PACIFIC REGION TECHNICAL NOTES

82-013
August 20, 1982

A Preliminary Verification of the Pacific Weather Centre's Probability of Precipitation Forecasts

David Grimes, Supervising Meteorologist
Pacific Weather Centre, Vancouver, B.C.

INTRODUCTION

Probability of Precipitation forecasts were introduced into the Canadian Public Weather forecasts on July 5th of this year. Since probabilities are subjectively produced and reflect the forecaster's degree or measure of belief in an expected event, verification is necessary to provide a meaningful feedback into the forecasting system. However, unlike other forecasts, the quality of any one probability forecast is not obvious by the next day and the accuracy of such forecasts can only be meaningfully determined over a collection of similar forecasts.

A preliminary examination using the Pacific Weather Centre's July probability forecasts has been undertaken in order to expose some of the early problems in our probability of precipitation program. The sample is rather small and therefore, one must take caution in making too many conclusive statements.

VERIFICATION METHOD

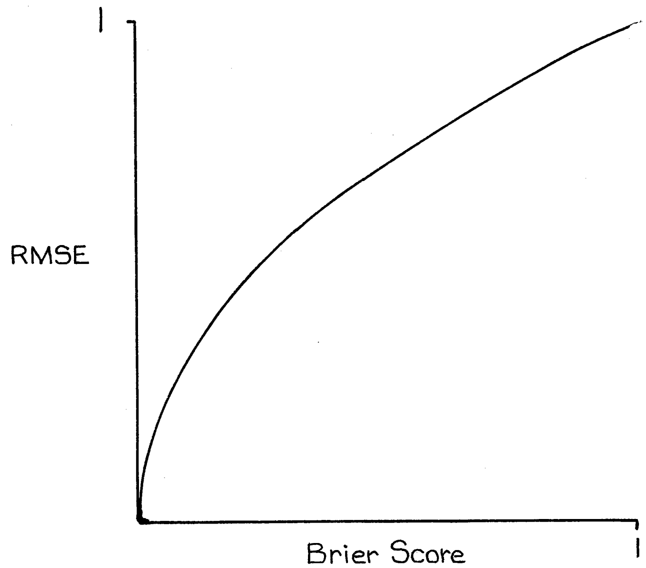
A universally accepted verification score for probability forecasts is the simplified Brier score, which for a two category event (either it rains or it does not) is expressed as follows:

$$P = \frac{1}{N} \sum_{i=1}^{N} (F_i - e_i)^2$$

N = total number of forecasts
F = probability expressed as a decimal
e = 1, if more than a trace of precipitation is observed
= 0, otherwise

The score is a measure of the mean square error, ranging between 0 and 1, and the lowest score reflects the best forecast.

When comparing forecast performance between stations or between months, the Root Mean Square Error (RMSE) may be a useful statistic, since it introduces more range in the error when the Brier score is low. The following graph illustrates the value of the RMSE.



Finally, in order to measure the forecaster's skill, the Brier score can be compared to a standard score such as climatology. The skill score can be computed as follows:

$$\text{SKILL} = \left[1 - \frac{\text{Brier Score}}{\text{Climatological Brier Score}} \right] 100$$

This score can be interpretted as the forecasters' percentage improvement or degradation over the climatological probability.

RESULTS

Table 1 shows the Brier, Root Mean Square, and Skill scores for 7 locations in British Columbia. The columns on the left hand side of the table list the verification statistics for the Today, Tonight, and Tomorrow forecasts that are issued at 5 AM, and the columns on the right side of the table list the verification statistics for the Tonight, Tomorrow, and the Third Day that are issued at 4 PM.

It would be difficult to make too many conclusions on these results because of the limited sample size, however some conclusions can be made about forecasters' tendencies. The office predicts "probabilities" most accurately during the first 12 hours, and the office's accuracy rapidly degenerates over the next two forecast periods. Comparing the two initial 12 hour forecast periods, the forecasters show more skill in predicting "probabilities" for Today than Tonight. Very little difference in skill can be detected over the next two periods with little improvement over the climatological probabilities. It also appears that Tomorrow's forecast is predicted with nearly the same level of skill no matter when the forecast is issued (5 AM or 4 PM).

The magnitude of the scores in Table 1 are not too impressive with the performance in the Today period close to what should be expected. However more comment about the magnitudes of the scores can be made when the forecast sample is larger.

In comparing the scores between stations, it appears that the forecasters did better for Victoria and Kelowna than for the other regions. However, one must be cautious in interpretting these statistics until a larger sample of forecasts are collected.

Figures 1 (5 AM forecast issue) and 2 (4 PM forecast issue) present Contingency Tables and Reliance Graphs for each time period for all the stations treated collectively. Precipitation probabilities are forecast in one of eleven different categories (0, 10%...90%, 100%). A forecast of 20 percent is said to be perfectly reliable if precipitation was observed (observed frequency) on 20 percent of the occasions when 20 percent was forecast. Therefore, if the observed frequency was higher, then forecasters would be underforecasting probabilities, and if the observed frequency was lower, then forecasters could be accused of overforecasting. In most cases a threshold of 50 percent could be used to make a categorical decision on the likelihood of precipitation.

Probability forecasts can be re-grouped into three categories (probabilities greater than 50 percent, probabilities equal to 50 percent, and probabilities less than 50 percent; A, B, and C respectively) to assess the forecasters' ability to predict precipitation based on this threshold. The Contingency Matrices in Figures 1 and 2 present the data in this fashion.

A quick glance through the Contingency Matrices on both figures illustrate that forecasters are underforecasting precipitation according to the probability statement in all the time periods. Table II illustrates the magnitude of the forecaster's biases. No detailed effort has been made to determine whether or not this large bias towards underforecasting is reflected in the wording of the forecasts. However, cursory review of the July forecasts suggests that forecasters may not be underforecasting precipitation to the extent shown by the probability scores, but that forecasters under-estimate the expected extent of showers or have introduced a "temporal" consideration into the definition of the Probability of Precipitation (e.g. one or two showers everywhere in the early morning does not constitute a 100 percent as the official definition would state, but rather some value around 40 percent because there would be no expected showers for the remainder of the day).

The graphs on Figures 1 and 2 also show a strong tendency to underforecast probabilities in almost all categories and time periods up to 50 percent. Above this threshold, the number of forecasts in each category is too few to show any meaningful trend.

SUMMARY

These initial findings indicate firstly that forecasters show more skill in predicting probabilities of precipitation in the first 12 hour period, with better accuracy during the Today period than during the Tonight period. Forecasters show very little skill over climatology in the second and third forecast periods, and therefore initially, forecasters should begin with climatology as a first guess and then deviate according to the weather pattern. Secondly, when using probability forecasts for deciding whether precipitation will or will not occur based on a threshold of 50 percent, the results show that forecasters are underforecasting the expected frequency of precipitation.

RECOMMENDATIONS

Based on these findings, forecasters should make an attempt to adjust for their tendency to underforecast. This could be achieved by making higher probability predictions, when precipitation is stated in the forecast. However, making adjustments according to the Reliance Graphs would be unwise until more forecasts have been verified. Some improvement in the probability forecasts could come by not allowing temporal modifiers for precipitation to cloud the forecaster's thinking when it comes to arriving at a probability value. Another improvement may result by producing a probability forecast which is more representative of the population centre than the surrounding area. This is especially true in British Columbia, where a great deal of the area is unpopulated.

TABLE I.

BRIER, ROOT MEAN SQUARE ERROR, SKILL SCORES FOR THE PROBABILITY OF PRECIPITATION FORECASTS ISSUED BY PWC IN JULY, 1982

TODAY

	5AM ISSUE			4 PM ISSUE		
	BRIER	RMS	SKILL(%)	BRIER	RMS	SKILL(%)
VANCOUVER	.10	.32	31			
VICTORIA	.07	.26	46			
PRINCE RUPERT	.16	.40	40			
KAMLOOPS	.10	.32	50			
PENTICTON	.11	.33	46			
CRANBROOK	.17	.41	-38			
PRINCE GEORGE	.24	.49	19			
AVERAGE	.15	.39	23			

TONIGHT

	5AM ISSUE			4 PM ISSUE		
	BRIER	RMS	SKILL(%)	BRIER	RMS	SKILL(%)
VANCOUVER	.13	.36	-34	.09	.30	8
VICTORIA	.15	.39	-14	.12	.35	11
PRINCE RUPERT	.28	.53	13	.23	.48	27
KAMLOOPS	.12	.35	-19	.13	.36	-25
PENTICTON	.23	.48	6	.16	.40	36
CRANBROOK	.10	.32	35	.15	.39	2
PRINCE GEORGE	.25	.50	5	.19	.44	25
AVERAGE	.19	.44	4	.15	.39	18

TOMORROW

	5AM ISSUE			4PM ISSUE		
	BRIER	RMS	SKILL(%)	BRIER	RMS	SKILL(%)
VANCOUVER	.19	.44	-31	.18	.42	-20
VICTORIA	.16	.40	-1	.22	.47	-34
PRINCE RUPERT	.24	.49	4	.27	.52	-8
KAMLOOPS	.16	.40	18	.21	.46	-1
PENTICTON	.25	.50	4	.23	.48	16
CRANBROOK	.19	.44	13	.24	.49	-5
PRINCE GEORGE	.24	.49	14	.23	.48	15
AVERAGE	.22	.47	4	.22	.47	2

3RD DAY

	5AM ISSUE			4PM ISSUE		
	BRIER	RMS	SKILL(%)	BRIER	RMS	SKILL(%)
VANCOUVER				.19	.44	-5
VICTORIA				.18	.42	4
PRINCE RUPERT				.27	.52	-6
KAMLOOPS				.22	.47	6
PENTICTON				.31	.56	-6
CRANBROOK				.17	.41	8
PRINCE GEORGE				.34	.58	-17
AVERAGE				.22	.47	6

TABLE II.

FORECASTERS' BIAS IN PREDICTING PRECIPITATION BASED ON A THRESHOLD POPS OF 50 PERCENT

BIAS (PERFECT BIAS = 0)

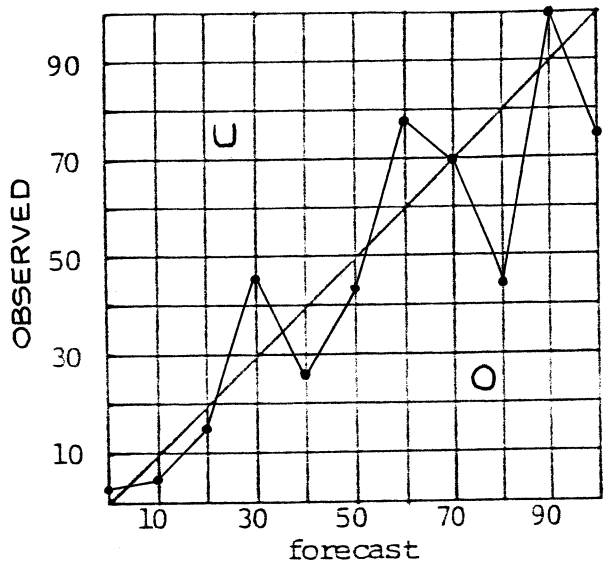
	5AM ISSUE	4PM ISSUE
TODAY	-31	
TONIGHT	-82	-55
TOMORROW	-67	-70
3RD DAY		-89

FIGURE 1.

STATION.. ____________ T O D A Y MONTH.. 5am/JULY

OBSERVED	forecast A	B	C	
RAIN	*25	3	24	52
NO RAIN	11	4	*122	137
	36	7	146	*189 T

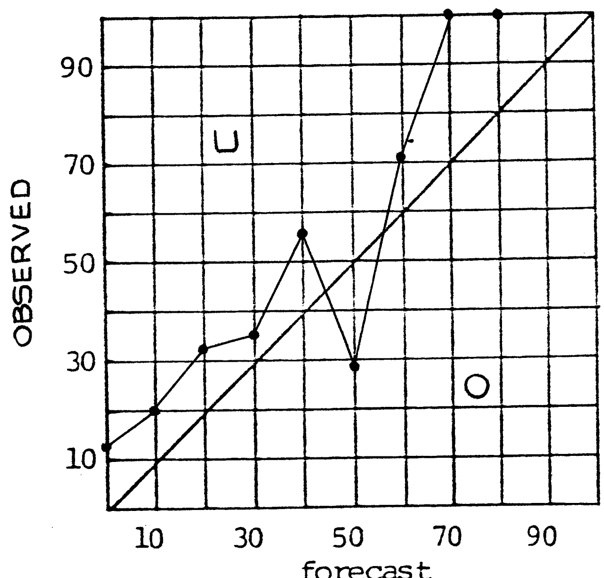
% correct = (A*+C*)/T*
= 78



T O N I G H T

OBSERVED	forecast A	B	C	
RAIN	*7	2	41	50
NO RAIN	2	5	*132	139
	9	7	173	*189 T

% correct = 74



T O M M O R R O W

OBSERVED	forecast A	B	C	
RAIN	*12	4	47	63
NO RAIN	9	0	*117	126
	21	4	164	*189 T

% correct = 68

A = prob > 50%
B = prob = 50%
C = prob < 50%

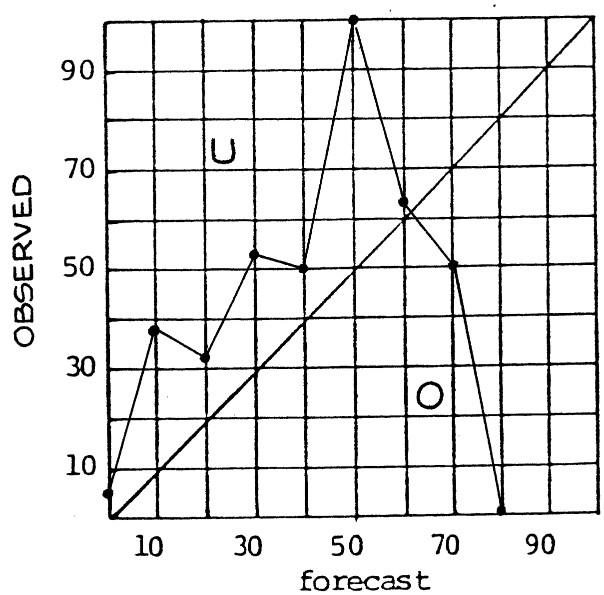


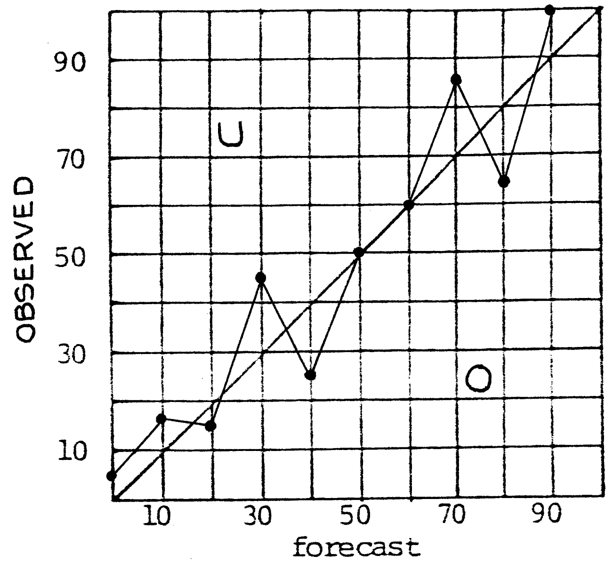
FIGURE 2.

STATION.. ______ TONIGHT MONTH.. 4PM/JULY

forecast

OBSERVED	A	B	C	
RAIN	*15	1	31	47
NO RAIN	6	1	*127	134
	21	2	158	*181 T

% correct = (A*+C*)/T*
= 78

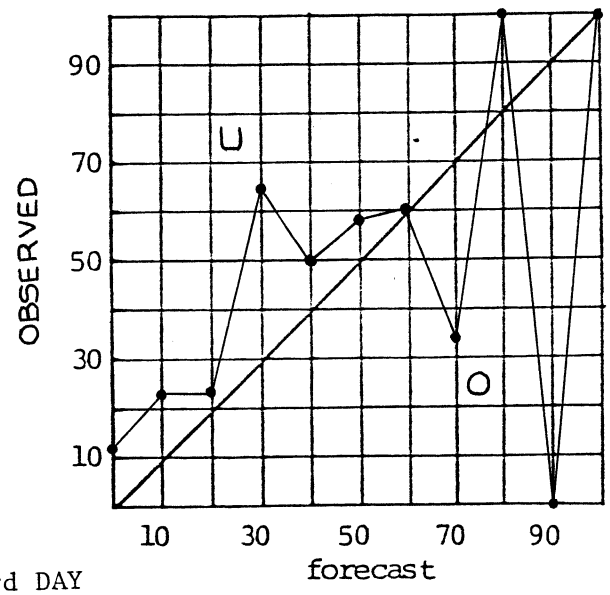


TOMORROW

forecast

OBSERVED	A	B	C	
RAIN	*8	7	45	60
NO RAIN	10	5	*106	121
	18	12	151	*181 T

% correct = 63

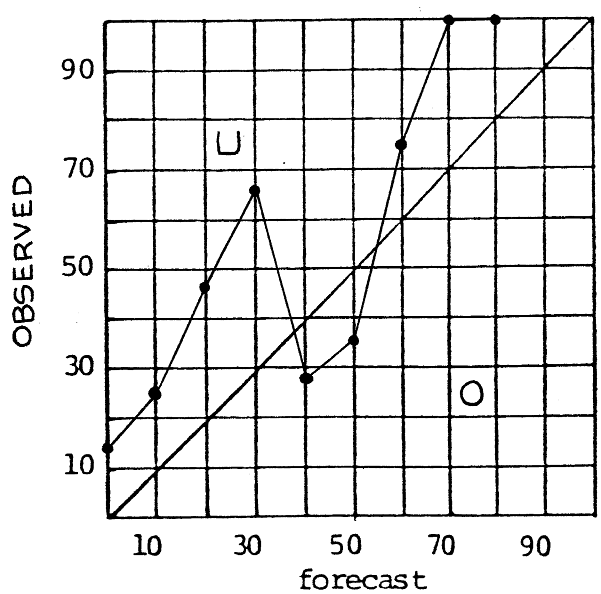


3rd DAY

forecast

OBSERVED	A	B	C	
RAIN	*6	4	54	64
NO RAIN	1	7	*109	117
	7	11	163	*181 T

% correct = 64



A = prob > 50%
B = prob = 50%
C = prob < 50%