Numerical prediction model performance summary April to June 1998

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

(Manuscript received September 1998)

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). Some reference is also made to the JMA (Japan Meteorological Agency) model.

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) and JMAGSM (JMA Global Spectral 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.

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

All statistics are a measure of the skill in forecasting geopotential height or mean sea level pressure, 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

ECSP

The model spectral resolution was increased from T213 to TL319 on 1 April 1998. The 'TL' notation refers to a linear Gaussian grid made possible by the use of a new two-level semi-Lagrangian numerical scheme. This means that despite the greater accuracy in representing atmospheric features with the higher spectral resolution, the grid calculations, including the physical parametrisations, are carried out on approximately the same grid as for the T213 calculations, with a consequent saving in computation.

Also on 1 April, improvements were made to the representation of orography which included corrections for some large errors over Antarctica, and improvements in the determination of sea-ice limits (Lalaurette 1998).

On 29 June, several other changes were made:

- the atmosphere and ocean wave models were coupled;
- significant level winds, temperatures and humidities were added to the standard levels used for radiosonde data;
- SSM/I total column water vapour data, analysed with a one-dimensional variational scheme, were added to the data used in the analysis.

USAVM

An upgrade of the NCEP global model took place from the 1200 UTC run on 15 June. Resolution was increased from T126 to T170 in the horizontal and from 28 levels to 42 levels in the vertical. This will affect only two weeks of the three months under consideration here. Unfortunately, performance has not been as anticipated and plans are to drop back to the lower resolution model at the end of September 1998. The modifications for physical parametrisation and for the analysis associated with the T170/L42 package are to be retained.

The new features which were implemented in June and are expected to be ongoing are:

- short wave radiation;
- changes to the land-surface package;
- new gravity wave drag scheme;
- improved convection;
- modifications to the analysis which make better use of satellite data.

Further details can be found on the web at http://sgi62.wwb.noaa.gov:8080/tpb97/TPB98/test.419.html http://www.ncep.noaa.gov/NCO/PMB/

UKGC

The UKGC global unified model was upgraded to higher resolution on 29 January 1998. The horizontal grid spacing was reduced from 90 to 60 km, and the number of vertical levels increased from 19 to 30. This upgrade follows the replacement of the C90 with the Cray T3E supercomputer as the operational machine. The assessment of the forecasts from the UKGC states that the amplitude and intensity of extratropical lows and troughs is greater in the new model, with stronger and more intense rainfall patterns. More accurate jet-stream forecasts are achieved. Tropical cyclones are considerably deeper with the enhanced resolution, and in better agreement with tropical cyclone advisories.

Improvements to the orography dataset over Antarctica were made on 12 May. The updated orography corrects errors in the order of 700 metres. The old orography was too high from 30°W to 30°E and too low from 30°E to 60°E. Trials showed that the largest impacts are in the surface winds and temperatures directly over Antarctica with improvements in both analyses and predictions. Away from Antarctica the impacts are small, although there is a reduction of the persistent westerly bias at 60°S.

JMAGSM

Medium-range predictions from the JMA Global Spectral Model have been verified within NMOC since October 1997. The products are available on a 2.5x2.5 degree grid and extend to a range of eight days at 1200 UTC. The resolution is T213 in the horizontal and 30 levels in the vertical. The analysis uses a multivariate three-dimensional statistical interpolation scheme except for the tropical region where a univariate scheme is applied.

April to June 1998 intercomparisons

Local models: (LAPS, TLAPS, GASP)

GASP continues to provide the best scores for MSLP and 500hPa and for both statistical types, skill score and rms-error (Fig. 1, Figs 3(a), 3(b)). LAPS outperforms TLAPS except at intervals +36 hours and +48 hours for 500 hPa rms error (Fig. 1(d)). The average plots for both 0000 UTC and 1200 UTC and all forecast intervals (Fig. 1) show this relative performance difference and it is confirmed by the 'snapshot' plots (Figs 3(a), 3(b)) which give the relationships for individual months at only one base and interval – 1200 UTC +24 hours. This second view of the scores reveals that the differences between models were small in June, particularly at 500 hPa (Fig. 3(b)). Figures 3(e) and 3(f) show a longer time-series of skill scores from July 1995. The relative performances of the three models is well-established over the longer period.

Synoptic overview for 24 h predictions

All models had difficulties at times with lows in easterly streams, on both sides of the continent. They usually underestimated the intensity of such systems and mistimed their capture in the westerly stream to the south. There were also, however, occasions when the models forecast similar situations well.

Examples of poor forecasts in West Australian latitudes occurred on 10 to 11 April and 29 to 30 June. On 17 June, 8, 25 and 31 May initial development predictions were poor but subsequent movement was quite well described. During 6 to 8 June, a similar situation was well handled.

On the eastern side of Australia, 23 to 27 April saw a deepening easterly dip off the coast move SW and develop into a cut-off low with subsequent easterly movement. All models had errors in the development phase, but showed movement of the mature systems accurately. A cut-off low moving SE across the Tasman towards New Zealand from 5 to 7 May produced poor scores although the shape and direction of movement were realistic. On 11 and 12 June a low in the Tasman was under-forecast and was subsequently moved too fast. Another cut-off, forming the cyclonic part of a blocking pair in the same area from 18 to 21 May was well handled as was another east-coast low system from 22 to 25 June. On 22 April, TLAPS and LAPS overdeveloped the easterly trough over northeastern Australia.

Fig 1(a) **Comparison for LAPS/TLAPS/GASP from April to June 1998. 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.**

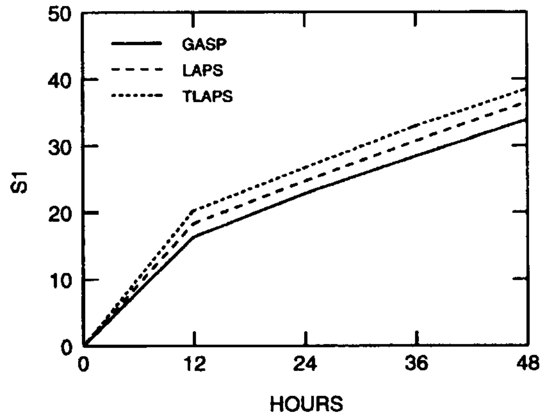


Fig 1(b) **Comparison for LAPS/TLAPS/GASP from April to June 1998. Root mean square errors of MSLP for combined base-times 0000 UTC / 1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.**

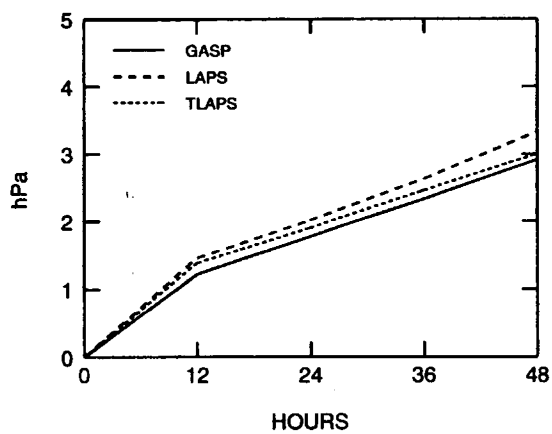


Fig 1(c) **Comparison for LAPS/TLAPS/GASP from April to June 1998. S1 skill scores of 500 hPa geopotential height (m) for combined base-times 0000 UTC / 1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.**

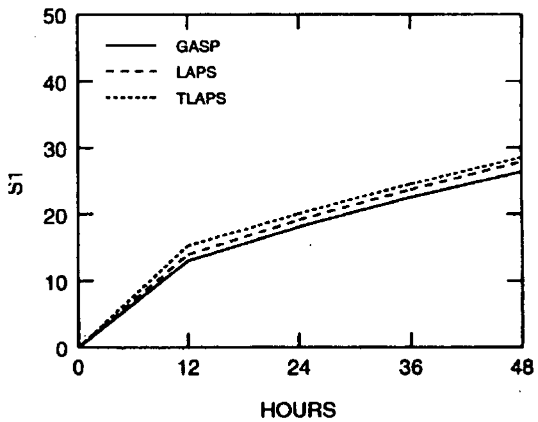
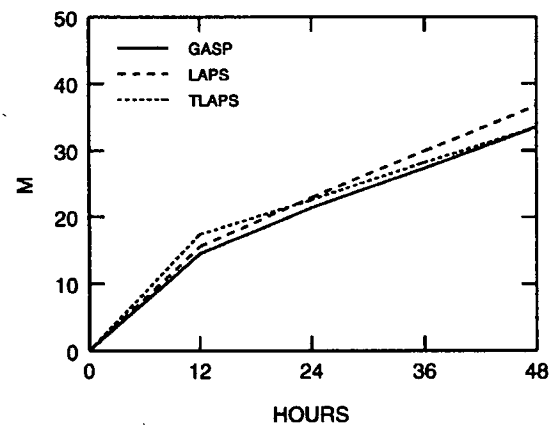


Fig 1(d) **Comparison for LAPS/TLAPS/GASP from April to June 1998. Root mean square errors of 500 hPa geopotential height (m) for combined base-times 0000 UTC / 1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.**



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

The skill-scores and rms errors (Fig. 2) clearly identify ECSP as the best and GASP as the poorest overall of the four models.

USAVM is slightly better than UKGC at MSLP (Figs 2(a), 2(b)) but it is difficult to detect any difference at 500 hPa. GASP does provide lower rms errors than UKGC for longer intervals, and for forecasts of +120 hours or more, it scores better than ECSP also.

The skill-scores at +72 hours for individual months (Figs 3(c), 3(d)), confirm ECSP and GASP in their relative positions, but show considerable variability between UKGC and USAVM. The time-series of these relationships (Figs 3(g), 3(h)) shows similar results.

The anomaly correlation (Fig. 4) includes plots of the JMA model and of persistence generated from GASP analyses. The JMAGSM scores were, on average, less skillful than GASP for this period. If the 60 per cent correlation line is taken as a measure of usefulness, all

Fig 2(a) **Comparison for GASP/EC/US/UK from April to June 1998. 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.**

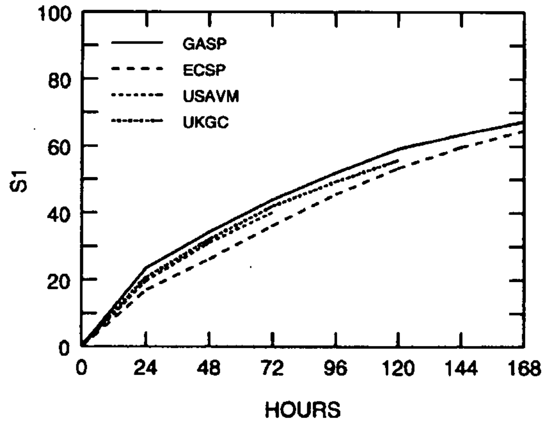


Fig 2(b) **Comparison for GASP/EC/US/UK from April to June 1998. Root mean square errors of MSLP for combined base-times 0000 UTC / 1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.**

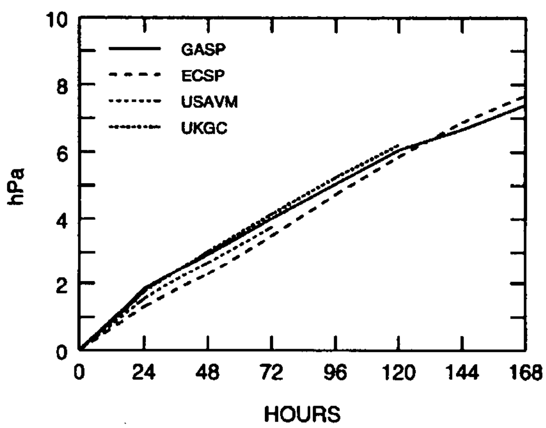


Fig 2(c) **Comparison for GASP/EC/US/UK from April to June 1998. S1 skill scores of 500 hPa geopotential height (m) for combined base-times 0000 UTC / 1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.**

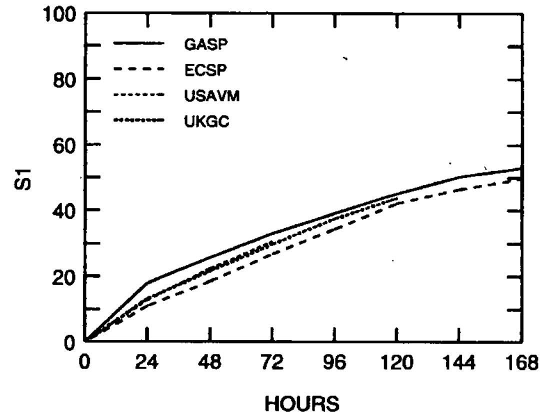
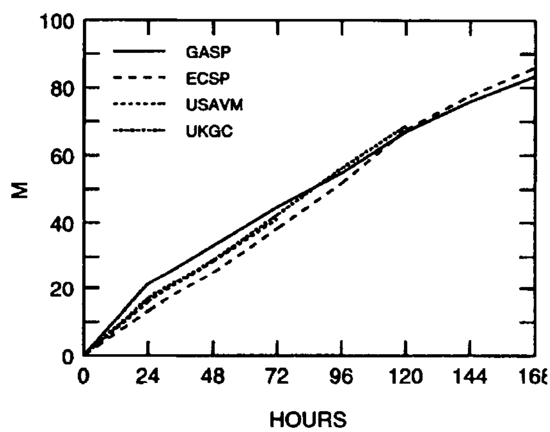


Fig 2(d) **Comparison for GASP/EC/US/UK from April to June 1998. Root mean square errors of 500 hPa geopotential height (m) for combined base-times 0000 UTC / 1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.**



models are useful at +96 hours and UKGC and ECSP are still useful at +120 hours. Longer interval forecasts are, on average, below this reference level although on an individual day, the forecast may be good.

A comparison of 1200UTC predictions from the global models

Some general comments applicable to the whole range of prediction times is followed by an assessment of the specific time, +120 hours. For +120 h the synoptic comparison is restricted to GASP, ECSP and UKGC as the USAVM model is only issued to +72 h. On this occasion, some references are also made to the Japanese model, JMAGSM.

A feature of the April to June period was the high number of cut-off lows and easterly troughs which formed off both the east and west coasts. Average prediction skill for the three-month period was actually lower than for the corresponding period last year despite model improvements. Skill scores for all models increased by 3-5 points for forecast ranges of 96 hours and longer. For shorter ranges all models showed some decrease in skill, except ECSP whose performance was equivalent to last year. Consequently, the gap between ECSP and others increased in forecast ranges up to 72 hours. However, in the five-day predictions the UKGC is relatively closer to ECSP compared with last year.

The initial assessment of a 'major error' was based on difference fields between the model's 120-hour MSLP forecast and its own analysis, with errors greater than 15 hPa between 45°S and 10°S, 110°E and 160°E.

The performance of the global model predictions at the five-day range was mixed with some very good forecasts indicating developments of lows or troughs if not the fine detail, but also some cases where there were very poor indications of such developments. For example, in the 5-day forecasts valid for 11-13 April the models all missed the initial development of a low over the Bight on 11 April and performed poorly during its subsequent movement eastwards. On 11 April forecast errors exceeded 20 hPa over the Bight for each model. Other examples of poor predictions for such synoptic situations were:

Fig 3(a) **Monthly S1 skill scores of MSLP for LAPS/TLAPS/GASP from April to June 1998 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.**

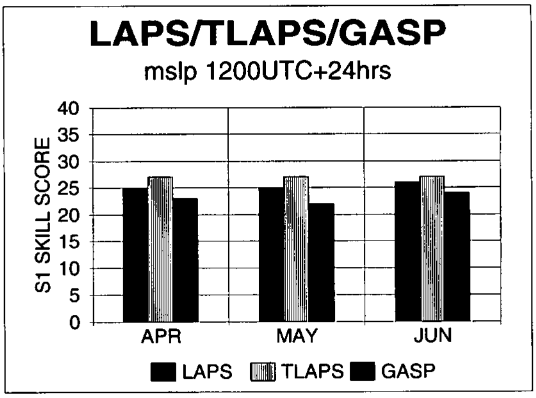


Fig 3(c) **Monthly S1 skill scores of MSLP for GASP/EC/UK/US from April to June 1998 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.**

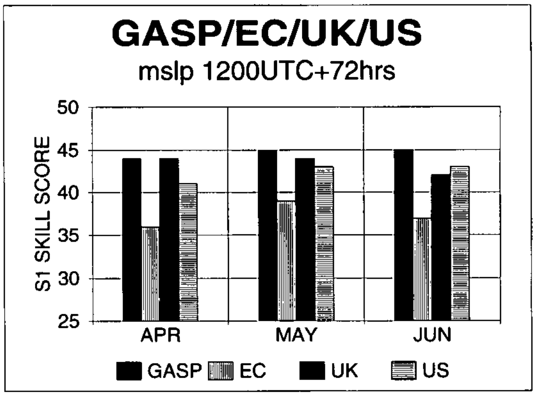


Fig 3(b) **Monthly S1 skill scores of 500 hPa geopotential height (m) for LAPS/TLAPS/GASP from April to June 1998 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.**

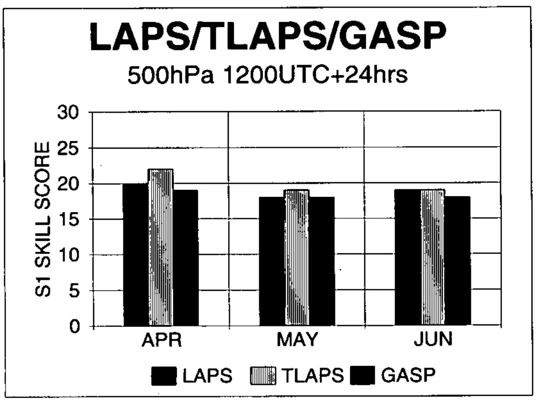


Fig 3(d) **Monthly S1 skill scores of 500 hPa geopotential height (m) for GASP/EC/UK/US from April to June 1998 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.**

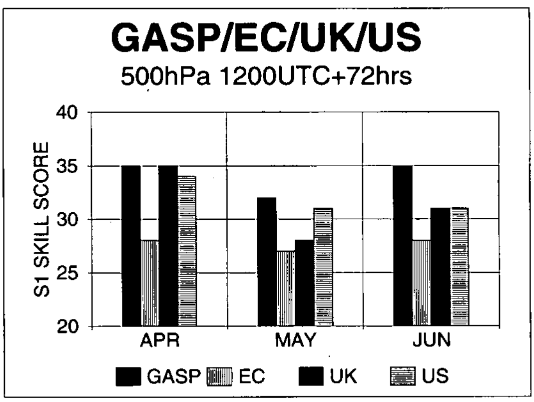


Fig 3(e) **Monthly S1 skill scores of MSLP for RASP/LAPS/TAPS/TLAPS from July 1995 to June 1998 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.**

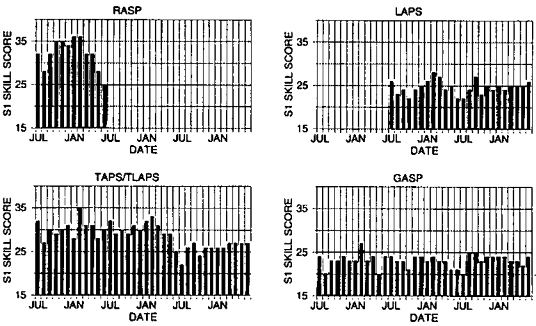
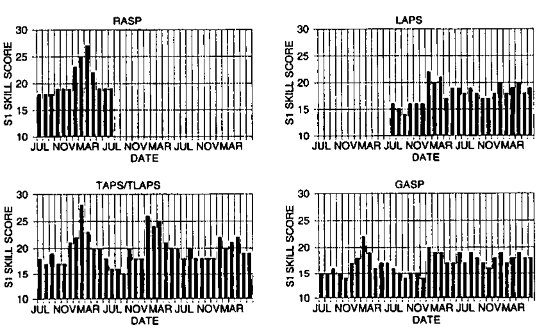


Fig 3(f) **Monthly S1 skill scores of 500 hPa geopotential height for RASP/LAPS/TAPS/TLAPS from July 1995 to June 1998 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.**



- an east coast low developing from an easterly dip during 23-26 April
- spurious lows by ECSP on 9-10 May and UKGC on 18 April
- a cut-off low near the southwest of Western Australia on 25 May for which all the available predictions were too weak and, except for JMAGSM, too linked to the westerly flow
- poor predictions by GASP and ECSP during 11-13 June for a low which developed over the Coral Sea and moved southeastwards.

As noted before, the GASP predictions were the most likely to miss developments and the least prone to generate spurious or intense systems. The probability of detection is higher with the ECSP and the UKGC systems but with an increase in the false alarm rate.

The very mixed performance in the five-day predictions, including those from ECSP, on average the best performer, suggests that five days is currently beyond the limit of reliable predictability in the Australian region for such complex synoptic situations. Table 1 compares the improvements as the outlook period is shortened in GASP and ECSP predictions which are poor at the five-day range. A threshold skill score of 60 was selected to define a 'poor' prediction and 1200 UTC predictions were used for each. There are several points to note.

Fig 3(g) **Monthly S1 skill scores of MSLP for GASP/EC/UK/US from July 1995 to June 1998 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.**

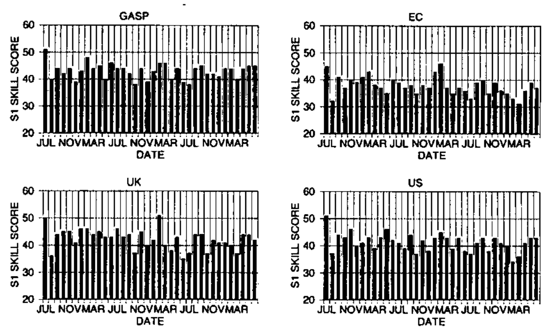


Fig 3(h) **Monthly S1 skill scores of 500 hPa geopotential height for GASP/EC/UK/US from July 1995 to June 1998 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.**

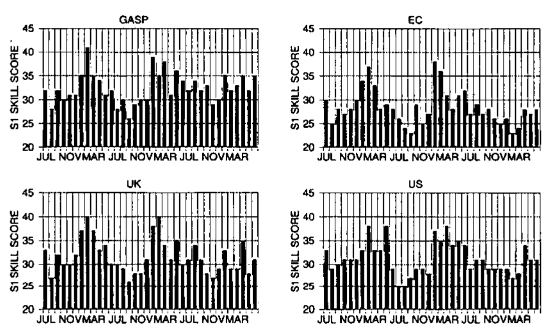
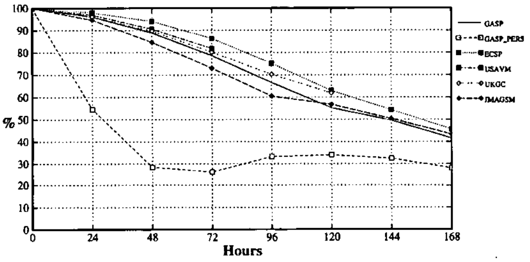


Fig 4 Anomaly Correlation of MSLP for GASP/EC/US/UK/JMA from April to June 1998 for combined base-times 0000 UTC / 1200 UTC over the irregular Australian verification grid.



(a) Most of the 'poor' predictions were related to developing low pressure systems; for GASP the errors were mainly due to under-development while for ECSP the errors were more in the details or involved spurious developments.
(b) There were roughly the same number of such 'poor' five-day predictions for GASP and ECSP, and the dates of most such cases were common to both GASP and ECSP
(c) Only two ECSP cases still fitted the "poor" criterion in the corresponding 72-hour predictions; 19 of the cases had a skill score exceeding 40 which is about the average score for 72hour predictions; for GASP, 7 out of the original 36 cases still had a skill score exceeding 60 in the 72-hour predictions and 28 had a skill score exceeding 40 at 72 hours.

Table 1. Comparison of the improvement in GASP and ECSP predictions for cases in which the five-day predictions for the same validity date had a skill score exceeding 60. The figures represent the number of predictions, out of the initial 'poor' five-day-predictions, with skill scores exceeding the specified values at shorter forecast intervals.

Skill score threshold	*GASP Forecast range*			*ECSP Forecast range*		
	48 h	72 h	120 h	48 h	72 h	120 h
60	0	7	36	0	2	32
50	4	19		0	7	
40	11	28		3	19	

(d) Most cases showed a greater drop in skill score than the slope of the average, but three ECSP cases showed little improvement. Two of these involved a low off the west coast, a situation which seems to present difficulties for current prediction systems.

UKGC

It has been noted previously that UKGC predictions have a tendency to underestimate lows. In the current period, the average five-day skill is much closer to that of ECSP and performance on individual cases appears to show a greater tendency for UKGC to over-intensify as much as underestimate development of systems. These changed characteristics could be attributed to the model upgrades implemented in late January this year.

JMAGSM

The verification scores (Fig. 4) show that the performance of the JMA global model predictions for the Australian region are comparable to GASP, although JMAGSM operates on a much shorter cutoff (2.5 to 3 hours). However, there were some excellent predictions, including cases of developing low pressure systems. A particular case was the five-day prediction for 30 June, when there were cut-off lows over the Tasman Sea and off the west coast. The JMAGSM prediction handled both systems well. UKGC, ECSP and GASP completely missed the western case. In the east, GASP showed only a very weak trough in the westerlies while ECSP and UKGC over-intensified the system. The JMA predictions appear to be a useful addition to the ensemble of available predictions.

Unusual error

A different type of error occurred in the five-day predictions for 16 to 18 June when each of the models significantly under-predicted pressure over northern Australia. Instead of strong ridging to the north, the models predicted easterlies with even a trace of cyclonic curvature.

Comparison of models against radiosondes

This comparison has been discontinued as the Australian global model (GASP) has not been verified against radiosondes since December 1997. It is hoped to resume the comparisons soon.

References

Lalaurette, F. 1998. Changes to the Operational Forecasting System. *ECMWF Newsletter Number 79* (Spring 1998).

Skinner, W. 1995. Numerical prediction model performance summary April to June 1995. *Aust. Met. Mag.*, *44*, 309-312.