qualities. First, defining the exponent  $k$  separately for the cool and warm seasons recognizes the greater tendency for dependence of precipitation events in the former. Also, the probability on the left-hand side of (5) is zero if either of the two individual probabilities are zero, and is equal to the smaller of the individual probabilities if the larger is unity. More generally, the procedure produces probabilities for the event  $\Pr\{A \cup B\}$  (6) which are at least as large as the larger of the two individual probabilities and no larger than their sum.

Using (6) will produce 24-h PoPs on the basis of two 12-h PoPs for consecutive forecast periods. For construction of 36-h PoPs, the H-S method does not use (2), but rather proceeds by two stagewise applications of (6):

$$\begin{aligned}
 \Pr\{A \cup B \cup C\} &= \Pr\{(A \cup B) \cup C\} \\
 &= \Pr\{A' \cup C\} \\
 &= \Pr\{A'\} + \Pr\{C\} - \Pr\{A' \cap C\}.
 \end{aligned}
 \tag{7}$$

Here, the probability of precipitation in at least one of the first two forecast periods,  $\Pr\{A'\} = \Pr\{A \cup B\}$ , is calculated first by using (6), after which (5) is applied to the two events  $\{A'\}$  and  $\{C\}$  to calculate the third term on the right-hand side of (7).

Initial examination of the performance of the H-S method, presented in more detail in the following section, indicated a consistent tendency for (6) to exhibit overforecasting in the probability range 0.1 to 0.4. That is, observed relative frequencies are consistently lower than forecast probabilities in cases when forecasts in this range are produced. In effect, the H-S dependence parameter,  $k$ , appears to be too large, but only when used with the smaller probabilities. Accordingly, a modification to the H-S method is also investigated. This is based on the modified dependence parameter

$$k^* = k(1 - \exp[-7 \Pr\{B\}]). \tag{8}$$

Again,  $k$  is the H-S dependence parameter and, as before,  $\Pr\{B\}$  is the smaller of the two forecast probabilities. The constant within the exponential was arrived at empirically, without a formal optimization. Instead, candidate integer values over a limited range were examined, with the intention of striking a balance between correcting the overforecasting of the lower probabilities while avoiding underforecasting of the larger ones. Thus, both the functional form and the specific value of the coefficient multiplying  $\Pr\{B\}$  in (8) have been chosen in an ad hoc manner, but with the hope that the equation will be general enough to

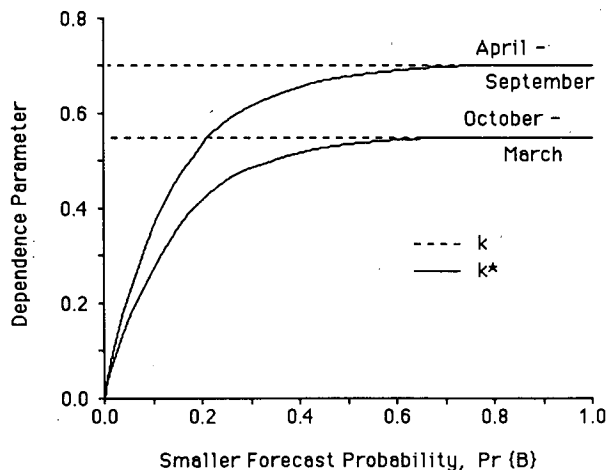


FIG. 2. Comparison of the original H-S dependence parameter,  $k$  (dashed lines), and the modified dependence parameter,  $k^*$ , defined by (8) (solid lines), as functions of the smaller of the two forecast probabilities,  $\Pr\{B\}$ .

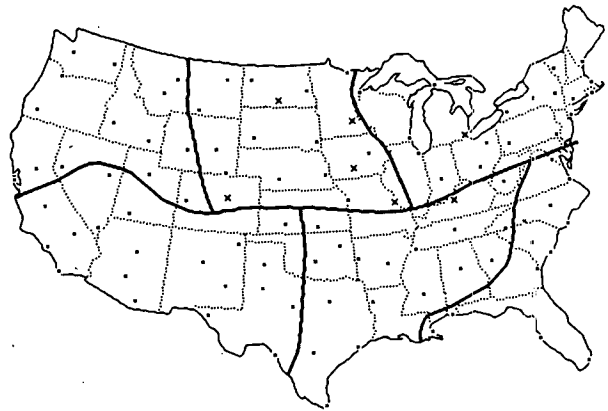


FIG. 3. Station locations for which verification data are available. Stations for which the original H-S formulation was derived are indicated by "x." Heavy lines separate geographic regions used for forecast stratification in section 3.

be applicable throughout the year at a wide variety of locations, as well as simple enough to use operationally.

The modified H-S procedure is implemented in exactly the same way as the H-S procedure, except that  $k^*$  (8) is used in (5) and (6) in place of  $k$ . Figure 2 illustrates the relationship between the H-S dependence parameter  $k$  and the modified H-S dependence parameter  $k^*$ . For values of the smaller forecast probability,  $\Pr\{B\}$ , greater than 0.5, the two formulations are essentially equivalent. As  $\Pr\{B\}$  decreases, the modified H-S formulation (8) effectively assumes a stronger degree of dependence between precipitation occurrence in consecutive forecast periods. Note that, while this apparent dependence becomes quite severe for  $\Pr\{B\} < 0.1$ , the combined forecasts are not sensitive to values of the dependence parameter in this range of  $\Pr\{B\}$ .

### 3. Verification results

Performance of 24- and 36-h PoP forecasts formulated according to both the original and modified H-S procedures are investigated in this section. For comparison, corresponding results produced by the assumption of independence are also presented. The data investigated are from the NWS Public Weather Verification Data Archive (VDA) (Carter and Polger 1986). This data set contains objective and subjective forecast probabilities for 12-h periods and corresponding observations for the 100 cities indicated in Fig. 3. The seven locations indicated by the symbol "x" (plus Topeka, Kansas, which is not part of the VDA) were also used in deriving the original H-S method. The heavy lines define somewhat arbitrary regional divisions, which will be used to investigate aspects of geographic variation in forecast performance. The period of record considered is January 1972 through March 1987. Data for both the 0000 and 1200 UTC cycles are combined.

Figure 4 presents reliability diagrams for the combined subjective forecasts at all 100 stations, stratified by cool (October–March) and warm (April–September) season. These diagrams show relative frequency of precipitation occurrence in the 24- or 36-h forecast periods as a function of the PoPs for those periods. Clearly, points falling close to the 45° line are to be desired. For clarity of presentation, forecasts have been rounded to the nearest 0.10. Together with the frequency of use of the forecasts (Table 1), this constitutes a graphical representation of the calibration-refinement

factorization of the data (Murphy and Winkler 1987), and as such presents a more complete representation of forecast quality than would a scalar score. Figures 4a and 4b contain results for both 24-h PoP forecast periods (the 12–36- and 24–48-h projections), with forecasts calculated using (6). Figures 4c and 4d present results for the 36-h period of the 12–48-h projection. In this latter pair of figures, the H–S and modified H–S forecasts are calculated using (7), while the forecasts assuming independence are calculated using (2).

The most striking feature of the panels in Fig. 4 is

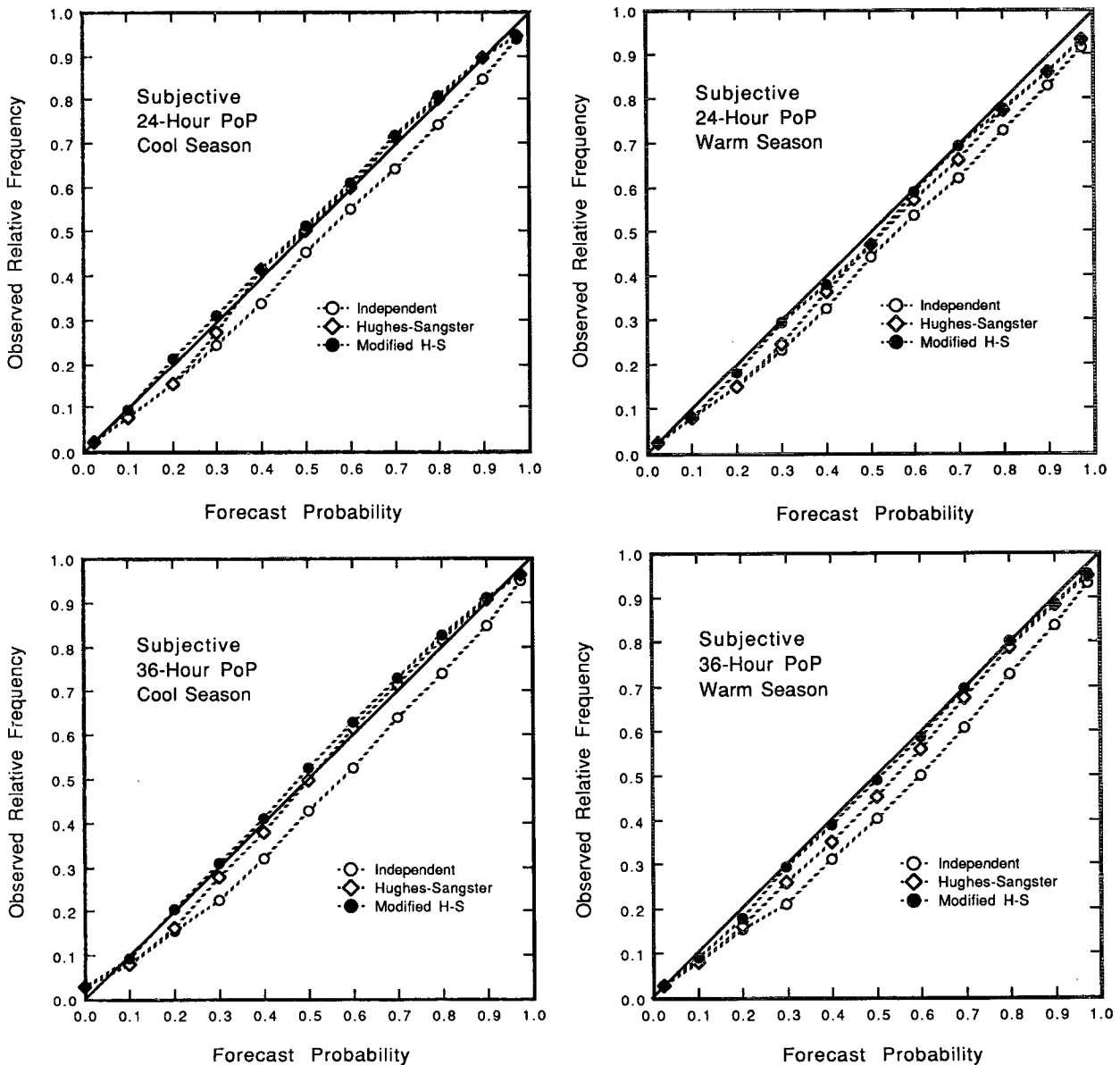


FIG. 4. Reliability diagrams (showing relative frequency of precipitation occurrence as a function of forecast probability) for 24- and 36-h forecasts based on 12-h subjective PoPs, aggregated over all stations shown in Fig. 3. Forecast probabilities have been rounded to the nearest 0.10. (a) 12–36- and 24–48-h forecasts for October–March, (b) 12–36- and 24–48-h forecasts for April–September, (c) 12–48-h forecasts for October–March, (d) 12–48-h forecasts for April–September.

TABLE 1. Sample sizes for reliability diagrams presented in Figs. 4 and 5 comparing matched samples of 24- and 36-h PoPs constructed from subjective and MOS 12-h PoPs under the assumption of independence, according to the H-S method, and using the modified H-S method. Also shown are overall forecast reliability (REL), Brier scores (BS), and skill in relation to the Brier scores of the sample climatological probabilities (SS<sub>sc</sub>).

Fcst.	Cool season						Warm season					
	24-h forecasts			36-h forecasts			24-h forecasts			36-h forecasts		
	Indep.	H-S	Mod. H-S	Indep.	H-S	Mod. H-S	Indep.	H-S	Mod. H-S	Indep.	H-S	Mod. H-S
Subjective (Fig. 4)												
0.00	194332	194332	194332	70870	70870	71576	146115	146115	146115	54845	54845	55706
0.10	147096	147096	206931	58562	59561	77789	122956	122956	122956	45907	45907	52975
0.20	86300	86300	71738	34544	47747	42857	88900	88900	149738	30759	49120	49603
0.30	56336	103078	57805	31809	36653	32299	67646	108752	74186	33240	32390	44152
0.40	72728	49406	49406	36523	32105	28950	109245	76923	50651	45412	40160	30062
0.50	39389	34157	43030	20735	25801	24054	53218	68962	68962	26775	32105	37255
0.60	39949	45047	36174	26087	25304	22777	46891	36093	50933	36066	37669	26757
0.70	34676	39236	43763	24546	22528	22992	37247	40378	31320	32882	27787	28804
0.80	43015	37188	32661	23159	25955	25202	36875	30135	24353	24951	24292	20532
0.90	41246	35208	35208	31573	27004	25260	23611	18636	18636	24986	17887	16570
1.00	36580	20599	20599	35018	19898	19670	12022	6876	6876	14345	8006	7752
Total	791647	791647	791647	393426	393426	393426	744726	744726	744726	370168	370168	370168
REL	.00160	.00053	.00034	.00281	.00067	.00050	.00260	.00117	.00046	.00461	.00163	.00046
BS	.1220	.1211	.1208	.1334	.1313	.1311	.1427	.1411	.1405	.1569	.1539	.1527
SS <sub>sc</sub>	40.9%	41.3%	41.4%	42.9%	43.8%	43.8%	28.9%	29.6%	29.9%	31.7%	33.0%	33.5%
MOS (Fig. 5)												
0.00	137657	137657	137657	42033	42033	43174	111401	111401	111401	35993	35993	37076
0.10	189536	189536	235240	67476	73386	90000	183544	183544	183544	59679	59679	72944
0.20	88153	88153	94019	44672	48246	50634	99123	99123	175866	46055	58854	56345
0.30	68185	115877	64307	34512	44963	37831	89351	129126	82845	38960	47585	62778
0.40	84552	60353	60353	41256	38849	35204	121517	93326	62864	56830	46149	34183
0.50	45091	44805	54944	26357	32845	27515	55533	72579	72579	31568	40862	43010
0.60	46982	52878	42739	30732	29834	28310	52446	38073	56196	40852	40576	27774
0.70	40321	41129	45869	28536	26226	26944	40668	42206	32069	34739	30415	31296
0.80	44087	36702	31962	26466	26800	25791	35419	30041	22055	26814	24969	21023
0.90	41672	33867	33867	32679	26322	24263	23027	16765	16765	24412	16690	15605
1.00	28880	14159	14159	30106	15321	15159	8285	4130	4130	11729	5859	5597
Total	815116	815116	815116	404825	404825	404825	820314	820314	820314	407631	407631	407631
REL	.00252	.00110	.00059	.00470	.00155	.00063	.00187	.00087	.00027	.00372	.00134	.00025
BS	.1270	.1258	.1251	.1407	.1375	.1366	.1456	.1445	.1439	.1607	.1583	.1573
SS <sub>sc</sub>	38.2%	38.9%	39.2%	39.6%	41.0%	41.4%	27.3%	27.8%	28.1%	30.0%	31.1%	31.5%

the overforecasting exhibited when independence of consecutive forecast periods is assumed (open circles), the data for which are consistently to the right of the diagonal line. The original H-S formulation (open diamonds) corrects this problem for the most part but, as mentioned in the previous section, exhibits a consistent tendency to overforecast in the probability range 0.1 to 0.4. This tendency is most pronounced for the 36-h PoPs and in the warm season, but is also evident to a small extent in the 24-h cool season PoPs. Overforecasting by the H-S method is nearly as strong as that when independence is assumed for the forecast range 0.1 to 0.2. The modification (8) to the H-S procedure (closed circles) improves reliability of the low probability values, although it slightly degrades reliability in the middle range, particularly for the 36-h

cool season PoPs. Similar results are seen for 24- and 36-h PoPs derived from the 12-h MOS PoPs (Fig. 5). Also presented in Table 1 is the summary measure of overall correspondence between forecast probability and observed relative frequency,

$$REL = \frac{1}{N} \sum_{s=1}^S N_s (f_s - \bar{d}_s)^2, \tag{9}$$

which has been referred to both as “reliability” (Murphy 1973) and “validity” (Sanders 1963). Here *S* is the number of possible forecasts (taken here to be 101 even though the reliability diagrams in Figs. 4 and 5 are plotted more coarsely), *f<sub>s</sub>* is the probability value of the *s*th of these forecasts (rounded to one of the

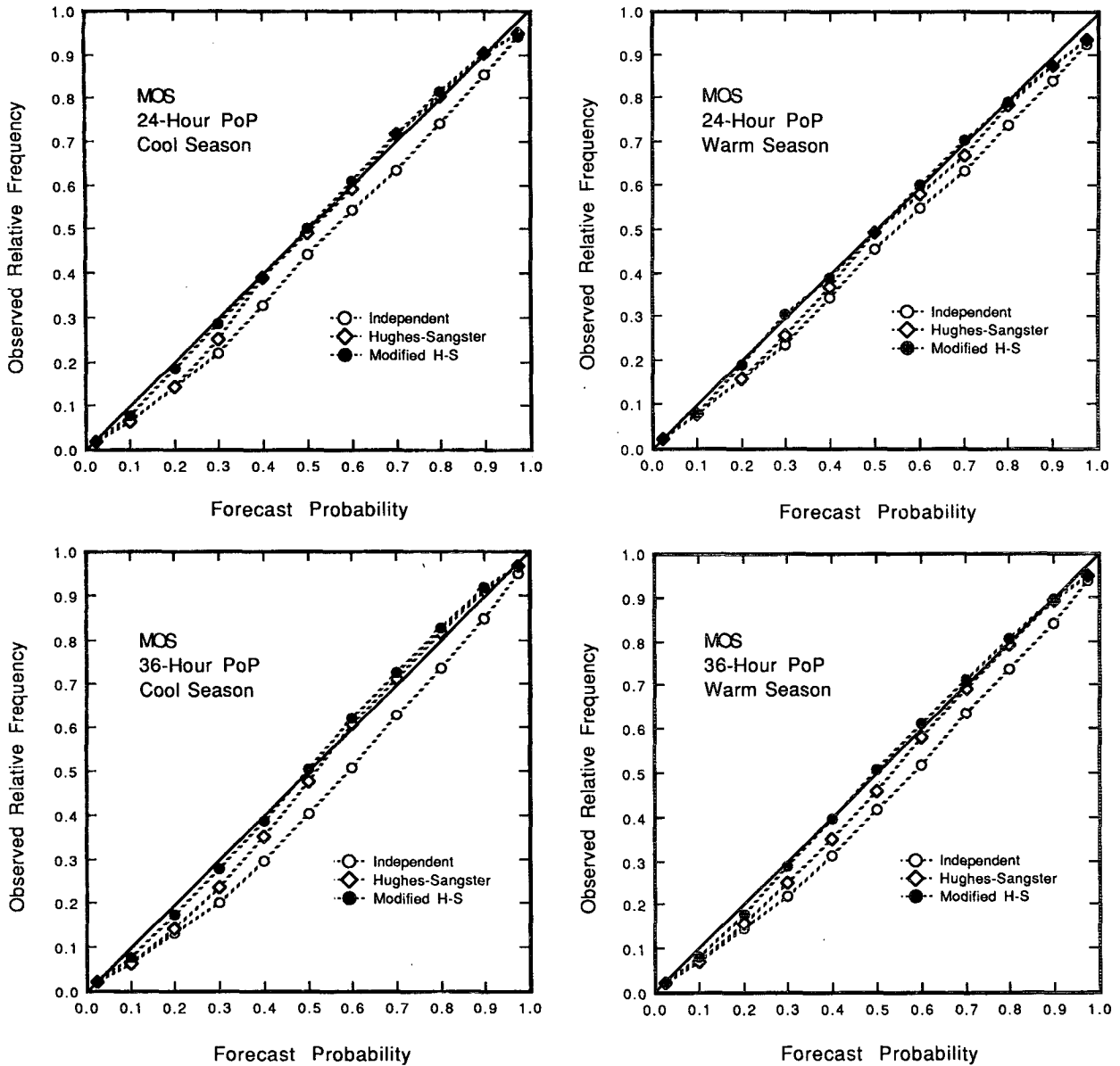


FIG. 5. As Fig. 4, but based on 12-h MOS PoPs.

values 0.00, 0.01, 0.02, . . . 0.99, 1.00),  $\bar{d}_s$  is the relative frequency of occurrence on the  $N_s$  occasions when  $f_s$  is forecast, and

$$N = \sum_{s=1}^S N_s \tag{10}$$

is the total sample size. The statistic (9) is thus nothing more than the weighted-average squared deviation of points on a reliability diagram from the 45° line representing perfect reliability, with better forecasts (other things equal) exhibiting smaller values. The REL values presented in Table 1 indicate that the modified H-S procedure consistently improves aggregate reliability

over the original H-S formulation, and that this improvement tends to be larger in the warm season, for both the subjective and MOS forecasts. Both the original and modified H-S procedures are substantial improvements over the assumption of independence.

Also presented in Table 1 are Brier scores (Brier 1950) for the combined forecasts [of which REL is a component (Murphy 1973)], and skill scores calculated with respect to Brier scores for the (sample) climatological probabilities. It can be seen that, in the course of improving the reliability of the combined PoPs, the modified H-S procedure also modestly improves overall skill over the original H-S procedure. Both methods exhibit somewhat higher overall skill

than do forecasts based on the assumption of independence.

The performance of the underlying original 12-h PoPs is of interest in relation to the H-S formulation and the modification suggested here. This performance is summarized graphically in Fig. 6, which shows reliability diagrams for subjective (Fig. 6a) and MOS-generated (Fig. 6b) 12-h PoPs, stratified by 6-month seasons. Here, data for both forecast cycles, all three projections, and all stations in the VDA have been combined. The most marked deviations from perfect

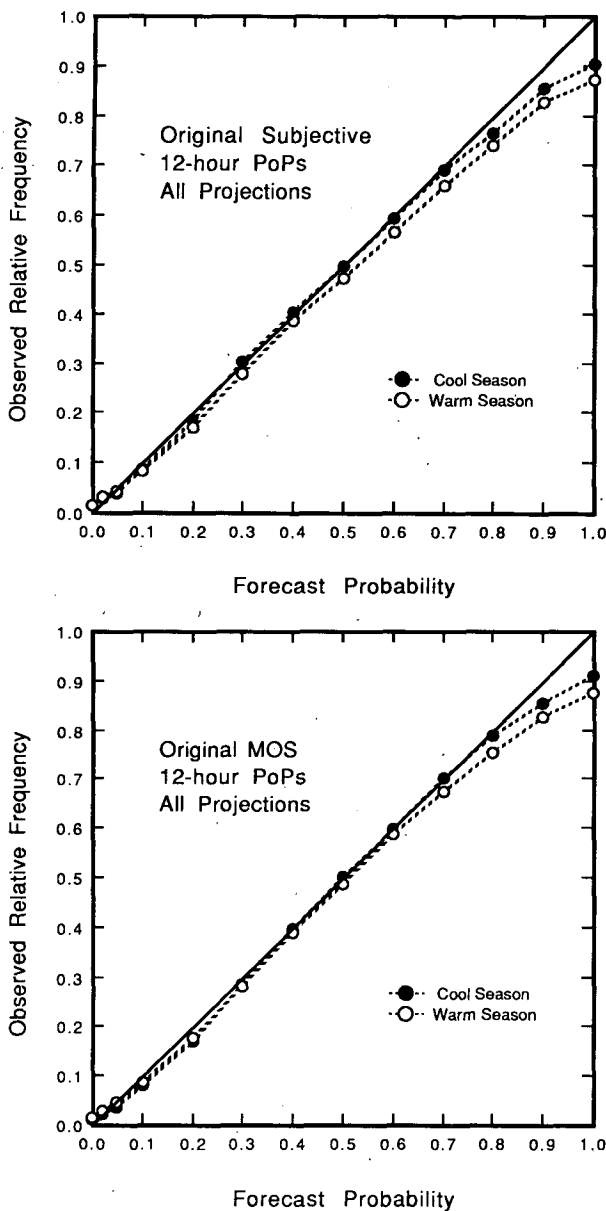


FIG. 6. Reliability diagrams for the underlying 12-h PoPs, with all stations and projections combined. Solid symbols, cool season (October–March); open symbols, warm season (April–September). (a) Subjectively formulated. (b) MOS-generated.

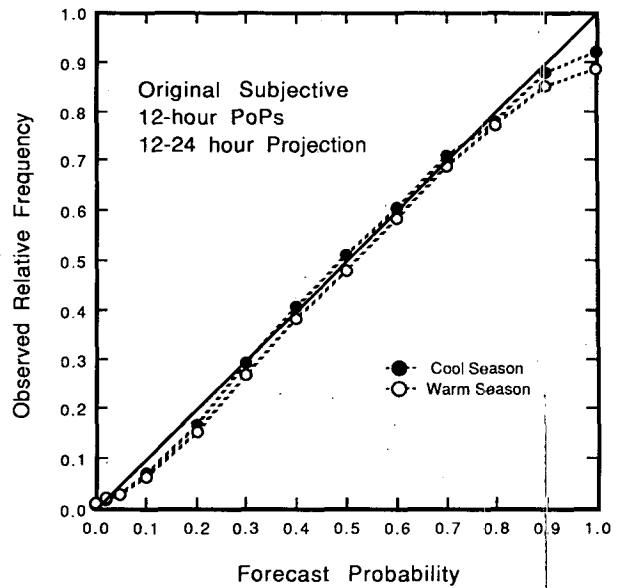


FIG. 7. As Fig. 6, for subjectively formulated 12–24-h PoPs only.

reliability in both panels are for the higher forecast probabilities, and for the warm season forecasts. The general pattern of Fig. 6 is representative of results for all three projections, except that a tendency exists for overforecasting in the 0.05 to 0.20 probability range in the 12–24-h projection. This is illustrated for the case of the subjective PoPs in Fig. 7. Corresponding results (not shown) for the most recent seven years of the data set (April 1980–March 1987) are not substantially different from Figs. 6 and 7 with respect to the reliability.

Table 2 contains sample sizes and summary REL values for the data plotted in Figs. 6 and 7. Higher REL values (inferior reliability) for the 12–24-h projections, as compared to those for all projections combined, apparently result from the overforecasting in the low probabilities mentioned above. However, the 12–24-h projection forecasts still exhibit higher skill (e.g., Carter and Polger 1986; Murphy and Sabin 1986). This comes as a consequence of the higher relative frequency of use of the zero forecast and of the largest probability values, evident in Table 2, so that the “resolution” component of the Brier score (Murphy 1973) more than compensates for the somewhat poorer reliability. Conversely, the slightly better REL score for the MOS forecasts in the warm season as compared to the cool season results from better reliability for the subsamples of forecasts concentrated in the 5% to 20% range, near the value of the overall event relative frequency. This characteristic also carries over to the combined PoPs (Table 1). The fairly major deviations from perfect reliability, shown in Figs. 6 and 7 for the largest probability values, degrade the overall REL in only a minor way, since relatively few of these forecasts are issued even for the 12–24-h projection.

TABLE 2. Sample sizes for reliability diagrams for the underlying original 12-h PoPs presented in Figs. 6 (subjective and MOS PoPs for all projections) and 7 (subjective PoPs for the 12–24-h projection). Also shown is the summary measure, REL, of overall forecast reliability.

Fcst.	Subjective, all projections		MOS, all projections		Subjective, 12–24-h projection	
	Cool season	Warm season	Cool season	Warm season	Cool season	Warm season
0.00	375137	276744	249467	192119	147215	107187
0.02	13136	12396	84964	81630	4348	4145
0.05	95426	84692	136033	135894	28788	25463
0.10	234840	238300	226931	262692	66556	68754
0.20	134949	182573	161154	218788	40265	56603
0.30	90155	126309	104605	134570	26819	40558
0.40	57967	70398	76029	84189	16520	22505
0.50	47729	49666	58658	55070	13856	16789
0.60	46806	37294	46112	34485	14461	13663
0.70	34166	21023	35987	20111	10802	8278
0.80	31310	14151	25301	11140	11964	6642
0.90	18249	5874	16957	5092	8394	3247
1.00	16775	4724	10718	2428	11190	3364
Total	1196645	1124144	1232916	1238208	401178	377198
REL	.00034	.00058	.00038	.00031	.00055	.00096

The aggregated reliability data for the original 12-h PoPs shown in Fig. 6 are also reasonably representative of the corresponding regional data. Table 3 presents the summary REL values for the underlying 12-h PoP forecasts, stratified by projection and the six geographic regions shown in Fig. 3. For perspective, Fig. 8 presents reliability diagrams for the original 12-h subjective PoPs for stations in the southeastern region. Table 3 indicates this to be the worst case for the warm season, and one of the worst cases for the cool season. Figure 8a shows data for all three projections pooled, and should be compared to Fig. 6a. Figure 8b shows the 12–24-h projection only, and should be compared to Fig. 7.

To some extent, the H–S dependence factor  $k$ , and (by extension) the modified dependence factor  $k^*$ , act to compensate for the overforecasting exhibited by the original 12-h PoPs. Similarly, some of the improvement resulting from the use of the modified H–S procedure is attributable to stronger compensation for the overforecasting of the smaller probabilities in the 12–24-h projection evident in Figs. 7 and 8b. An idea of the extent to which this is the case can be obtained by examining the performance of the three PoP combination procedures when the underlying 12-h PoP forecasts have been artificially recalibrated, and therefore exhibit perfect reliability. An example reliability dia-

TABLE 3. Overall reliability, REL, for the original underlying 12-h PoPs, stratified by projection and the geographic regions shown in Fig. 3.

Region	Cool season				Warm season			
	12–24	24–36	36–48	All	12–24	24–36	36–48	All
Subjective								
NW	.00061	.00043	.00056	.00032	.00087	.00055	.00080	.00056
NC	.00039	.00036	.00039	.00027	.00052	.00040	.00053	.00028
NE	.00126	.00077	.00057	.00072	.00106	.00073	.00071	.00063
SW	.00029	.00019	.00045	.00021	.00101	.00083	.00101	.00083
SC	.00065	.00048	.00069	.00046	.00136	.00091	.00080	.00086
SE	.00089	.00064	.00054	.00058	.00201	.00115	.00107	.00122
All	.00055	.00037	.00040	.00034	.00096	.00063	.00056	.00058
MOS								
NW	.00081	.00089	.00067	.00065	.00048	.00037	.00084	.00045
NC	.00044	.00035	.00026	.00030	.00043	.00021	.00024	.00020
NE	.00068	.00073	.00058	.00061	.00062	.00058	.00046	.00048
SW	.00046	.00032	.00025	.00027	.00049	.00024	.00033	.00031
SC	.00077	.00045	.00050	.00051	.00056	.00043	.00063	.00045
SE	.00068	.00093	.00074	.00069	.00069	.00061	.00056	.00057
All	.00051	.00040	.00038	.00038	.00042	.00030	.00036	.00031



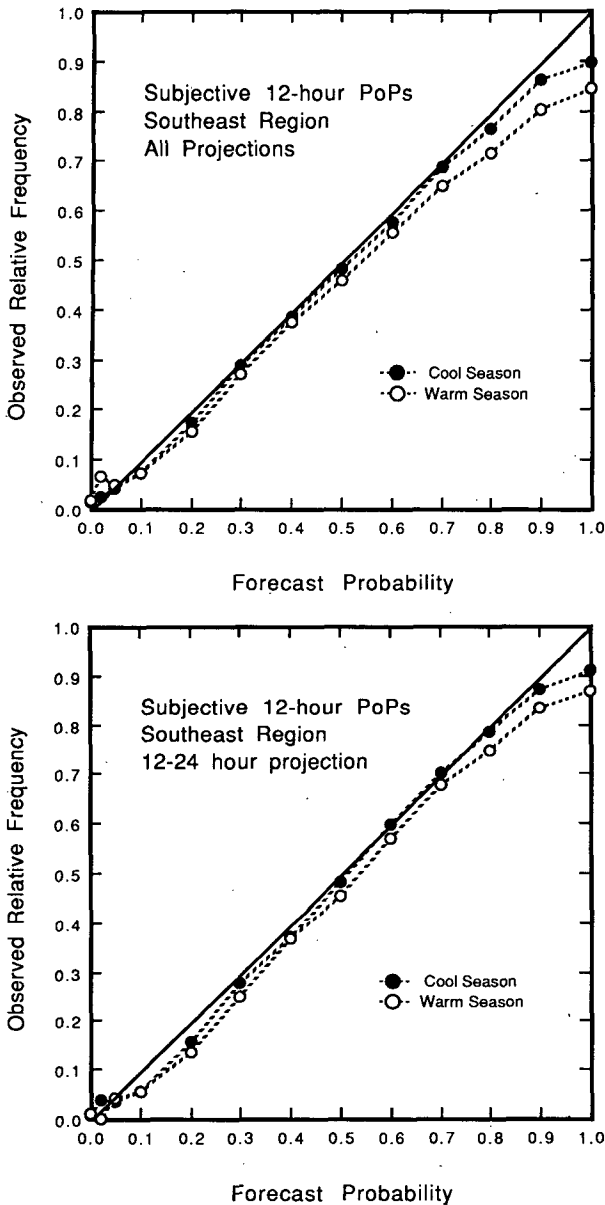


FIG. 8. Reliability diagrams for the underlying 12-h PoPs, for the stations in the southeastern region of Fig. 3 only. (a) All projections combined. (b) 12-24-h projection only.

gram for such combined forecasts, for the warm season 36-h PoPs, is shown in Fig. 9. (The corresponding figure for the uncalibrated forecasts is Fig. 4d.) Here, the three PoP combination procedures have been applied exactly as before, but using actual observed relative frequencies tabulated separately for each 12-h forecast value and warm-season projection. That is, the ordinates in Fig. 7 for the 12-24 projection, and the corresponding observed relative frequencies for the 24-36- and 36-48-h projections, have been used in place of the actual forecasts to construct the 36-h PoPs.

Both the H-S and the modified H-S method exhibit

underforecasting (points to the left of the diagonal) for the larger probabilities, and the extent to which this occurs is a measure of the compensation for the overforecasting exhibited by the underlying 12-h PoPs. The H-S procedure overforecasts the smaller combined PoPs even when the 12-h PoPs are perfectly calibrated; while the modified H-S procedure continues to underforecast a bit. If the real 12-h PoPs were perfectly reliable, the appropriate H-S and modified H-S dependence factors would be larger. This would effectively result in shifting points in Fig. 9 to the right, so that the H-S overforecasting of the smaller probabilities would be more pronounced, and the modification (8) would still constitute an improvement. The source of this characteristic of the forecasts is not clear, although it is as if periods of enhanced persistence of precipitation occurrence are identified by consecutive 12-h periods for which low PoPs are issued.

Finally, it is of interest to investigate the applicability and performance of the forecast combination schemes with finer time and geographic resolution. Rather than present a large number of reliability diagrams, this more detailed forecast performance is summarized using the REL statistic (9). Figure 10 presents REL values describing performance of combined PoP forecasts based on subjective 12-h PoPs assuming independence, and using the original and modified H-S formulations. Here, the verification data are stratified in space using the geographic regions shown in Fig. 3, and stratified in time by calendar month. Figure 10a shows results for the 12-36-h projection, and Fig. 10b shows results for the 12-48-h projection. Note that, for compactness, the bars in Fig. 10 are drawn as if standing behind one

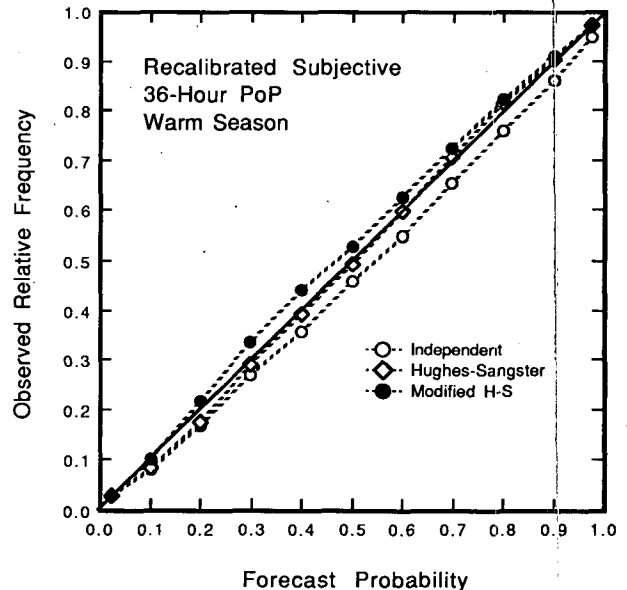


FIG. 9. Reliability diagram for the three PoP combination procedures operating on artificially recalibrated subjective 12-h PoPs (which exhibit perfect reliability) to produce warm season 36-h PoPs.

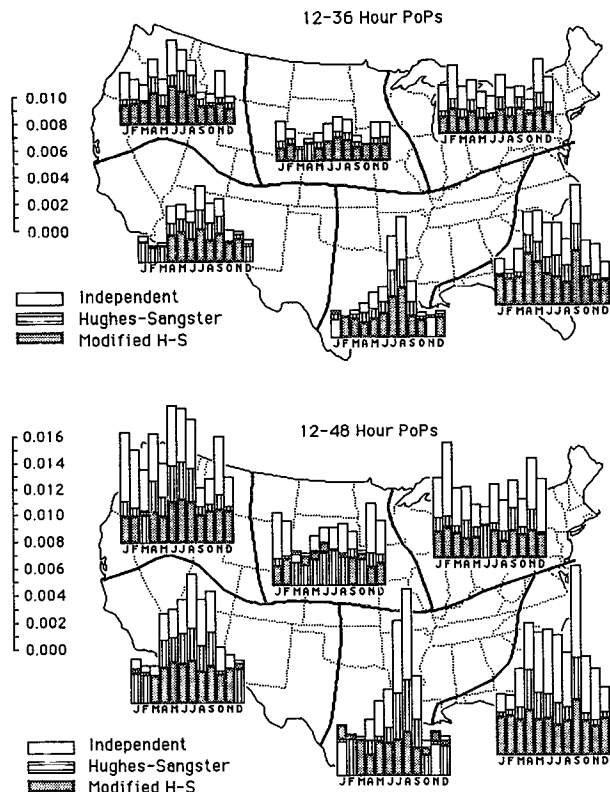


FIG. 10. Overall reliability, REL, of (a) 24- and (b) 36-h forecasts derived from subjective 12-h PoPs stratified by month and geographic region. Bars are drawn as if standing behind one another with the smallest in front and the largest behind.

another with the smallest in front and the largest behind. Thus, for example, in Fig. 10a, the REL value for the modified H-S forecasts in January for the northwest region is slightly smaller than for the original H-S formulation, which, in turn, is about half that for forecasts consistent with the assumption of independence.

Again, the largest REL values in Fig. 10 are exhibited by the forecasts generated by assuming independence, reflecting the substantial overforecasting this method generally produces. It can be seen that the original H-S formulation performs best for locations in the north-central region (where most of the stations used in its derivation are located) but that it is remarkably robust with respect to both a finer time resolution and a broader spatial application than those for which it was originally developed. The modified H-S procedure improves on the original in most cases, and performs nearly as well as the original in the remainder. The differences are most striking for the 36-h PoPs, shown in Fig. 10b. Again, examination of reliability diagrams for the more finely stratified data (not shown) indicates that the overforecasting of the original H-S formulation in the lower probabilities is fairly general, and the improvements effected by the modification (8) result pri-

marily from amelioration of this problem. Similar results (also not shown) are obtained when using the MOS forecasts.

4. Summary and conclusions

The formulation of Hughes and Sangster (1979) has been investigated for a longer period of record, a wider geographic domain, and a finer time stratification than that for which it was originally derived. It is seen to produce 24- and 36-h PoP forecasts which improve substantially on those formulated in a manner consistent with the assumption of independence of precipitation in consecutive 12-h periods. As expected, these latter forecasts consistently produce overestimates of precipitation probabilities through failure to account for persistence.

The H-S formulation is seen to be remarkably robust, but exhibits a general tendency to overforecast the smaller probabilities. The poorest performance in this regard is for the forecast range 0.1 to 0.4. A simple correction for this problem is suggested, which effectively amounts to an assumption of stronger dependence of precipitation occurrence for consecutive 12-h periods on occasions when low PoPs are issued. This modification of the H-S procedure is seen to produce forecasts exhibiting good reliability throughout the range of possible forecasts, and the improvement is most marked for the warm season. It may be that the modified H-S method constitutes an improvement because periods of enhanced persistence are identified by consecutive low 12-h PoPs, although determination of whether regimes of greater or lesser dependence of 12-h precipitation occurrence are actually discriminated by the PoP forecasts remains for future investigation.

The suggested modification to the H-S method has been deliberately kept simple in order to enhance ease of application. Better fits to verification data could undoubtedly be achieved if both the H-S dependence factor and the empirical constant in (8) were allowed to vary seasonally and geographically. Finally, it should be emphasized that the method presented here was developed for use specifically with the 12-h PoP forecasts and is thus "tuned" to the statistical characteristics of those forecasts. It should not be used with other types of probability forecasts without previous demonstration of its validity.

PoP forecasts for periods other than 0000-1200 and 1200-0000 UTC are useful for a variety of decision problems, many of them in agriculture. The procedure presented here will enable the operational provision of reliable 24- and 36-h PoPs by using routinely generated forecast products. Furthermore, it should be possible to use the procedure together with a simple scheme to generate probability forecasts of precipitation amount, involving PoPs and climatological distributions of precipitation amount (Wilks 1990). In this case, the com-

bined PoPs would be used with climatological probabilities for precipitation in selected amount classes over 24- or 36-h periods. The resulting 24- or 36-h probability forecasts for precipitation amount would be useful in the construction of formal decision making models (e.g., Murphy et al. 1985) for realistic problems sensitive to precipitation amounts over periods longer than 12-h.

*Acknowledgments.* The verification data set was kindly provided by Paul Polger and Gary Carter. This research was supported by the USDA under regional project NE-132.

#### REFERENCES

- Allen, W. H., and J. R. Lambert, 1971: Application of the principle of calculated risk to scheduling of supplemental irrigation. II. Use on flue-cured tobacco. *Agric. Meteor.*, **8**, 325–340.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3.
- Carter, G. M., and P. D. Polger, 1986: A 20-year summary of National Weather Service verification results for temperature and precipitation. NOAA Tech. Memo NWS FCST 31, U.S. Dept. Of Commerce, 50 pp. [NTIS PB88 235353/AS.]
- , J. P. Dallavalle and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting*, **4**, 401–412.
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.
- Hughes, L. A., and W. E. Sangster, 1979: Combining precipitation probabilities. *Mon. Wea. Rev.*, **107**, 521–524.
- Lindgren, B. W., 1976: *Statistical Theory*. Macmillan Publishing Co., 614 pp.
- Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595–600.
- , R. W. Katz, R. L. Winkler and W.-R. Hsu, 1985: Repetitive decision-making and the value of forecasts in the cost-loss ratio situation: A dynamic model. *Mon. Wea. Rev.*, **113**, 801–813.
- , and T. E. Sabin, 1986: Trends in the quality of National Weather Service Forecasts. *Wea. Forecasting*, **1**, 42–55.
- , and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330–1338.
- Sanders, F., 1963: On subjective probability forecasting. *J. Appl. Meteor.*, **2**, 191–201.
- Wilks, D. S., 1990: Probabilistic quantitative precipitation forecasts derived from PoPs and conditional precipitation amount climatologies. *Mon. Wea. Rev.*, **118**, 874–882.