551.509.333(931)

NEW ZEALAND METEOROLOGICAL SERVICE

TECHNICAL INFORMATION CIRCULAR NO. 189

REPORT ON THE EXTENDED RANGE WEATHER FORECAST PROJECT - PHASE I

A.B. Mullan, A.A. Neale and C.G. Revell

J S Hickman

J.S. Hickman
Director

New Zealand Meteorological Service
P.O. Box 722
WELLINGTON

28 April 1983

REPORT ON THE EXTENDED RANGE WEATHER FORECAST PROJECT - PHASE I

A.B. Mullan, A.A. Neale & C.G. Revell

Abstract

The objective of Phase I of the Extended Range Weather Forecast Project was to test the feasibility of predicting the general character of the weather during the three-day period (D+3, D+4 and D+5) beyond that for which detailed forecasts are normally issued. This report explains the prediction method, and reviews the accuracy of the 54 forecasts prepared during Phase I.

Forecasts were assessed both subjectively by verifying the broad-scale surface flow pattern in the New Zealand region, and objectively by verifying categories of rainfall amount and maximum temperature prepared for six regions of New Zealand. Both verification methods showed evidence of skill in the first two-day period (D+1, D+2) and reduced but still useful skill in the extended outlook three-day period. The categorical forecasts were generally too conservative, with heavy rain and temperature extremes being underforecast, and there was no strong relationship between accurate categorical forecasts and accurate flow-pattern forecasts.

SECTION I: INTRODUCTION AND METHOD

Introduction

It is unnecessary to reiterate the potential value of skilful extended outlook forecasts to many sectors of the community - farmers, fishermen and contractors, to mention a few.

For many years extended range forecasts have been issued daily to offices of the Meteorological Service. The accuracy of these in terms of 'rain/no rain' in the various districts has been evaluated (Laidlaw, 1983). Results indicate that the overall skill is not high for forecasts for three days ahead. By comparison, high skill has been achieved only when predicting negligible rainfall over a number of four-day periods which happen to follow the occurrence, in summer of specified criteria related to the synoptic situation. However, the latter method does not predict all dry spells which occur.

In the majority of cases, poor accuracy can be attributed primarily to the inability of forecasting techniques currently in use to predict the behaviour of the smaller scale weather systems (e.g. fronts, vorticity centres) - both in terms of location and intensity - beyond the span of the next two days.

With the establishment of the satellite observing sub-system, which led in particular to the routine acquisition of data indicating the thermal structure of the troposphere, analyses of the Southern Hemisphere have become much more accurate than hitherto. The receipt, from the World Meteorological Centre in Melbourne, of grid data for the 500 hPa* level from these analyses, and the subsequent processing of the data within our own computer, has provided the means of monitoring the behaviour of wave patterns in the flow round the hemisphere. This, in turn, has enabled extended outlook forecasts to be tackled from a new point of view.

Wave pattern at 500 hPa level

The pattern of waves (trough/ridge pairs), which occurs in the wind flow at the 500 hPa level, can be decomposed by Fourier analysis into components having different wavelengths and amplitudes. The longest wave has a wavelength of 360 degrees of longitude (i.e. around the hemisphere there is one trough/ridge pair). The next longest wave has a wavelength of 180 degrees (two trough/ridge pairs around the hemisphere), and so on.

It is simpler to refer to waves by the number which occur around the hemisphere than by the wavelength. For example, wave 4 refers to those with a wavelength of 90 degrees, there being four waves around the hemisphere. In the absence of an accepted terminology to describe various groups of waves, the following has been adopted:

Waves 1 to 3: Planetary waves

Waves 4 to 9: Medium waves

Collectively these waves are commonly called 'long waves'. They have a number of characteristics which prove useful when making predictions about their future behaviour:

(i) Shorter waves (higher numbers) tend to move faster than longer ones (lower numbers), and consequently they overtake and move through the longer waves;

(ii) The behaviour of the shorter waves tends to be influenced by the longer wave pattern through which they move; and

(iii) There is a tendency for the behaviour of the longer waves to change more slowly than that of the shorter waves.

In practice the situation is complicated by waves interacting with each other or with the zonal mean flow (wave 0). Nevertheless, reasonably coherent trough/ridge patterns are obtained by considering the combined contribution from a group of waves, and the partition into quasi-stationary 1-3 and (usually) mobile 4-9 seems the most appropriate.

---

* 1 hectopascal (hPa) = 1 millibar (mb)

Relationship between waves and weather

The 500 hPa data is considered from two different points of view. What we would prefer to predict are the physical synoptic features which appear on weather charts. What we are generally dealing with are the Fourier components of the wave field, which individually are not physical entities. Taking all components together would recreate the original synoptic field, but by retaining only the long-wave components we arrive at a smoother pattern which hopefully is more predictable.

Weather conditions (wind, cloudiness, rain, etc.) are closely related to the synoptic systems. The latter can undergo large variations of intensity and location within a few days. Forecasting techniques available until recently have often failed to adequately predict such changes beyond about the next two days. This has been the reason for the generally poor performance of extended range forecasts.

The greater predictability of the Fourier components out to about five days ahead offers a means of being able to infer the general character of synoptic systems also. This results from the tendency for certain synoptic weather systems to develop and to be associated with the wave components; in other words with the long-wave pattern.

An example illustrates the type of relationship which exists between synoptic systems and long waves. When a synoptic low pressure trough overtakes a long-wave trough, the following sequence of events is commonly observed (Fig.1). As the synoptic scale trough approaches the longer-wave trough (located over the central Tasman Sea in Fig.1), convection becomes enhanced in the region of cyclonic vorticity associated with the synoptic trough. Commonly, on the southeastern side of this enhancement, deep convection leads to a merging of anvil cloud tops to form a dense cloud mass. Once the synoptic trough passes through to the eastern side of the longer-wave trough, the area of dense multilayered cloud becomes progressively more extensive.

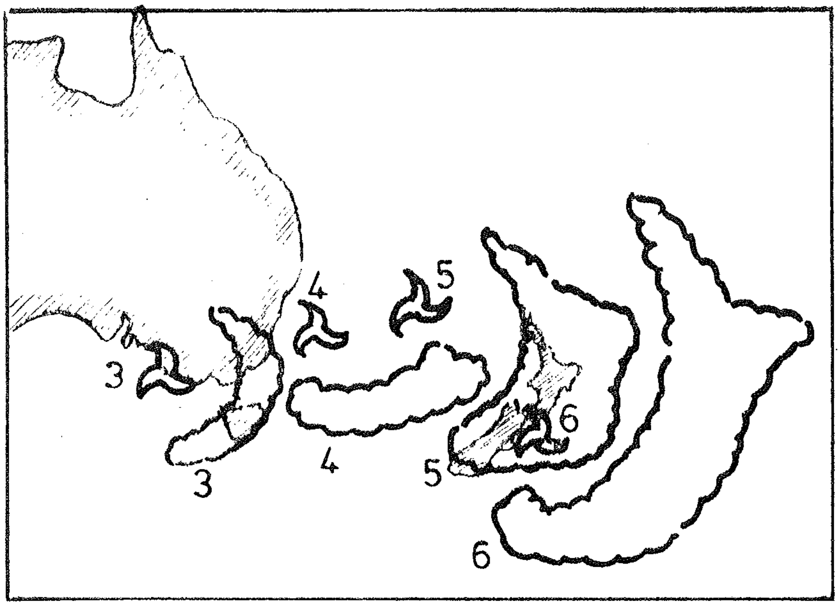

Fig.1.

Main components of the cloud system associated with a synoptic scale wave at 1200 GMT on the four days 3-6 July 1981. Positions of the vorticity centre (spiral symbol) and the area of dense multi-layered cloud are shown for successive days.

Looking at it from another point of view, regions located within and immediately downstream from a long-wave trough are favourable for the development of active cloud systems and cyclonic synoptic features. However, for these to occur, moisture must be present as well as an ascent mechanism (normally cyclonic vorticity advection and/or warm advection associated with a synoptic scale trough). While the existence of a long-wave trough favours the occurrence of active cyclonic synoptic systems (troughs, lows), anticyclonic synoptic systems (ridges, highs) can exist there also, but it would be unusual for them to be more than transient systems in that location. Similarly, long-wave ridges favour anticyclonic synoptic systems, but can permit the occurrence of transient troughs or lows.

The intensity of the long waves is important too. For instance, an intense long-wave ridge will strongly favour anticyclonic synoptic systems, and cause the decay of cyclonic systems which move into its region. Conversely weak long waves exercise no more than a slight influence on synoptic systems.

Sometimes conditions become so favourable for a synoptic scale system (say, a low) to intensify, that a very deep circulation develops with a greater than normal wavelength. It then takes on some of the characteristics of longer waves - greater persistence and often slower movement. Such developments cause sudden (and unexpected) changes to the overall long-wave pattern.

Analysis of the long waves

A number of methods are available for analysing the long waves.

(i) Time smoothing.

The height of the 500 hPa level is averaged over a 5-day period at each of a series of points spaced at 10 degrees longitude around a specified latitude circle. This technique filters out the shorter, faster-moving waves which influence a given grid point for no more than two or three days.

Figure 2 shows an example using this method. (For further information, consult Trenberth et al.(1978)). Because each 5-day running mean is credited to the middle day of the five, the most recent data on the diagram refers to two days ago. To try and overcome this deficiency, information for the last two days is provided by using the 3-day mean for the first of these, and the 1-day value for the last day. As illustrated in Fig.2, this device results in contamination of the longer waves, shown in the majority of the diagram, with the shorter waves. This leads to uncertainties when interpreting the most recent data.

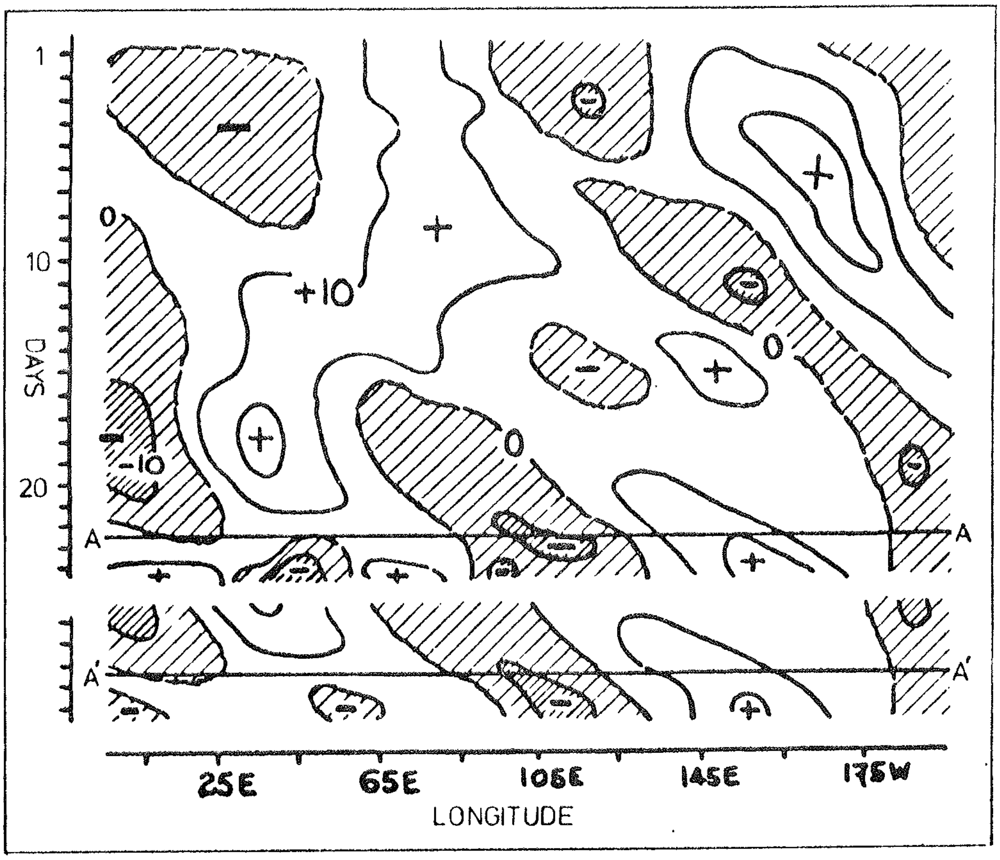

Fig.2. Hovmoller diagram of 500 hPa level at 55°S showing 5-day running mean heights given as departures (in decametres) from the zonal-average height for that day. Negative values hatched. Data for the two days below the line A-A come from 3-day mean, and 1-day values. The effect of including the shorter waves is seen by comparing the part of the diagram immediately below A-A with the repeat of that part of the diagram (shown at the bottom of the figure), where 5-day means are used for the portion below the line (shown as A'-A' in the repeated segment).

(ii) Fourier analysis.

The disadvantage just referred to can be overcome by performing a Fourier analysis. This leads to identification of the various waves, their location and amplitude, at a series of latitudes (in practice 25°S, 35°S, 45°S and 55°S). It then becomes possible to trace the behaviour of each wave, though such a procedure can be laborious. Hovmoller diagrams*, one showing the effect of the planetary waves (waves 1, 2 and 3 combined), the other showing the effect of the medium waves (waves 4 to 9 combined), are a convenient means of tracing the behaviour of the longer waves. This method of analysis was the one used in this project.

(iii) Filtered wave components.

By applying a filter to the wave component field, the effect of the shorter waves can be removed, leaving only the longer waves. This technique, employed by the Australian Bureau of Meteorology, results in a hemispheric chart of the long wave component field. It provides a simple means of visualising variations in the combined long waves over all latitudes and longitudes from day to day.

Method used to predict the long waves

Prediction of the long waves relied principally on the Fourier analysis technique. A variety of data displays were used, including the two Hovmoller diagrams which show the behaviour of the planetary and medium waves at four latitudes. Another gives the sequence of positions of each wave trough at these four latitudes - a portion of this display is given in Fig.3. In this figure, lines have been drawn to show the movement of waves 3 and 4; the former tending to be stationary from day 10, while the latter progresses eastwards at an average rate of eight degrees of longitude per day. Other material consisted of tabulations of the contributions of each wave to the total variance of the 500 hPa height at each latitude; the strength of the westerly flow over various latitude bands and longitude segments; and hemispheric charts showing the sequence of daily positions and strengths of the planetary and medium-wave troughs.

The prediction technique used was similar to that practised by synopticians in pre-computer days: extrapolation of preceding trends while making adjustments to take climatological data into consideration. Since reliable hemispheric data has become available only within the last ten years, a climatology of the long waves is just being developed. Nevertheless some useful data are becoming available. For example, Fig.4 shows that the speed of the

---

* The term 'Hovmoller diagram' is applied to any diagram showing the behaviour of troughs and ridges on a time-longitude cross-section. It is applied irrespective of the method of smoothing used (time-average, latitude-average, Fourier analysis).

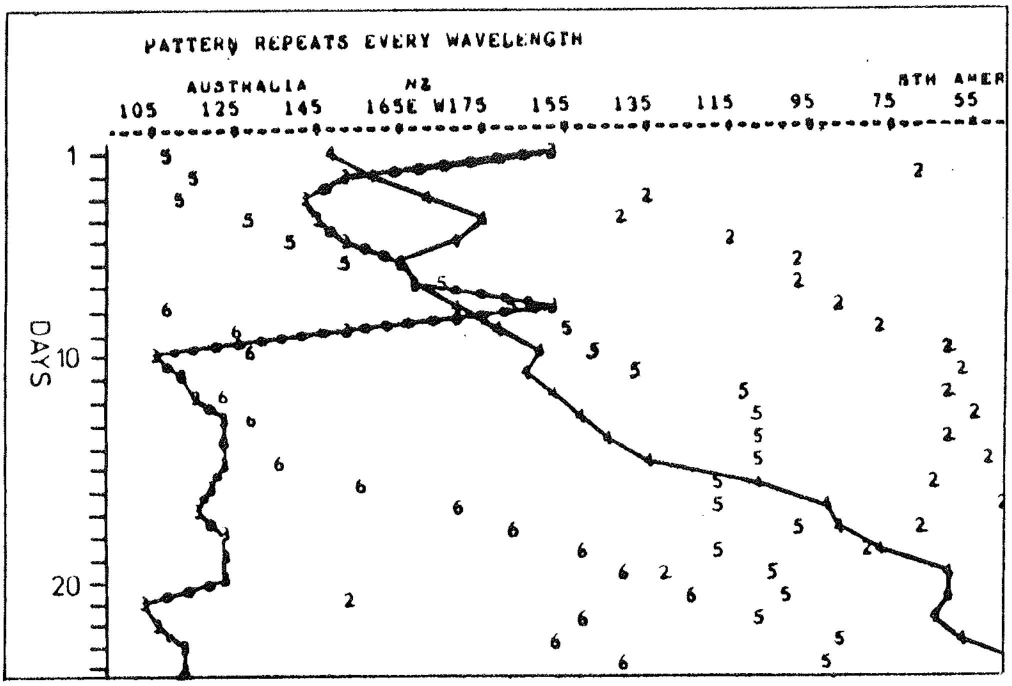

Fig. 3. Daily positions of the 500 hPa waves (at a given latitude) between 100°E and 50°W. From day ten, wave 3 remains stationary while wave 4 progresses eastwards at an average rate of eight degrees of longitude per day.

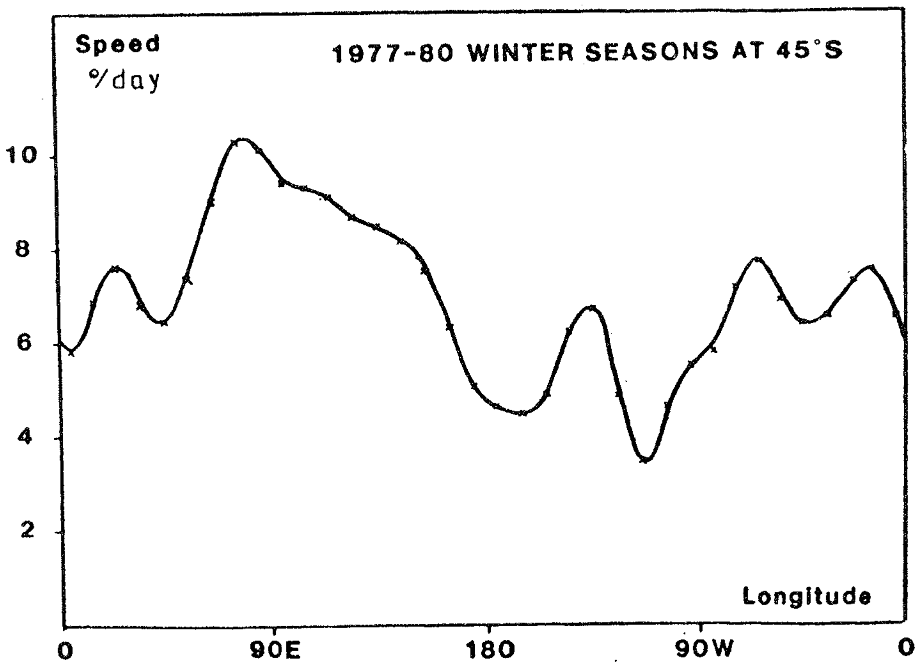

Fig. 4. (From Mullan, 1983). Mean speed (in degrees of longitude per day) of the medium waves at 45° south.

medium waves at 45°S in winter is likely to be high in the Indian Ocean (10 degrees of longitude per day on average), but to decrease eastwards to a low value of about 5 degrees per day in the western Pacific.

Because the wave number which is dominant at one latitude can be different from that dominant in an adjacent latitude, and because their speed of movement can vary too, troughs in the various latitudes do not always 'line up' coherently, nor do they always move as a stable pattern. While on some occasions the overall wave pattern is relatively simple, more often than not it is quite complex. In the complex situations it is found helpful to locate the most coherent and dominant features of the long-wave pattern - the anchor troughs and ridges - and, having decided the most likely future behaviour of these systems, extend the prediction into the more difficult areas of the hemisphere.

In practical terms, it is a case of determining the future position and strength, at various latitudes, of both the planetary and medium waves. Coincidence of like features of both wave trains gives the most favourable conditions for predicting synoptic weather systems. For instance, when a planetary wave ridge coincides with a medium wave ridge, there is a high probability that there will be an anticyclonic synoptic system in that region. However, occasionally such a forecast fails because a cyclonic system temporarily transits the region. It is difficult to cope with such transient features.

The juxtaposition of the ridges of the planetary waves with troughs of the medium waves, or vice versa, leads to uncertainty over which will dominate when wave strengths are comparable. In such circumstances it is uncertain what type of synoptic system is favoured to dominate the region.

Certain long-wave patterns tend to be stable; that is, they tend to persist for a number of days with little change. Forecasting is more reliable during the early part of such periods; also when the pattern is evolving into a stable mode. As might be expected, uncharacteristic changes to long-wave patterns are poorly forecast, and usually result in poorly predicted synoptic systems.

Satellite pictures, hemispheric 36-hour prognoses (surface and 500 hPa), and extended prognosis statements - the latter two received from Melbourne - were used to assist with the interpretation of, and provide confirmation of, Fourier analyses.

Timetable of forecasts and verifications

The 500 hPa hemispheric data used in the Fourier analyses, on which forecasts were based, is for 1200 GMT (denoted as D = 0). The timetable of activity is shown in Fig.5. Forecasts were prepared 12 hours after this data time, i.e. at 0000 GMT. Hence the forecast for the first 24 hours (or day) is centred on 1200 GMT on day 1 (D + 1). Numbering of days proceeds on this basis. Operational forecasts are issued an hour or so prior to 0000 GMT, and have a validity to 1200 GMT on D = 2, plus an outlook for a further 24 hours.

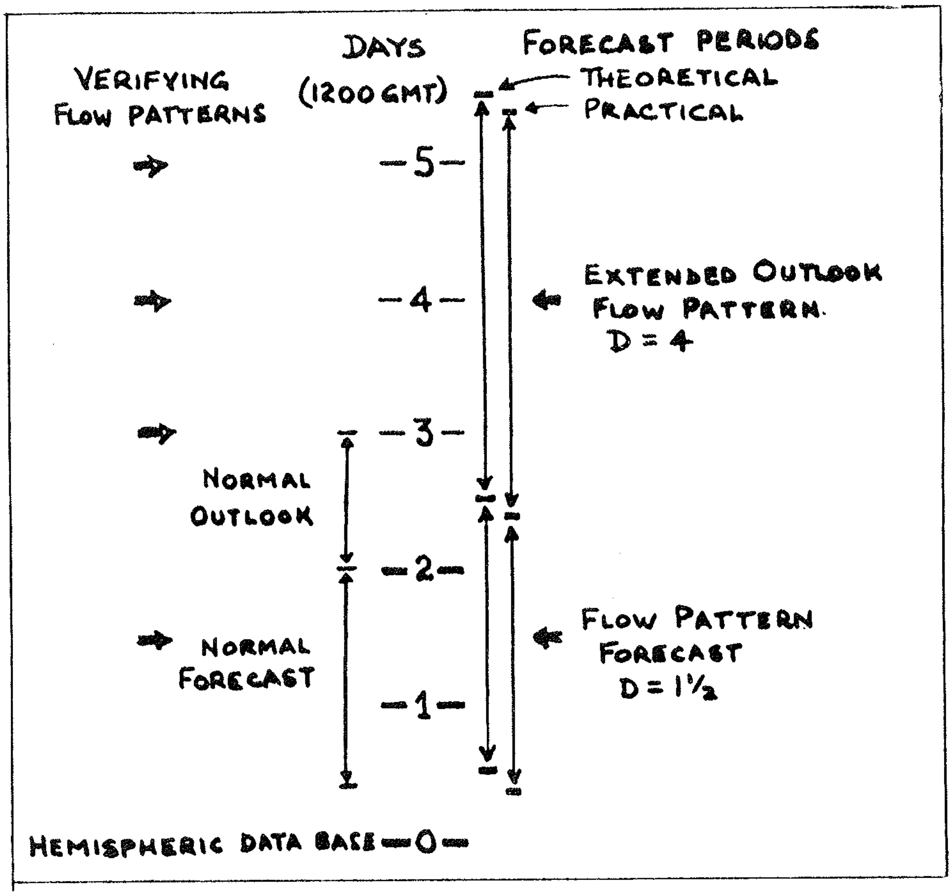

Fig.5. Timetable of activity showing normal operational forecast validity in addition to that of the extended outlook project. See text for additional explanation.

Project forecasts covered two periods, one for days 1 and 2, and the other days 3, 4 and 5. The first period corresponds to that for which forecasts are issued operationally, and so provides a bench mark of skill against which to gauge skill displayed in forecasts for the second period (which is largely beyond the period for which operational forecasts are normally issued). The second period is termed the 'extended outlook' period.

Forecasts were prepared on Mondays, Wednesdays and Fridays (Tuesdays and Fridays initially), and consisted of:

(i) A statement of the expected behaviour of the long waves out to D = 5.

(ii) Flow patterns in the New Zealand region for:

(a) D = 1.5 (0000 GMT). The mid-point of the first period. This was normally identical to the operational MSL prognosis for that time; and

(b) D = 4. The mid-point of the extended outlook period. An example of forecast flow patterns is given in Fig.6.

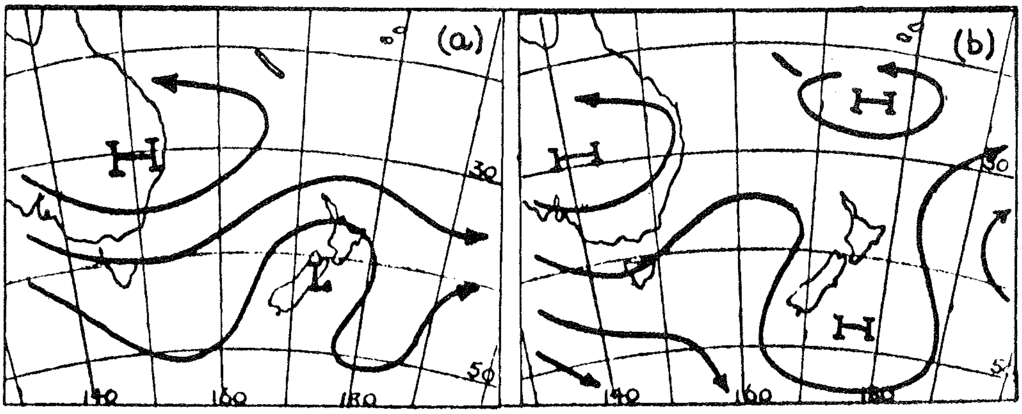

Fig.6. Forecast flow patterns for (a) D = 1.5; and (b) D = 4.

(iii) Categorical forecasts of rainfall in six regions of New Zealand for both periods; regions are shown in Fig.7. Categories are: 1= up to 3mm; 2 = more than 3 mm and up to 15 mm; 3 = more than 15 mm. The rainfalls forecast are averages of the total rainfall recorded during the period at the set of stations shown in Fig.7. The 'rainfall day' used in verifying forecasts ends at 2100 GMT, whereas the 'forecast day' ends at 0000 GMT (shown as practical/theoretical in Fig.5). Categorical forecasts of mean daily maximum temperature, following a procedure similar to that for rainfall. The categories are: 1 = more than 2°C below normal; 2 = no more than 2°C different from normal; 3 = more than 2°C above normal. Forecast values represent averages over the forecast period, and over the set of stations in each area.

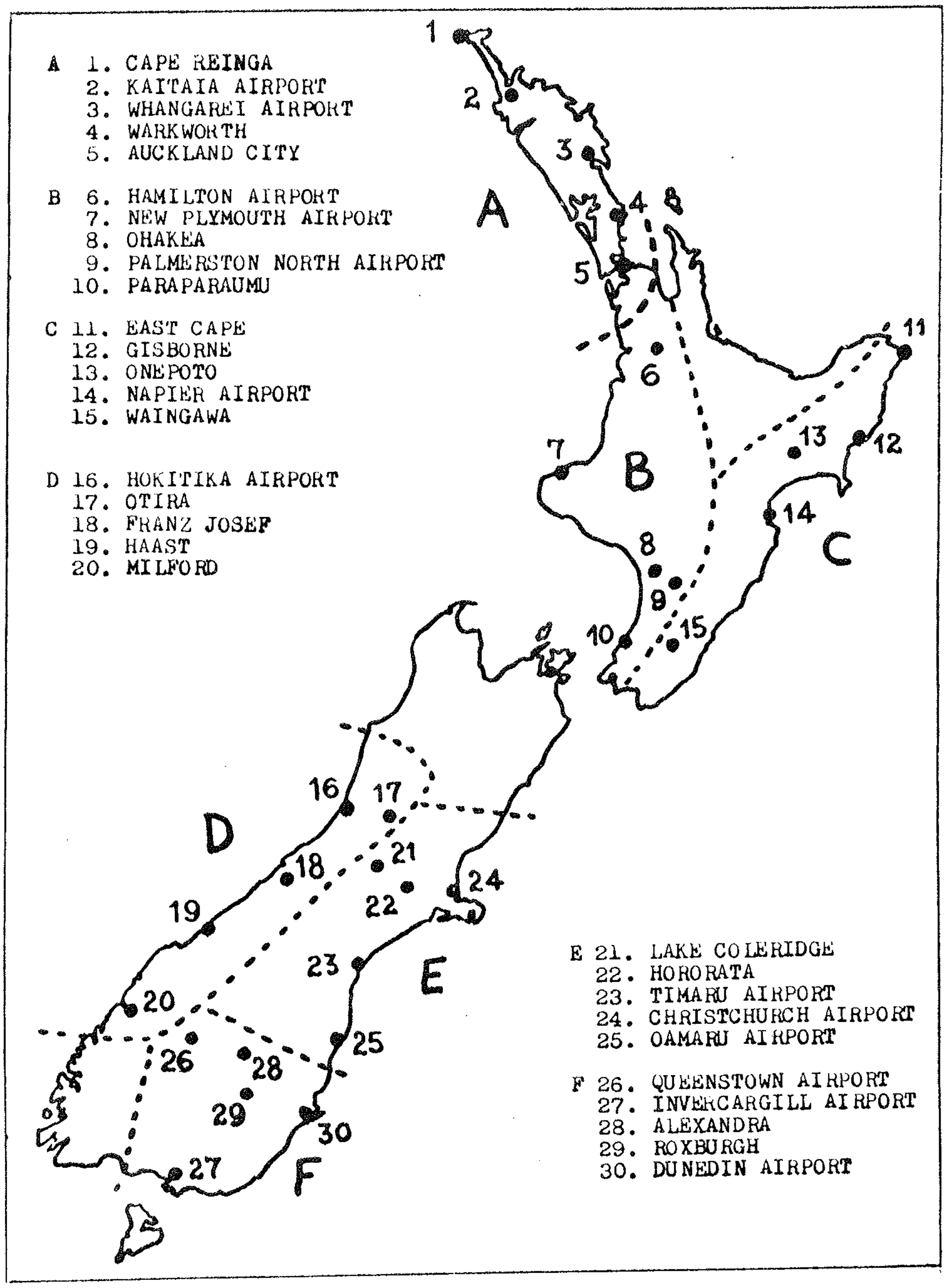

Fig.7. Dots mark the location of reporting stations used for verifying categorical forecasts for the six regions (5 stations per region). A, B and C are northern, western and eastern North Island; D, E and F are western, eastern and southern South Island.

(iv) An experimental statement describing the expected weather pattern during the next five days. The following is one of these statements.

> "Period 22-26 July 1981. A strong long-wave trough is located near west Australia, and a strong long-wave ridge occupies the Tasman Sea-New Zealand region. Slow eastward movement of these features appears likely. The low and frontal trough over east Australia are likely to decay as they move slowly east into the long-wave ridge, but a temporary northerly over southern New Zealand ahead of the trough could bring some rain there as the cloud decays. Ridging would be favoured behind this trough. The long-wave ridge is stronger at 45°S and 55°S than at 35°S, hence there is a possibility that pressures will fall by the end of the period over the north Tasman, while the high pressure axis lies southwest-northeast not far east of New Zealand."

Whereas the flow pattern for the first period ($D = 1.5$) is quite specific and is intended to represent conditions at that time, the pattern for the second period ($D = 4$) is intended to represent the broader scale flow, the short waves (which it has been noted cannot be forecast in detail this far ahead) having been ignored. Therefore, it is more likely that this pattern will represent the average flow over the three days (3, 4 and 5) than the instantaneous flow at $D = 4$.

During this phase of the project, fifty-four forecasts were prepared. The results of verification of these forecasts follow.

SECTION II: RESULTS AND CONCLUSIONS

Verification of forecasts

Although verifications of statements (i) and (iv), and flow patterns (ii) are subjective, they are nevertheless worth doing since any operational extended range forecast scheme will involve this type of material. By contrast, categorical forecasts (iii) can be verified objectively.

(i) Forecasts of the 500 hPa long-wave pattern were verified using a three step marking scheme:

A = substantially correct - good guidance (18 forecasts, 33%)

B = some errors but not seriously misleading - useful guidance (27 forecasts, 50%)

C = completely misleading - (9 forecasts, 17%).

Table 1, showing association of accurate categorical forecasts with verification marking of 500 hPa long-wave forecasts, indicates that there is only a weak relationship linking good long-wave forecasts with large numbers of correct categorical forecasts, and poor long-wave forecasts with small numbers of correct categorical forecasts. Indeed within each group of long-wave verifications, there are large differences in the number of correct categorical forecasts from one case to another. Hence, even long-wave forecasts which are completely misleading can be associated occasionally with a large number of accurate categorical forecasts.

Table 1 Association of accurate categorical forecasts with long-wave forecast marking in three steps, A, B and C (see text for explanation). Figures in brackets refer to rainfall forecasts only. On each day there are 24 categorical forecasts in total, 12 each for rainfall and maximum temperature.

Long wave verification	No. of cases	No. of categorical forecasts correct	
		Mean	Range
A	18	16.0 (7.5)	6 - 21 (4 - 10)
B	27	14.3 (6.5)	8 - 21 (4 - 9)
C	9	13.2 (5.8)	7 - 19 (2 - 9)

(ii) Verification of flow patterns:

A little objectivity was introduced into the verification by marking the accuracy of the flow direction and flow curvature at three points (North Cape, Cook Strait and Stewart Island). Directions were taken to the nearest of the eight main compass directions (e.g. N, NE,etc.), and curvature was taken as either nearly straight, or as cylonic or anticyclonic. Verifications were scored numerically in the following manner:

Direction:	Exact correspondence:	4 points;
	Error of 45 degrees:	2 points;
	Error of 90 degrees or more:	0 points;
Curvature:	Exact correspondence:	2 points;
	Maximum error (cyclonic when anticyclonic forecast or vice versa):	0 points;
	All other cases:	1 point.

In this scheme direction is given twice the weight given to curvature, and the pattern is verified only over New Zealand.

Verification of first period forecasts involved a simple comparison of the forecast flow with that which subsequently occurred. The verification of extended outlook forecasts was done in two ways. In one - the 'three-day method' - the forecast was verified separately against actual flows at D = 3, D = 4 and D = 5. The individual daily scores S_D were averaged to give $S = \frac{1}{3.5}(S_3 + 1.5\ S_4 + S_5)$. (The score for D = 4 was increased by 50% to take account of the fact that the forecast chart should be more applicable to that day than those immediately before and after it). In the other - the 'mean method' - the three verifying directions (for D = 3, 4 and 5) and the three curvatures were meaned prior to scoring against the forecast. This latter method should represent a more realistic measure of skill than the former because the forecast flow is intended to ignore short-wave features, and hence is akin to a chart of mean flow. Scores are converted to percentages of maximum possible score.

Results of these verifications are shown in Fig.8. This shows that, whereas almost all first period forecasts scored 60% or more, giving a skewed pattern, the 'three-day method' gave a nearly symmetrical pattern with only a small number of forecasts scoring less than 20% or more than 80%, and almost half scoring about 50%. However the 'mean method' gave scores intermediate between these two, and indicating that more than half the forecasts scored over 60%.

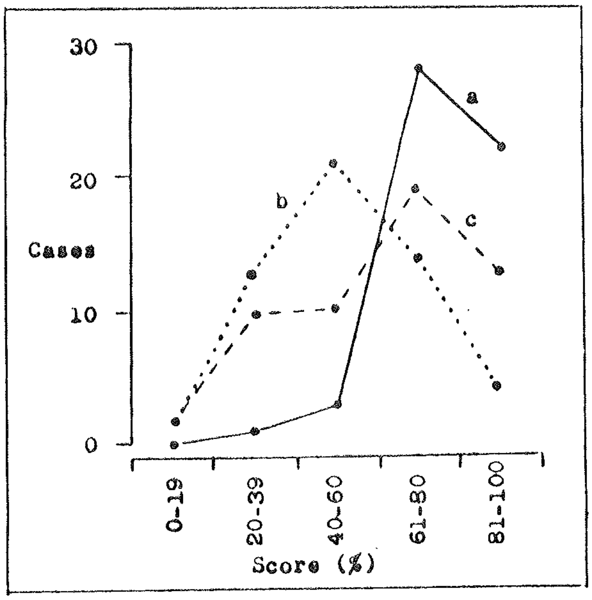

Fig.8. Frequency distribution of flow forecast accuracy.

(a) first period forecasts (D = 1.5);
(b) extended period forecasts (D = 3, 4 and 5) using three-day method;
(c) extended period forecasts using mean method.

The accuracy of the forecast flow (over New Zealand) for the first period does not appear, on average, to influence the accuracy of that for the extended outlook period. This fact is revealed in Table 2, which gives the average accuracy of the extended outlook flow for various ranges of accuracy score pertaining to the corresponding first period forecast.

Table 2. Average accuracy scores (%) of extended outlook flow pattern forecasts over New Zealand for given accuracy score ranges of preceding first period forecasts. (a) from using the 'three-day method'; (b) the 'mean method'.

	Accuracy (%) of flow pattern forecasts for the first period (D = 1.5)		
	0 - 60	61 - 80	81 - 100
(a)	57	51	52
(b)	63	62	59

Evaluation of the accuracy of the extended outlook flow pattern forecasts was extended beyond the confines of New Zealand by using the following two schemes. One considered the location of the main trough, or ridge, axis within the westerly flow. Having identified the position of the axis which lay closest to New Zealand in each forecast, an estimate was made of the lag (in days) until the configuration occurred. This gives an indication of timing errors involved in forecasting the eastward progression of the broad-scale troughs and ridges. The sequence of forecasts has been divided into four equal quarters, and a histogram of the lags is given in Table 3. Although there are only 13 or 14 cases per quarter, Table 3 shows that at the inception of the project the long waves tended to be moved too quickly eastwards, but that from the second quarter onwards there was a small average bias only, and large individual lags became less frequent.

Table 3. Frequency distribution of the lag between the forecast of a broad-scale trough or ridge axis in the westerlies near New Zealand and its occurrence in that position.

	Lag (days)								
Quarters	-2	-1	0	1	2	3	4	> 4	Average
First	0	0	3	2	3	1	1	1	+1.9
Second	1	4	4	2	1	0	2	0	+0.5
Third	1	2	5	1	3	1	0	0	+0.5
Fourth	1	2	9	1	0	0	0	0	-0.2

In some of the verifications, forecasts tended to be more accurate toward the end of the project period - this shows up in the figures for the fourth quarter in Table 3, as well as in other results. Since there was a change in situation type (to more dominant westerlies) about this time, the improvement may not be due entirely to increasing skill on the part of the project team.

The second evaluation was the westerly index. Both forecast and verifying flow patterns in the New Zealand area (see map, Fig.6) were placed in one of three index categories: (a) High - pressures high to the north and low to the south of New Zealand; (b) Low - pressures low to the north and high to the south; or (c) 'Half Index' - either a high or low centre near the latitude of central New Zealand.

Results are given in Table 4. From this it is evident that a greater proportion of high index situations (65%) were correctly forecast than was the case with the other two categories (57% for 'half index', and 40% for low index situations). Overall 31 of the 54 forecasts were correct - 57%.

Table 4. Number of cases of forecast and subsequently observed westerly index estimates, in three categories: High, Half and Low. (For further details see text).

		Observed			
		High	Half	Low	All
Forecast	High	15	5	0	20
	Half	7	12	6	25
	Low	1	4	4	9
	All	23	21	10	54

(iii) Verification of categorical forecasts of rainfall and maximum temperature.

The rainfall and maximum temperature forecasts* were verified in one of three categories, and the results placed in a 3 x 3 contingency table as indicated in Table 5a. A skill score was calculated for each district and forecast type separately, using the formula

$$S = \frac{1}{N} \sum_{i,j=1}^{3} n_{ij} \; s_{ij}$$

---

* These forecasts are referred to in the text as Rain and Temp. respectively.

where $N = \Sigma n_{ij}$ is the total number of forecasts. The skill score weights S_{ij} (see Table 5b) involve the climatological probabilities p_i, and in practice these were approximated by the sample observed probabilities,

$$\text{i.e., } p_i = \frac{1}{N}(n_{i1} + n_{i2} + n_{i3}) \quad ; \; i = 1,2 \text{ or } 3.$$

These skill score weights are chosen in such a way that the maximum possible score is +1, and a score of 0 indicates 'no skill' which would be the long-term result of picking categories at random or always forecasting the same category. For further discussion on the selection of skill score weights see Gordon (1982).

Table 5

(a) Forecast/Verification contingency table. Element n_{ij} is the number of forecasts of category j that subsequently verify as category i.

	Forecast Category		
Verif.Categ.	1	2	3
1	n_{11}	n_{12}	n_{13}
2	n_{21}	n_{22}	n_{23}
3	n_{31}	n_{32}	n_{33}

(b) Skill score weights S_{ij}, where p_i is the climatological probability of category i occurring.

	Forecast Category		
Verif.Categ.	1	2	3
1	1	$1 - \frac{1}{2p_1}$	$1 - \frac{2}{3p_1}$
2	$1 - \frac{1}{3p_2}$	1	$1 - \frac{1}{3p_2}$
3	$1 - \frac{2}{3p_3}$	$1 - \frac{1}{2p_3}$	1

Table 6 shows the resulting scores for the sample of 54 forecasts, along with the correspondong scores for persistence forecasts. Averaging over the six districts A to F we find skill scores of 0.39 (Rain) and 0.35 (Temp) for the first 2-day period, decreasing to 0.14 (Rain) and 0.19 (Temp) in the extended outlook 3-day period. A large part of this decrease is attributable to reduced persistence in the extended outlook period. Thus, comparison with persistence is a more stringent test of forecast skill than comparison with climatology (the 'no skill' case), at least for the first 2-day period. In the extended outlook period there is some indication of 'anti-persistence' in the rainfall amount.

The standard deviations of the scores are quite high, typically about 0.15 for each district separately. This is mainly a consequence of small sample size and means we can make few statements about the statistical significance of the results, apart from the claim that many of the 2-day forecasts are significantly (95% level, or two standard deviations) better than random. A few districts have exceptionally large standard deviations: in particular, 2-d Rain District E, 2-d Temp District A and 3-d Rain District F. These are due to very unbalanced climatological distributions for the category definitions we adopted (e.g., in District A, the north of the North Island, there were very few occasions when the maximum temperature departed from category 2). Hence there is less confidence in the skill scores for these few cases.

Considering the inter-district variation in Table 6, we see the skill of 2-d Rain forecasts is much the same for all districts (with the possible exception of District E), in spite of considerable variation in the persistence score. This might indicate some other limiting factor in predicting areal rainfall amounts (such as intrinsic variability). For the extended outlook, the improvement over persistence in rainfall forecasts is not too different between districts with the notable exception of the east of the North Island. This poor performance in District C is due to consistently underestimating the rainfall amounts there, particularly in westerly situations. In fact, there turned out to be an overall bias towards underforecasting heavy rain in all districts.

In the maximum temperature predictions, we show more skill in eastern districts (C, E) for both forecast periods, even after allowing for greater persistence in the first 48 hours. This is probably related to the success in forecasting high index situations, and therefore the high temperatures on the east coast associated with strong north-westerlies. Perhaps surprisingly, the 3-day Temp forecasts for the west of the South Island (District D) also score highly.

Table 6: Skill Scores (S) for Actual forecasts, Persistence forecasts and Actual-Persistence (shown for each district and forecast type separately). Standard deviation σ(S) of Actual scores also given.

2-day Rain District:

	A	B	C	D	E	F	A-F Average
S: Actual	.43	.42	.39	.47	.17	.43	0.387
Persistence	.02	.28	.04	.35	.23	.03	0.159
Actual-Per.	.42	.14	.35	.12	-.06	.40	0.228
σ(S) Actual	.14	.17	.12	.13	.31	.26	

3-day Rain

	A	B	C	D	E	F	A-F Average
S: Actual	.20	.19	.01	.26	.14	.06	0.140
Persistence	-.01	-.05	-.01	-.10	-.13	-.11	-0.069
Act.-Per.	.20	.24	.03	.36	.27	.16	0.209
σ(S) Actual	.12	.13	.08	.13	.20	.46	

2-day Temp

	A	B	C	D	E	F	A-F Average
S: Actual	.23	.21	.48	.24	.57	.35	0.345
Persistence	.11	.10	.24	.08	.34	.17	0.173
Act.-Per.	.12	.11	.24	.16	.22	.18	0.171
σ(S) Actual	.51	.23	.13	.20	.14	.14	

3-day Temp

	A	B	C	D	E	F	A-F Average
S: Actual	.47	.04	.17	.23	.21	.04	0.194
Persistence	.51	.02	-.04	-.01	.01	-.04	0.075
Act.-Per.	-.04	.02	.21	.24	.19	.08	0.118
σ(S) Actual	.20	.25	.15	.27	.15	.16	

The maximum temperature in District D only occasionally gets out of the 'normal' category, and our rather conservative forecasts were never more than one category in error. Districts A and B show no improvement over persistence in the extended outlook and the south of the South Island (District F) scores poorly because of our bias towards underestimating the maximum temperature.

Skill scores were also calculated for various subsets of the 54 forecasts. The smaller sample sizes limit our confidence in the conclusions; so the results are not presented in any detail here. We will restrict ourselves to a comment of the variation of forecast skill with time. It was noted in the section on verification of flow patterns that forecasts tended to be more accurate toward the end of the project period. This statement is supported by a comparison of categorical skill scores for the first half of the project (24/4/81 to 10/7/81) with those for the second half (13/7/81 to 11/9/81). There is little change in the skill of rainfall forecasts, but the temperature forecasts show considerable improvement (e.g., from 0.26 to 0.43 for 2-day Temp averaged over all districts). However, persistence scores for both rainfall and maximum temperature are also distinctly higher in the second half of the project period which was characterised by more dominant westerlies.

Forecasting Bias

Figure 9 shows the distribution of observed and forecast categories for each forecast type, averaged over all districts. The type of shading indicates whether a particular category was forecast more or less often than actually occurred.

Underforecasting rain in District C has already been noted, but the same error occurred in all other districts too, with the exception of the south of the South Island (District F). Here, rainfall over 15 mm was observed on only 2 out of 54 occasions in both the 2-day and 3-day periods and this event was slightly overforecast (3 times and 4 times for 2 and 3 respectively). However, in general, category 3 rain was predicted fewer times in the extended outlooks than in the first 2 days in spite of the fact that it is climatologically more likely over the longer periods.

The maximum temperature histograms show the preponderance of near normal temperatures. These figures are for all districts combined, and the distribution was even more unbalanced for District A (87% chance of category 2 in 2-day period) and for Districts B (76%) and D (74%).

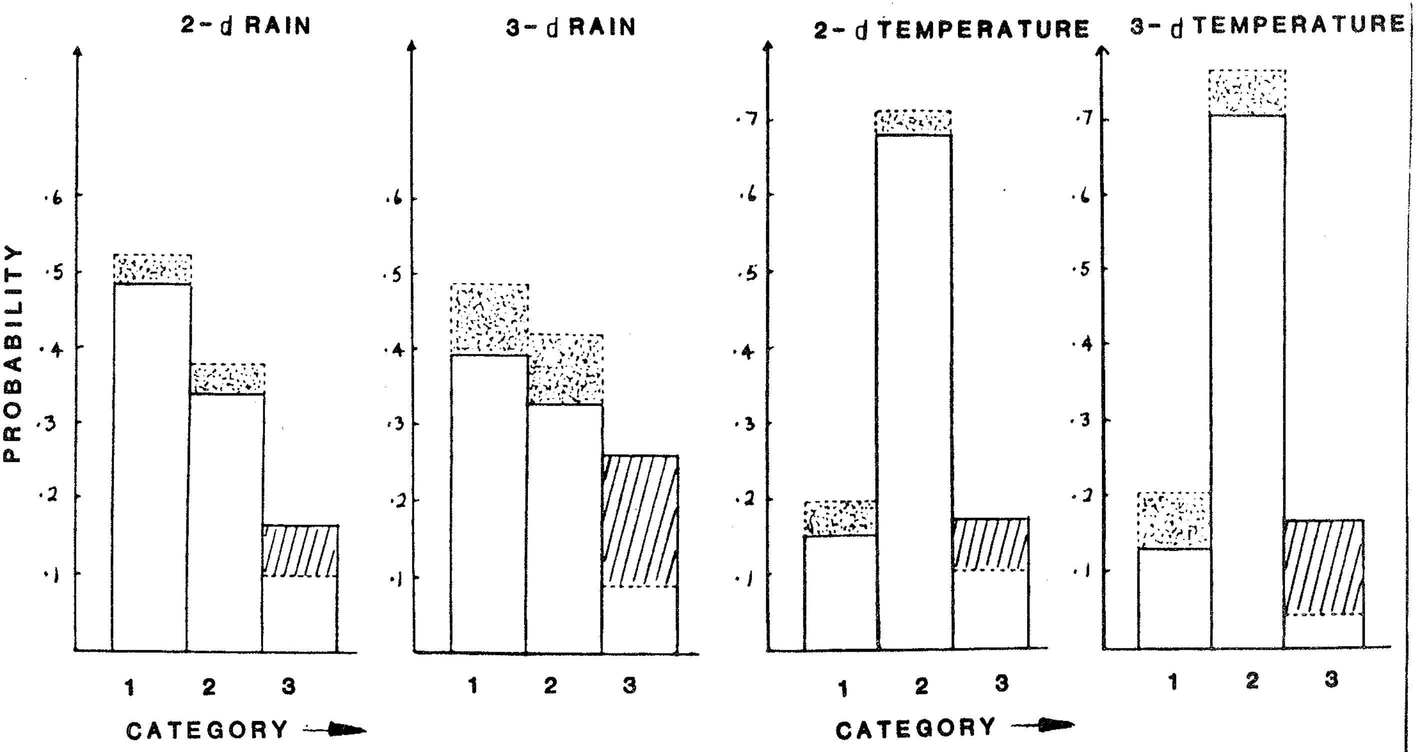

Fig. 9. Probability distribution of observed and forecast categories for 2-day Rain, 3-day Rain, 2-day Temperature and 3-day Temperature forecasts. Observed probabilities p_i (for all 6 districts combined) are shown as solid lines and the forecast category distribution as dashed. Differences are indicated as stippled (overforecast) or hatched (underforecast).

This had the effect of making the skill score very sensitive to the failure to predict an extreme temperature. The other notable feature of the Temp distributions is the underforecasting of warmer than normal temperatures. The Extended Outlook Project spanned the winter half-year and whether we considered the colder spells to be more important to forecast or more likely to occur is open to speculation. Certainly the climatology indicates that temperature departures are near normally distributed, and hence category 3 events (temperatures more than 2 degrees above normal) are at least as likely as category 1.

(iv) Experimental statements

These were subjectively evaluated by dividing each into a number of individual verifiable elements. This number ranged from as few as four on some days to ten or more on others. Each element was marked either correct or incorrect, and the accuracy taken as the ratio of correct items to the total number, expressed as a percentage.

The frequency distribution in Fig.10 shows the number of forecasts which achieved given categories of accuracy. Twenty-eight percent of forecasts could be regarded as of no value (accuracy about 50% or less), while nearly 40% of forecasts gave good guidance (accuracy 80% or greater). Average accuracy was 71%.

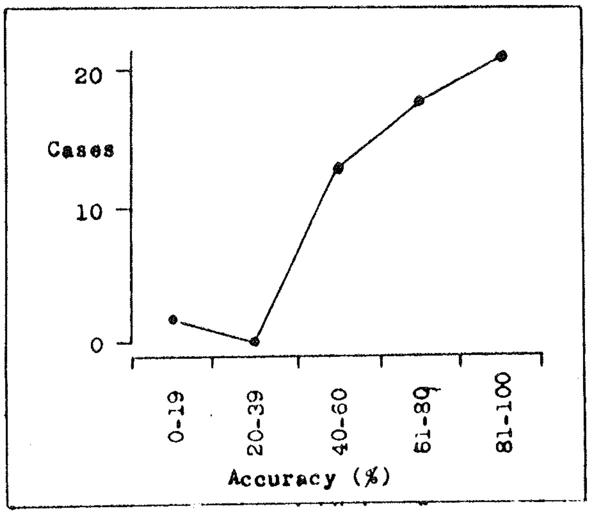

Fig.10. Frequency distribution of experimental statement accuracy.

Items included in the statements could be chosen to suit the circumstances of the situation, and the confidence placed on the particular prognosis. The fact that the average accuracy of statements increased progressively over the four quarters of the project (62, 66, 76 and 80 percent respectively), could indicate increasing skill at selecting what items to include, as much as increasing skill of extended outlook prediction.

The accuracy of statements was compared with the accuracy of the extended outlook flow pattern (mean method). Results, shown in Fig.11, indicate that when the quality of one is poor (accuracy 50% or less), that of the other tends to be equally poor. However, as the quality of the statement becomes progressively better, the accuracy of the predicted flow also improves, but at a slower rate. Consequently, a statement accuracy of 90-100% corresponds, on average, to a mean flow accuracy over New Zealand of 77%.

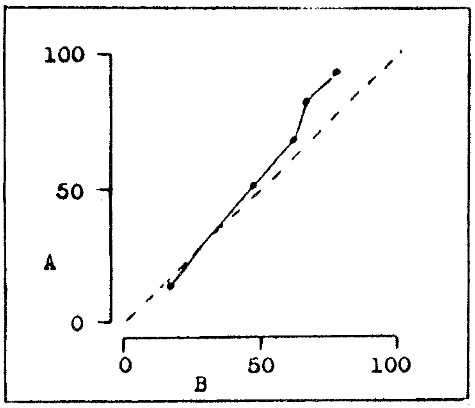

Fig.11. Accuracy (%) of the experimental statements (A) versus accuracy of the corresponding outlook flow patterns - mean method (B). The diagonal is shown by a pecked line.

In other words, statements regarding the behaviour of pressure systems and fronts, which can be quite definite and verify well, are more difficult to convert into equally accurate estimates of the flow pattern over the New Zealand area. This fact emphasises the impossibility, with present techniques, of being confident about the effect that the smaller scale weather systems will have on particular districts in New Zealand in the extended outlook period.

Conclusions

To test the feasibility of issuing useful extended range forecasts, 54 forecasts were prepared, based primarily on Fourier analyses of the hemispheric wave patterns at 500 hPa. Subjective verifications of the predictions of the 500 hPa long waves, and the broad-scale flow in the New Zealand region, as well as objective verifications of categorical forecasts of rainfall and maximum temperature in six regions of New Zealand, led to the following conclusions.

Skill is evident in predictions of both the behaviour of the long-wave pattern, and of the broad-scale surface flow near New Zealand. However correlation between these two factors has not yet been investigated.

Skill scores for categorical forecasts are significantly better than chance for the first 2-day period, but show a considerable decrease in performance for the following 3-day period. However, a large part of this reduction can be attributed to reduced persistence: this is particularly noticeable for rainfall prediction.

Rainfall was underpredicted in all districts, but especially in the east of the North Island where more rain fell in westerly conditions than was anticipated. The heaviest rainfall category was predicted less often for the longer 3-day period than for the 2-day period, despite climatological reasons for its being likely to occur more often.

Results imply that successful predictions of the long-wave pattern and broad-scale surface flow do not always lead to successful categorical forecasts. The relationship between weather elements and broad-scale synoptic patterns needs further study.

Results indicate an increase in skill during the period the project was in operation. Westerly wind regimes occurred more frequently in the latter part of the project, and this result may indicate that prediction is easier during this weather type. It is possible also that later in the project there was greater astuteness when deciding what sort of information should be included in the forecast statements.

The method used for recording forecasts of maximum temperature was found to be incapable of reflecting temperature variations which were expected during the 3-day period. This may have resulted in a lower skill score than would have been achieved with a different scoring scheme.

Results from Phase I of the project suggest that it would be feasible to issue extended range forecasts to the public on a routine basis, provided these are carefully worded to avoid the use of terms which, while being meaningful to the meteorologist, are obscure to the layman.

References

Gordon, N.D., 1982: Evaluating the skill of categorical forecasts. *Monthly Weather Review 110*, 657-661.

Laidlaw, M.R., 1983: Skill scores for extended range forecasts. N.Z. Met.S. Research Report (Unpublished).

Mullan, A.B., 1983: A long-wave climatology of the Southern Hemisphere. (To be published).

Trenberth, K.E., Neale, A.A. and Browne, M.L., 1978: Hovmoller diagrams. *N.Z. Met.S.Tech. Inf.Circ.* No. 161.