Aust. Met. Mag. 46 (1997) 319-324

Numerical prediction model performance summary April to June 1997

W. Skinner and T. Hart
National Meteorological Centre, Bureau of Meteorology, Australia

(Manuscript received September 1997)

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 Centre (NMC) Melbourne and from ECMWF (European Centre for Medium-range Weather Forecasts), NCEP (National Centers for Environmental Prediction) and UK (United Kingdom Meteorological Office).

The three models considered from NMC Melbourne are: LAPS (Limited Area Prediction System), TAPS (Tropical Analysis and 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) and UKGC (United Kingdom Meteorological Office Grid PE model).

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

For Figs 1, 2 and 3 results have been calculated within NMC Melbourne, where the models were verified against their own analyses for the irregular Australian verification area only. Figure 4 shows verification data supplied by ECMWF and NCEP for models verified against radiosondes in the southern hemisphere.

These statistics are in accordance with the recommendations of the World Meteorological Organization's Commission for Basic Systems. In this context the southern hemisphere is 20° to 90° S and models are verified against a list of 68 stations. An updated version of this list with 66 stations was implemented operationally from 3 June 1997.

All statistics are a measure of the skill in forecasting geopotential height or MSLP. Other field types are not included in these summaries.

LAPS and TAPS 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 TAPS grid. TAPS scores are calculated without these points and are therefore not strictly comparable with those from other models.

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

Notes on NWP systems

TAPS/TLAPS

The local model Tropical Analysis and Prediction System (TAPS) was replaced from 0000 UTC on 20 June 1997. The new model uses the same code as the Limited Area Prediction System (LAPS) and so has been assigned the acronym TLAPS. However since this summary is for the three-month period April to June inclusive and the new system was available for only ten days of that period, it has not been included in this summary. Subsequent articles will compare the TLAPS performance with other local models, but early results are promising.

ECMWF

Several modifications to the three-dimensional variational data assimilation system were implemented on 14 May 1997. These included modifications to the balance constraint giving a better fit of the analysis to observations especially in the tropics and amendments in the error growth factor for asynoptic data. Other changes were modifications in the use of TOVS data including

Fig. 1(a) Comparison for LAPS/TAPS/GASP from April to June 1997. 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.

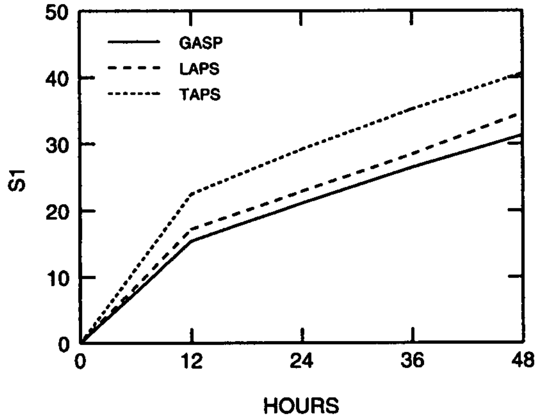
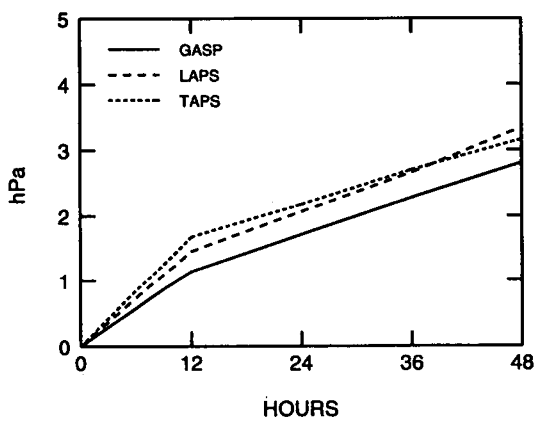


Fig. 1(b) Comparison for LAPS/TAPS/GASP from April to June 1997. Root mean square errors of MSLP for combined base-times 0000 UTC/1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.



revised tropical bias correction and restriction of the initialisation process to small scales (Lalaurette 1997).

April to June 1997 intercomparisons

Local models: (LAPS, TAPS, GASP)

The local models performed in descending order GASP, LAPS and TAPS (Fig. 1). This order is maintained for skill scores at MSLP and 500 hPa. Root mean square (rms) errors also show the same order out to +24 h but TAPS performs better than LAPS at longer intervals. The individual monthly scores (Figs 3(a), 3(b)) also retain this performance order. All the models had lower skill scores at both levels for the three months April to June than for January to March as the subtropical anticyclone became firmly established in the centre of the verification area.

Since this is the last summary for TAPS, two extra graphs (Figs 3(e), 3(f)) are supplied showing the time series of all local models' skill scores since July 1995. This period covers the change from RASP to LAPS,

Fig. 1(c) Comparison for LAPS/TAPS/GASP from April to June 1997. 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.

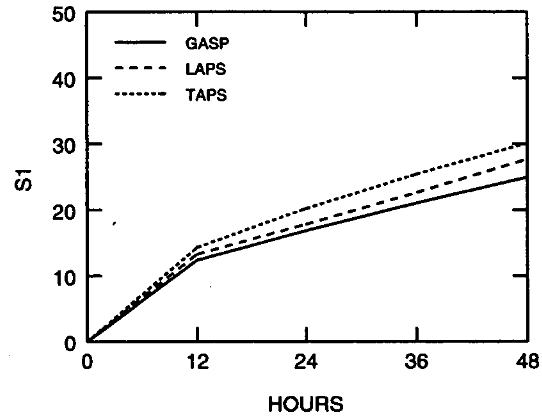


Fig. 1(d) Comparison for LAPS/TAPS/GASP from April to June 1997. Root mean square errors of 500hPa geopotential height (m) for combined base-times 0000 UTC/1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.

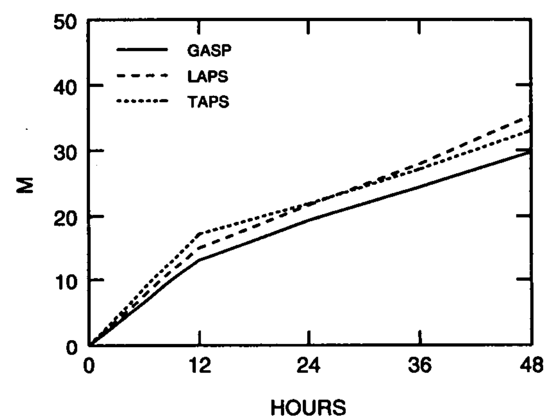


Fig. 2(a) **Comparison for GASP/EC/US/UK from April to June 1997. S1 skill scores of MSLP for combined base-times 0000 UTC/1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.**

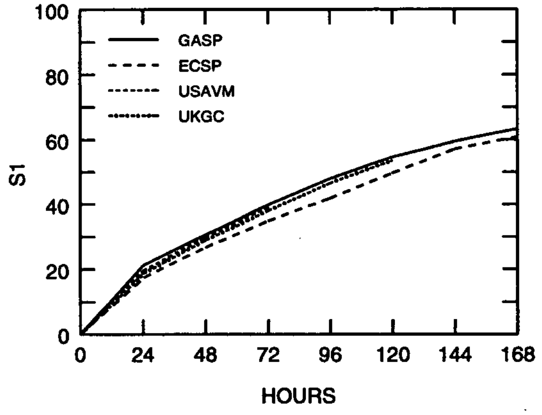


Fig. 2(b) **Comparison for GASP/EC/US/UK from April to June 1997. Root mean square errors of MSLP for combined base-times 0000 UTC/1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.**

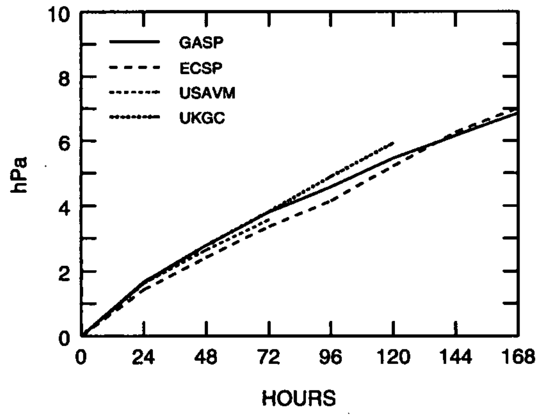


Fig. 2(c) **Comparison for GASP/EC/US/UK from April to June 1997. S1 skill scores of 500 hPa geopotential height (m) for combined base-times 0000 UTC/1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.**

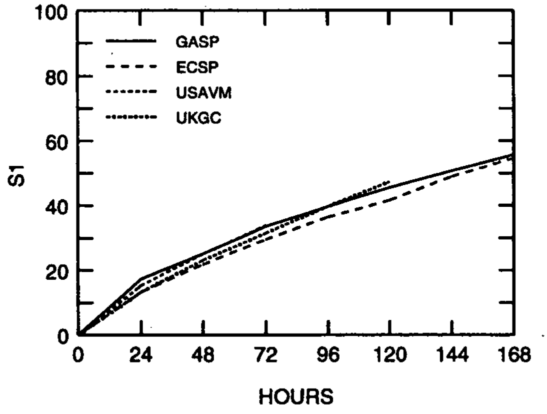
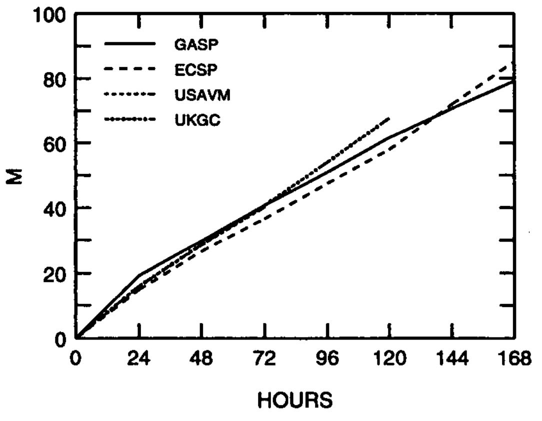


Fig. 2(d) **Comparison for GASP/EC/US/UK from April to June 1997. Root mean square errors of 500 hPa geopotential height (m) for combined base-times 0000 UTC/1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.**



which together span the interval, and clearly indicates the improvement in the later model. At both MSLP and 500 hPa the familiar order of GASP, LAPS, TAPS in decreasing skill is evident. However the period April to June 1997 was better for TAPS than the same three months in 1996. Indeed, June 1997 was the best recorded score for TAPS at MSLP.

Examination of the daily MSLP synoptic charts for the period shows a number of synoptic situations were poorly handled by the models. These situations strongly parallel those noted below where the global models performed poorly at +120 h. Most bad errors occurred where systems with strong gradients were incorrectly positioned, however the worst errors tended to fall outside the verification area (e.g. in fast-moving or amplifying systems in the Southern Ocean south of 50°) and eluded verification. During April, TAPS tended to consistently show errors in the Coral Sea reflecting relatively poor skill in forecasting the onshore gradient of the southeast trades.

On 1 and 2 April, all models overforecast a trough development in the Bight and its eastward movement. On this occasion, LAPS performed best.

Fig. 3(a) Monthly S1 skill scores of MSLP for LAPS/TAPS/GASP from April to June 1997 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.

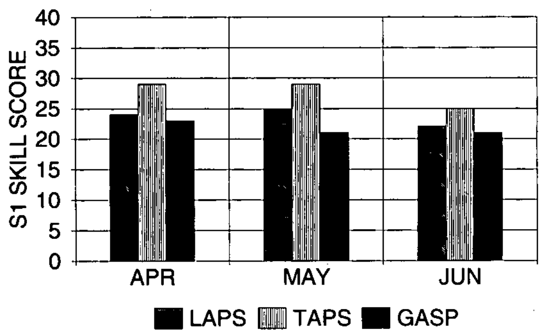


Fig. 3(b) Monthly S1 skill scores of 500 hPa geopotential height (m) for LAPS/TAPS/GASP from April to June 1997 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.

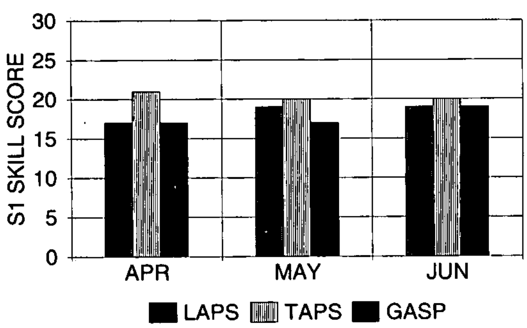


Fig. 3(c) Monthly S1 skill scores of MSLP for GASP/EC/UK/US from April to June 1997 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.

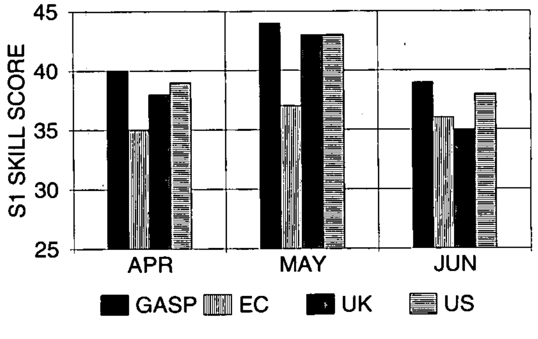


Fig. 3(d) Monthly S1 skill scores of 500 hPa geopotential height (m) for GASP/EC/UK/US from April to June 1997 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.

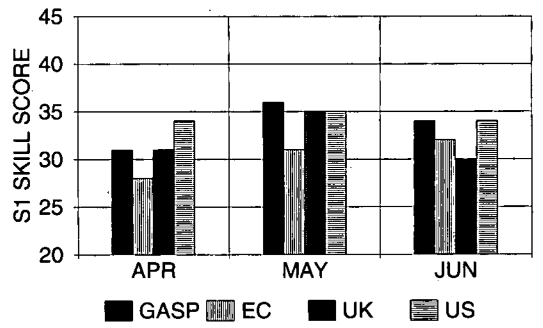


Fig. 3(e) Monthly S1 skill-scores of MSLP for RASP/LAPS/TAPS/GASP from July 1995 to June 1997 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.

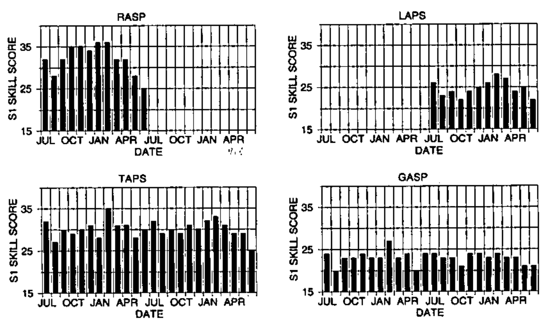
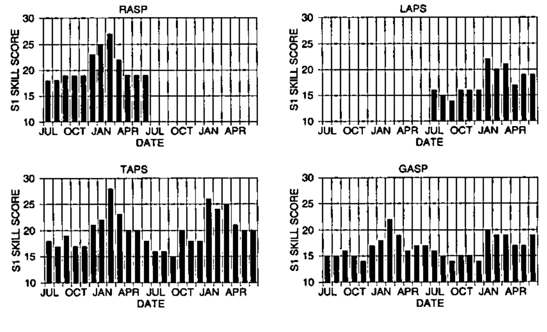


Fig. 3(f) Monthly S1 skill scores of 500 hPa geopotential height for RASP/LAPS/TAPS/GASP from July 1995 to June 1997 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.



On 3 to 5 May, an easterly trough over the northern Tasman Sea was deepened too quickly causing significant errors for TAPS. Later in the same month, (8th to 11th) an east coast low caused severe weather along the New South Wales coast. It was poorly estimated and positioned, particularly by TAPS. From 15 to 17 May, an easterly trough which formed over inland New South Wales and then moved out to sea, again caused errors by over-intensification. This time it was LAPS which had spectacular errors extending over southeast Australia and all of the south Tasman Sea on the 15th and 17th. Also, from 17 May, a cut-off low developed over southwest Western Australia and moved into the Bight. This too was poorly captured by all three models with LAPS producing the worst scores.

In June, starting on the 8th, a trough approaching Western Australia was thoroughly underestimated by TAPS and initially by LAPS. TAPS also showed poor performance on 13 and 14 June when it underestimated the speed and strength of the establishment of a high in the Bight.

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

Figure 2 shows that in general ECSP continues to outperform all other models. It is followed typically, by UKGC, USAVM and GASP. This order is slightly altered for skill scores at 500 hPa where GASP performance edges ahead of USAVM at +72 h and ahead of UKGC from +96 h (Figs 2(a), 2(c)). Root mean square errors show USAVM performing slightly better than UKGC at MSLP and almost identically at 500 hPa. GASP performs better than UKGC at both levels from +72 h and better than ECSP from +144 h (Figs 2(b), 2(d))

Scores for individual months (Figs (3(c), 3(d)) show GASP as the worst performer except at 500 hPa in April when both UKGC and USAVM scored relatively badly. ECSP is overall the best, but UKGC scores better in June at both MSLP and 500 hPa.

A comparison of global models at 1200 UTC +120h

The use of +120 h for this synoptic comparison has to be restricted to GASP, ECSP and UKGC as the USAVM model is only issued to +72 h. The region examined is 10-50°S and 100-160°E. Errors exceeded 10 hPa in magnitude on about 90 per cent of occasions for ECSP predictions and on nearly every day for those of GASP and UKGC. The frequency of errors exceeeding 20 hPa ranged from about 35 per cent for GASP and ECSP to 48 per cent for UKGC. These values have increased significantly from the low values of the previous predominantly summer period. GASP continued to be dominated by positive errors, ECSP errors were clearly skewed to negative and those of UKGC were fairly evenly distributed in sign. These aggregate figures suggest a continuing tendency for over-activity in the ECSP predictions, a tendency illustrated in the specific cases below.

Most of the large errors were located at high latitudes in the strong westerly regime. Over mainland Australia itself, the frequency of errors exceeding 10 hPa in magnitude was considerably less, ranging from 23 per cent for ECSP to 35 per cent for UKGC. There were only four cases where the error over the mainland exceeded 20 hPa; all were negative in sign associated with overdevelopment of cyclonic systems. These cases, discussed in more detail below, were the UKGC prediction for 9 May and the ECSP predictions for 19 and 20 May and 14 June.

The predictions for the period 7-11 May varied considerably among the models and from day to day for each model. This period contained a severe weather situation for the New South Wales coast. A mobile trough had developed into a cut-off low in the west Tasman Sea on 9 May. This low moved westwards and intensified to become an intense system close to the central NSW coast on 10 May producing heavy rainfall, and leading to extensive wind and wave damage. During 11 May the system commenced moving away, initially to the east-northeast and then southeastwards. The UKGC five-day predictions valid in the lead-up period (i.e. 7, 8 and 9 May) were premature and tended to over-predict development. For 10 May, when the system was at its most intense, the corresponding predictions were, by contrast, a significant under-prediction although better in location than the GASP and ECSP forecasts, both of which also failed to capture the initial development of the system. UKGC also gave the best 72-hour predictions for the mature stage of this situation.

Another situation poorly handled by the models was a cut-off low over southwest Western Australia and the Bight from 17-20 May. The initial development was handled best by ECSP, and missed completely by GASP. Subsequently all predictions linked this feature with a higher latitude trough and moved it too rapidly away to the southeast. The ECSP predictions spuriously intensified this complex trough as it moved eastwards, producing the large forecast errors noted above.

The final case of 20 hPa error, that of 14 June, where the ECSP predicted an over-active westerly trough crossing southeastern Australia. A spurious secondary circulation was depicted at the apex of the trough whereas in reality the flow behind the trough was anticyclonic. The 72-hour prediction for the same day was considerably better.

Another case of cyclonic development over the Tasman was poorly handled by the models at the five-day range. This occurred during the period 24-30 June when a low associated with a cut-off cold pool developed in easterly flow over the north Tasman region. Predictions during the period were quite variable. GASP predictions mostly failed to capture the system at all, showing a strong anticyclone until near the end of the event. The ECSP and UKGC initially developed an over-intense system much further to the southeast. Towards the later stages the predictions were better although those of ECSP were again too strong and too slow in movement once the system began to migrate.

A small cut-off low over southwest Western Australia on 8 June caused wind damage in the area. This system was missed by the three models at the five-day range.

Comparison of models against radiosondes

Figure 4 shows rms errors from a comparison of model forecasts against a list of 68 WMO-nominated radiosonde stations (66 since 3 June). The level is 500 hPa and intervals +24 h, +72 h and +168 h. Only GASP and ECSP can be shown at +168 h but USAVM is available up to +72 h. The verification area is from 20°S to

Fig. 4(a) Monthly root mean square errors of 500 hPa geopotential height (m) against radiosondes for GASP/EC/US from January 1996 to June 1997 for base-time 1200 UTC and interval +24 h over the southern hemisphere area 20° S to 90° S.

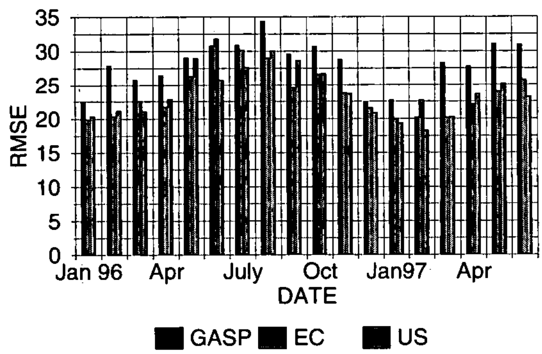


Fig. 4(b) Monthly root mean square errors of 500 hPa geopotential height (m) against radiosondes for GASP/EC/US from January 1996 to June 1997 for base-time 1200 UTC and interval +72 h over the southern hemisphere area 20° S to 90° S.

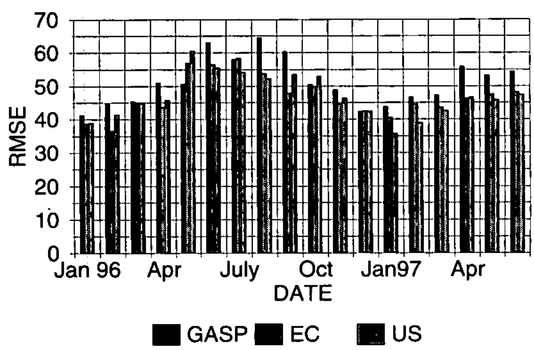
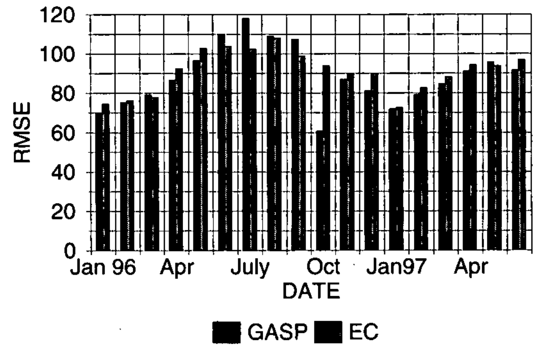


Fig. 4(c) Monthly rms errors of 500 hPa geopotential height (m) against radiosondes for GASP/EC from January 1996 to June 1997 for base-time 1200 UTC and interval +168 h over the southern hemisphere area 20° S to 90°S.



90°S and 0° to 360°E. The southern hemisphere area is used here, rather than the Australian region only, as it has a better coverage of radiosondes at 1200 UTC.

GASP's performance improves with the longer interval and by +168 h is slightly ahead of ECSP for eight of the last nine months (Fig. 4(c)). At +24 h and +72 h GASP is consistently the worst performer although the margin is worse at +24 h. This verification against radiosondes does not establish ECSP so clearly as the best performed global model. USAVM has lower rms errors about half the time (Figs 4(a), 4(b)).

References

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

Lalaurette, F. 1997. Changes to the operational forecasting system. *ECMWF Newsletter Number 75* (Spring 1997), 1-2.