Aust. Met. Mag. 49 (2000) 149-157

Numerical prediction model performance summary October to December 1999

K. Mulder and T. Hart
National Meteorological Operations Centre, Bureau of Meteorology, Australia

(Manuscript received March 2000)

Introduction

This summary continues the series comparing the performances of Numerical Weather Prediction (NWP) models.

Models and methods

A description of the Australian verification methods can be found in a previous article (Skinner 1995).

Models are from the National Meteorological Operations Centre (NMOC), Melbourne, and from ECMWF (European Centre for Medium-range Weather Forecasts), NCEP (National Centers for Environmental Prediction) and UKMO (United Kingdom Meteorological Office). Reference is also made to the JMA (Japan Meteorological Agency) model.

The three models considered from NMOC, Melbourne, are: LAPS_PT375 (Limited Area Prediction System); TLAPS_PT375 (Tropical Limited Area Prediction System); and GASP (Global Assimilation and Prediction).

Overseas global models included in the comparisons are: ECSP (ECMWF Spectral Assimilation); USAVM (NCEP Washington Spectral model for aviation); UKGC (UK Meteorological Office Grid PE model); and JMAGSM (JMA Global Spectral Model).

Very short summaries of the models can be found in the initial article (Skinner 1995) with references to model updates in subsequent issues.

All results have been calculated within NMOC, Melbourne, where the models were verified against their own analyses. Results are presented for the irregular Australian verification area only.

Corresponding author address: Ms K. Mulder, National Meteorological Operations Centre, Bureau of Meteorology, GPO Box 1289K, Melbourne, Vic. 3001, Australia.

The statistics are a measure of the skill in forecasting geopotential height at 500 hPa or mean sea-level pressure, MSLP. Other field types are not included in these summaries.

The limited area models are run several hours earlier than GASP and this premature data cut-off, particularly for satellite information, adversely affects their skill compared to GASP.

Note that the Australian region verification grid has southerly points which are outside the TLAPS_PT375 grid. TLAPS_PT375 scores are calculated without these points and are therefore not strictly comparable with those from other models.

Notes on NWP systems

ECSP

The ECSP increased the number of vertical levels in the model from 50 to 60 on October 12. The horizontal resolution is unchanged at TL319. The extra levels provide finer resolution in and just above the boundary layer.

October to December 1999 intercomparisons

Local models: (LAPS_PT375, TLAPS_PT375, GASP)

The skill-score at MSLP (Fig. 1(a)) shows that GASP still outperforms the other models although at 500 hPa (Fig. 1(c)) LAPS_PT375 and GASP scores are identical for the first 12 hours, after which GASP outscores the other model again. The rms errors at MSLP (Fig. 1(b)) show TLAPS_PT375 outscoring GASP slightly after 24 hours.

Fig. 1(a) Comparison for LAPS_PT375/TLAPS_PT375/GASP from October to December 1999. S1 skill-scores of MSLP using combined base-times 0000 UTC/1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.

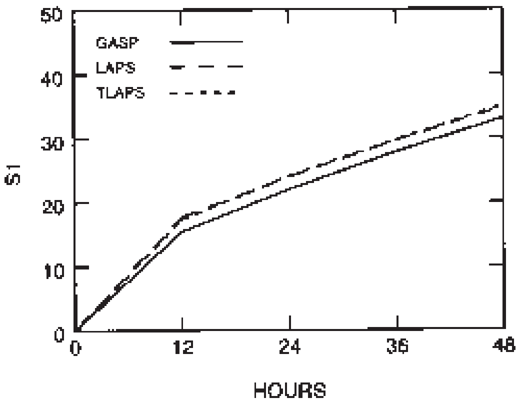


Fig. 1(c) Comparison for LAPS_PT375/TLAPS_PT375/GASP from October to December 1999. S1 skill-scores of 500 hPa geopotential height(m)for combined base-times 0000 UTC/1200 UTC and intervals 12, +24, +36, +48 h over the irregular Australian verification grid.

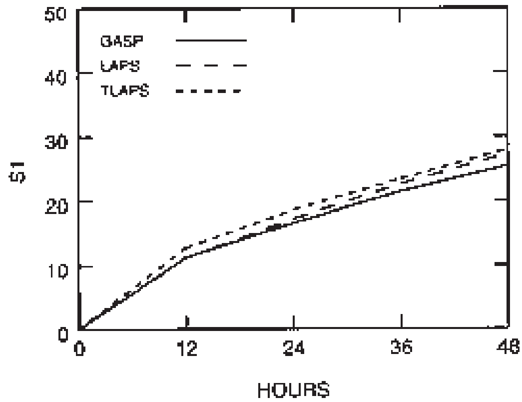


Fig. 1(b) Comparison for LAPS_PT375/TLAPS_PT375/GASP from October to December 1999. Root mean square errors of MSLP for combined base-times 0000 UTC/1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.

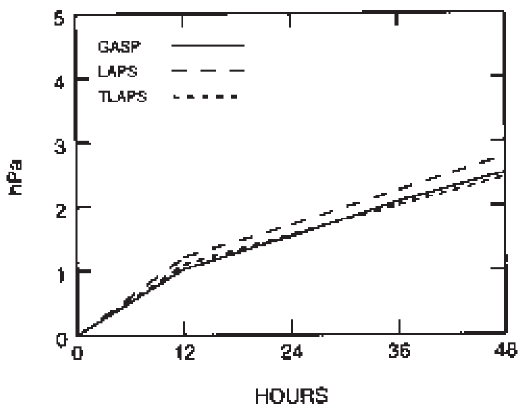
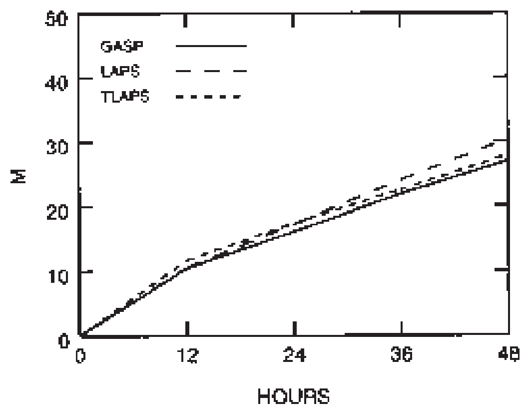


Fig. 1(d) Comparison for LAPS_PT375/TLAPS_PT375/GASP from October to December 1999. Root mean square errors of 500 hPa geopotential height (m) for combined base-times 0000 UTC/1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.



The individual month's results for MSLP (Fig. 3(a)) show that LAPS_PT375 and TLAPS_PT375 performed identically at +24 hours in November and December, but were still a little behind GASP (in December only by one skill-score point). All models improved their scores from October to December. At 500 hPa the reverse was true, skill-scores increased over the three-month period with GASP performing best again.

Synoptic overview for 24-hour MSLP predictions

On 2 October a complex low pressure system over South Australia was well captured by LAPS_PT375-the associated unstable conditions caused significant damage to crops in areas east of Adelaide.

On 5 October all models failed to forecast the intensity of the low well to the southwest of Australia although it was outside the verification grid. They all scored well over the remainder of the forecast area. All the models continued to underestimate the depth of the cut-off low as it approached the southwest of Western Australia during the following two days with GASP performing the best, TLAPS_PT375 forecasting a trough only. However, on 8 October TLAPS_PT375 performed better than GASP to pick up the low and trough, now south of Western Australia, fairly accurately.

On 12 October GASP and LAPS_PT375 forecast secondary ridging in the Bight from a high centred off New South Wales. A weak low was also forecast to the south of the ridge. The verifying analyses showed a deeper low further to the northwest. A trough extended from it to the low over southeastern Western Australia. TLAPS_PT375 forecast the situation correctly. All local models over-forecast the secondary trough in the Bight on 13 October but moved the primary trough west of Western Australia too slowly. On 14 October all models overestimated the depth of the Southern Ocean low, GASP by up to 15 hPa. It however was the only model to pick up the low to the west well. The low in the Southern Ocean continued to cause problems as it moved southeast but it was out of the verification area.

LAPS_PT375 captured the depth of the low in the Tasman Sea on 19 October well but had the position wrong, so did not score well. GASP the following day positioned the low extremely well while the other models all had it too far to the north.

The deepening of a trough over coastal Western Australia was very well handled by all the models from 20 to 22 October. They continued to predict the eastward movement of the southern portion of the trough well on 24 October, while holding the northern part back to a low over the Pilbara. During this time the models also handled the deepening of a trough over Queensland and the development of a low east of New South Wales very well, GASP generally a little better than the rest. During this time southeastern Queensland and northeastern New South Wales experienced severe thunderstorms with hail and strong winds that caused structural damage. The models continued to handle the situation very well until 27 October.

A broad trough in the Southern Ocean on 28 October was over-forecast by all the models.

On 29 October all models failed to pick up the amplitude of the trough south of the continent. By the next day they had captured the situation well, although a low well south of Australia was not picked up. All models performed well until 3 November.

All models handled the development of a low over southeast Australia on 5 November well and also its subsequent southeastward movement although they all overestimated the depth of the low on 7 November when it was moving away to the south between Tasmania and New Zealand.

On 8 November all models developed a low in the easterlies over the Queensland/New South Wales border area although all models predicted the low to form further west than it did. GASP and TLAPS_PT375 did a particularly good forecast for the next day when the low moved off towards New Zealand. It had deepened considerably and all models predicted the central pressure well. LAPS_PT375 forecast the intensity well but misplaced the centre slightly, leading to a poor score.

All models scored extremely well (skill-scores 17-19) on 13 November. There was a broad low pressure area covering half of Australia with another broad area of low pressure over the Southern Ocean. In between, ridging from the west covered coastal areas of Western Australia and South Australia, and Queensland from the east. The forecasts continued to be good until 15 November while the trough moved northeast across the southern States and then New South Wales. GASP and TLAPS_PT375 performed a little better than LAPS_PT375. New South Wales and Queensland experienced severe wind gusts and hail storms associated with this trough.

On 16 November all models extended a trough in the easterlies over all of Western Australia while the ridge to the south was kept too far south. TLAPS_PT375 captured it best. All models overcompensated the next day by not extending the trough far enough to the south.

All models performed well from 18 to 20 November when a broad low pressure area covered the continent, extending a trough into South Australia. The trough subsequently moved east slowly while a low formed in it near the Queensland/South Australia border. The development of a low in this trough south of Victoria continued to be handled well on 21 November although both LAPS_PT375 and TLAPS_PT375 positioned the ridge south of Western Australia too far to the north in the 24-hour forecasts. The continued southeast movement and collapse of the low was well represented.

The big difference in the performance of the models on 25 November was due to the positioning of the ridge to the southwest of Western Australia. LAPS_PT375 and TLAPS_PT375 located the ridge too far to the north.

On 8 December all models forecast tight gradients ahead of a surface trough linked to a low south of Tasmania. South Australia, New South Wales and Victoria experienced severe thunderstorms and wind gusts as the system moved east. Considerable damage was caused to buildings, trees, crops and power lines.

Tropical cyclone *John* developed on 12 December. On 13 December all models under-forecast the central pressure by approximately 6 hPa. The cyclone intensified to category 5 on the 14th and threatened the Pilbara coast with winds estimated at up to 290 km/h. The trough from tropical cyclone *John* extended south into the Bight. By 15 December a deep low was forecast to develop, and did develop well to the west of Tasmania. As the associated trough moved east, heavy rain and thunderstorms were reported in Victoria in the convergence area ahead of the trough.

On 22 December all models performed very well with the trough in inland Western Australia and the developing trough from the centre of the continent to the southeast corner. The synoptic situation over the continent remained similar on the 23rd and 24th with the high to the south moving rapidly east. From 25 to 27 December the trough in the east sharpened with resulting thunderstorms, strong winds and flash-flooding in South Australia on 25 December and Victoria on 27 December, when a low formed off the southeast corner of Victoria. Some locations recorded record highest daily rainfalls for December. All models had predicted the low.

On 27 December a low also developed over north-east Queensland. From then until 31 December all three models predicted its rapid movement southeast to New Zealand well, TLAPS_PT375 a little better than the other two.

Table 1 shows the (best or equal best)/(worst or equal worst) occurrences for each model during the period October to December 1999. For example, GASP in November had the best score 18 times and the worst score only four times. GASP was generally the most reliable model and always had the highest number of days with the best scores and the lowest number of days with the worst scores. GASP is run several hours later than the other two models and this is reflected in the scores.

Table 1. (best or equal best)/(worst or equal worst) occurrences for each model during the period October to December 1999.

	GASP	*TLAPS_PT375*	*LAPS_PT375*
October	15/7	13/10	5/22
November	18/4	6/15	8/16
December	15/7	9/18	10/13
Quarter	48/18	28/43	23/51

Global models: (GASP, ECSP, UKGC, USAVM, JMAGSM)

Figure 2 shows the comparisons of all the above models at intervals up to +192 h. ECSP still outscores the other models at both MSLP and 500 hPa and for skill-scores and rms errors. UKGC and USAVM were similar at MSLP out to +72 h while UKGC slightly outperformed USAVM at 500 hPa for the same period. GASP and JMAGSM had similar scores at 500 hPa for the whole period and at MSLP out to +48 h, after which JMAGSM outperformed GASP slightly.

Figure 3(c) confirms that ECSP scored better at MSLP at 1200 UTC +72 h each month, improving markedly in December. UKGC and USAVM were inconsistent in their relationship but both models also improved over the three months. GASP outperformed JMAGSM in October and November but the positions were reversed in December. Both models recorded higher scores in November than October but improved significantly in December to be much closer to the scores of the other models.

ECSP recorded the lowest scores at 500 hPa at 1200 UTC + 72 h (Fig. 3(d)) in October and December while UKGC performed best in November. Fig. 3(d) also confirms that UKGC outperformed USAVM at this level. GASP and JMAGSM had the same relationship at 500 hPa as they did at MSLP, with GASP scoring better than JMAGSM in October and November and the positions reversing in December.

Figure 3(g) and Fig. 3(h) show the performance of four of these models over time. The improvement of ECSP and USAVM is evident at MSLP although the trend at 500 hPa has been more variable over the previous three months.

If the 60 per cent level of the anomaly correlation (Fig. 4) is taken as the mark of usefulness, then ECSP was providing good forecasts out to nearly 7 days, JMAGSM to nearly 6 days and GASP to 5+ days.

Synoptic intercomparison of global models

How good is the prediction of synoptic features in the medium-range? This section is a very crude attempt to begin answering this question, at least for five-day MSLP predictions from 1200 UTC global model predictions for the Australian region (loosely defined as between 10° and 40°S from 100° to 140°E and between 10° and 45°S from 140° to 160°E). We have attempted to classify errors into broad categories with the aim of identifying the types of synoptic situations which the models handle either poorly or well as a guide to the use of the model output in forecasting. Although based on forecast-analysis difference charts, the classification is subjective, and the cate-

Fig. 2(a) Comparison for GASP/ECSP/USAVM/UKGC/JMAGSM from October to December 1999. S1 Skill-scores of MSLP for combined base-times 0000 UTC/1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.

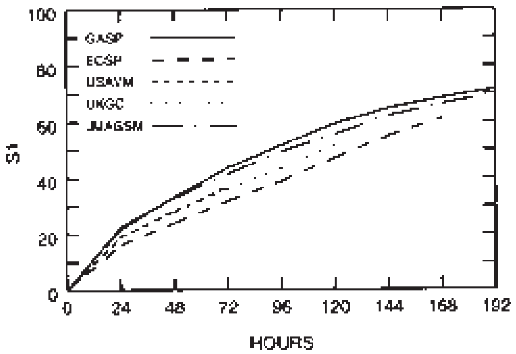


Fig. 2(b) Comparison for GASP/ECSP/USAVM/UKGC/JMAGSM from October to December 1999. Root mean square errors of MSLP for combined base-times 0000 UTC/1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.

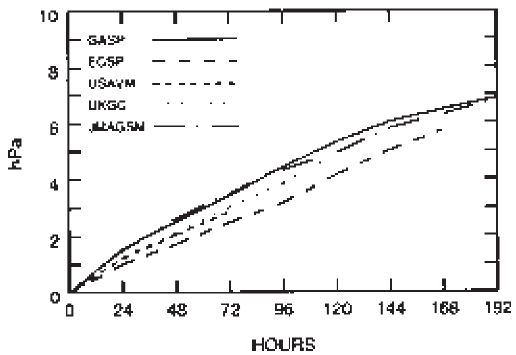


Fig. 2(c) Comparison for GASP/ECSP/USAVM/UKGC/JMAGSM from October to December 1999. S1 skill-scores of 500 hPa geopotential height (m) for combined base-times 0000 UTC/1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.

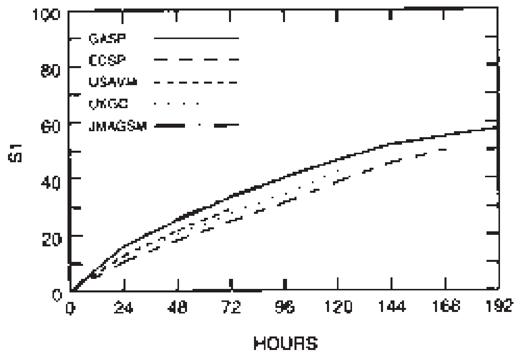
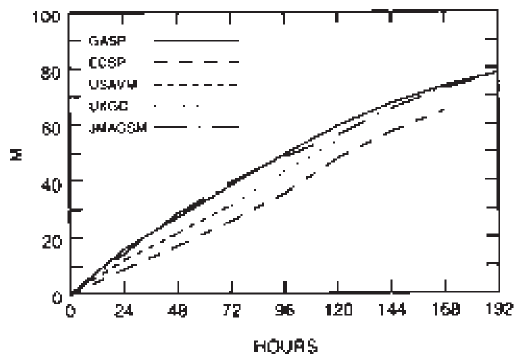


Fig. 2(d) Comparison for GASP/ECSP/USAVM/UKGC/JMAGSM from October to December 1999. Root mean square errors of 500 hPa geopotential height (m) for combined base-times 0000 UTC/1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.



gorisation ignores many subtle differences. Four global predictions routinely monitored in NMOC were assessed – GASP, ECMWF, JMA and UKMO.

Table 2 lists the error patterns used in the classification and the number of cases identified in each category for each global NWP system. The numbers should be taken as indicative, as the number of error patterns identified for a given day was not fixed.

As a general statement, it can be said that, outside the tropics, the five-day predictions at least showed a reasonable correspondence for the existence of synoptic features between the predicted and observed patterns. This is not to say that there were not large errors, but simply that in most cases it was possible to match features in the predicted and observed patterns. In 17 cases for ECMWF and 13 cases for UKMO there was no major difference between the predicted and observed pattern. (i.e. the forecast errors were

Fig. 3(a) **Monthly S1 skill-scores of MSLP for LAPS_PT375/TLAPS_PT375/GASP from October to December 1999 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.**

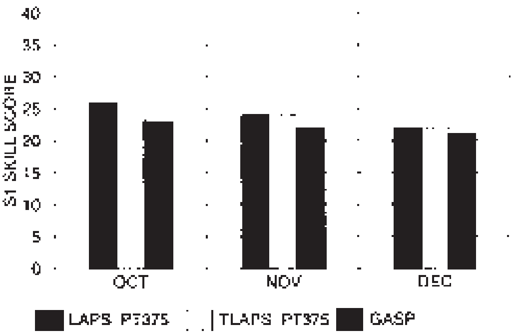


Fig. 3(b) **Monthly S1 skill-scores of 500 hPa geopotential height (m) for LAPS_PT375/ TLAPS_PT375/GASP from October to December 1999 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.**

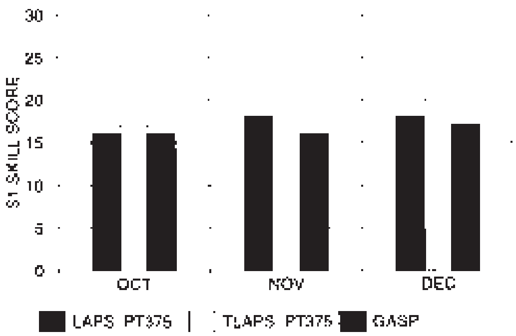


Fig. 3(c) **Monthly S1 skill-scores of MSLP for GASP/EC/UK/US/JMA from October to December 1999 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.**

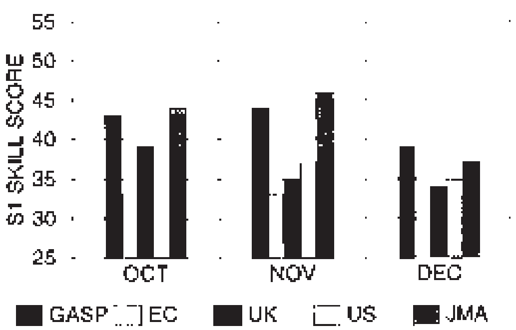


Fig. 3(d) **Monthly S1 skill-scores of 500 hPa geopotential height (m) for GASP/ EC/UK/US/JMA from October to December 1999 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.**

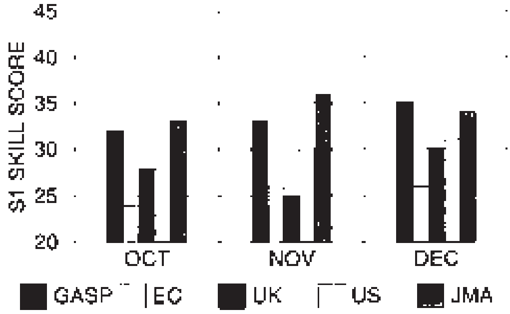


Fig. 3(e) **Monthly S1 skill-scores of MSLP for LAPS/LAPS_PT375/TLAPS/TLAPS_PT375 from January 1997 to December 1999 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.**

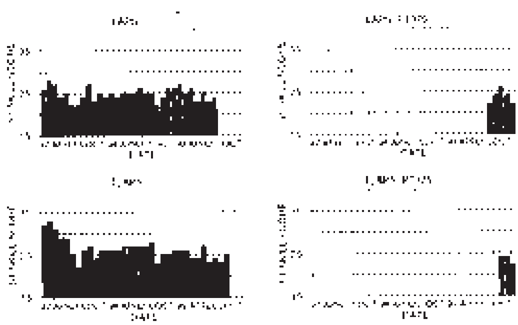


Fig. 3(f) **Monthly S1 skill-scores of 500 hPa geopotential height for LAPS/ LAPS_PT375/ TLAPS/TLAPS_PT375 from January 1997 to December 1999 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.**

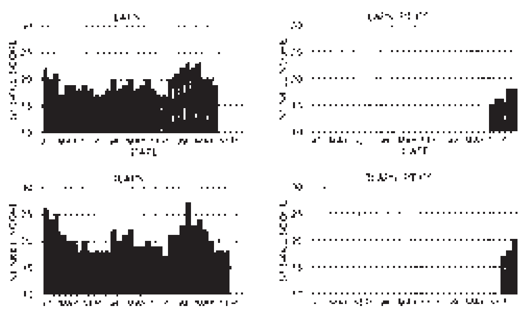


Fig. 3(g) **Monthly S1 skill-scores of MSLP for GASP/EC/UK/US from January 1997 to June 1999 for base-time 1200 UTC and interval +72h over the irregular Australian verification grid.**

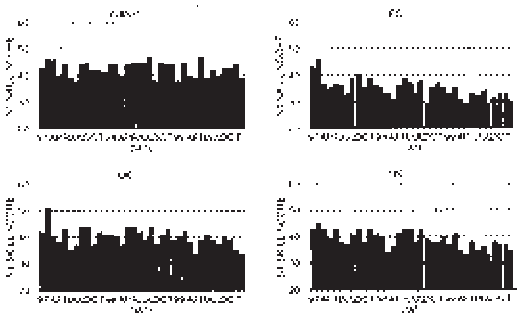


Fig. 3(h) **Monthly S1 skill-scores of 500 hPa geopotential height for GASP/ EC/UK/US from January 1997 to December 1999 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.**

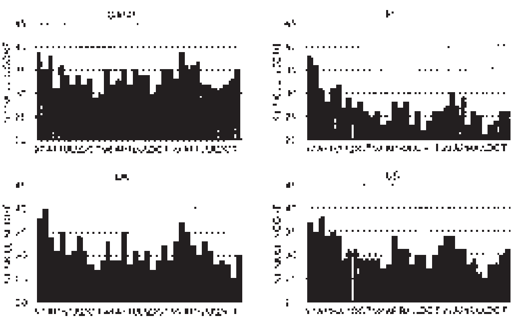


Fig. 3(i) **Monthly S1 skill-scores of MSLP for GASP/EC/UK/JMA from January 1997 to December 1999 for base-time 1200 UTC and interval +120 h over the irregular Australian verification grid.**

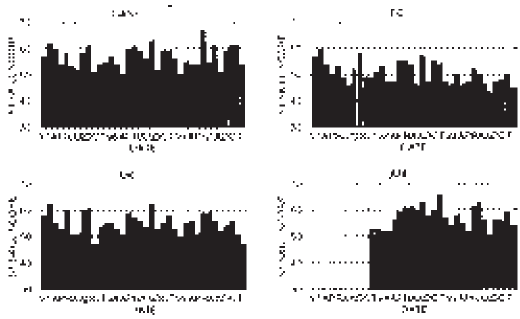
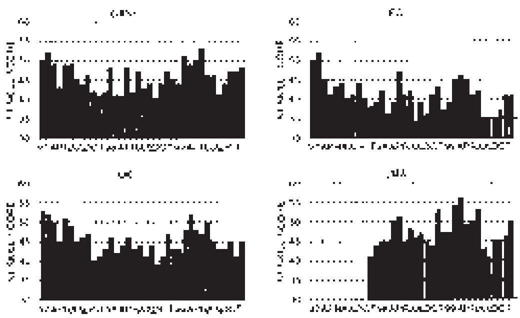


Fig. 3(j) **Monthly S1 skill-scores of 500 hPa geopotential height for GASP/ EC/UK/JMA from January 1997 to December 1999 for base-time 1200 UTC and interval +120 h over the irregular Australian verification grid.**



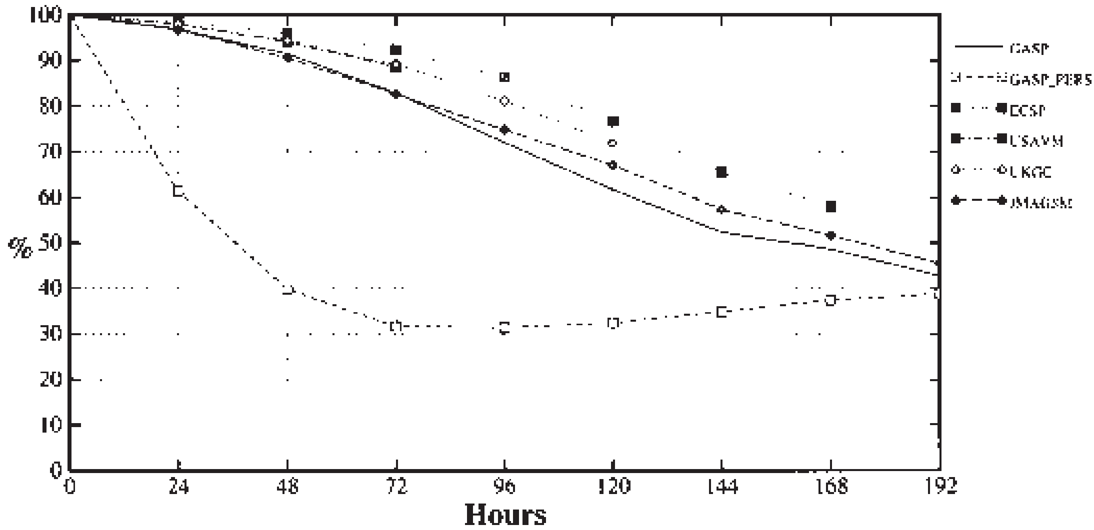
less than about 5 hPa over the domain). JMA and GASP had fewer such cases. In GASP's case this was due to some consistent biases evident in the predictions as discussed below. The cases of low prediction error generally involved an extensive strong high pressure system over the region or the combination of a strong high and a mobile trough in the westerlies.

On the other hand, the Table 2 shows that, except for the ECMWF model, in four or five cases for each of the other models it was not possible to match the observed and predicted features. Several of these cases involved slow-moving, decaying cut-off low pressure systems or deep lows in the westerlies. The UKMO had a run of such occurrences from 28-30 October. The analyses for this period showed strong westerly flow to the south of the continent with the migration of a strong trough. The starting analyses for 23-25 October showed a cut-off cyclone/anticyclone pair over eastern Australia and the Tasman Sea. Each of the four models, but the UKMO system in particular, had problems during this period of flow transition.

The following are some comments on particular synoptic features:

(a) Troughs and lows in easterly flow. Cut-offs and easterly dips were common during the October-December period. GASP and JMA showed a marked tendency to make such features too deep. UKMO showed the same characteristic to a lesser extent, while ECMWF had fewer problems overall with the prediction of intensity, and was fairly balanced between over and under-prediction. All except GASP were more likely to link such fea-

Fig. 4 **Anomaly Correlation of MSLP for GASP/ECSP/US/UKGC/JMAGSM from October to December 1999 for combined base-times 0000 UTC/1200 UTC over the irregular Australian verification grid.**



tures too much or too early to the westerlies. GASP tended to move these features too slowly by a clear margin, while the other models showed no strong bias on mobility.

For the particular case of troughs over Western Australia, GASP has a clear problem, showing 37 cases of the trough being too deep to only three cases of too much ridging. The error may only be several hPa but it is pervasive and often distorts the predicted pattern. Conversely, UKMO showed a propensity for too strong ridging over Western Australia and under-prediction of the strength of easterly troughs in that region. JMA and ECMWF had errors that were more evenly spread in sign.

(b) Extratropical lows or troughs in the westerlies. During the three-month period there were 12-14 cases for each model of spurious development of a low or at least marked over-development. In a few cases (such as 5 October, 3 and 26 November) the over-development was evident in three or four of the models. In most cases, however, these developments appeared in only one or two of the models. However, the models missed low pressure developments at about the same frequency. About half of these occasions tended to be common across all, or at least three of the models.

Particular cases were a low over the Bight on 8 and 12 October, over the Tasman on 19 October and over Victoria on 5 November. Predicted lows had about the correct strength but the location was poor in 5-10 other cases. A conclusion is that the ensemble of predictions may give some warning of the development of a low or trough, but absence of such a feature does not guarantee non-occurrence. The models do not give reliable prediction of cyclogenesis at the five-day range.

For mobile westerly troughs where the structure and mobility were acceptable, the intensity was underestimated rather than over-estimated by a factor of two or three to one, with this characteristic most marked in GASP.

(c) Ridging and high pressure systems. All models, but particularly UKMO and to a lesser extent ECMWF, tended to over-predict the strength of high pressure systems, and the extent of ridging. As noted above, this affected the performance of the UKMO system for the trough in the easterlies over Western Australia.

(d) Tropical systems. GASP showed a clear propensity to be less active in the tropics, in the sense of generating fewer discrete lows, than the other models. There were many occasions of poor prediction of tropical features, but apart from GASP, there was a fair spread between under- and over-prediction, as well as cases of mislocation.

In the case of systems which could be classed as tropical cyclones there were some excellent forecasts of genesis at the five-day range, but the performance was generally inconsistent even within the predictions from one model. As was the case with more general tropical lows, GASP was the least active of the group in predicting tropical cyclone genesis.

Table 2. Number of occasions during October-December 1999 when error occurred in each synoptic pattern category in the five-day MSLP prediction from each of the global models.

Error pattern	*GASP*	*ECMWF*	*JMA*	*UKMO*
1 No significant errors in pattern	2	17	8	13
2 Unable to match features	4	0	5	5
3 Low in easterlies or cut-off low				
A too linked to westerlies	4	8	6	9
B too much in easterlies	3	3	3	4
4 Trough in easterlies:				
A too deep	22	7	17	11
B too weak	2	5	2	4
5 Trough in easterlies over WA:				
A too deep	37	8	10	2
B too weak	3	5	6	10
6 Trough in easterlies:				
A too mobile	2	4	4	3
B too slow	7	1	4	4
7 Ridging or high pressure system:				
A too strong	7	10	5	14
B too weak	3	0	3	0
8A Spurious low or marked over-development	14	12	14	12
8B Missed development of a low or trough	13	9	12	11
9 Retaining a low or trough too long	1	-	-	-
10 Trough in westerlies basically correct in structure but;				
A too strong	5	10	7	6
B too weak	20	18	16	15
11 Low predicted well in intensity but location poor	8	5	10	5
12 Trough in westerlies:				
A too mobile	1	1	2	0
B too slow	2	5	3	2
13 Tropical lows:				
A over-developed	0	5	5	7
B missed or too weak	10	5	5	6
C mis-located	-	4	1	3

References

Skinner, W. 1995. Numerical prediction model performance summary April to June 1995. *Aust. Met. Mag., 44,* 309-12.