Aust. Met. Mag. 47 (1998) 63-69

Numerical prediction model performance summary July to September 1997

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

(Manuscript received December 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 Operations Centre (NMOC), 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 NMOC, Melbourne, are: LAPS (Limited Area Prediction System), TLAPS (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) 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 Figures 1, 2 and 3 results have been calculated within NMOC 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 66 stations.

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

Notes on NWP systems

TAPS/TLAPS

The local model Tropical Analysis and Prediction System (TAPS) was replaced operationally from 0000 UTC on 20 June 1997. The new model has been assigned the acronym TLAPS (Tropical Limited Area Prediction System). Analysis (Seaman et al. 1995) and forecast model (Puri et al. 1996) features from the Limited Area Prediction System (LAPS) are used. TLAPS runs over essentially the same domain as the former TAPS system and retains the tropical specific components of diabatic nudging (Davidson and Puri 1992), synthetic GMS moisture observations and tropical cyclone bogusing. The horizontal resolution has been increased to 0.75°, improvements have been made to the numerics and convection parametrisation, and a soil moisture analysis scheme introduced. The operational configuration is for 19 vertical levels, horizontal resolution of 0.75° and the domain from 44.25°N to 45.0°S and 70.0°E to 189.25°E. This represents a 120x160 latitude/longitude grid.

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

July to September 1997 intercomparisons

Local models: (LAPS, TLAPS, GASP)

The new model TLAPS is assessed for the first time in this summary. Comparisons with this three-month period last year show the gap in skill scores at MSLP between the tropical models and LAPS have significantly reduced (Figs 1(a) and (c)). TLAPS shows a similar improvement over the previous TAPS model since the April to June period. At MSLP, TLAPS now outperforms LAPS for rms errors (Fig. 1(b)). This reverses the situation from both the April to June period and the July to September 1996 period. At 500 hPa, TLAPS had the highest rms errors by a slim margin (Fig. 1(d)). GASP performed best overall (Fig. 1).

Skill scores for the three single months (Fig. 3) at +24 h show the same trends. TLAPS is now scoring

Fig. 1(a) Comparison for LAPS/TLAPS/GASP from July to September 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.

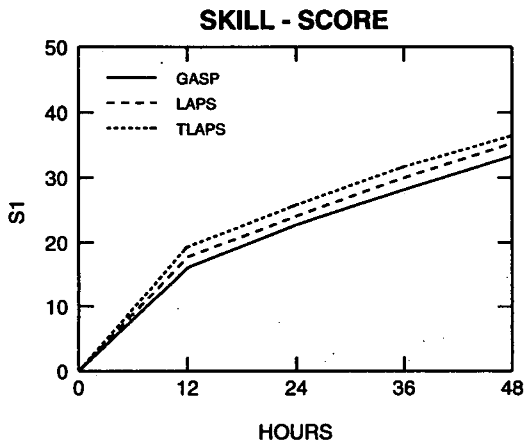


Fig. 1(b) Comparison for LAPS/TLAPS/GASP from July to September 1997. Root mean square errors of MSLP for combined base-times 0000 UTC / 1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.

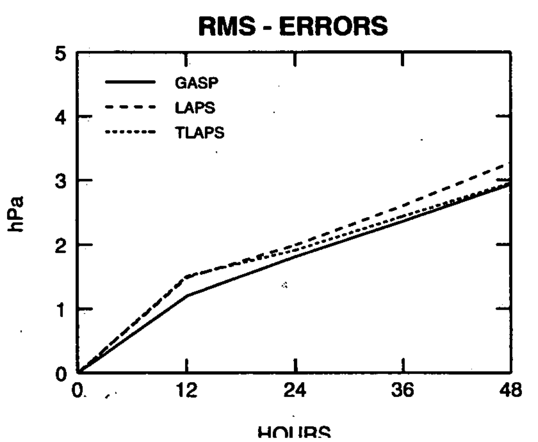


Fig. 1(c) Comparison for LAPS/TLAPS/GASP from July to September 1997. S1 skill scores of 500 hPa geopotential height (m) for combined base-times 0000 UTC / 1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verifica-

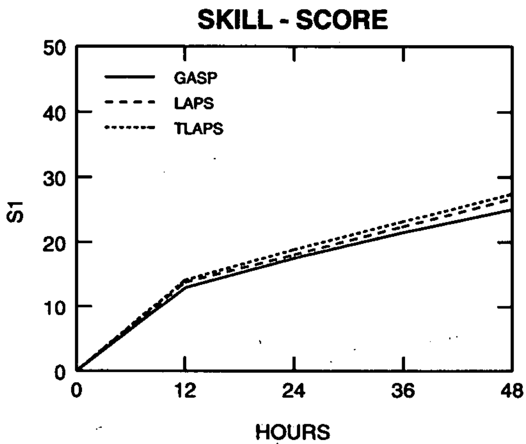


Fig. 1(d) Comparison for LAPS/TLAPS/GASP from July to September 1997. Root mean square errors of 500 hPa geopotential height (m) for combined base-times 0000 UTC /1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.

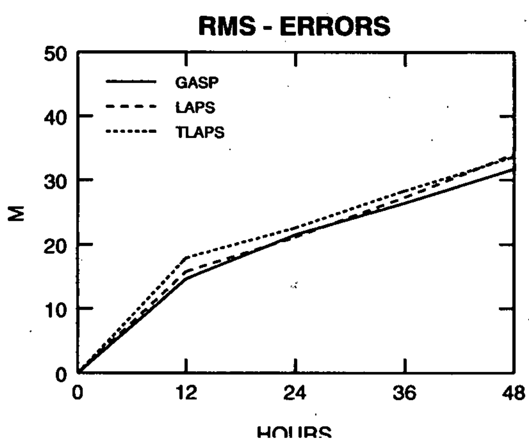


Fig. 2(a) Comparison for GASP/EC/US/UK from July to September 1997. S1 skill scores of MSLP for combined base-times 0000 UTC / 1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.

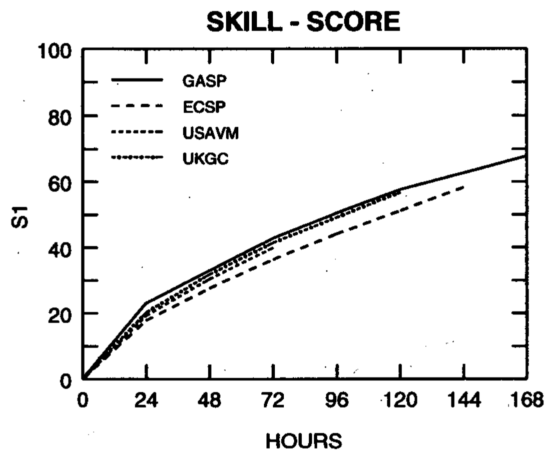


Fig. 2(b) Comparison for GASP/EC/US/UK from July to September 1997. Root mean square errors of MSLP for combined base-times 0000 UTC / 1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.

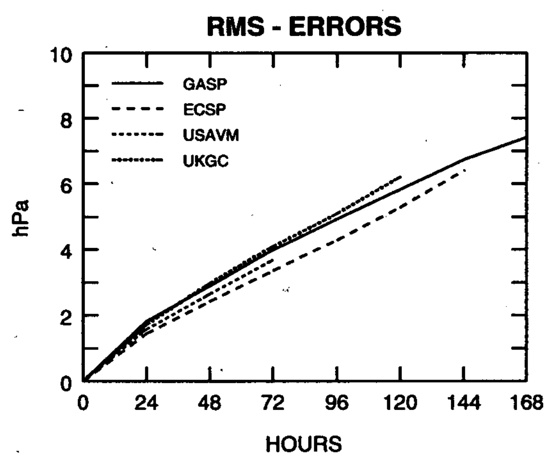


Fig. 2(c) Comparison for GASP/EC/US/UK from July to September 1997. S1 skill scores of 500 hPa geopotential height (m) for combined base-times 0000 UTC / 1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.

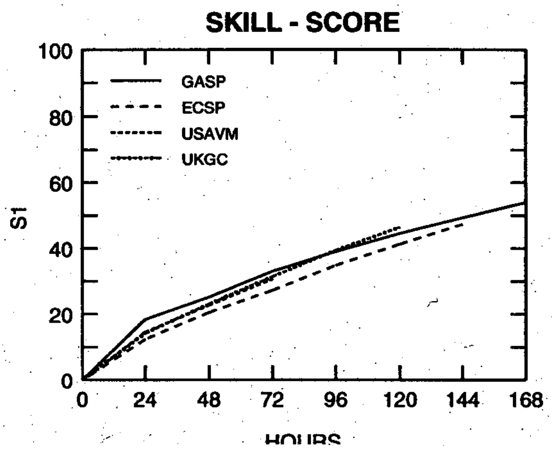
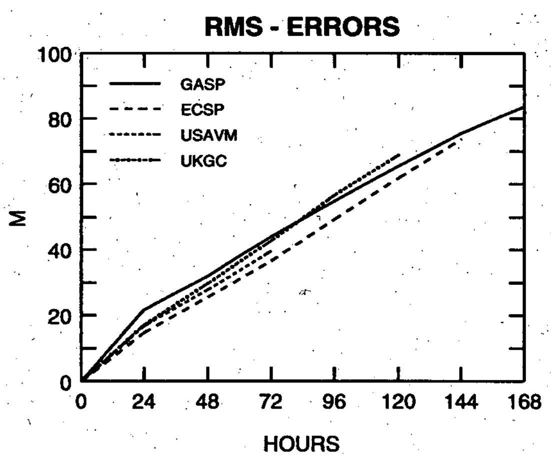


Fig. 2(d) Comparison for GASP/EC/US/UK from July to September 1997. Root mean square errors of 500 hPa geopotential height (m) for combined base-times 0000 UTC / 1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.



only slightly worse than LAPS. At MSLP (Fig. 3(a)) it averages five skill-score points better than TAPS did for this period last year. At 500 hPa (Fig. 3(b)) the relative differences between the models have not changed much in twelve months. However, persistence in the same three-month period last year scored four to five skill-score points higher than this year so the model scores are also higher. Comparison with the period April to June 1997 shows an improvement in score for TLAPS of two to three skill-score points over TAPS. A longer time series (Figs 3(e) and (f)) from July 1995 shows the improvement of TLAPS over TAPS and the Australian region LAPS over its predecessor RASP.

Synoptic overview for 24 h predictions

In the first three weeks of July the Australian verifica-

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

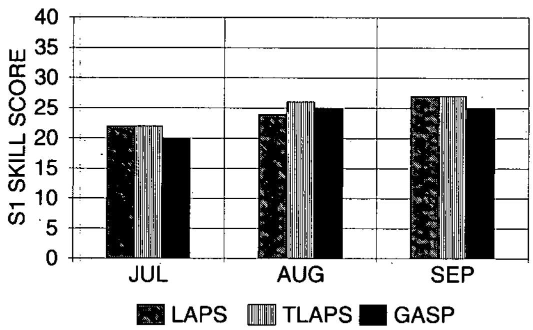


Fig. 3(b) Monthly S1 skill scores of 500 hPa geopotential height (m) for LAPS/TLAPS/GASP from July to September 1997 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.

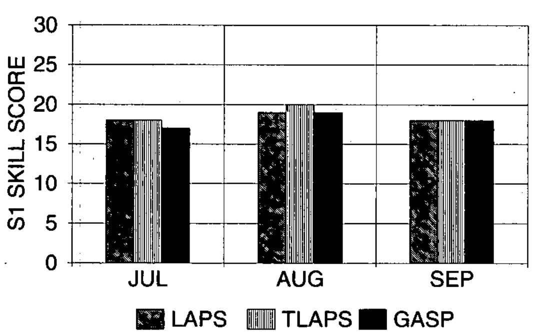


Fig. 3(c) Monthly S1 skill scores of MSLP for GASP/EC/UK/US from July to September 1997 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.

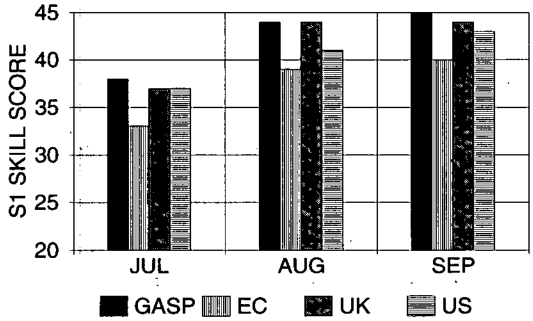


Fig. 3(d) Monthly S1 skill scores of 500 hPa geopotential height (m) for GASP/EC/UK/US from July to September 1997 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.

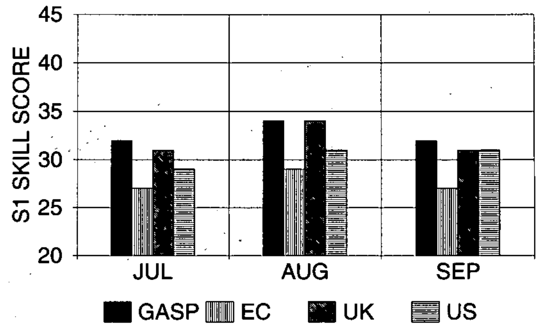


Fig. 3(e) Monthly S1 skill scores of MSLP for RASP/LAPS/TAPS/TLAPS from July 1995 to September 1997 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.

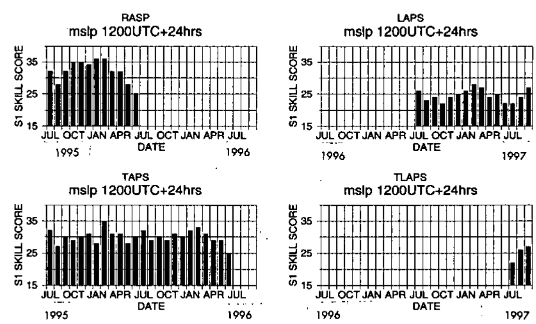
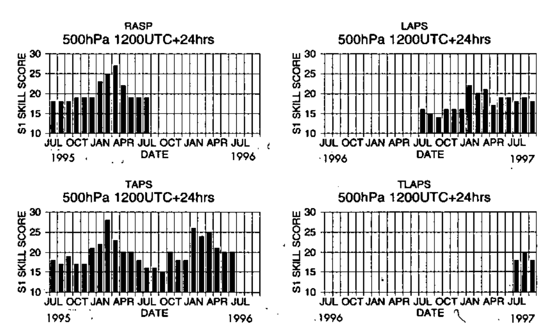


Fig. 3(f) Monthly S1 skill scores of 500 hPa geopotential height for RASP/LAPS/TAPS/TLAPS from July 1995 to September 1997 for base-time 1200 UTC and interval +24h over the irregular Australian verification grid.



tion area was dominated by large, slowly moving anticyclones. The local models handled these fairly well with occasional errors in the depiction and movement of the upstream troughs. Large errors were nearly always in the Southern Ocean and outside the verification area. By the end of August the pattern was more meridional and a slow-moving cut-off in the Tasman Sea caused problems. GASP scored a massive 38 on the 24th and 35 on the 25th of August; scores which were nearly ten points above the monthly average. LAPS handled this situation best. By the 30 August, an easterly dip over the Bight caused all models to score badly. A complex east-west trough resulted, and the subsequent period to 5 September produced consistent skill scores over 30 for all models. On 9 September none of the models picked the development of the Western Australian trough and the cradling effect of the eastern Indian Ocean and the Bight highs. September showed a marked increase in the blocking index over longitudes 90° to 180°E with cut-off easterly troughs being common. The discussion of the global model comparison at +120 h mentions some of these situations.

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

The series of graphs (Fig. 2) demonstrate the on-going superiority of ECSP at both levels and for both skill-score and rms error. The USAVM model was overall the second best, although UKGC had better skill scores at 500 hPa (Fig. 2(c)) and hence reversed the situation from April-June. GASP tended to outperform the UKGC model at longer forecast ranges, but was, in general, the least skillful.

The skill scores at +72 h for individual months (Figs 3(c) and (d)) show a similar pattern with the order of decreasing skill being ECSP, USAVM, UKGC and GASP.

A comparison of 1200 UTC 120-hour predictions from the global models

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 +72h. Over the region 10° to 50°S and 100° to 160°E errors exceeded 10 hPa, somewhere within the area, on nearly every day for each of the models considered. The frequency of errors exceeding 20 hPa ranged from 41 per cent for ECSP to 46 per cent for UKGC. This was quite a contrast to the same period last year for UKGC when the comparable figure was 68 per cent. This improvement is also reflected in the skill scores for the Australian region where the UKGC performed better than GASP this year, the reverse of last year. These large errors were fairly evenly distributed in sign for UKGC, strongly dominated by positive errors in GASP, while ECSP errors were skewed to negative. The high frequency of these large errors confirms that the models continue to have major problems at the five-day range in the prediction of the individual cyclonic circulation features in the strong westerly regime at high latitudes.

Over Australia itself, the frequency of errors exceeding 10 hPa in magnitude varied from 38 per cent for ECSP to about 50 per cent for GASP and UKGC. There are nine cases where the error exceeded 20 hPa mainly associated with failure to adequately predict the intensity of low pressure systems.

Skill and the changing flow pattern

The synoptic pattern for the first half of the period showed a preponderance of high pressure systems over the southern Australian region interspersed with mobile troughs in the westerlies. In the second half there was a change in the pattern with a high frequency of slow-moving cut-off low pressure systems over southern Australia. The skill scores showed a strong variation with this change in pattern. The average MSLP skill score over the Australian region for GASP was 52 in July but 61 in September. This variation during the period was typical of the other models. During the second half of September the five-day persistence value was of comparable skill to the model predictions on many occasions although this appeared to be an exceptional period for the skill of persistence, and hints at a small component of a five-day cycle during this period.

The worst skill scores for the period for each of the models occurred during 26 to 30 August. The synoptic situation showed a cut-off low developing near the southwest of Western Australia and moving slowly eastwards to the eastern Bight. The models initially failed to capture this system.

GASP and UKGC subsequently absorbed the low into a trough in the easterlies over the continent, while ECSP moved it southeastwards too quickly. ECSP also developed a spurious new low near Perth in the five-day forecast for 29 August. Overall, however, the ECSP predictions were best. A similar case occurred on 24 July when again GASP and UKGC missed the initial development.

The significant variation of skill during the period and the specific case noted above, provide a guide to the dependence of model performance on the synoptic situation. In particular, it confirms that at the five-day range the level of skill is much lower for cases involving slow-moving cut-off low pressure systems, particularly those which originate over the eastern Indian Ocean.

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

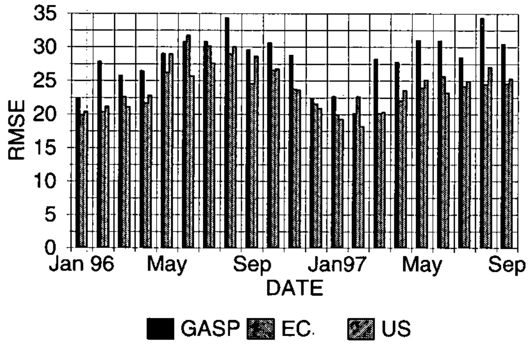


Fig. 4(b) **Monthly rms errors of 500 hPa geopotential height (m) against radiosondes for GASP/EC/US from January 1996 to September 1997 for base-time 1200 UTC and interval +72 h over the southern hemisphere area 20°S to 90°S.**

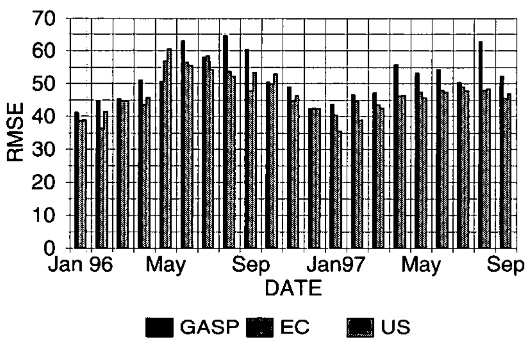
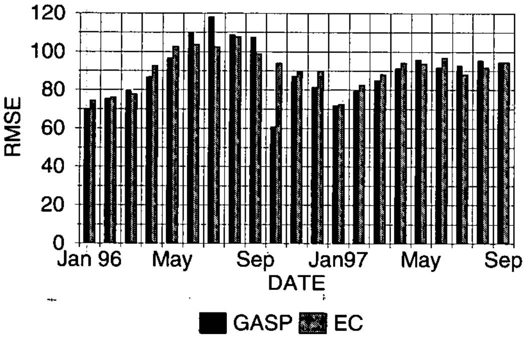


Fig. 4(c) **Monthly rms errors of 500 hPa geopotential height (m) against radiosondes for GASP/EC/US from January 1996 to September 1997 for base-time 1200 UTC and interval +168 h over the southern hemisphere area 20°S to 90°S.**



Notable major errors

The development of a cut-off low east of Tasmania on 23 to 24 August was handled quite poorly in some of the five-day predictions. The low developed from a cyclonic vorticity advection centre on the west of a mobile trough in the westerlies, bringing cold outbreak conditions to southeastern Australia. UKGC in particular moved the trough too rapidly and failed to show any development of a secondary centre. Again ECSP provided the best predictions.

There are several examples of spurious developments of low pressure centres. One example occurred over the Tasman Sea on 11 to 13 July when UKGC developed a strong low and ECSP a marked easterly dip. GASP had a spurious easterly dip on 11 July. Another spurious easterly dip was predicted by ECSP east of Brisbane on 19 August. On 28 July all models predicted a trough over the eastern Bight region with ECSP showing a marked front. The verifying analysis had a strong high over the area.

Prediction of individual low pressure centres in the strong westerlies continued to be a problem at the five-day range even though the phase of the broader scale troughs may be well handled. ECSP in particular occasionally spins up a spuriously deep centre. One example occurred on 5 July when the ECSP prediction showed a low of 976 hPa centred near 40°S 105°E and that of UKGC a centre of 990 hPa further east. The verifying analysis showed a broad trough in the area without these intense centres.

GASP

With the seasonal heating of the continent, GASP's problem of over-developing inland troughs re-appeared. This feature compromises predictions as weaker troughs become merged with these over-developed heat troughs. Examples occurred from mid-September on, in particular 22 to 25 September.

GASP was again the most likely to miss developments of small-scale lows, and least likely to over-develop systems.

UKGC

UKGC showed more variability than the other systems with some very good forecasts, but also stood out from the others on the less skillful side on many occasions. It should be remembered that UKGC is running with a considerably shorter data cut-off than the other two systems (three hours compared with seven hours for GASP and nine hours for ECSP).

One noticeable feature during the period was a tendency for UKGC to over-predict the central pressures of highs. Predictions during the second half of July were quite consistent in showing this characteristic.

Comparison of models against radiosondes

Figure 4 shows rms errors from a comparison of model forecasts against a list of 66 WMO-nominated radiosonde stations. 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 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 at the longer ranges and by +168 h is equal to or slightly better than ECSP for nine of the last twelve months (Fig. 4(c)). At +24 h and +72 h GASP consistently has the highest errors, particularly 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 of the time (Figs 4(a) and (b)).

References

Davidson, N.E. and Puri, K. 1992. Tropical prediction using dynamical nudging, satellite-defined convective heat sources and cyclone bogus. *Mon. Weath. Rev., 120*, 2501-22.

Puri, K., Dietachmayer, G., Mills, G., Davidson, N., Bowen, R., Logan, L. and Leslie, L. 1996. The new BMRC Limited Area Prediction System: A discussion of the numerics. *BMRC Research Report No. 54*, Bur. Met., Australia.

Seaman, R., Bourke, W., Steinle, P., Hart, T., Embery, G., Naughton, M. and Rikus, L. 1995. Evolution of the Bureau of Meteorology's global analysis and prediction system. Part 1:analysis and initialisation. *Aust. Met. Mag., 44*, 1-18.

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