Objective and Subjective Precipitation Probability Forecasts: Statistical Analysis of Some Interrelationships

ROBERT T. CLEMEN

College of Business Administration, University of Oregon, Eugene, OR 97403

ALLAN H. MURPHY

Department of Atmospheric Sciences, Oregon State University, Corvallis, OR 97331

(Manuscript received 13 December 1985, in final form 3 March 1986)

ABSTRACT

This paper addresses two specific questions related to the interrelationships between objective and subjective probability of precipitation (PoP) forecasts: Do the subjective forecasts contain information not included in the objective forecasts? Do the subjective forecasts make full use of the objective forecasts? With respect to the first question, an analysis of more than 11 years of data indicates that the subjective PoP forecasts add information above and beyond that contained in the objective PoP forecasts for all combinations of geographical area, lead time, and season investigated in this study. For longer lead times, this conclusion appears to contradict the results of earlier studies in which the two types of PoP forecasts were compared using aggregate skill scores. With regard to the second question, the statistical results demonstrate that the subjective forecasts generally do not make full use of the objective forecasts. However, these latter results are not as strong, in a statistical sense, as the results related to the first question; moreover, they indicate that it is primarily in the vicinity of the climatological probability (i.e., 0.10 to 0.40) that better use could be made of the objective forecasts. This conclusion suggests that it may be possible to combine the objective and subjective forecasts to produce a PoP forecast with even greater information content.

1. Introduction

Forecasts of the probability of measurable precipitation have been formulated on a nationwide basis by National Weather Service (NWS) forecasters since 1965. These subjective probability of precipitation (PoP) forecasts are generally quite reliable and skillful (e.g., see Murphy, 1985), and they are now an important and integral part of the weather forecasts routinely issued to the general public in the United States. Since 1972, NWS forecasters have received objective PoP forecasts as guidance in the preparation of their subjective forecasts. These numerical-statistical forecasts are produced by the model output statistics (MOS) system (Glahn and Lowry, 1972), and they are available for the same locations and lead times for which the subjective PoP forecasts must be prepared. Traditional measures of the performance of such forecasts (e.g., skill scores) indicate that the quality of the objective forecasts is now nearly comparable to that of the subjective forecasts (e.g., see Murphy, 1985; Murphy and Sabin, 1986). As a result, NWS forecasters have been strongly encouraged in recent years to give careful consideration to the objective forecasts when formulating their subjective forecasts.

In view of the general interest in the respective roles of man and machine (i.e., models) in weather forecasting, it should not be surprising that the relative performance and information content of the objective and subjective PoP forecasts have generated considerable debate and controversy in the operational meteorological community (e.g., see Murphy and Brown, 1984; Snellman, 1977). For example, some individuals evidently believe that the relatively high quality of the subjective forecasts is due largely to the reliability and skill of the objective forecasts and that NWS forecasters are no longer able to improve consistently upon the guidance forecasts. Other individuals maintain that the forecasters can still improve upon the guidance given adequate motivation and time and that undue reliance on a single source of information such as the objective PoP forecasts can adversely affect the quality of the official subjective forecasts. The fact that very little evidence is currently available to support these or other positions on such important issues is indicative of the need to undertake in-depth studies of the interrelationships between these two types of PoP forecasts. In this regard, two basic questions suggest themselves: Do the subjective PoP forecasts contain information (regarding the occurrence of precipitation) not included in the objective PoP forecasts? Do the subjective forecasts make full use of the information contained in the objective PoP forecasts? The primary purpose of this paper is to use a methodological framework recently developed by Clemen (1986) to answer these two questions.

In section 2 we briefly describe the data—a sample of 11½ years of objective and subjective PoP forecasts—upon which this study is based. The approach and methodological framework used to investigate the interrelationships between the two types of forecasts are outlined in section 3. Sections 4 and 5 contain the principal results of the study, with the former concerned with whether or not the subjective forecasts contain information not included in the objective forecasts and the latter concerned with the extent to which the subjective forecasts make full use of the objective forecasts. Section 6 consists of a discussion of the results of this study, and section 7 contains a short summary and conclusion.

2. Data

The data analyzed in this paper consist of objective and subjective PoP forecasts for 16 NWS offices in various sections of the United States. These data, covering the period from April 1972 through September 1983, were provided by the NWS Techniques Development Laboratory. In order to obtain sample sizes large enough to permit a reasonable statistical analysis, the data for the 16 offices were aggregated into four "areas" consisting of four offices in each area. The areas and offices that comprise each area are identified in Table 1, together with the climatological probabilities of measurable precipitation in the respective areas for the warm (April–September) and cool (October–March) seasons. Since forecasts from individual offices are generally made by many different forecasters, the aggregation of the data over areas with relatively homogeneous meteorological and climatological regimes would not appear to compromise the results of this study in any material way.

TABLE 1. NWS offices and areas defined in terms of those offices for which objective and subjective PoP forecasts were analyzed in this study. Also shown are the overall sample climatological probabilities of measurable precipitation in the cool and warm seasons in the respective areas.

Area	NWS office	Climatological probabilities: Cool season	Climatological probabilities: Warm season
Northeast (NE)	Albany, NY Boston, MA New York, NY Philadelphia, PA	0.22	0.22
Southeast (SE)	Asheville, NC Atlanta, GA Birmingham, AL Columbia, SC	0.20	0.23
Southern Plains (SP)	Amarillo, TX Dallas/Fort Worth, TX Oklahoma City, OK Wichita, KS	0.11	0.13
Rocky Mountain (RM)	Boise, ID Denver, CO Great Falls, MT Salt Lake City, UT	0.22	0.17

Objective and subjective PoP forecasts are made twice each day, in conjunction with the so-called 0000 and 1200 GMT cycle times. Moreover, for each cycle time, forecasts are formulated for three consecutive 12-h periods, with valid times corresponding to 12–24 h, 24–36 h and 36–48 h. In addition, these data traditionally are separated into warm and cool seasons (as defined above). Thus, for each area and type of forecast (objective/subjective), we analyzed 12 different combinations of results involving two cycle times, three lead times, and two seasons.

3. Approach and methodology

Objective and subjective PoP forecasts can be viewed as two sources of information regarding the likelihood of occurrence of measurable precipitation in the period of interest. Of course, these two types of forecasts are based on overlapping sets of information. Moreover, the forecasters usually consult the objective PoP forecasts in the process of formulating their subjective PoP forecasts. If we adopt the perspective of a potential user of these forecasts, it seems reasonable to ask the following question: Is it necessary to consult both types of forecasts? That is, do the objective (subjective) forecasts contain all of the information included in the subjective (objective) forecasts? For the purposes of the present paper, this question can be separated into two parts: Do the subjective PoP forecasts contain information not included in the objective PoP forecasts? Do the former make full use of the information contained in the latter? Clemen (1986) has formulated a methodological framework for answering such questions (given the relevant data), and we briefly describe and illustrate the use of this methodology here.

The methodological framework developed by Clemen is based upon the evaluation of the individual and joint calibration functions of the various experts (for convenience, we will sometimes refer to both the subjective forecasters and the MOS system as experts or forecasters). In this context, the term "calibration function" refers to the probability of measurable precipitation given the forecast probability; thus, the calibration function consists of conditional probabilities. In the meteorological literature, this function has generally been referred to as the reliability curve (or function), and it has frequently been displayed in the form of a reliability diagram (e.g., see Murphy, 1985). Individual calibration functions relate to individual experts, whereas joint calibration functions involve two or more experts. For example, a joint calibration function involving both types of PoP forecasts would indicate the probability of measurable precipitation given the objective *and* subjective forecast probabilities.

For illustrative purposes, we present a summary of the data for the Southeast (SE) area, warm season, 0000 cycle time, and 12–24 h lead time in Table 2. The table contains estimates of both individual and joint calibration functions as well as frequencies of the various individual and joint forecasts. Even though a total of 6411 cases exists, the majority occurred when the objective and subjective forecasts were close to each other, resulting in many cells far from the main diagonal having few (if any) cases. It is worth noting that Table 2 permits an almost complete reconstruction of the data for the specified combination of area, season, cycle time, and lead time. The only information not recoverable is (i) the actual sequence of forecast events and (ii) the identification of forecasts with particular offices. Similar tables are available on request from the authors for all 48 combinations studied in this paper.

The individual calibration functions (marginal entries in Table 2) indicate that, in general, both objective and subjective forecasters tend to overestimate the probability of precipitation. This result is consistent with previous calibration studies (e.g., Murphy and Winkler, 1977). To use the joint calibration figures, we can read from Table 2, for example, that precipitation occurred on 37.3% of the 150 12 h periods when the objective forecast was 0.30 and the subjective forecast was 0.40. For those combinations of forecast probabilities involving few historical cases, this joint calibration function has virtually no meaning. However, when many cases are available, it may be useful to consider the forecasters' joint performance. Indeed, the joint calibration functions would appear to indicate that both of the forecasts add information; for any given forecast value from either forecaster, the joint calibration function generally increases with increasing forecast values from the other forecaster. Figures 1 and 2 demonstrate this effect graphically for the SE area. The curves in Fig. 1 can be thought of as the subjective forecaster's conditional calibration curves. If the subjective forecast added no information, then the curves would be flat. On the other hand, Fig. 2 shows conditional calibration curves for the objective forecast, given the subjective forecast. The upward trend is much less pronounced and is only obvious when the subjective forecast is 0.40 or 0.60. This latter result suggests that the subjective forecast makes complete (or nearly complete) use of the information contained in the objective forecast.

TABLE 2. Individual and joint calibration functions for the SE area, warm season, 0000 cycle time, and 12–24 h lead time. In each pair, the top value represents the relative frequency of occurrence of precipitation and the bottom value represents the sample size.

Probability of precipitation														
Subjective PoP forecast	Objective PoP forecast													
	0.000	0.020	0.050	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	1.000	
0.000	0.005	0.005	0.007	0.018	0.014	0.133	0.125	0.000	—	—	0.000	—	—	0.010
	655	211	279	279	69	15	8	4	—	—	1	—	—	1521
0.020	—	0.000	0.000	0.000	0.000	—	—	—	—	—	—	—	—	0.000
	—	4	4	4	1	—	—	—	—	—	—	—	—	13
0.050	0.000	0.000	0.000	0.046	0.167	0.000	—	—	0.000	—	—	—	—	0.032
	8	3	29	44	6	3	—	—	1	—	—	—	—	94
0.100	0.000	0.035	0.053	0.061	0.059	0.053	0.000	0.000	0.500	—	—	—	—	0.052
	65	57	133	395	238	75	23	2	2	—	—	—	—	990
0.200	0.115	0.071	0.088	0.146	0.131	0.167	0.174	0.389	0.429	—	—	—	—	0.146
	26	14	57	212	389	168	46	18	7	—	—	—	—	937
0.300	0.000	0.143	0.091	0.160	0.220	0.299	0.293	0.215	0.381	0.571	0.500	—	—	0.251
	7	7	22	94	268	314	116	65	21	7	2	—	—	923
0.400	0.000	—	0.000	0.233	0.284	0.373	0.366	0.338	0.314	0.400	0.250	1.000	—	0.340
	1	—	6	30	74	150	183	74	35	5	4	2	—	564
0.500	—	—	0.333	0.222	0.267	0.392	0.470	0.454	0.486	0.630	0.500	—	1.000	0.446
	—	—	3	9	30	74	83	183	70	27	2	—	1	482
0.600	—	—	—	0.667	0.214	0.500	0.511	0.623	0.585	0.676	0.500	0.000	1.000	0.572
	—	—	—	6	14	28	47	106	130	37	16	1	1	386
0.700	—	—	—	—	0.750	0.667	0.773	0.724	0.698	0.763	0.842	0.333	1.000	0.744
	—	—	—	—	4	6	22	29	43	76	19	3	1	203
0.800	—	—	—	1.000	1.000	0.500	0.500	0.824	0.640	0.844	0.757	1.000	0.667	0.756
	—	—	—	1	1	2	8	17	25	32	37	6	6	135
0.900	—	—	—	—	1.000	0.500	—	0.800	0.700	1.000	0.954	0.917	1.000	0.904
	—	—	—	—	2	2	—	5	10	11	22	12	9	73
1.000	—	—	—	0.000	1.000	0.500	1.000	0.667	0.909	1.000	0.950	0.812	0.909	0.878
	—	—	—	1	1	4	2	3	11	10	20	16	22	90
	0.008	0.017	0.032	0.085	0.150	0.279	0.364	0.466	0.552	0.751	0.772	0.825	0.900	0.228
	762	296	533	1075	1097	841	538	506	355	205	123	40	40	6411

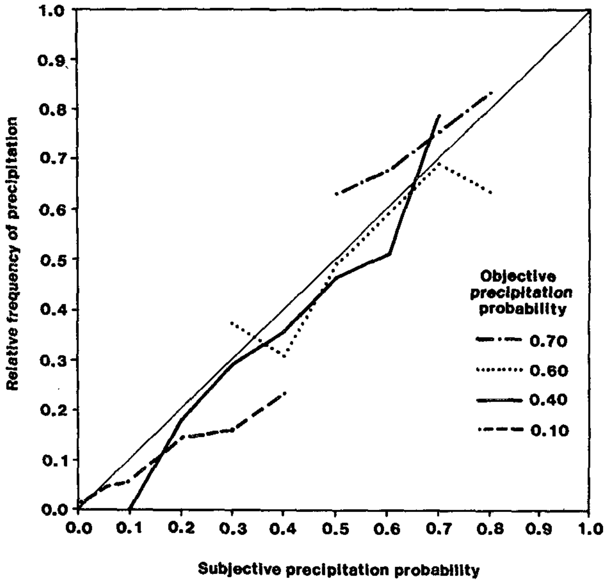

FIG. 1. Some conditional calibration curves for the subjective forecast, given the objective forecast, in the SE area, warm season, 0000 cycle time, and 12–24 h lead time. Calibration figures based on less than twenty observations have been excluded.

Conditional calibration curves, such as those depicted in Figs. 1 and 2, could be quite useful to forecasters in improving their forecasts. However, in order to enhance the utility of the information, these diagrams should also contain confidence intervals indicating the bounds on the errors of estimating the probability of precipitation. In a forthcoming paper, we will investigate these and other issues related to the potential usefulness of combining objective and subjective PoP forecasts using joint calibration procedures.

A formal statistical procedure can be developed to test the null hypothesis that one of the forecasts incorporates all of the information contained in the other forecast. Suppose we ask whether the subjective forecast adds information not contained in the objective forecast. If not, then once the objective forecast is received, the probability of precipitation should not change regardless of the value taken by the subjective forecast. In probabilistic terms, we would summarize the situation as follows: If the subjective forecast adds no information, then the occurrence of precipitation, given the objective forecast, would be independent of the subjective forecast. This situation would have to hold for all possible values of the objective forecast. Alternatively, the calibrated probability of precipitation based only on the objective forecast must be the same as the jointly calibrated probability of precipitation, regardless of the values for both the objective and subjective forecasts. For example, if the subjective forecaster added no information to the objective forecast, we would find that each column of Table 2 would be constant, and that constant value would, in turn, be equal to the calibrated probability of precipitation based only on the objective forecast.

Since the null hypothesis is equivalent to the hypothesis of independence, we can use a contingency table approach and test for independence in the contingency table with the familiar χ^2 test. Using this approach, we create a series of 2×2 contingency tables, one for each value of the objective forecast; in these tables, one dimension relates to the precipitation event (occurrence vis-à-vis nonoccurrence), whereas the other dimension relates to the level of the subjective forecast (less than the objective forecast vis-à-vis greater than or equal to the objective forecast). (When the objective forecast involves a probability of zero, the dichotomy is whether the forecasts are equal or not.) Since the contingency table is 2×2, the χ^2 statistic has one degree of freedom. A significant χ^2 value leads to rejection of the null hypothesis that the subjective forecast adds no information for the specified value of the objective forecast.

Table 3 shows an example from the SE area, 0000 cycle time, 12–24 h lead time, and warm season. The value of the objective forecast probability is 0.30. In this situation, the conditional probability of precipitation given that the subjective forecast is 0.20 or less is estimated as $34/261 = 0.13$. On the other hand, the conditional probability of precipitation given a subjective forecast greater than or equal to 0.30 is approximately $192/580 = 0.33$. This evidence is clearly inconsistent with the hypothesis of independence, as reflected in the drastically different conditional probabilities and the very large value of the χ^2 statistic.

To test the hypothesis that the subjective forecast adds no information for all values of the objective fore-

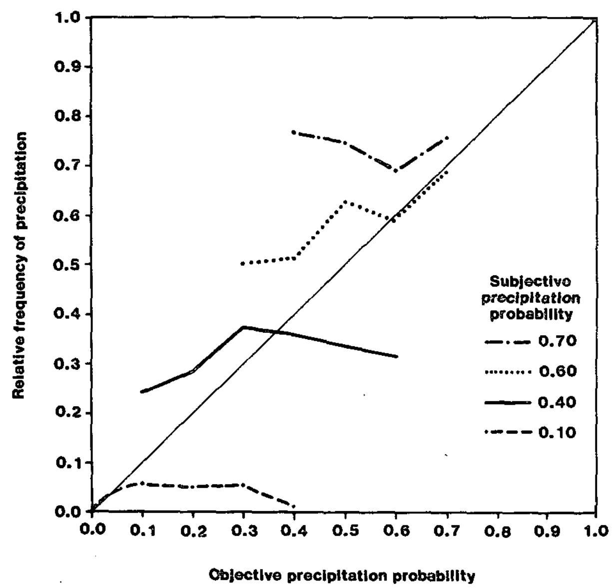

FIG. 2. As in Fig. 1 but for the objective forecast, given the subjective forecast.

TABLE 3. A χ^2 test for independence between the occurrence of measurable precipitation and the subjective PoP forecast. This table contains data for the SE area, warm season, 0000 cycle time, and 12–24 h lead time. The value of the objective PoP forecast is 0.30. Note that, since measurable precipitation occurred on 226/841 = 26.9% of the forecasting occasions, the objective forecast is quite well calibrated.

Subjective forecast	Event: Precipitation	Event: No precipitation	Total
0.20 or less	34	227	261
0.30 or more	192	388	580
Total	226	615	841

$\chi^2 = 41.82$ ($p = 0.000$).

cast, we can sum the individual χ^2 statistics for the various values of the objective forecast. If there are κ_O different values of the objective forecast, then κ_O individual χ^2 statistics exist, each possessing one degree of freedom, and the sum of these statistics is χ^2 with κ_O degrees of freedom. If this "overall" χ^2 statistic is significant, we reject the null hypothesis that the subjective forecast adds no information.

Some comments about this procedure are in order. First, the use of the χ^2 statistic with κ_O degrees of freedom is very powerful; it will detect very slight deviations from independence and can be substantially affected by a significant result in any one of the individual χ^2 terms. Offsetting the power of the test is the fact that a good deal of information is lost in compressing the data into a 2×2 table. If κ_S different values for the subjective forecast exist, we could create $\kappa_S \times 2$ contingency tables and perform the χ^2 test of independence on that table with $\kappa_S - 1$ degrees of freedom. The motivation for compressing the data into a 2×2 configuration is that the objective and subjective forecasts are usually similar. Using the $\kappa_S \times 2$ contingency table would, in most situations, result in a substantial number of cells with very small expected values, thus rendering the χ^2 approximation inappropriate.

4. Subjective forecasts: Do they contain information not included in objective forecasts?

In this section, we use the methodological framework described in section 3 to investigate whether the subjective PoP forecasts contain information not included in the objective PoP forecasts. If the subjective forecasts do not possess any new information, then the precipitation probabilities derived from the *joint* calibration function should be independent of these forecasts. Construction of a series of 2×2 contingency tables and calculation of the individual and overall χ^2 statistics, as described in the previous section, provides a basis for testing the hypothesis of independence. In this situation, independence is equivalent to the hypothesis that the subjective forecasts do *not* contain any new information over and above that included in the objective forecasts.

Tables 4 and 5 present the results of the χ^2 analysis of the 0000 cycle time forecasts for the 12–24 h and 36–48 h lead times, respectively, stratified by season and geographical area. Separate analyses were undertaken for the 1200 cycle time forecasts as well, but no obvious differences between the results for the two cycle times were discovered. We also found that the results for the 24–36 h lead time did not differ appreciably from the results for the 36–48 h lead time. Thus, we have omitted the results for the 24–36 h lead time and the 1200 cycle time to conserve space.

Examination of the overall χ^2 values in Tables 4 and 5 reveals that they are all significant at the 0.001 level. Overall, then, the hypothesis that the subjective forecasts do not contain new information must be rejected for all combinations of season and area. It is of interest to note that the overall χ^2 values for the 12–24 h lead time are generally considerably larger than the corresponding values for the 36–48 h lead time (the only exception occurs for the SE area in the cool season). To the extent that a higher χ^2 value can be interpreted as stronger evidence against the hypothesis of independence, this result suggests that, in an overall sense, the subjective forecasts added more information at the shorter lead time than at the longer lead time.

Inspection of the individual χ^2 statistics in Table 4 (12–24 h lead time) reveals that large values of these statistics—or, equivalently, small p-values—occur for almost all combinations of season, area, and objective probability value. In fact, if p-values greater than 0.05 are considered to be indicative of independence, then only four combinations satisfy this condition (ignoring combinations for which χ^2 may be inappropriate or for which such statistics cannot be calculated). The individual χ^2 statistics in Table 5 (36–48 h lead time) are generally smaller than the corresponding values in Table 4. Nevertheless, a substantial majority of the p-values associated with these statistics are less than 0.05, indicating a lack of independence for most combinations of season, area, and objective probability value. However, p-values larger than 0.05 occur much more frequently for the 36–48 h forecasts than for the 12–24 h forecasts, a result which is consistent with the overall χ^2 statistics. Combinations with small individual χ^2 values and large p-values—and therefore indicative of independence—appear to be associated primarily with relatively large and small values of the objective forecast probability.

The results related to the individual χ^2 statistics suggest that the subjective PoP forecasts contain additional information vis-à-vis the objective PoP forecasts for (i) almost all combinations of season, area, and objective PoP value in the case of the 12–24 h lead time and (ii) most such combinations in the case of the 36–48 h lead time. However, at the longer lead time, the sub-

TABLE 4. Chi-square analysis to determine whether subjective forecasts add information. Values shown are for 0000 cycle time and 12–24 h lead time. A large χ^2 value (or small p-value) indicates that the subjective forecast adds information for the specified value of the objective forecast.

Area		Objective precipitation probability													Overall χ^2 (dof)
		0.000	0.020	0.050	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	1.000	
								Cool season							
NE	χ^2 (dof = 1)	7.09*	1.55*	18.56	11.91	22.53	26.33	37.54	9.47	19.36	8.32	17.41	9.13	22.40	202.96
	p-value	0.008	0.213	0.000	0.001	0.000	0.000	0.000	0.002	0.000	0.004	0.000	0.003	0.000	(11)
	n	1616	330	627	717	543	418	296	227	231	203	169	210	233	
SE	χ^2 (dof = 1)	15.05*	25.38*	14.92	26.58	25.70	16.69	8.65	18.33	18.73	8.43	7.25	9.03	9.97	164.28
	p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.004	0.007	0.003	0.002	(11)
	n	1686	719	641	799	426	330	250	278	224	221	211	162	188	
SP	χ^2 (dof = 1)	19.24*	23.18*	10.75	39.19	37.02	20.83	23.04	1.33	17.02	16.31	1.75	3.55*	0.55*	167.24
	p-value	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.248	0.000	0.000	0.186	0.060	0.457	(9)
	n	1998	762	791	1083	509	269	227	181	179	104	106	68	28	
RM	χ^2 (dof = 1)	19.27*	12.74*	7.38	51.65	52.21	27.32	23.48	27.84	11.23	6.40	1.66	0.11*	0.75*	209.17
	p-value	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.000	0.001	0.011	0.198	0.743	0.385	(9)
	n	1193	282	600	1171	889	589	430	367	259	151	100	71	60	
								Warm season							
NE	χ^2 (dof = 1)	41.69*	14.73*	22.96	26.28	17.42	50.73	22.11	25.62	25.72	24.98	8.37	13.69	6.30*	237.87
	p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.012	(10)
	n	864	431	671	1120	702	626	367	279	249	184	177	138	69	
SE	χ^2 (dof = 1)	6.49*	6.53*	12.05	24.26	34.05	41.82	26.06	32.78	15.77	9.25	9.82	0.67*	0.05*	205.85
	p-value	0.011	0.011	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.414	0.832	(9)
	n	762	296	533	1075	1097	841	538	506	355	205	123	40	40	
SP	χ^2 (dof = 1)	20.56*	8.60*	12.89	27.33	49.48	11.87	26.47	3.38	16.82	6.54	11.00	1.76*	#	165.76
	p-value	0.000	0.003	0.000	0.000	0.000	0.001	0.000	0.066	0.000	0.011	0.001	0.185	#	(9)
	n	1121	459	890	1738	1190	616	386	200	116	81	55	15	3	
RM	χ^2 (dof = 1)	25.53	10.83	20.10	38.79	57.30	44.32	27.99	13.55	22.01	10.56	8.66	5.85*	#	279.66
	p-value	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.016	#	(11)
	n	1245	354	800	1596	1136	669	408	268	199	133	92	44	19	

* One or more cells had an expected frequency less than 5. Chi-square may be inappropriate.
# Chi-square statistics cannot be calculated.

jective forecasts appear to add information primarily when the objective forecasts take on probability values in the vicinity of the sample climatological probability (say values between and including 0.10 to 0.30 or 0.40), and they add less information when the objective forecasts take on more extreme probability values. The individual χ^2 statistics also reveal that independence occurs more frequently in the Southern Plains (SP) and Rocky Mountain (RM) areas than in the Northeast (NE) and Southeast (SE) areas in both seasons for the 36–48 h lead time.

5. Subjective forecasts: Do they make full use of objective forecasts?

We now turn to the question of whether the subjective forecasts make full use of the objective forecasts. If the former do make full use of the latter, then, for a specified value of the subjective forecast, the level of the objective forecast will be independent of the occurrence of precipitation. Thus, we create a series of 2 $\times$ 2 contingency tables, this time for each value of the subjective forecast. In this situation, one of the dimensions is the occurrence of precipitation and the other dimension is the level of the objective forecast (less than the subjective forecast vis-à-vis greater than or equal to the subjective forecast). Calculating individual and overall χ^2 statistics permits us to test the hypothesis of independence in the contingency table, which is equivalent here to the hypothesis that the subjective forecast makes full use of the information contained in the objective forecast.

Table 6 shows the results of the χ^2 analysis for the 0000 cycle time forecasts with a 12–24 h lead time for both cool and warm seasons for each area. Table 7 contains corresponding results for 36–48 h lead time. (Again, we also separately analyzed the 1200 cycle time forecasts but found no obvious effects due to this factor. Moreover, the results for the 24–36 h forecasts did not differ substantially from those for the 36–48 h forecasts.) The most obvious difference between these tables

TABLE 5. As in Table 4 but for 36–48 h lead time.

		Objective precipitation probability													Overall χ^2
Area		0.000	0.020	0.050	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	1.000	(dof)
								Cool season							
NE	χ^2 (dof = 1)	5.02*	4.92	8.22	23.85	16.12	15.17	19.60	16.33	18.41	10.83	6.34	6.15	0.02*	145.95
	p-value	0.025	0.027	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.012	0.013	0.896	(11)
	n	868	341	535	1079	766	491	378	338	293	266	202	147	38	
SE	χ^2 (dof = 1)	28.02*	8.87*	5.43	31.39	20.79	29.65	39.75	20.86	22.15	7.67	0.30	1.02*	4.17*	177.99
	p-value	0.000	0.003	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.584	0.313	0.041	(9)
	n	1170	216	687	1044	838	601	482	360	276	218	83	46	13	
SP	χ^2 (dof = 1)	2.79*	23.31*	5.36	24.49	21.84	15.31	4.28	0.79	1.85	0.49	0.28*	#	#	74.42
	p-value	0.095	0.000	0.021	0.000	0.000	0.000	0.039	0.373	0.174	0.482	0.598	#	#	(8)
	n	1479	513	961	1307	749	514	275	183	139	66	10	1	0	
RM	χ^2 (dof = 1)	4.23*	6.52*	2.86	34.25	22.50	27.56	10.56	3.71	0.49	1.09	0.26*	2.94*	#	103.00
	p-value	0.040	0.011	0.091	0.000	0.000	0.000	0.001	0.054	0.486	0.297	0.613	0.087	#	(8)
	n	601	248	597	1302	1165	876	501	390	199	111	45	16	3	
								Warm season							
NE	χ^2 (dof = 1)	4.49*	8.48	2.35	9.65	21.77	16.09	0.68	11.36	8.59	11.91	3.56	1.15*	#	94.44
	p-value	0.034	0.004	0.125	0.002	0.000	0.000	0.409	0.001	0.003	0.001	0.059	0.283	#	(11)
	n	685	269	494	1078	1065	675	522	394	297	175	74	42	4	
SE	χ^2 (dof = 1)	4.04*	2.18*	7.00	7.90	23.74	53.73	19.69	11.07	12.15	5.62	1.20*	1.91*	#	140.91
	p-value	0.044	0.140	0.008	0.005	0.000	0.000	0.000	0.001	0.001	0.018	0.274	0.168	#	(8)
	n	462	206	399	1031	1161	1165	771	537	350	147	49	8	3	
SP	χ^2 (dof = 1)	0.55*	2.12	6.50	17.90	26.67	20.52	9.09	7.93	0.16	1.61*	#	#	#	90.88
	p-value	0.457	0.146	0.011	0.000	0.000	0.000	0.003	0.005	0.688	0.205	#	#	#	(8)
	n	677	375	869	2128	1442	635	356	159	80	15	0	0	0	
RM	χ^2 (dof = 1)	15.22	6.44*	1.29	25.35	26.06	4.29	2.77	0.00	0.56	2.63*	0.32*	#	#	75.54
	p-value	0.000	0.011	0.256	0.000	0.000	0.038	0.096	0.961	0.456	0.105	0.571	#	#	(8)
	n	746	323	831	1822	1416	814	467	224	118	51	9	0	0	

* One or more cells had an expected frequency less than 5. Chi-square may be inappropriate.

Chi-square statistics cannot be calculated.

and Tables 4 and 5 is that the overall χ^2 values, while still significant at the 0.001 level, are much smaller than the corresponding values in Tables 4 and 5. This result is especially noticeable in the warm season for the 12–24 h forecasts. Inspection of the individual χ^2 terms reveals that many of these terms are relatively small. As in section 4, the small terms tend to occur for more extreme values of the subjective forecast probability. Although this tendency is apparent for all situations, it is best exemplified by the SE area in the warm season for the 36–48 h lead time. These results would suggest (i) that in the aggregate the subjective forecast is not making complete use of the objective forecast and (ii) that it is particularly in the vicinity of the sample climatological probability (0.10 to 0.40) that better use could be made of the objective forecast.

Although Tables 6 and 7 exhibit no clear patterns in the individual terms between seasons or lead times, the overall χ^2 values do show some interesting tendencies. The overall values are lower for the 12–24 h forecasts than for the 36–48 h forecasts in every situation except for the NE area during the cool season. This pattern suggests that the subjective forecast makes better use of the objective forecast for shorter lead times (again to the extent that larger overall χ^2 values can be interpreted as stronger evidence against independence). Similarly, we can compare overall χ^2 values for the warm and cool seasons. The results are mixed for the 36–48 h forecasts, but for the 12–24 h forecasts the values for the warm season are uniformly smaller than the corresponding values for the cool season, and the differences are substantial for the NE, SE and RM areas. With the same caveat as above, we would interpret these results to indicate that the subjective forecasts make the most complete use of objective forecasts during the warm season and for the shortest time horizon.

6. Discussion of results

The results presented in Tables 4 through 7 reveal that in general (i) subjective forecasts add information above and beyond that contained in the objective forecasts and (ii) subjective forecasts make less-than-perfect use of the objective forecasts in a statistical sense. These

TABLE 6. Chi-square analysis to determine whether subjective forecasts make full use of objective forecasts. Values shown are for 0000 cycle time and 12–24 h lead time. A large χ^2 value (or small p-value) indicates that the subjective forecast does not make full use of the information contained in the objective forecast for the specified value of the objective forecast.

		Subjective precipitation probability													
Area		0.000	0.020	0.050	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	1.000	Overall χ^2 (dof)
								Cool season							
NE	χ^2 (dof = 1)	7.69	#	#	7.58	13.09	9.82	0.95	3.64	2.00	0.81	4.40	0.13	10.51	60.60
	p-value	0.006	#	#	0.006	0.000	0.002	0.331	0.056	0.157	0.369	0.036	0.723	0.001	(11)
	n	1923	0	0	1130	742	379	242	246	202	155	211	197	393	
SE	χ^2 (dof = 1)	11.31*	#	0.17*	7.11	14.09	5.46	5.05	4.52	1.87	2.69	6.58	0.54	1.23	49.13
	p-value	0.001	#	0.684	0.008	0.000	0.020	0.025	0.034	0.171	0.101	0.010	0.464	0.267	(10)
	n	2808	5	104	853	497	345	223	210	209	168	229	172	312	
SP	χ^2 (dof = 1)	3.52	0.43*	0.84*	10.46	5.75	1.88	4.34	0.01	6.84	3.31	0.49	3.74	1.37*	40.34
	p-value	0.061	0.512	0.359	0.001	0.017	0.171	0.037	0.916	0.009	0.069	0.484	0.053	0.241	(10)
	n	3588	24	214	1020	462	246	160	128	139	80	109	74	61	
RM	χ^2 (dof = 1)	8.50	0.92*	1.78*	9.32	9.09	5.79	1.09	1.79	1.32	4.34	1.70	0.22*	0.96*	42.94
	p-value	0.004	0.338	0.183	0.002	0.003	0.016	0.296	0.181	0.251	0.037	0.192	0.640	0.327	(9)
	n	1671	65	638	1094	691	528	325	320	291	165	180	100	94	
								Warm season							
NE	χ^2 (dof = 1)	4.09	#	#	5.94	1.36	7.49	0.15	1.52	1.92	0.85	2.86	0.57	5.22*	26.73
	p-value	0.043	#	#	0.015	0.244	0.006	0.695	0.218	0.166	0.357	0.091	0.449	0.022	(10)
	n	1752	0	0	1078	811	587	371	332	253	175	185	130	202	
SE	χ^2 (dof = 1)	3.29	#	0.41*	2.05	1.04	12.46	0.75	3.30	0.71	0.58	0.17	0.79*	0.27*	24.33
	p-value	0.070	#	0.522	0.152	0.308	0.000	0.387	0.069	0.401	0.448	0.684	0.373	0.606	(9)
	n	1521	13	94	990	937	923	564	482	386	203	135	73	90	
SP	χ^2 (dof = 1)	6.42	#	0.22*	8.16	18.92	3.10	1.57	0.10	0.40	1.25	0.23*	0.86*	0.31*	39.92
	p-value	0.011	#	0.641	0.004	0.000	0.078	0.210	0.755	0.529	0.263	0.634	0.353	0.579	(8)
	n	2066	30	266	1759	1260	634	326	190	130	88	70	29	22	
RM	χ^2 (dof = 1)	1.42*	1.58*	0.02*	0.48	6.92	6.43	0.00	4.32	0.35	1.37	0.38	1.49*	0.65*	20.25
	p-value	0.234	0.208	0.886	0.488	0.009	0.011	0.955	0.038	0.554	0.242	0.538	0.223	0.420	(8)
	n	1618	112	697	1397	1128	795	370	289	250	118	90	46	53	

* One or more cells had an expected frequency less than 5. Chi-square may be inappropriate.
# Chi-square statistics cannot be calculated.

results make good sense given the current forecasting process in which the forecasters consider the objective forecast as only one of many inputs in arriving at the subjective forecast. The procedure through which a subjective forecaster integrates these inputs to form his PoP forecast is not well understood, undoubtedly varies among forecasters, and typically involves informal consideration of the importance of the various inputs. To make perfect use of the objective forecast in a probabilistic sense would require a careful statistical analysis of the interaction between the objective and subjective forecasts. One such approach is exemplified by Table 2. Since subjective forecasters usually do not perform such analyses, it is perhaps surprising that they are able to use the objective forecasts as well as they do.

On the other hand, our results are striking when considered in the light of previous studies in which the comparative performance of objective and subjective PoP forecasts has been evaluated by calculating Brier scores or skill scores (e.g., Charba and Klein, 1980; Murphy and Brown, 1984; Murphy, 1985). Typical results of such studies have shown that relatively little difference exists between the scores for objective and subjective forecasts, especially for longer lead times. As a consequence, many individuals have concluded that the subjective forecasts contain very little information above and beyond that contained in the objective forecasts. Our results certainly appear to contradict this conclusion. It should be noted that previous studies have focused on which forecast performs best by itself, whereas we have investigated the interaction of the two forecasts. In this regard, the results in sections 4 and 5 suggest that combining objective and subjective forecasts (e.g., through a joint reliability table such as Table 2) should lead to better performance than either forecast alone. However, the results also indicate that most of the information in the combined forecast is already contained in the subjective forecast. Thus, the performance of the *calibrated* subjective forecast should be almost as good as the combined forecast.

TABLE 7. As in Table 6 but for 36–48 h lead time.

Area		Subjective precipitation probability 0.000	0.020	0.050	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	1.000	Overall χ^2 (dof)
								Cool season							
NE	χ^2 (dof = 1)	288	#	#	11.80	8.82	4.57	0.24	0.60	1.39	5.19	9.27	7.57	0.71*	52.33
	p-value	0.090	#	#	0.001	0.003	0.033	0.627	0.438	0.238	0.023	0.002	0.006	0.399	(10)
	n	1259	0	0	1409	1002	424	317	278	295	302	237	154	64	
SE	χ^2 (dof = 1)	11.21	#	1.12*	18.01	0.99	18.15	4.49	0.50	2.59	0.15	0.56	0.16*	1.54*	56.63
	p-value	0.001	#	0.289	0.000	0.321	0.000	0.034	0.482	0.108	0.703	0.456	0.691	0.215	(9)
	n	2202	14	124	1194	623	493	328	309	290	218	155	58	26	
SP	χ^2 (dof = 1)	13.38	1.17*	5.70	26.56	7.79	2.86	0.87	3.96	9.72	0.43	1.99*	#	#	71.26
	p-value	0.000	0.280	0.017	0.000	0.005	0.091	0.352	0.047	0.002	0.513	0.159	#	#	(9)
	n	2846	43	384	1655	405	329	188	116	131	67	27	5	0	
RM	χ^2 (dof = 1)	2.03	0.35*	3.82	18.00	2.05	7.45	2.92	10.14	0.92	2.39	0.85	0.47*	#	50.57
	p-value	0.154	0.555	0.051	0.000	0.153	0.006	0.087	0.002	0.338	0.122	0.355	0.495	#	(10)
	n	924	69	697	1389	902	732	434	375	300	122	79	29	2	
								Warm season							
NE	χ^2 (dof = 1)	6.79	#	#	8.63	9.17	8.22	2.42	19.26	5.34	5.90	1.98	4.71	0.23*	72.40
	p-value	0.009	#	#	0.003	0.003	0.004	0.120	0.000	0.021	0.015	0.159	0.030	0.630	(10)
	n	1176	0	0	1346	1051	579	475	408	326	197	123	70	22	
SE	χ^2 (dof = 1)	7.33	#	0.64*	0.35	7.72	8.36	5.42	0.62	3.44	0.87	1.71*	0.02*	#	34.10
	p-value	0.007	#	0.422	0.557	0.006	0.004	0.020	0.432	0.064	0.351	0.191	0.880	#	(8)
	n	1194	9	93	1217	1020	1029	655	504	342	143	65	10	8	
SP	χ^2 (dof = 1)	4.67	0.81*	3.02*	8.96	12.83	11.08	4.28	1.67	2.26	0.51*	#	#	#	45.75
	p-value	0.031	0.367	0.082	0.003	0.000	0.001	0.039	0.197	0.133	0.475	#	#	#	(7)
	n	1477	36	350	2541	1199	597	274	143	85	24	8	0	2	
RM	χ^2 (dof = 1)	6.16	0.09*	0.00	11.25	23.56	21.54	19.90	8.92	7.06	2.98	0.81*	#	#	101.37
	p-value	0.013	0.764	0.956	0.001	0.000	0.000	0.000	0.003	0.008	0.084	0.367	#	#	(9)
	n	920	100	834	1873	1290	856	381	268	179	67	37	11	5	

* One or more cells had an expected frequency less than 5. Chi-square may be inappropriate.

Chi-square statistics cannot be calculated.

Careful examination of Tables 4 through 7 reveals a problem that could indicate a weakness in the analysis. We noted that the very high χ^2 values tend to occur in the vicinity of the sample climatological probability; that is, from about 0.10 to 0.40. The largest sample sizes also occur in this range. It could be argued that we obtain a significant test statistic primarily because of the large sample size; the two random variables may be very nearly independent, if not perfectly so, but the test performed with the large sample size is sufficiently powerful to detect very small deviations from independence. Observation of this effect dates back at least to Berkson (1938) and leads to the typical statement that "any null hypothesis can be rejected if the sample size is large enough." Thus, the argument might be made that it is only such statistical artifacts of the analysis that lead to our conclusions, especially the (possibly incorrect) conclusion in section 5 that subjective forecasts do not make full use of the objective forecasts.

We are sensitive to the statistical argument presented above, and it cannot be dismissed lightly. However, we offer an alternative interpretation of these results. Consider the subjective forecaster's problem. He obtains the objective forecast that is, say, in the range from 0.10 to 0.40. In virtually every situation, this forecast is not far from the climatological probability of precipitation. We can argue that such an objective forecast provides relatively little help in deciding whether or not precipitation will occur. If, on the basis of additional inputs, the forecaster becomes relatively certain of the outcome (with respect to the objective forecast and climatology), it would appear that little could be gained by returning to the objective forecast and trying to extract more information from it. The subjective forecaster apparently finds the situation a relatively easy one in which to formulate a forecast. On the other hand, if the subjective forecaster's additional inputs provide mixed signals, he may conclude that the situation is relatively difficult to forecast, with the result that his subjective forecast is also fairly close to climatology. If this situation arises, then by returning to the objective forecast he may be able to extract the statistical information remaining in it to improve upon the subjective forecast. This interpretation is also consistent with the tendency in Tables 6 and 7 for smaller χ^2 values to occur when the objective forecast is further away from climatology. In these situations, it is not so much that the subjective forecaster makes better use of the information contained in the objective forecast; rather, the information contained in the objective forecast is less important in what is a relatively straightforward forecasting situation.

We can make a similar argument for the same pattern of occurrences of small χ^2 values in Table 5, particularly for the SP and RM areas. When the objective forecast reflects relative certainty about the outcome (that is, relative to climatology), the situation apparently is "easy" to forecast and the subjective forecaster can add little to the objective forecast. When the objective forecast is close to climatology, the subjective forecaster has the opportunity to add substantially more information to the forecast.

Finally, this interpretation is consistent with (i) our understanding of forecasting differences between cool and warm seasons and between shorter and longer lead times and (ii) the pattern in Tables 6 and 7 of generally smaller χ^2 values for the warm season and for the 12–24 h lead time. The subjective forecaster has a definite advantage when it comes to forecasting for shorter lead times; he is able to adapt more easily to rapidly changing local conditions. As the lead time increases, this advantage is reduced. Similarly, subjective forecasters have an advantage over the objective forecasting procedure during the warm season when precipitation events tend to depend more directly on local conditions. In both situations we would expect that little useful information would remain in the objective forecast after the subjective forecaster has considered the objective forecast along with local developments.

7. Conclusion

This paper has addressed two specific questions related to the interrelationships between objective and subjective precipitation probability forecasts: Do the subjective forecasts add information over and above that included in the objective forecasts? Do the subjective forecasts make full use of the information contained in the objective forecasts? The overall results, based on a data set containing more than ten years of PoP forecasts, indicate that (i) the subjective forecasts add information above and beyond that included in the objective forecasts for all combinations of geographical area, season, and lead time considered in this study and (ii) the subjective forecasts generally do not make full use of the objective forecasts. It should be noted that the results related to the second question are not as strong statistically as the results related to the first question; in this regard, it is primarily in the vicinity of the climatological probability (say 0.1 to 0.4) in which the forecasters could make better use of the objective forecasts.

Two important conclusions can be drawn from these results: (i) the forecasters are making a significant contribution to the official subjective precipitation probability forecasts issued by the National Weather Service and (ii) the forecasters are not making full use of the objective forecasts in formulating their forecasts in many situations. This latter conclusion suggests that it may be possible to combine the objective and subjective forecasts to produce forecasts with even greater information content. It also may be possible to improve the performance of the official forecasts through individual and/or joint calibration. In a forthcoming paper, we will investigate these and other issues related to the performance of individual and combined precipitation probability forecasts.

Acknowledgments. The objective and subjective precipitation probability forecasts analyzed in this paper were graciously provided by Gary M. Carter of the NWS Techniques Development Laboratory. This work was supported, in part, by the National Science Foundation (Division of Atmospheric Sciences) under Grant ATM-8507495.

REFERENCES

Berkson, J. A., 1938: Some difficulties of interpretation encountered in the application of the chi-square test. *J. Amer. Stat. Assoc.*, **33**, 526–536.

Charba, J. P., and W. H. Klein, 1980: Skill in precipitation forecasting in the National Weather Service. *Bull. Amer. Meteor. Soc.*, **61**, 1546–1555.

Clemen, R. T., 1986: Extraneous expert information. *J. Forecasting*, **5**, in press.

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.

Murphy, A. H., 1985: Probabilistic weather forecasting. *Probability, Statistics, and Decision Making in the Atmospheric Sciences.* A. H. Murphy and R. W. Katz, Eds., Westview Press, 337–377.

——, and R. L. Winkler, 1977: Reliability of subjective probability forecasts of precipitation and temperature. *Appl. Stat.*, **26**, 41–47.

——, and B. G. Brown, 1984: A comparative evaluation of objective and subjective weather forecasts in the United States. *J. Forecasting*, **3**, 369–393.

——, and T. E. Sabin, 1986: Trends in the quality of National Weather Service forecasts. *Wea. Forecasting*, **1**, 42–55.

Snellman, L. W., 1977: Operational forecasting using automated guidance. *Bull. Amer. Meteor. Soc.*, **58**, 1036–1044.