PACIFIC REGION TECHNICAL NOTES

83-026
August 16, 1983

Have the Fire Weather Precipitation Forecasts Improved? - A Case Study for Terrace, B.C.

Bob Beal, Meteorologist
Scientific Services Division
Atmospheric Environment Service, Vancouver, B.C.

INTRODUCTION

Fire Weather Forecasts have been issued for British Columbia since 1958. These forecasts have always included the point forecast probability of the occurrence of rain for a number of selected locations throughout the province. Twelve years ago, Nikleva (1971) verified the forecast probabilities issued for six stations over the fire weather seasons (May to September) from 1965 to 1969 inclusive. A verification scheme initially proposed by Brier (1950) was used. The main objective of this note will be to compare Nikleva's scores, for Terrace, against those derived from the forecast probabilities issued during the 1982 fire weather season (May to September). In view of the number of years elapsed, the results should prove to be of interest.

It should be noted that a direct comparison, of the scores, was only possible at Terrace. Over the years changes were made for the locations for which specific point forecasts were issued, until today only Terrace remains common to the set of stations employed by Nikleva and for which point forecasts are issued currently.

DATA

Nikleva's study included all rain probability forecasts, for Terrace, issued daily at noon, five days a week, from May to September for the years 1965 to 1969. The forecasts were valid for 24 hrs, beginning the following day at 1 PM. For verification purposes the forecast verification period was taken to be from 10 AM (on the next day after the forecast was issued) to 10 AM the following day. The forecasts were partitioned into twelve forecast percentage probability classes (0,5,10,20,...,90, 100%).

The 1982 Fire Weather Forecasts are also issued near noon, five days a week, and, in addition, an early morning (1345 G.M.T.) forecast is issued as well for seven days a week. Both forecasts are valid for 24 hrs, with the noon forecast valid at 1 PM the next day, and the early morning one valid at 1 PM on the same day. To evaluate the 1982 season forecast for Terrace, both the noon and morning forecasts were used and two forecast verification periods were examined. The 24 hr period, mentioned above, and a shorter 12 hr period from 10 AM to 10 PM.

The verification scheme employed was identical to the one used by Nikleva.

To compare the scores with climatology the long-term climatological relative frequency of the number of days with measureable rain (C), from May to Sept., was derived from the Canadian Climate Normals (1951-1980). The value C = 0.43 was identical to the one obtained earlier by Nikleva, using the 1931-1960 Normals.

Verification - Brier Score

The Brier score (B) may be expressed by (see Sanders, 1963):

$$B = \frac{1}{N}\sum_{k=1}^{T} M_k (F_k - \bar{\phi}_k)^2 + \frac{1}{N}\sum_{k=1}^{T} M_k \bar{\phi}(1 - \bar{\phi}_k) \qquad (1)$$

where,

N = total number of probability rain forecasts.

T = total number of forecast categories considered.

F_k = probability forecast value of the k th category.

M_k = total number of forecasts (and observations) assigned to the k th forecast category.

$\bar{\phi}_k$ = relative frequency of occurrence of "rain" events corresponding to the k th forecast category.

A perfectly correct set of forecasts would score zero (B=0) and a completely unskillful set would score one (B=1). A perfect score (B=0) is theoretically possible but rather unlikely in practice as it requires an absolutely correct and confident set of predictions.

Reliability - The first term on the right-hand side of (1) is often called the "reliability" or "validity" of the forecasts. It is essentially the mean square error of the difference between the forecast and observed relative frequencies of rain events. A value of zero would imply that all forecasts were perfectly reliable, the maximum value of one would be attained by an abysmally unreliable forecast set.

Resolution - The "resolution" or "sharpness" is given by the second term on the right-hand side of (1). For a sufficiently large sample size, it is a measure of the forecasters' ability to correctly separate nearly certain instances (of rain/no rain) as often as possible. Perfect resolution (i.e. zero) would be attained if the observed relative frequency was zero or one for each forecast category. Alternatively if all forecast probabilities were 50% (tossing a coin!), no resolution is exhibited and the resolution term reaches its maximum value of 0.25.

Verification - Skill Score

To relate the basic Brier score to a non-skill forecast, to facilitate comparisons, a Skill score (S) was devised (Sanders, 1963). It is of the form,

$$S = 100\ (B_c - B)/B_c \qquad (2)$$

Here B_c represents the climatological Brier score, it is identical to the Brier score except that the long-term climatological relative frequency of days with measureable rain (C) replaces the forecast probabilities. The greatest skill, S=100%, is reached with a Brier score (B) of zero. A negative or zero Skill score reflects the fact that the forecasts performed worse or no better than climatology.

RESULTS

The first two columns of Table I summarize the verification scores from Nikleva and those from the 1982 Fire Weather Forecasts, issued by the Pacific Weather Centre. Quite a significant improvement is evident as the Skill score rose from 16% to a substantial 45% for the 1982 season. It is interesting to note that the overall reliability remained constant. This is indicated by Figures 1 and 2, giving the forecast probabilities against the actual observed relative frequencies of rain occurrences. For the 1982 season, we may state that the accuracy of the forecasters to assign forecast probability levels to future events has not significantly improved. However, more important, the ability to recognize nearly certain instances, of rain or no rain, more often has definitely improved. In other words, the forecasters were more often able to express with a greater degree of confidence the occurrence or non-occurrence of rain, and this is reflected by the lower value of the resolution term for the 1982 season. Figure 4 portrays this trend graphically. One may note that the number of forecasts expressing a greater degree of confidence has increased for the 1982 season as opposed to the 1965-69 seasons.

Returning to Table I, various scores (1982) were also summarized for the early morning forecasts (column 3), and recalculated for the noon and morning forecasts using the shorter 12 hr verification period (columns 4 & 5 respectively). First, no significant improvement in the resolution occurs between the noon and morning forecasts, which is somewhat unexpected. One would surmise the morning forecasts to more frequently recognize near certain instances. The reliability for the morning forecasts improves dramatically however. An example is shown in Fig. 3. Most points are within 15% or less from the line of perfect reliability (diagonal). Employing a contracted verification period (12 hrs.), to perhaps examine a forecast tendency to do better in the immediate future, yielded no significant differences in the scores over those from the 24 hr. verification period.

For the 1982 fire weather season, the effect of partitioning the forecast probabilities into a number of fixed categories was examined, see Table II. The worst score was obtained by simply partitioning each

forecast probability to one of two categories (0 and 100%). All forecasts of 50% or greater were assigned to the 100% class. As the number of forecast classes increase, the scores, in general, improve. It is evident that the number of classes cannot be increased ad infinitum, and likely an optimum number, depending upon the level of expertise exists which would yield the best overall score. The results from this study are not conclusive, but one could tentatively infer that forecasting probabilities in steps of either 25 or 10% would exhibit good reliability and skill. The best overall score was obtained for the case with 5 forecast classes, illustrated in Fig. 5.

SUMMARY

From the study one may tentatively conclude:

1) The fire weather precipitation forecast probabilities, 1982 season, have significantly improved over those issued in the late 1960's. The reliability of the forecast probabilities has remained constant, while there was an excellent improvement in the forecast resolution.

2) The early morning forecasts were found to be more reliable than the ones issued near noon the previous day but, of course, valid for the same time period. However, unexpected was the result that the resolution (measure of the ability to separate events into YES/NO categories) did not significantly improve for the morning over the noon forecasts.

3) The noon forecasts reveal a distinct tendency to over-forecast for the higher probabilities and under-forecast for those around 30 to 50%.

4) A decrease in the verification forecast period (to 12 hrs) did not alter the scores significantly over the 24 hr. verification period.

5) Likely there is an optimum number of forecast probability values which should be issued, contingent upon the state-of-the-art.

The above conclusions must be considered preliminary due to the limited sample size employed. However, we expect to continue this analysis for the 1983 season to solidify the conclusions presented.

REFERENCES

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

Nikleva, S., 1971: Verification of probability of precipitation occurrence forecasts. Environment Canada, Technical Memoranda, TEC 749, 19 pp.

Sanders, F., 1963: On subjective probability forecasting. J. Appl. Meteor., 2, 191-201.

Table I. TERRACE, B.C.

Comparison of the verification scores from the 1982 fire weather season and the seasons over the period 1965-69. The forecast point rain probabilities verified were taken from the "morning" and "noon" Fire Weather Forecasts. Furthermore, two forecast verification periods were employed, a 24 hr (10 AM - 10 AM) and a 12 hr (10 AM - 10 PM) period.

Scores	Fire Weather Season (May - Sept.)				
		1982			
	1965 - 69 Noon 24 hr	Noon 24 hr	Morning 24 hr	Noon 12 hr	Morning 12 hr
Brier Score	0.20	0.13	0.12	0.13	0.11
Skill Score (%)	16	45	49	44	51
Climatology Score	0.24	0.23	0.23	0.22	0.22
Reliability	0.015	0.015	0.006	0.016	0.008
Resolution	0.19	0.11	0.11	0.11	0.10
# of Forecasts	534	108	153	108	153
# of Precipitation Events	213	33	46	30	40

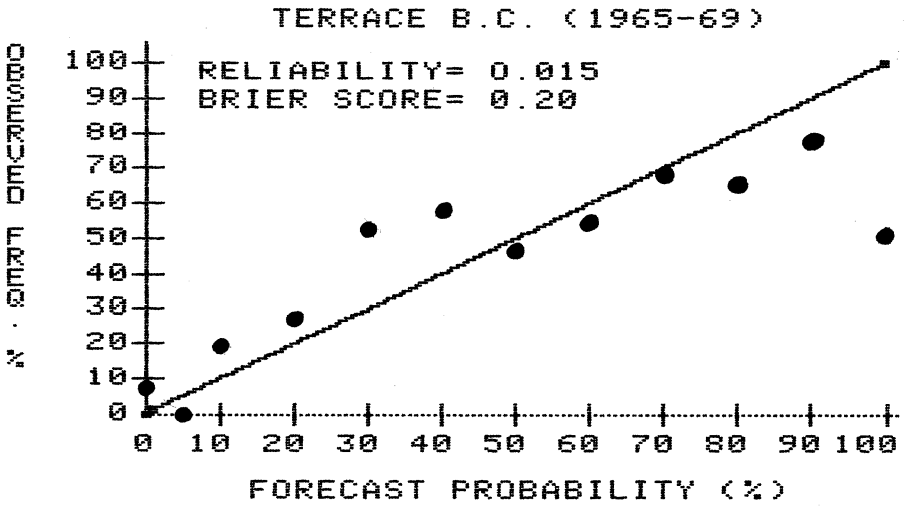

Fig. 1.
Forecast probability, from the noon Fire Weather Forecasts, versus the observed relative frequency of occurrence, for the 1965-69 fire weather seasons. The forecast verification period was 10 AM - 10 AM.

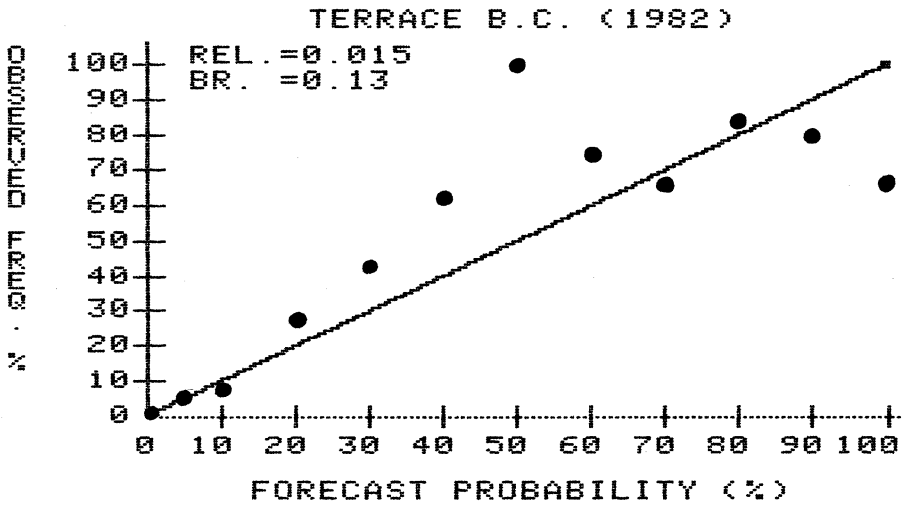

Fig. 2.
Forecast probability, from the noon Fire Weather Forecasts, versus the observed relative frequency of occurrence, for the 1982 fire weather season. The forecast verification period was 10 AM - 10 AM.

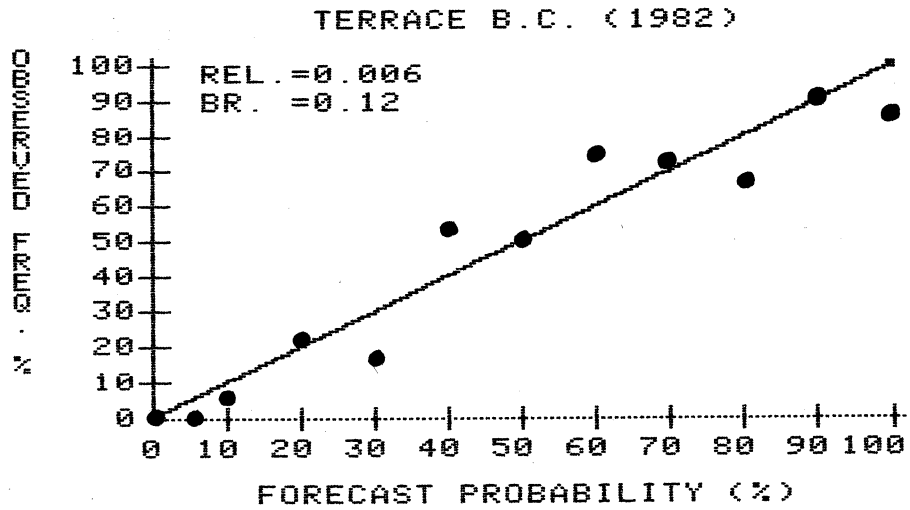

Fig. 3.
Forecast probability, from the morning Fire Weather Forecasts, versus the observed relative frequency of occurrence, for the 1982 fire weather season. The forecast verification period was 10 AM - 10 AM.

Fig. 4.

Frequency distribution of the forecast point rain probabilities from the noon Fire Weather Forecasts issued during the fire weather seasons (May - Sept.) over 1965 - 69 (solid bars) and for the 1982 season (shaded or dashed bars). The forecast verification period was 10 AM - 10 AM. The resolution (RES.) or sharpness for both distributions is shown.

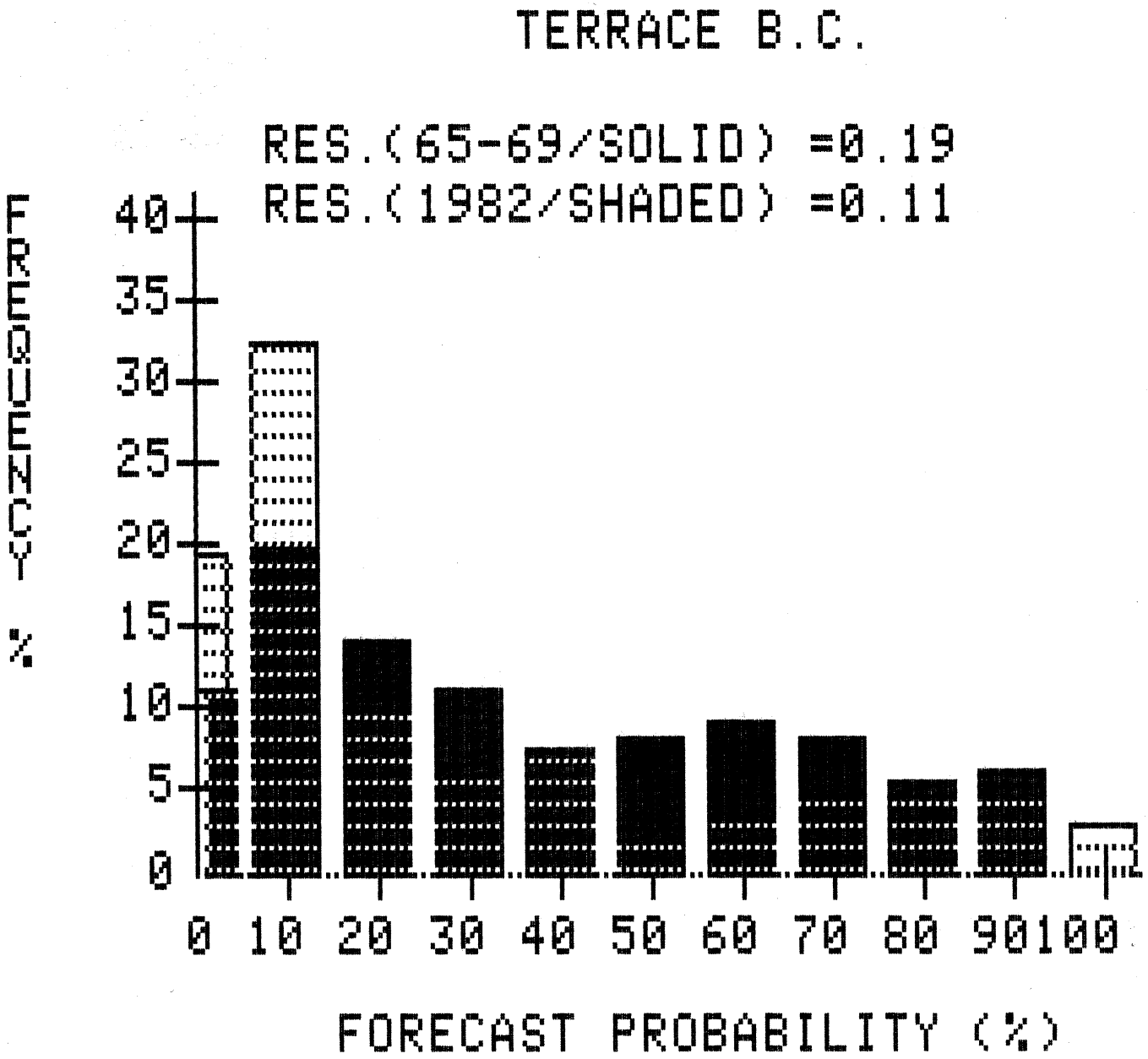

Table II. TERRACE, B.C.

Effect on the verification scores by using a number of different forecast probability categories, for the 1982 Fire Weather Forecasts issued in the morning from May to September. A 24 hr. forecast verification period (10 AM - 10 AM) was used.

Score	Forecast Probability Classes (%)				
	0 & 100	0, 50, 100	0, 25, 50, 75, 100	0, 10, 20..., 90, 100	0, 5, 10, 15, 20..., 95, 100
Brier Score	0.18	0.13	0.11	0.12	0.11
Skill Score (%)	22	41	51	49	50
Reliability	0.030	0.010	0.002	0.006	0.020
Resolution	0.14	0.13	0.11	0.11	0.10
Total number of forecasts used 153					
Total number of precipitation events 46 (or 30.1%)					

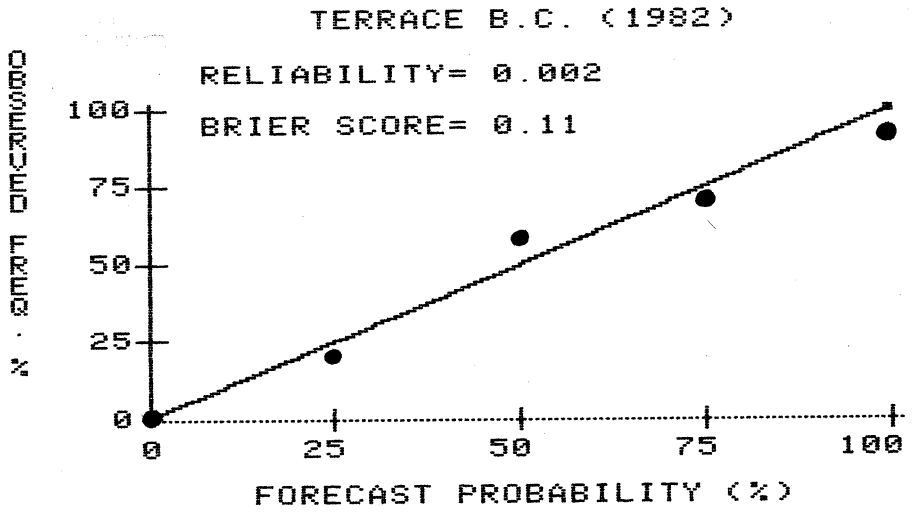

Fig. 5. Forecast probability, from the morning Fire Weather Forecasts, versus the observed relative frequency of occurrence, for the 1982 fire weather season. The forecast verification period was 10 AM - 10 AM.
Note, in this case only 5 forecast categories were employed.