Numerical prediction model performance summary January to March 1998

W. Skinner and G.A. Mills*
National Meteorological Operations Centre, Bureau of Meteorology, Australia

(Manuscript received July 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 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 Figs 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

ECSP

A number of changes to the physical parametrisation scheme were introduced on 16 December 1997 (Lalaurette 1997; Lalaurette 1997/98a).

These included a modification in the treatment of the water vapour absorption in the long-wave part of the radiation scheme, a new method of moisture convergence closure and a new treatment of the ice fall-out in the cloud scheme.

The four-dimensional variational data assimilation system (4D-Var) implemented on 25 November 1997 is described and includes an assessment of the impact on operational performance (Lalaurette 1997/98b). There has been a small improvement in forecast scores, a better fit of background fields to observations and a reduction in the short-range spin-up of precipitation.

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

* Permanent affiliation: Bureau of Meteorology Research Centre, Australia.

Fig. 1(a) Comparison for LAPS/TLAPS/GASP from January to March 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.

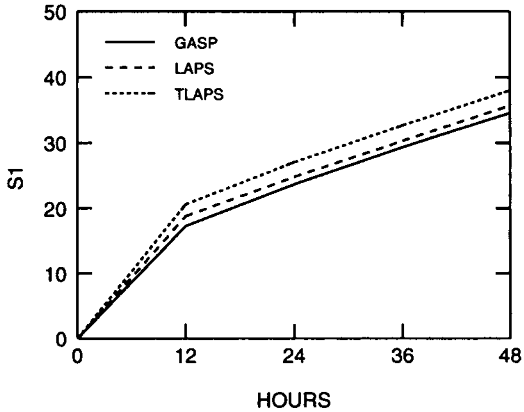


Fig. 1(b) Comparison for LAPS/TLAPS/GASP from January to March 1998. Root mean square errors of MSLP for combined base-times 0000 UTC/1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.

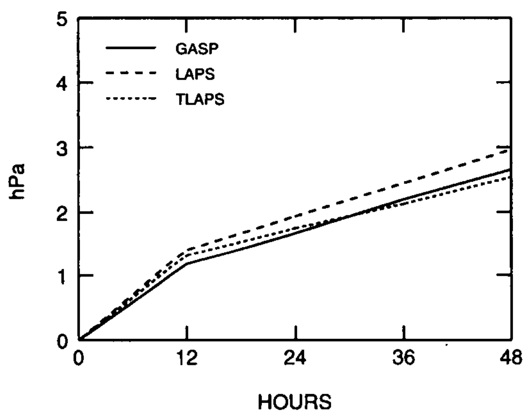


Fig. 1(c) Comparison for LAPS/TLAPS/GASP from January to March 1998. S1 skill-scores of 500 hPa geopotential height (m) for combined base-times 0000 UTC/1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.

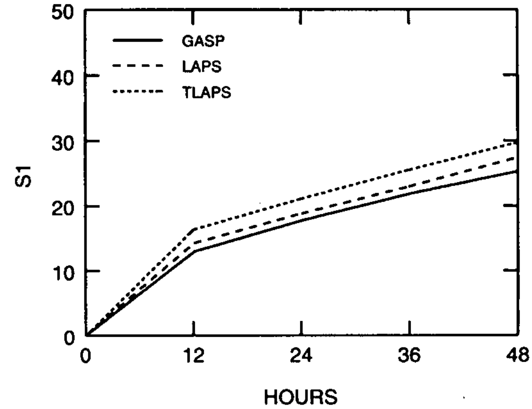
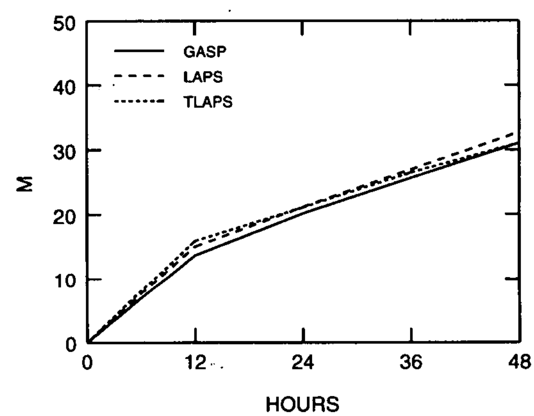


Fig. 1(d) Comparison for LAPS/TLAPS/GASP from January to March 1998. Root mean square errors of 500 hPa geopotential height (m) for combined base-times 0000 UTC/1200 UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.



January to March 1998 intercomparisons

Local models: (LAPS, TLAPS, GASP)

The best performed model averaged over the three monthly period and as measured by skill-scores continues to be GASP, followed by LAPS and TLAPS. This applies at both MSLP and 500hPa (Figs 1(a) , 1(c)) with the improvement being most evident at 500 hPa. The same pattern shows up in all the monthly averages (Figs 3(a), 3(b)). Root mean square errors, however, indicate a greater skill for TLAPS where it out-performs LAPS for all intervals at MSLP and GASP for intervals greater than +36 h (Fig 1(b)). At 500 hPa (Fig 1(d)), TLAPS does not do better than LAPS until +24 h and only catches up to GASP at +48 h. However, as has been mentioned before, the verification points absent from the TLAPS scores come from the mid-latitude region where variance is greatest.

Fig. 2(a) **Comparison for GASP/EC/US/UK from January to March 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.**

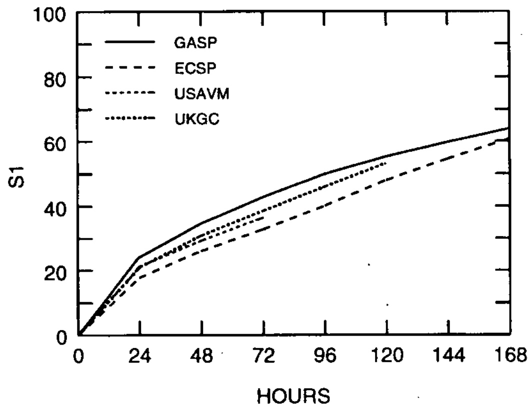


Fig. 2(c) **Comparison for GASP/EC/US/UK from January to March 1998. S1 skill-scores of 500 hPa geopotential height (m) for combined base-times 0000 UTC/1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.**

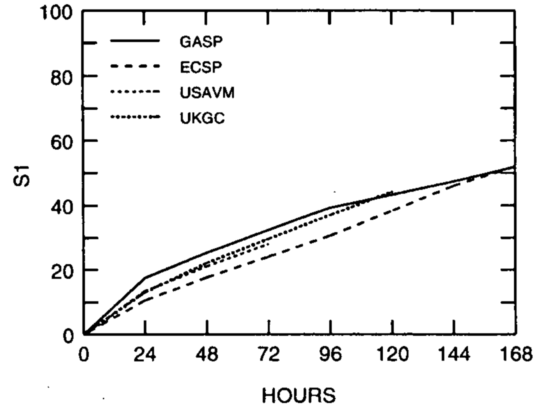


Fig. 2(b) **Comparison for GASP/EC/US/UK from January to March 1998. Root mean square errors of MSLP for combined base-times 0000 UTC/1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.**

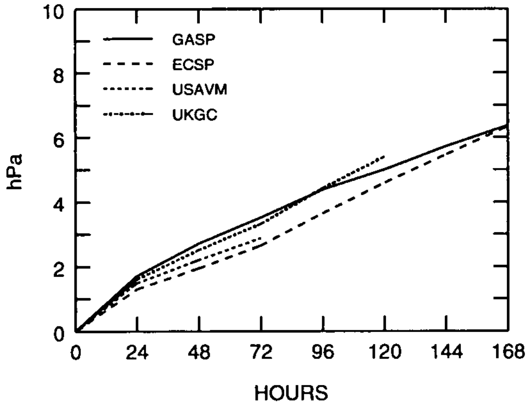
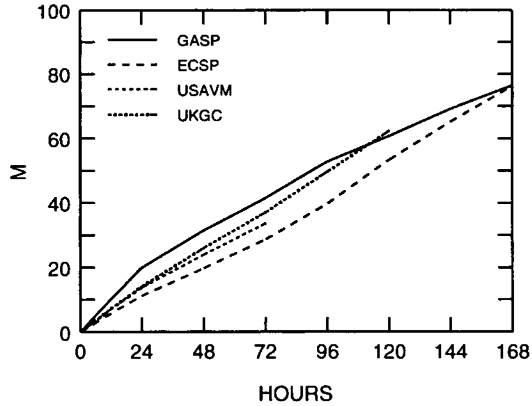


Fig. 2(d) **Comparison for GASP/EC/US/UK from January to March 1998. Root mean square errors of 500 hPa geopotential height (m) for combined base-times 0000 UTC/1200 UTC and intervals +24 h to +168 h over the irregular Australian verification grid.**



A longer time-series of skill-scores (Figs 3(e), 3(f)) from July 1995 show the improvement of two regional models over their predecessors i.e. LAPS over RASP and TLAPS over TAPS. The comparison between LAPS and TLAPS on a monthly basis overlaps the information plotted in Figs 3(a) and 3(b) but the trend over time is more apparent and shows that the skill-score advantage of LAPS over TLAPS is well established.

Synoptic overview for 24 h predictions

Tropical cyclones *Katrina, Les* and *Tiffany* during January, *Victor* and *May* in February and *Nathan* in March all caused some forecast errors with the usual problem of a small displacement of a small intense system producing large errors. GASP was least affected as the systems were often below the resolution of the model which is approximately 120 km (for resolution equivalent to grid-point models) and hence the cyclone

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

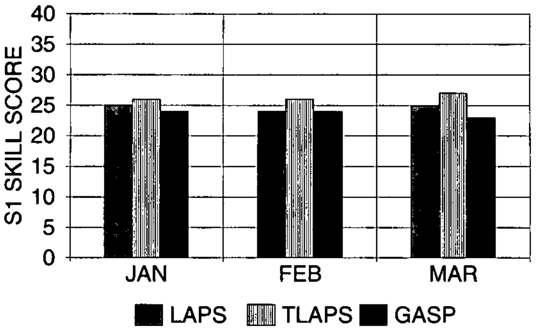


Fig. 3(b) Monthly S1 skill-scores of 500 hPa geopotential height (m) for LAPS/TLAPS/GASP from January to March 1998 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.

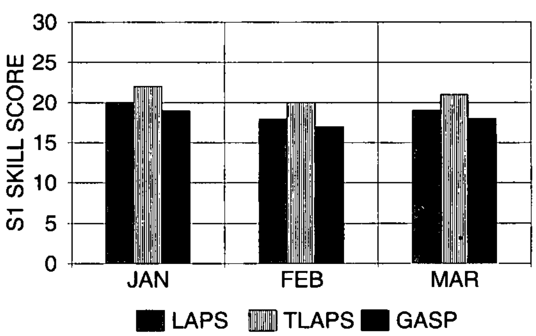


Fig. 3(c) Monthly S1 skill-scores of MSLP for GASP/EC/UK/US from January to March 1998 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.

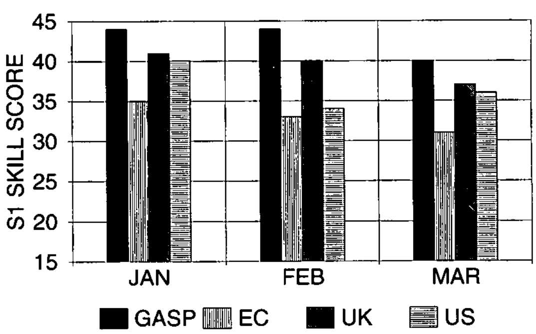


Fig. 3(d) Monthly S1 skill-scores of 500 hPa geopotential height (m) for GASP/EC/UK/US from January to March 1998 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.

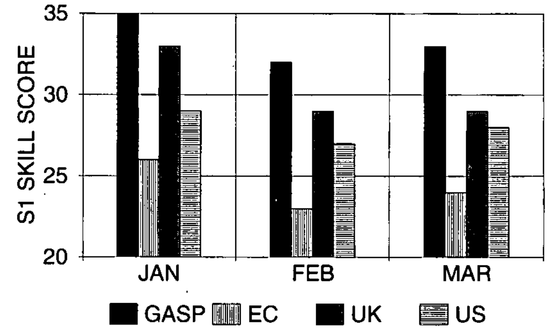


Fig. 3(e) Monthly S1 skill-scores of MSLP for RASP/LAPS/TAPS/TLAPS from July 1995 to March 1998 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.

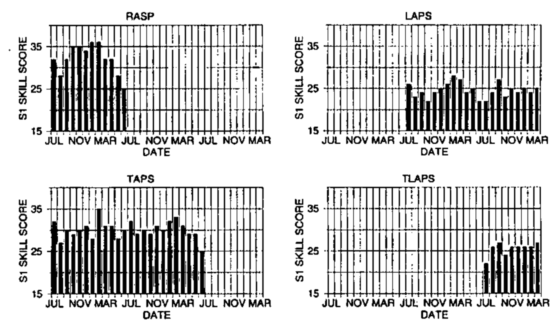
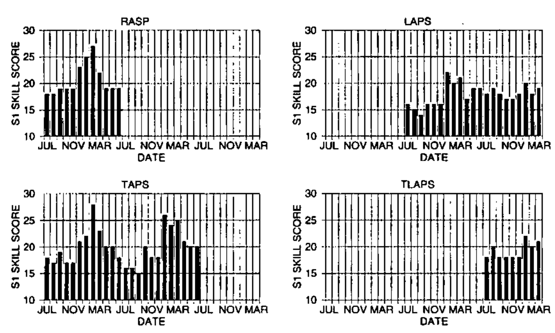


Fig 3(f) Monthly S1 skill-scores of 500 hPa geopotential height for RASP/LAPS/TAPS/TLAPS from July 1995 to March 1998 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.



was neither analysed nor forecast. LAPS and TLAPS were at a disadvantage if either the base or verifying analyses depicted the small scale feature poorly. This can cause large verifying errors. Both these models currently have a horizontal resolution of 75 km.

Other model errors for +24 h forecasts were valid for:

14 Jan	LAPS moved a deep low south of Tasmania too slowly.
15 Jan	LAPS and TLAPS were both too slow in moving the Bight high to the southeast.
16 Jan	GASP and TLAPS both failed to forecast the development of a secondary low in the western Bight; TLAPS had 15 hPa errors. The following day, LAPS over-intensified the system and moved it too slowly resulting in errors of 17 hPa.
05 Feb	LAPS failed to deepen an easterly trough over Victoria.
15 Feb	TLAPS had a major under-forecast for a deepening easterly trough over southeast Australia. It also over-estimated the depth of the WA trough.
17 Feb	GASP misjudged the speed of two Southern Ocean troughs.
05/06 Mar	All models underestimated cyclogenesis southwest of Tasmania and TLAPS missed the depth of the subsequent development of a cut-off in the Tasman Sea.
07 Mar	LAPS and GASP over-forecast the WA trough. LAPS and TLAPS were too fast with the Bight high.
09 Mar	LAPS and TLAPS over-forecast the WA trough.
10 Mar	TLAPS linked the WA trough to the westerlies too quickly.
19 Mar	TLAPS moved the trough south of WA too slowly.
20 Mar	All models had problems with the speed of systems in westerlies in the Southern Ocean.
27 Mar	All models underestimated the development of a cut-off in the Tasman Sea.

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

The skill-scores and rms errors (Fig. 2) all show ECSP clearly performed best, followed in order by USAVM, UKGC and GASP. The differences between UKGC and USAVM are not large but the models have reversed relative position since the October to December quarter when UKGC had the better skill-scores and the July to September quarter when USAVM was superior. At the longer intervals, USAVM is not represented and GASP starts to outperform UKGC and converges to the ECSP scores.

The skill-scores at +72 h for individual months and two levels (Figs 3(c), 3(d)), again show the skill order as ECSP, USAVM, UKGC and GASP. The superiority of ECSP is particularly evident.

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

The use of +120 h restricts this synoptic comparison to GASP, ECSP and UKGC as the USAVM model is only issued to +72 h.

The initial assessment of a major error was based on difference fields between a 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 being examined.

Small-scale errors resulting from the spurious prediction of tropical lows were identified 8, 1 and 13 times for the ECSP, GASP and UKGC models respectively. Interestingly, all these were negative, perhaps indicating not only a tendency for the higher resolution models to spuriously develop tropical lows, but also that the tropical analyses are failing to adequately resolve lows which are present in the initial state. There remained 27 cases when at least one of the models had at least one area of error greater than 15 hPa. This is again around 30 per cent of the forecasts, as in previous quarters.

In nine of these cases, all the models had similar errors. Five of these (forecasts valid on 12 and 15 January and 6, 12 and 17 February) were major phase errors of similar character for the major mid-latitude trough-ridge systems. Of the other four cases, all three models failed to predict a secondary cyclone in the Bight on 13 January, all three failed to resolve a small-scale thickness trough over southeastern Australia on 24 January, all three failed to develop a deep easterly trough over southern New South Wales on 12 February, and all three models forecast a spurious low just south of WA on 20 February. Thus three of these cases showed a failure to develop small-scale systems, while the fourth case was one of consistent (between the models) spurious development.

In six cases (forecasts valid 16, 21 and 25 January, 16 February and 6 and 7 March), the ECSP forecast produced greater errors in forecasting MSLP than either of the other models. However, in each of these cases it had correctly forecast the development of a low pressure system, but with a small latitude or longitude error. The other models generally only predicted open or no troughs in the area of the verifying low. In these cases this form of assessment of a 'forecast bust' is misleading, as these can be considered excellent forecasts.

There were four cases (26 January, 8, 9 and 27 March) when GASP failed to forecast significant lows in the Tasman Sea. The UKGC model also failed to forecast the first of these.

GASP overdeveloped the heat low/continental easterly trough on 2 and 14 February. This is consistent with previous experience.

The other errors were:

14 Jan UKGC model failed to forecast a secondary trough over Tasmania.

17 Jan UKGC model forecast rapid eastward ridging across the Bight, and so failed to capture the extratropical cyclone southwest of Tasmania.

04 Feb UKGC model overdeveloped a low just southwest of WA.

05 Feb ECSP and GASP models had a major phase error for a mid-latitude cyclone south of the Bight.

07 Feb ECSP and UKGC models had a ridge through Bass Strait where there should have been a small, significant low. GASP did predict an open trough here.

21 Mar GASP had a large phase error for a trough south of Adelaide.

In summary, the models in general show a remarkable ability to resolve the evolution of the main features of the atmosphere, with two-thirds of the forecasts not showing major pressure errors at five days. The fact that the ECSP model in particular can predict Tasman Sea cyclogenesis five days in advance shows a large measure of skill. At this time-frame, an error of a few degrees latitude or longitude can show in forecast-analysis difference fields as a major error, yet should really be seen as an excellent forecast.

A cautionary note should be sounded, though. In approximately ten per cent of cases, all three models produced consistent, though erroneous, forecasts. Forecasters who use consistency of model solution ('convergence of the ensemble') as a measure of confidence in the medium range forecast should be aware of this possibility.

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 this program in the near future.

References

Lalaurette, F. 1997. Changes to the Operational Forecasting System. *ECMWF Newsletter Number 77*(Autumn 1997).

Lalaurette, F. 1997/98a. Changes to the Operational Forecasting System. *ECMWF Newsletter Number 78* (Winter 1997/98), 1.

Lalaurette, F. 1997/98b. The operational implementation of 4-D-Var. *ECMWF Newsletter Number 78* (Winter 1997/98), 2.

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