NEW ZEALAND METEOROLOGICAL SERVICE

TECHNICAL NOTE NO. 177

SKILL SCORES ACHIEVED IN N.W.F.C. PRECIPITATION FORECASTING

Donald C. Thompson

Issued by:

The Director,
New Zealand Meteorological Service,
P.O. Box 722,
WELLINGTON

22 July 1968

NEW ZEALAND METEOROLOGICAL SERVICE

SKILL SCORES ACHIEVED IN N.W.F.C. PRECIPITATION FORECASTING

Donald C. Thompson

Abstract

The skill in predicting climatological raindays for the periods 12-36 hours and 36-60 hours from forecast time has been evaluated from the 12.30 p.m. "Dominion Forecasts" for the cities Auckland, Wellington, Christchurch and Invercargill for the years 1963-1966. Using an index suggested by Hanssen and Kuipers it was found that the skill was highest for forecasts for Auckland, lowest for Christchurch, and higher in winter than in summer. Raindays were forecast rather more often they actually occurred for the 12-36 hour period, particularly for Christchurch. Skill in forecasting deteriorations in weather was nearly equal to that for the continuation of wet or dry spells, but changes from wet to dry were not well predicted at all. Skill for the 36-60 hour forecasts was lower than that for the 12-36 hour forecasts but was still of a usefully high level. It was concluded that although N.W.F.C. precipitation forecasts have a high potential economic utility to users, this utility cannot be fully realised without the introduction of more precise terminology into the forecasts.

...

Introduction

Weather forecasts are sometimes incorrect, and studies of the basic predictability of the weather (Lorenz 1963) indicate that to some extent this may always be so. In spite of this limitation it is still possible for many users to obtain considerable economic gains from intelligent use of the forecasting services. This study was made as part of an attempt to evaluate the utility of public forecasts in influencing economic decisions, since intelligent use implies a knowledge of forecasting skill. Although the study was conducted primarily from the point of view of the user of weather forecasts it was expected that the results would be of interest and help to the forecasters themselves. With this in mind the study was extended to see if the forecasts were influenced by conditions prevailing at the time they were issued, and to compare the skill scores with those obtained with climatological and persistence methods.

Precipitation was chosen because it is the most important weather element in most activities in New Zealand, although it is also one of the hardest to forecast. There are other important elements in a forecast so this study does not claim to be a complete evaluation of forecasting skill, except in so far as one would expect the skill for these other elements to be highly correlated with that for precipitation.

A major obstacle in obtaining a quantitative measure of forecasting skill (and indeed in using public forecasts in any systematic way) is the fact that these forecasts are expressed in undefined, semi-qualitative terms. These general terms are used because the forecasts apply to relatively large areas and/or cover relatively long periods of time. There is also the need to cope with the natural inhomogeneity and uncertainty of the weather. To obtain quantitative values from the forecasts for verification against data actually recorded a certain amount of interpretation is necessary. This same kind of interpretation must also be carried out by the user before he makes his decision. Any systematic treatment of the problem would boil down to the setting up of arbitrary

...

quantitative equivalents of the various qualitative terms used in the forecasts. Since the interpretation could differ from user to user, and from the actual intention of the forecaster, it was considered desirable in this study to choose a predictand which involved as little of this interpretation as possible.

Thus this study was concerned only with the skill in forecasting climatological raindays (0.01 inch or more), rather than a more practical example such as the forecasting of days with more than say 10 points of rain. It is felt, however, that this compromise is not a serious restriction on the usefulness of the results.

Interpretation and Verification of Forecasts

In line with the introductory remarks, the following problem was chosen for study. Hypothetical individuals in each of the four cities Auckland, Wellington, Christchurch and Invercargill wish to know if the next day (commencing at midnight) will be a rainday and also if the day after (commencing midnight) will be a rainday. The individuals listen to the 12.30 p.m. "Dominion Forecast", or equivalently read the press forecast in the evening paper, since this is based on practically the same information as the radio forecast. We wished to evaluate the forecasting skill which these individuals may expect.

The "Dominion Forecast" is based on analyses and prognoses prepared during the latter part of the morning hours, from about 0800 hours onwards, but a number of weather reports as late as 1200 hours are available before broadcast time. Therefore the range of the forecasts to be verified was 12-36 hours in advance for the main forecast, and 36-60 hours in advance for the "further outlook". It will be noted that there is a gap of some 12 hours between the time of issue of the forecast, and the time when verification is to commence. This gap lessens the chance that rain falling at or near the time of the forecast would have continued to fall in the early part of the verification period.

Although whether or not a forecast was for a rain day was one of the simplest questions to be answered, there were

...

still difficulties. For example suppose the forecaster expected showers in the general area but could not promise a shower in any specific location. He might have said "isolated showers", "scattered showers" etc. to convey this meaning. Since "scattered showers" conveys a fairly high degree of probability of having one or more points of rain in a 24 hour period, and "isolated showers" a fairly low probability, it was assumed that in this case the individual would make the arbitrary decision to count the former as a forecast of rain and the latter as a forecast of no rain.

Similar arbitrary decisions were made for other non-categorical statements commonly occurring in the forecasts. Fortunately there were only a small number of borderline cases and once these decisions were made it was fairly simple to interpret the forecasts objectively. The resulting set of rules used is given in the Appendix. In making these decisions, and in generally interpreting the forecasts, the author attempted to adopt the viewpoint of an ordinary member of the public rather than that of a meteorologist.

Since weather fluctuates a great deal from season to season and from year to year it is necessary to consider a fairly large sample to get a valid measure of forecasting skill. In this study, four years' data were used, 1963-1966 inclusive, containing over 1450 forecasts for each of the four cities. Of these only about 5% of the 12-36 hour forecasts and 14% of the "further outlooks" could not be clearly interpreted. Some of these, particularly the outlooks, were excessively uncertain or ambiguous, but a large number were rejected because there was uncertainty as to whether or not the rain was expected in the forecast period. For example consider "a period of rain overnight" when the forecast period ends at midnight.

For each day and for each city the interpreted forecasts and further outlooks were expressed as a binary number, with 1 representing a forecast of rain and 0 a forecast of no rain. The same terminology was used for the verification data, which was obtained by referring to tabulated hourly raingauge data for Albert Park, Kelburn,

...

Christchurch Airport and Invercargill Airport. As an index of whether or not it was raining at the city in question when the forecast was being prepared a third bit was introduced into the verification which was 0 if there was less than 1 point of rain between 0600 and 1200 NZST on the day of issue and 1 if there was 1 point or more. Thus each city was assigned a five digit number each day in which the first two digits referred to the forecast and outlook respectively, the third digit to the weather on the morning the forecast was prepared and the last two to the actual rainfalls in the 12-36 and 36-60 hour forecast periods. Where the forecast or outlook could not be interpreted according to the pre-established rules an X was used instead of 0 or 1.

Different permutations of these five digits represent different combinations of correct and incorrect forecasts. By finding totals of the number of occurrences of each of these permutations and then combining certain of these totals according to their meanings it was possible to compute the required skill scores, together with much additional information.

Results and Discussion

In the following sections the 12-36 hour forecasts will be referred to simply as "forecasts" and the 36-60 hour forecasts as "outlooks".

(a) Percentage correct

The simplest measure of forecasting skill is the percentage of forecasts which proved correct. Table 1 shows this quantity for the four stations for all the forecasts analysed, together with the total number of forecasts.

...

Table 1

Percentage of Forecasts which Proved Correct
(total number of forecasts in parentheses)

Auckland(AK)	Wellington(WN)	Christchurch(CH)	Invercargill(IN)
72.5 (1403)	71.0 (1342)	67.0 (1333)	68.9 (1369)

(b) Hanssen Index

One disadvantage of the simple "percentage correct" figure as a measure of skill is that it can be misleading if, for example, the number of raindays is small compared with the number of fine days. In this case little is said about the accuracy of forecasting these raindays. Also one must know, for comparison, the number expected to be correct by chance. To avoid these difficulties a number of different "skill scores" have been devised by various authors, each with its particular advantage. In this work a skill score devised by Hanssen and Kuipers (1965) was adopted. (See also Gringorten, (1967), who presents a closely related index).

The "Hanssen Index" I is defined by

1 + I = Fraction of raindays correctly forecast (R)
+ Fraction of non-raindays correctly forecast (F)

Hanssen and Kuipers show that the expected value of I is 0 when the forecast and the actual conditions are statistically independent. For perfect forecasts I = 1. It will be seen that such

...

unskilful strategies as always forecasting rain, tossing a coin, or even drawing black and white balls from an urn containing these balls in the proportion of the climatological frequency of rain will expect to be rewarded with a score of 0. Note that the fraction of raindays correctly forecast is given equal weight to the fraction of non-raindays (referred to hereafter as fine days) correctly forecast. For specific economic operations this equality might not be ideal; for example if there is a heavy loss associated with an unforecast rainday compared with an unforecast fine day one might wish to give the forecasting of raindays a higher weight. On the other hand some users might be more interested in the forecasting of fine days so it could be unfair for a meteorologist to impose any arbitrary weighting. The Hanssen Index does, however, weight successful individual forecasts of rain in inverse proportion to the frequency of occurrence of rain. Thus, if rain were a comparatively rare event the index would gain more from a successful forecast of rain than from a successful forecast of a fine day. Many writers on the economic utility of weather forecasts have asserted that in general the public adjusts its activities to the climatological frequencies of occurrence of the various weather phenomena. If this is so then it is more important to forecast the rare events so use of the Hanssen index as a measure of forecasting skill appears reasonably appropriate for the activities of the public as a whole.

The two components, R and F, are interesting in themselves and are included below in all cases.

...

Table 2

Hanssen Index for the Years 1963-1966, 12-36 hr Forecasts

		Summer	Autumn	Winter	Spring	Year
AK	R	.805	.742	.850	.800	.802
	F	.600	.682	.704	.620	.648
	I	.405	.424	.554	.420	.450
WN	R	.643	.652	.750	.689	.692
	F	.735	.734	.733	.707	.726
	I	.378	.386	.483	.396	.418
CH	R	.636	.500	.670	.676	.621
	F	.671	.721	.725	.719	.693
	I	.307	.221	.395	.395	.314
IN	R	.771	.834	.788	.829	.820
	F	.508	.478	.573	.496	.500
	I	.279	.312	.361	.325	.320

By comparison with Table 1 the higher degree of skill of forecasts for Auckland compared with those for Christchurch is much more apparent. For Auckland and Invercargill raindays are relatively more often correctly forecast than fine days, but the reverse is the case for Christchurch and Wellington for the year as a whole. The rather poor forecasting of fine days at Invercargill should be noted.

The variation of forecasting skill with season is also shown in Table 2. For all four centres the skill score was highest in winter and,

...

except for Christchurch, lowest in summer. The skill for Christchurch was lowest in autumn, mostly due to the poor forecasting of raindays in that season. As noted above, fine days were relatively more correctly forecast than raindays for Wellington for the year as a whole, but this was reversed in winter because of a considerable improvement in the forecasting of raindays in that season.

(c) Climatological Frequencies

In the absence of any forecasting skill, one strategy is to always predict the climatologically most probable event. Table 3 shows the relative frequencies of raindays and fine days for each season for the four year period.

Table 3

Relative Frequency (per cent) of Raindays and Non-Raindays

		Summer	Autumn	Winter	Spring	Year
AK	Rain	43.0	45.0	61.3	48.8	49.6
	Fine	57.0	55.0	38.7	51.2	50.4
WN	Rain	37.6	45.2	63.8	46.0	48.4
	Fine	62.4	54.8	36.2	54.0	51.6
CH	Rain	29.2	32.0	36.3	33.2	32.5
	Fine	67.8	68.0	63.7	66.8	67.5
IN	Rain	56.6	61.6	58.9	59.2	58.6
	Fine	43.4	38.4	41.1	40.8	41.4

...

The larger of the two numbers would give the percentage of forecasts correct using this method, although of course the Hanssen Index would be zero. It will be seen that this percentage for Christchurch for the year is slightly in excess of the percentage of successful 12-36 hr forecasts using conventional methods but it can hardly be deemed a better method, because no raindays would have been forecast.

(d) Persistence Methods

Another "strategy" is to make use of the fact that the conditional probability of a day being a rainday, given that the previous day was a rainday, is greater than the climatological probability of rain. This "persistence" of raindays has been discussed for some New Zealand stations by Finkelstein (1967). In the present application it is assumed that a forecaster using this method would predict a rainday if one point or more of rain had fallen at the city in question between 0600 and 1200 NZST on the morning that the forecast was prepared. Because of the 12-hour gap between this time and the time when verification starts one would expect the degree of persistence to be somewhat less than that discussed for consecutive days by Finkelstein.

The use of persistence is not an entirely unskilful strategy in that a prior knowledge of the existence of this property is required, although it hardly justifies the need for a professional meteorologist. A Hanssen Index greater than zero is therefore expected and this is verified in Table 4 which shows the index which would have been achieved by this method of forecasting over the four years, and also the percentage of days correctly forecast.

...

Table 4

Hanssen Index and Percentage Correct by Persistence Method

		Summer	Autumn	Winter	Spring	Year
AK	R	.297	.314	.493	.296	.362
	F	.950	.811	.820	.819	.846
	I	.247	.125	.313	.115	.208
	%					60.5
WN	R	.350	.382	.514	.404	.426
	F	.867	.869	.810	.859	.855
	I	.217	.251	.324	.263	.281
	%					64.7
CH	R	.242	.324	.288	.219	.266
	F	.904	.875	.923	.880	.879
	I	.146	.199	.211	.099	.145
	%					68.0
IN	R	.410	.414	.429	.367	.405
	F	.843	.750	.732	.738	.771
	I	.253	.164	.161	.105	.176
	%					55.7

...

Comparison with Table 2 shows that the skill by persistence methods was considerably less than that of the conventional forecasts. Except for Wellington the skill of "persistence forecasting" was least in the spring. As might be expected, fine days were relatively more often correctly forecast by persistence methods than raindays. The percentage of forecasts correct was slightly higher for Christchurch than that achieved by the conventional forecasts, but for the other three centres it was considerably lower.

(e) Bias in Forecasting

A question of interest which arises is whether there was any tendency to forecast raindays more or less often than they actually occurred. Table 5 shows the ratio of the number of raindays forecast to the number of actual raindays for the four year period, as a percentage.

Table 5

Ratio of Forecast to Actual Raindays (expressed as a percentage)

	Summer	Autumn	Winter	Spring	Year
AK	134	113	104	120	116
WN	115	105	98	111	106
CH	142	109	129	124	126
IN	120	116	116	118	117

It is seen that there was in fact a bias towards the forecasting of raindays. This was rather small for Wellington except in the summer but it was moderate for Auckland and Invercargill and rather marked for Christchurch. This bias was greatest in the summer and least in winter.

...

(f) Forecasting Changes

The forecasting of changes of weather is generally regarded as more difficult than the forecasting of the continuation of dry or wet spells. It was possible in this study to consider the cases where the weather at forecast preparation time (as defined by whether or not one point or more of rain had fallen between 0600 and 1200 NZST) differed from the weather in the verification period. Table 6 shows the total number of such changes that occurred over the four year period and, of these, the number and percentage that were correctly forecast.

Table 6

Forecasting of Weather Changes

	Type of Change	Total Number	Fcsts Correct	Ratio(%)
AK	Fine to rain	444	329	74
	Rain to fine	108	41	38
WN	Fine to rain	373	257	69
	Rain to fine	150	66	44
CH	Fine to rain	318	184	58
	Rain to fine	109	42	39
IN	Fine to rain	477	375	79
	Rain to fine	129	44	34

...

With the exception of Christchurch a reasonably high percentage of the changes to rain were correctly forecast, but changes from rain to fine were not well forecast at all. Analysis of the forecasts showed that when a forecast for a change from rain to fine was issued the chance of it proving correct was about 65% for all four centres. When forecasts of changes from fine to rain were issued the chances of them proving correct were about 65% for Auckland and Invercargill but only 50% for Wellington and 47% for Christchurch.

It was possible that there may have been a systematic bias by forecasters for or against changes and to study this the material presented in Table 7 was extracted. This table shows the number of changes which actually occurred compared with the number of forecasts issued for the same types of changes in the four-year period. The corresponding figures when there was no change are also given. A significant difference in the two numbers would indicate a bias for or against that particular sequence on the part of the forecasters.

There seems no evidence for any significant tendency towards simple "persistence forecasting", for although rain was forecast to continue somewhat more often than actually happened the reverse was the case with the continuation of fine weather. Forecasts of changes from fine to rain were issued rather more often than such conditions occurred, particularly for Christchurch. For all cities there is evidence for a marked reluctance on the part of forecasters to predict an improvement, this tendency being greatest for Auckland.

...

Table 7

Frequencies of Different Weather Sequences which were Forecast, compared with the Frequencies which Actually Occurred

	Sequence	Fcst	Actual	Ratio (%)
AK	No Change, (N.C.) Rain	297	252	118
	" " , Fine	532	599	89
	Change (C) to Rain	511	444	115
	" " Fine	63	108	58
WN	N.C., R	277	277	100
	N.C., F	553	592	93
	C to R	412	373	110
	C to F	100	150	67
CH	N.C., R	152	115	132
	N.C., F	716	791	90
	C to R	393	318	124
	C to F	73	109	66
IN	N.C., R	367	325	113
	N.C., F	342	438	78
	C to R	573	477	120
	C to F	87	129	67

...

(g) 36-60 hour "Further Outlooks"

Similar analyses were possible for the outlooks. In view of the generality of the wording used in this part of the "Dominion Forecast", however, a more abbreviated treatment was considered in order. The number of forecasts was slightly less because as mentioned above, a rather greater number could not be interpreted in the desired manner.

Table 8 shows the Hanssen Index, the percentage of predictions which proved correct, and the ratio of the number of predictions of raindays to the actual number of raindays.

Table 8

36-60 hour "Further Outlooks"

	Hanssen Index			Per Cent Correct	Forecasts Rain / Actual Raindays (%)
	R	F	I		
AK	.686	.586	.272	62.5	114
WN	.575	.675	.250	63.0	100
CH	.360	.814	.174	67.1	76
IN	.685	.545	.230	62.5	104

As would be expected, the Hanssen Indices were considerably smaller than those for the 12-36 hour forecasts, but the skill is still appreciable. For example, the indices were generally higher than those obtained by "persistence forecasts" for the 12-36 hour period. Thus, the 'outlooks' contain significant, useful information. It is interesting to note that the forecasting of fine

...

days for Christchurch and Invercargill was actually improved with respect to the 12-36 hour forecasts. For these two stations in particular there was considerably less tendency for the forecasts to be biassed in favour of raindays, in fact the bias apparent in the 12-36 hour forecasts was heavily reversed for Christchurch. The improvement in fine day prediction was undoubtedly at the expense of the prediction of raindays.

Conclusions

Before discussing these results further it should be pointed out that many of the forecasts would have been counted as incorrect in this study when it could be argued that they were in fact satisfactory. For example we might legitimately expect a small percentage of the cases where "isolated showers" were forecast to produce rain, whereas in this study we have always taken "isolated showers" to mean a forecast of no rain. A detailed analysis of this has not been made but study of a few months' forecasts for Wellington suggests that the number of correct forecasts could be increased by between 4 and 8% to take these situations into account. However, as we have pointed out any individual wishing to make a decision based on the forecasts must adopt a set of arbitrary interpretation rules of the type used above, so that he would expect effective skill levels similar to the ones obtained.

It has also been argued that the forecasts as issued refer to extensive areas and that it is unfair to verify them at single stations. This may be so, but most people can only be in one place at a time, so the great majority of users are only concerned with conditions in a very limited area. Only a very few users, such as catchment authorities, are interested in true area forecasts of precipitation. It is perhaps appropriate to remark here that after having read a large number of forecasts it is clear that the larger the area that the forecaster attempts to include in a single forecast statement the more difficult it will be

...

to deduce the conditions expected at a single point within the area.

It would be of interest to compare the skill scores with those obtained by other Meteorological Services. There is however very little published information available and nothing directly comparable. It is only possible to make the qualitative statement that while there is obviously room for improvement the quality of the precipitation forecasts is of the same order as obtained overseas.

The skill scores show the precipitation forecasts to be potentially useful to the public from an economic point of view. Because of the difficulties in interpreting the forecasts in their present form it is evident however, that their utility would be greatly enhanced for certain users if a more precise terminology were introduced.

The results of this study suggest that a useful quality control service would be provided by a continuous monitoring of forecasting skill. Such a service would be more meaningful if applied to forecasts framed in quantitative terms. Previous attempts at measuring forecasting skill for this function have involved too much subjectivity and have used arbitrary indices of skill which had only limited relevance to the needs of individual users.

There is evidence of a bias towards forecasting rain more often than it actually occurs. This may be partly due to the fact that many of the forecasts were conceived as "area forecasts", which would naturally cause precipitation to be mentioned more often than if only individual points were considered. Another explanation is that some forecasters consider it desirable to be a little "pessimistic" in precipitation forecasts, to minimise the number of unforecast raindays.

Weather changes from fine to wet were forecast about equally as well as were the continuation of wet or fine spells, but changes from wet to dry weather were not well forecast at all. Further study should be made of this latter aspect of the forecasting problem.

...

The accuracy of precipitation forecasts for Christchurch was poor, particularly in autumn. This fact was evident in earlier unpublished estimates of forecasting skill and further study must obviously be devoted to the problem.

Acknowledgement

The author wishes to thank Mr. E.W. Laws for his assistance in the rather considerable task of reading, interpreting and verifying the forecasts.

References

Finkelstein, J., 1967: Persistence of Daily Rainfall at some New Zealand Stations. N.Z.J. Hydrology, 6, 33-45.

Gringorten, I.I., 1967: Verification to Determine and Measure Forecasting Skill. J. Appl. Meteor., 6, 742-747.

Hanssen, A.W., and W.J.A. Kuipers, 1965: On the Relationship between the Frequency of Rain and Various Meteorological Parameters. Mededelingen en Verhandelingen No. 81, Netherlands Meteorological Institute.

Lorenz, E.N., 1963: The Predictability of Hydrodynamic Flow. Trans. N.Y. Academy of Sciences, Series II, 25, 409-432.

APPENDIX : RULES USED IN INTERPRETING FORECASTS

1. Isolated showers $\equiv$ no rain (0) .

2. Some/occasional/scattered, showers $\equiv$ rain (1) except where scattered showers are specified as principally confined to areas which do not include the station in question. In the latter event put 0 for scattered showers.

3. "Areas of drizzle" or "extensive drizzle" expected to last a few hours $\equiv$ rain (1).

4. Scattered drizzle/patches drizzle/isolated drizzle $\equiv$ no rain (0) except if there is some additional precipitation mentioned, e.g. "patches morning drizzle clearing but isolated afternoon showers" $\equiv$ rain (1).

5. Some/areas of/scattered/patches of, rain $\equiv$ rain (1), unless specifically confined to areas not including the station in question.

6. "Probability" of rain or showers taken as a positive indication unless there is further qualification.

7. Use "X" when

   (a) Wording too vague or imprecise. (e.g. "unsettled conditions").

   (b) A low, poorly defined probability of rain is implied (e.g. "possibly a period of showers with the wind change"). See, however, 6 above.

   (c) The timing of the precipitation relative to the forecast period is in doubt. (e.g. "a period of rain tonight" may be before or after midnight).

   (d) There is excessive doubt about intensity, location or duration of precipitation.

8. In general, attempt to evaluate the impression conveyed to a member of the public rather than to a meteorologist.