551.509.25: 551.509.5

NEW ZEALAND METEOROLOGICAL SERVICE

TECHNICAL INFORMATION CIRCULAR NO. 162

AN OBJECTIVE ASSESSMENT OF THE ACCURACY OF PROGNOSTIC MSL CHARTS

The following notes were written by Mr A.A. Neale and are circulated for general information.

J.S. Hickman
Director

New Zealand Meteorological Service
P.O. Box 722
Wellington.

26 July 1978

AN OBJECTIVE ASSESSMENT OF THE ACCURACY OF PROGNOSTIC MSL CHARTS

A. A. Neale

Abstract

The accuracy of 404 manually produced 24-hr prognostic MSL charts are assessed by calculating S1 skill scores over an 8-point grid in the New Zealand region. Results are consistent with long-held beliefs that accuracy depends, inter alia, on the following factors: the type of weather situation, the length of experience of the prognosticator, his ability to consult colleagues about the prognosis, and the length of time available for its preparation.

1. Introduction

A variety of methods have been employed to estimate the accuracy of prognostic charts that depict isobars at MSL or height contours of constant pressure surfaces. These range from purely subjective eyeball comparisons between forecast patterns and those of verifying analyses to objective relationships between forecast and verifying values (of pressure or height) at a set of grid points.

Unless the grid length is sufficiently small, objective methods cannot adequately take account of small scale features in the prognosis (or analysis). Hence, checking the accuracy of charts covering a big area necessitates the use of such a large number of grid points that, in the absence of computer facilities for extracting both the forecast and verifying values and for processing them, the task becomes too time consuming for manual operation.

At present in New Zealand data from predicted and analysed charts are not available as input to a computer, and hence the reliability of 24-hr prognostic MSL charts prepared operationally in the National Weather Forecasting Centre (NWFC) cannot be checked objectively over the whole chart area - approx 25S to 55S between 140E and 165W. However it is possible to operate manually on a small grid in the immediate vicinity of New Zealand. Despite the small area covered, this check is desirable for a number of reasons.

Within the next several months, output from a numerical weather prediction (NWP) model will be introduced into the operational work of the Forecasting Branch. Its impact on the accuracy of MSL prognoses will be impossible to assess without some objective means of comparing present performance to that after NWP output is available. In addition there is a continuing need to monitor the quality of output in order to be able to answer questions such as; what weather systems are associated with the poorest prognoses? Do prognoses tend to be better one season by comparison with another? Are prognoses getting better/worse?

Working conditions in the NWFC vary diurnally in a fairly regular manner. Does that affect the quality of the product to the extent that prognoses prepared at one time of day tend to be better than those prepared at another?

2. S1 Skill Score

Of the various methods available for comparing predicted and verifying pressures at a set of grid points, that proposed by Teweles and Wobus (1954) was used in this study. A skill score (S1) is calculated for each prognosis using predicted values (P_i, P_{i+1}, ...) at adjacent grid points and subsequent analysis values (A_i, A_{i+1}, ...) at the same points (i, $i+1$, ...). The skill score is defined by;

$$S1 = 100 \frac{\sum_i \left| (P_i - P_{i+1}) - (A_i - A_{i+1}) \right|}{\sum_i MAX\left\{ |A_i - A_{i+1}|, |P_i - P_{i+1}| \right\}}$$

where the denominator is the sum of the observed or forecast pressure difference between grid points, whichever is largest.

The use of this method is attractive for several reasons; (i) by verifying pressure gradients, it is assessing, in effect, the accuracy with which wind components have been forecast and these, in turn, are related to the distribution of cloud and precipitation; (ii) it employs a simple (though time-consuming) procedure that can be handled manually; (iii) the use of the difficulty coefficient $1/\sum_i MAX\left\{ |A_i - A_{i+1}|, |P_i - P_{i+1}| \right\}$ penalises attempts to hedge and forecast weak pressure systems (in this respect it overcomes the disadvantage of verification methods that use root mean square errors, and in which prognosticators are encouraged to be conservative thereby avoiding the heavy penalties arising from large forecast errors); and (iv) it is the method used by the Australian Bureau of Meteorology to assess the performance of prognoses they prepare.

Skill scores have been calculated for all 24-hr prognostic MSL charts issued by the NWFC in the 4-month period from June to September 1977. The 8-point grid used in the New Zealand area is shown in Fig. 1. Predicted pressures at these points are compared with pressures subsequently observed at all but the point located in the Tasman Sea (TS).

It is highly desirable that as many observed values as possible be used for verification, otherwise unwanted uncertainties are introduced into the score because of subjective interpretations made by the analyst in areas devoid of observations. The presence, from time to time, of ship observations in the vicinity of TS kept subjectivity in this area to an acceptable level. Because few ship observations are available south of Chatham Islands (CI), no grid point was selected in that area.

Although the area covered by the grid is small, skill scores obtained from pressure differences between points linked by solid lines in Fig. 1 will give an estimate of the accuracy with which pressure gradients are predicted in the New Zealand area. The grid is coarser than that used over the U.S.A. by Teweles and Wobus, the distance between points averaging 490 miles with a range from 360 to 600 miles, compared to their grid length averaging 350 miles with a range from 270 to 525 miles.

In addition to calculation of skill scores, errors of predicted pressure at grid points were studied briefly.

3. Results

During the period from June to September 1977 the NWFC prepared 404 twenty-four hour prognostic MSL charts. Mostly these were for 0000, 1200 or 1800[1], but thirty-two were for 0600 and six amended prognoses were issued.

Skill scores were calculated for all charts. When interpreting results, remember (i) that the smaller the score, the greater the accuracy of the prediction, and (ii) that the accuracy of a prediction is influenced by many factors, the most important being the skill of the meteorologist, the work environment (including availability of data) and the type of weather situation involved.

Average skill scores on a monthly basis were: June 54.3; July 53.0; August 47.9; September 49.0; indicating that prognoses were most accurate in August and least accurate in June. Such a result relates logically to average pressure distribution over New Zealand (taken from monthly notes on the weather prepared by the Climatology Branch) indicating that skill score does depend on situation type. Pressures in June, although near average over the country, were higher both to the north and to the south, indicating disturbed conditions with low pressure systems affecting New Zealand; at this time prognoses were least accurate. During the other three months pressures were higher than normal south of the country, there was a marked absence of westerly winds, and probably few fast-moving weather systems. Notably pressures were higher than average over New Zealand in August indicating more settled conditions, and at this time prognoses were most accurate.

Fig. 2a gives the average skill scores four times daily for each month and indicates erratic fluctuations in July and September. The relatively high skill scores at 1200 in those

---

1. All times are GMT

months should be considered in relation to the frequency distributions of skill scores given in Table 1. The highest scores obtained (100-109) apply to 1200 charts in July and September and, in addition, six charts at 1200 in July scored 80 or higher - at least three times more than at any other time in any month. Since weather systems are unlikely to be less predictable at 1200 GMT than at other times, the lower accuracy shown in prognoses at this time most likely is attributable to the work environment and/or the lower skill displayed by the prognosticator(s) concerned.

Scores at 0600 are difficult to interpret because the small data sample could be unrepresentative, and this possibility is reinforced by the fact that the lowest skill score (in September) occurs in the month containing the smallest number of charts (5) issued at that time of day. For this reason, although 0600 data are presented, generally they are not discussed.

Fig. 2b shows the average skill scores for prognoses valid at different times of the day. The most accurate were those prepared for 1800 (score 49.3), followed by those for 1200 (51.5) and, least accurate, 0000 (53.0). Although tests extending over a longer or a different period may not confirm this ranking, there are good reasons why prognoses for 1800 should be more accurate than the others. The former are prepared (i) by the more experienced meteorologists, (ii) with the assistance provided by recently acquired satellite imagery, and (iii) often after consultation with a panel of meteorologists drawn from all sections at Kelburn.

With regard to the latter point, the panel is not available Saturdays, Sundays and public holidays and consequently its influence can be gauged. Table 2 gives the distribution of skill scores attained without (row a) and with (row b) panel participation. Although the former group contains only 35 cases, and consequently deductions are tentative, prognoses in general were more accurate with than without assistance from the panel - average skill scores being 49.1 and 49.9 respectively.

The poorer accuracy rating of prognoses for 0000 could result from the shorter period available for their preparation ($3\frac{1}{2}$ hours from data time) compared with that for 1200 and 1800 prognoses (4 to $4\frac{1}{2}$ hours). However, although the overall accuracy of 0000 prognoses was less than that at other times, large errors were less numerous than at 1200.

Gauntlett (1975) notes that Shuman suggests the following criteria for MSL prognoses over the continental U.S.A. An S1 skill score of 30 or less corresponds to a perfect forecast whilst one of 80 or more may be equated with a worthless forecast. Applying almost the same criteria (29 or less, and 80 or more) to our result leads to the conclusion that 7% of forecasts were 'perfect' and 4% 'worthless'.

Table 3 lists the number of 'perfect' and 'worthless' forecasts relative to time of day and month, and also the ratio (perfect to worthless) of the two. The figures confirmation of the higher quality of prognoses for 1800. However worthless forecasts are markedly more numerous at 1200 than 0000, and in addition Table 1 shows that very large individual skill scores (in excess of 99) only occur at 1200. It is believed this can be attributed to the small number of meteorologists on duty when the prognosis for 1200 is prepared. Consultation is limited or absent and consequently important errors of judgement are not detected.

On six occasions amended prognoses were issued that gave, in general, more accurate predictions than the original charts. In one case the improvement was dramatic, as indicated by a change of skill score from 82 to 55. In another the improvement was more modest (64 to 55) but still important. One amendment was significantly poorer than the original (69 to 76). Skill score changes in the remaining three cases were small - 44 to 40, 32 to 30, and 43 to 45.

The accuracy with which pressures at individual grid points were predicted (in contrast to pressure gradients between points) is shown in Fig. 3 where the distribution of average daily errors of 24-hr forecast pressures (in mb) for the eight grid points is given relative to time of day. While distributions for prognoses valid at 0000 and 1200 are symmetrical about zero error, that for 1800 is displaced and indicates a bias towards forecasts of pressure too high.

Table 4, giving the distribution of the size of forecast errors at each point, shows that the bias of forecasting pressures too high at 1800 occurs at six of the eight points. No bias is apparent at Macquarie Island (MI) or at 40S 160E (TS). The component of the rate of change of pressure due to the semi-diurnal pressure oscillation has its greatest positive value at the time when prognoses for 1800 are prepared, and is small (prior to being strongly negative soon afterwards) when those for 0000 and 1200 are issued. Although the existence of this semi-diurnal pressure oscillation is known, it is apparently still influencing pressure values forecast for 1800.

In general pressure fluctuations are greater and more rapid in higher than in lower latitudes. Consequently errors in forecast pressures become larger with increasing latitude, a fact that is clearly shown in Table 4.

For each prognostic chart, the largest error of forecast pressure difference between pairs of grid points was tabulated. Every combination of points was considered, not only those joined by bold lines in Fig. 1. The distribution of the magnitude of this error relative to time of day is shown in Fig. 4, and has a flat-topped peak that extends from 8 to 19mb in prognoses for 0000 and 1200, but from 8 to 15 mb only in those for 1800. Further confirmation of the smaller overall errors in estimating future pressure differences at 1800 compared to other times.

4. Conclusions

Calculations of S1 skill scores from 24-hr prognostic MSL charts prepared during a 4-month period in 1977, using an 8-point grid in the New Zealand area, are the first of their kind. Consequently there can be no comparison with earlier years to determine whether accuracy has changed. Nevertheless deductions (sometimes tentative because of the small data sample) can be made on certain factors that influence the accuracy of MSL prognostic charts produced manually in the NWFC.

1. The type of weather situation. Comparison on a monthly basis of skill scores with weather summaries, and the monthly ratio of perfect or worthless forecasts, both support the following long-held belief. Prognostic charts tend to be more accurate when the westerly index is low (and weather systems often move slowly), and when anticyclonic systems predominate, than when the index is high and when cyclonic systems predominate.

2. Forecasting experience of the prognosticator. The belief that, on average, forecasters of longer experience will produce more accurate prognoses than those of shorter experience was confirmed. Forecast charts prepared for 1800 (when staff with longer experience were involved) were more accurate than those prepared for other times (when staff with less experience were involved).

3. Consultation. Discussions of prognoses with colleagues before issue appears to be beneficial. (i) Because the prognosticator has little if any opportunity to discuss the forecast chart prepared for 1200, some of these contain large errors. This is shown both by the occurrence of the highest skill scores recorded and by the high proportion of worthless forecasts issued compared to perfect ones. (ii) Prognoses prepared for 1800 received a higher skill score when a panel of Kelburn meteorologists did not discuss their validity than when they did.

4. Time available for prognosis. The higher average skill score of prognoses prepared for 0000 may be caused by the shorter time available for their production by comparison with that available for preparation of prognoses for other times.

5. Future Research

A start has been made on the calculation of skill scores for prognoses valid during the same four months - June to September - of 1978. This will be the last opportunity to monitor the accuracy of 24-hr MSL prognoses prepared without assistance from output from a NWP model. In addition results will help to confirm (or confound) deductions made in this report.

Acknowledgement Observing Officers in the NWFC extracted and tabulated forecast and actual pressure values at the eight grid points. Their assistance is gratefully acknowledged.

References

Gauntlett, D.J., 1975: The application of numerical models to the problems of meteorological analysis and prognosis over the Southern Hemisphere. Meteorological study No. 28. Bureau of Meteorology, Canberra, 127-129.

Teweles, S. and H.B. Wobus, 1954: Verification of prognostic charts. Bull. Amer. Met. Soc., 35, 455-463.

S1 Skill Score

Time of validity GMT		0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99	100-109
0000	Jn	.	1	1	2	5	9	5	5	1	1	.
	Jl	.	.	3	2	7	10	5	3	1	.	.
	Au	.	.	1	6	12	3	5	5	.	.	.
	Se	.	.	1	7	7	8	5	1	1	.	.
0600	Jn	.	.	.	2	1	5	1	.	.	1	.
	Jl	.	.	.	3	1	3	1	.	1	.	.
	Au	.	1	.	2	2	2	1	.	.	.	.
	Se	.	.	1	2	1	1	.	.	.	.	.
1200	Jn	.	.	2	5	7	7	7	1	1	1	.
	Jl	.	.	2	6	6	3	6	3	5	.	1
	Au	.	1	2	6	10	7	5	.	.	.	.
	Se	.	.	3	6	8	6	4	2	.	.	1
1800	Jn	.	.	2	5	9	4	5	3	2	.	.
	Jl	.	.	2	6	10	7	6	2	.	.	.
	Au	.	1	4	4	9	8	2	3	.	1	.
	Se	.	1	1	9	8	4	5	2	.	.	.

Table 1. Frequency distribution (number of cases) of S1 skill scores, for 24-hour prognostic MSL charts, related to month and time of day.

S1 Skill Score

	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99	100-109
a	.	.	9	23	17	26	14	9	.	3	.
b	.	2	7	18	33	16	14	8	2	.	.

Table 2. Frequency distribution (%) of S1 skill scores for 24-hour prognostic MSL charts valid at 1800 GMT (June - September 1977 incl.).

a - charts prepared on Saturdays, Sundays and public holidays (35 cases);

b - all others (90)

	$S1_{mi}$	$S1_{ma}$	$S1_{mi}/S1_{ma}$	Total Cases
0000	7	4	1.8	123
0600	2	2	1.0	32
1200	10	9	1.1	124
1800	11	3	3.7	125
Jun	6	7	0.9	101
Jul	7	8	0.9	105
Aug	10	1	10.0	103
Sep	7	2	3.5	95
All	30	18	1.7	404

Table 3. Number of prognoses with S1 skill scores of less than 30 ($S1_{mi}$) and of 80 or more ($S1_{ma}$), and the ratio of these two quantities.

24-hr forecast pressure minus verifying pressure (mb)

	negative					positive				
	18-22	13-17	8-12	3-7	2-2	3-7	8-12	13-17	18-22	23-27
NI 00	.	1	2	19	59	19	2	.	.	.
06	.	.	3	12	48	36	.	.	.	.
12	.	.	1	20	59	19	1	1	.	.
18	.	.	2	12	68	17	1	.	.	.
RI 00	.	.	1	15	58	25	1	.	.	.
06	.	.	.	31	59	9	.	.	.	.
12	.	.	.	21	59	20	.	.	.	.
18	.	.	2	16	58	22	2	.	.	.
KT 00	.	1	3	20	51	22	2	.	.	.
06	.	.	.	25	53	19	3	.	.	.
12	.	2	2	17	56	15	7	2	.	.
18	.	.	4	17	50	24	6	.	.	.
TS 00	.	.	3	27	44	20	6	.	.	.
06	.	.	3	25	38	28	6	.	.	.
12	1	.	2	19	49	24	4	.	.	.
18	.	1	6	18	54	17	4	1	.	.
KL 00	.	1	7	19	48	20	6	1	.	.
06	.	.	3	31	50	9	6	.	.	.
12	.	2	6	21	44	21	5	2	.	.
18	.	1	5	16	47	25	6	1	.	.
CI 00	.	.	7	25	37	22	8	1	.	.
06	3	.	6	22	41	22	3	3	.	.
12	.	.	9	26	30	23	10	2	.	.
18	.	.	9	17	43	24	6	.	1	.
NV 00	.	2	11	21	35	23	5	2	2	.
06	.	.	9	38	53	3	3	3	.	.
12	1	1	7	23	39	24	2	3	1	.
18	1	1	7	18	42	22	5	3	1	.
MI 00	1	5	15	18	25	10	11	11	3	1
06	.	9	9	25	25	19	3	3	.	6
12	1	4	10	23	22	15	13	9	3	1
18	1	3	14	19	22	19	13	5	4	.

Table 4.

Frequency distribution (%) of errors in 24-hr forecast MSL pressures at eight points (shown in Fig. 1) related to time of validity of forecasts.
Data period June - September 1977 incl.

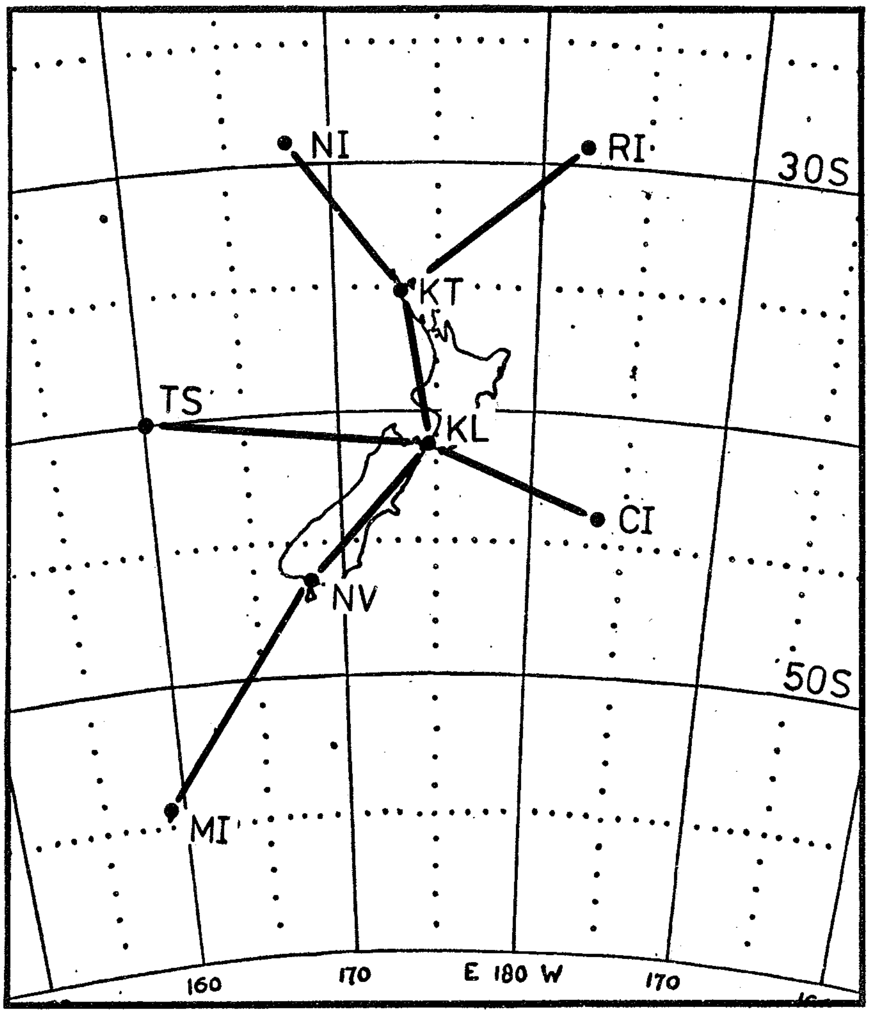

Fig.1. Grid of points at which sea level pressures were read off for use in verification. Pressure differences were computed between points joined by bold lines.

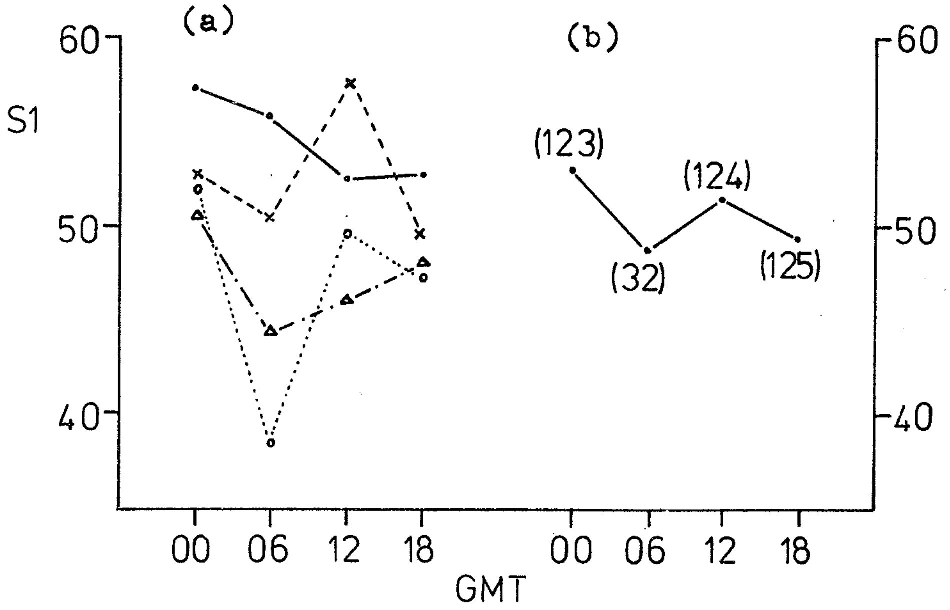

Fig.2. Average S1 skill scores for 24-hr prognostic MSL charts prepared for 0000, 0600, 1200 and 1800 GMT. (a) June •———• ; July ×- - - - -× ; August △·—·—·—△ ; September ∘..........∘ . (b) all four months combined - number of cases in brackets.

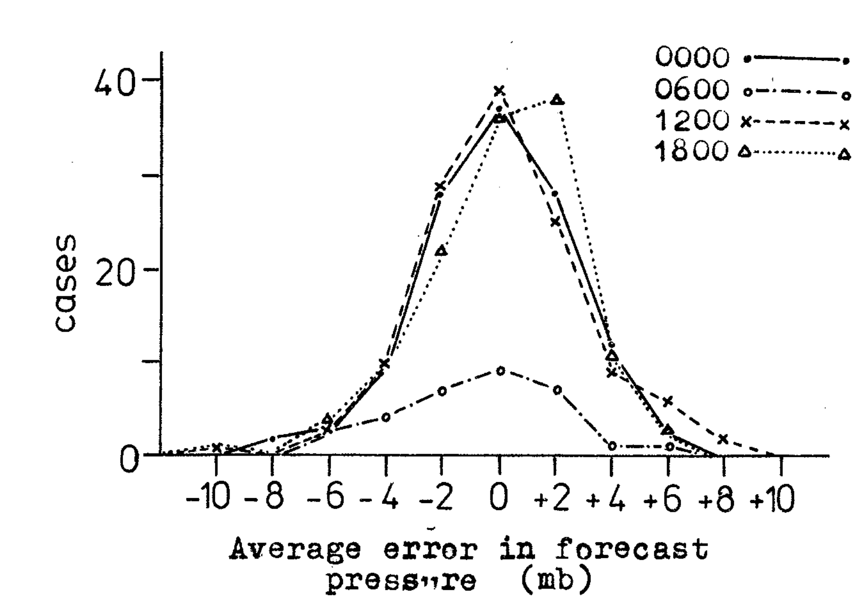

Fig.3. Frequency distribution of average daily error in 24-hr forecast MSL pressures (forecast minus actual pressure) at the eight grid points shown in Fig.1, related to time of validity of forecast.

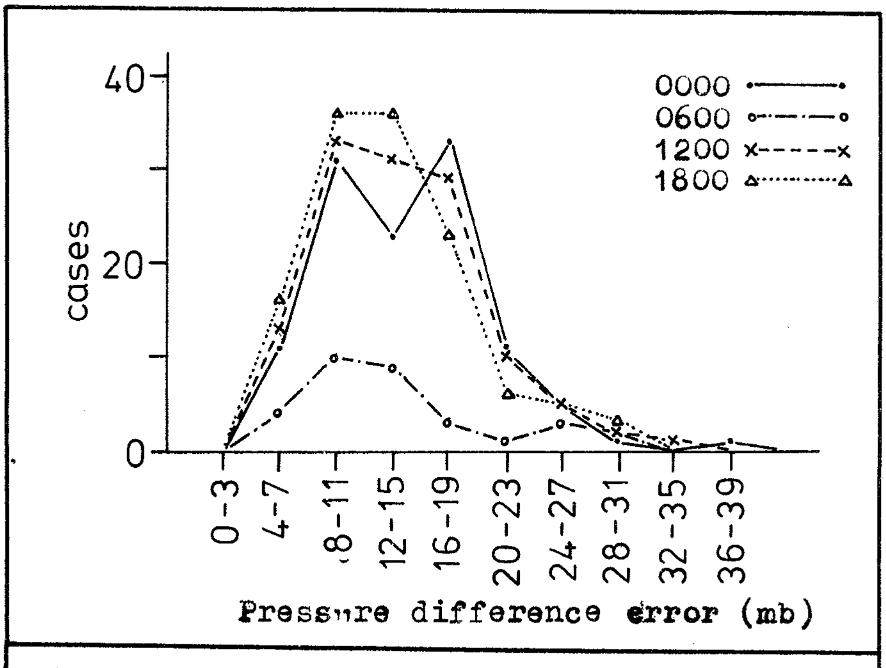

Fig.4. Frequency distribution relative to time of day, of the maximum error in 24-hr forecast pressure differences found on each prognostic chart. The selection is made from pressure differences between every pair of grid points shown in Fig.1, not only those linked by bold lines.