Numerical prediction model performance summary January to March 2000

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

(Manuscript received July 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 Oceanographic Centre (NMOC), Melbourne, and from ECMWF (European Centre for Medium-range Weather Forecasts), NCEP (National Centers for Environmental Prediction, USA) 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).

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

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.

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

US global model

The resolution of the US global model was upgraded on 24 January. The new version is triangular wavenumber T170 in which the smallest half-wavelength resolved is about 116 km. The number of levels has also been increased from 28 to 42. No other changes were made to the analysis or forecast components.

A useful running history of recent modifications to the US global forecast/analysis system can be viewed at the web page of the Environment Modeling Centre, the centre within the National Centers for Environmental Prediction (NCEP) which develops and implements the NWP systems. The URL is http://www.emc.ncep.noaa.gov/index.html.

UKMO

The UKMO has resumed publishing its very helpful NWP Gazette. An on-line version is available at http://www.met-office.gov.uk/sec5/NWP_Gazette/index.html.

JMA

The physics package of the Global Spectral Model (GSM) was updated on 7 December 1999. The new package includes a prognostic cloud water scheme, effects of orographic updraft and turbulence in the boundary layer on cumulus convection and direct effects of aerosols on radiation.

WMO Annual World Weather Watch (WWW) Technical Progress Report

Reports on progress during the past year from many NWP centres are published annually by WMO in the Annual WWW Technical Progress Reports. The reports cover computing infrastructure, observations used, details on the suite of numerical analysis and prediction systems operated, any statistical applications or forecasting techniques used with the NWP output, performance information and future plans. The most recent report covering activities in 1999 can be viewed on the WMO web site under World Weather Watch, then Global Data Processing System. The URL is: http://www.wmo.ch/web/www/DPS/APRGDPS99/Cover-9-2000.html.

January to March 2000 intercomparisons

Local models: (LAPS_PT375, TLAPS_PT375, GASP)

The skill score at MSLP (Fig. 1(a)) shows GASP still outperforms the other models and TLAPS is next best. At 500 hPa (Fig. 1(c)) however, LAPS_PT375 scores are closer to GASP scores, and the performance of TLAPS_PT375 falls off. Individual month's scores (Fig. 3(a), (b)) confirm this with TLAPS having high scores in both February and March at 500 hPa. The rms errors at MSLP (Fig. 1(b), (d)) show TLAPS_PT375 outscoring LAPS_PT375 at mslp and, after 24 hours, at 500 hPa.

Synoptic overview for 24 h MSLP predictions (local models)

On 4 January, a small cut-off low formed just east of Bass Strait. At +24 h the feature was predicted but LAPS and TLAPS had the system too far east and too shallow by 12 to 14 hPa. The GASP forecast was good at this interval. The subsequent east-northeast movement of the low was well handled by all models.

On 12 and 13 January, LAPS underestimated the depth of a trough in the easterlies over western and central Australia. By 14 January all models had errors as pressures fell over Western Australia. The system was also underestimated as it moved slowly south-southeast and became complex on 16 January.

Another trough, amplifying as it approached New Zealand was under-forecast on 16 and 17 January. Some of the poorest prediction scores for the month were associated with this situation.

By 19 January, a complex cut-off low system was established in the eastern Indian Ocean which LAPS moved too rapidly east, resulting in poor scores. TLAPS southern boundary was too far north to capture the system properly but GASP handled it quite well.

A deepening Tasman trough was missed on 23 January, with LAPS producing the worst prognosis.

TC *Kirrily*, off the northwest coast of Western Australia on 28-30 January produced some errors in all model forecasts.

During this period, another system, a Tasman Sea cut-off low, moved southeast. GASP and TLAPS initially overestimated the depth of this system, but although LAPS was best initially, it subsequently moved the low too rapidly southeast.

On 4 February, a trough approaching Tasmania was underestimated by all models. On 11 February in the same area, LAPS and TLAPS underestimated another trough while GASP was more accurate. LAPS also moved this system too quickly south on 12 February.

By 14 February, a blocking system had become established in the southern Tasman Sea. LAPS and TLAPS both underestimated the depth of the associated cut-off low although all models captured the general pattern.

A trough in the easterlies over Western Australia deepened on 18 February and formed a low in the Bight which moved southeast on 19 February and another easterly trough was established over Eastern Australia. This in turn was captured by the westerly stream by 22 February.

All models underestimated the formation of the initial low over the Bight and then moved it too fast. The prediction of the formation of the second cut-off on 19 February was marred by the miscalculation of the cradling ridge to the south. Models all moved the

Fig. 1(a) **Comparison for LAPS_PT375/TLAPS_PT375/GASP from January to March 2000. S1 skill-scores of MSLP using combined base-times 0000UTC/1200UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.**

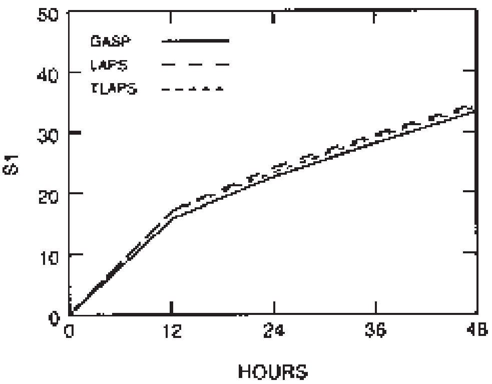


Fig. 1(b) **Comparison for LAPS_PT375/TLAPS_PT375/GASP from January to March 2000. Root mean square errors of MSLP for combined base-times 0000UTC/1200UTC and intervals +12, +24, 36, +48 h over the irregular Australian verification grid.**

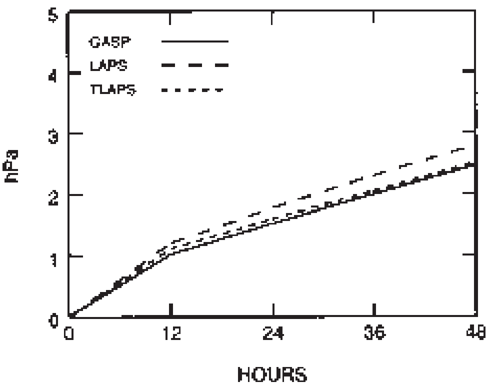


Fig. 1(c) **Comparison for LAPS_PT375/TLAPS_PT375/GASP from January to March 2000. S1 skill-scores of 500 hPa geopotential height (m) for combined base-times 0000UTC/1200UTC and intervals 12, +24, +36, +48 h over the irregular Australian verification grid.**

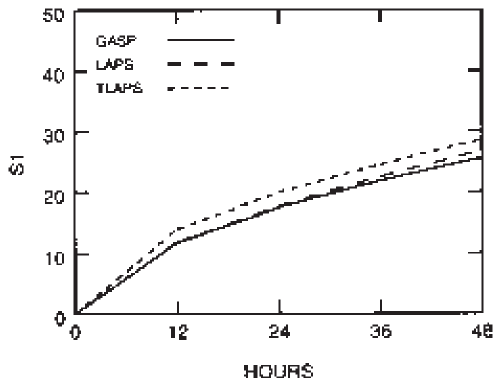
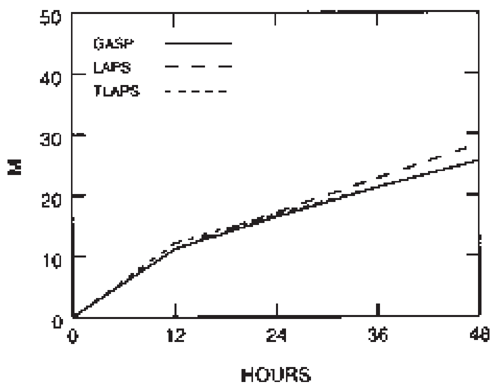


Fig. 1(d) **Comparison for LAPS_PT375/TLAPS_PT375/GASP from January to March 2000. Root mean square errors of 500 hPa geopotential height (m) for combined base-times 0000UTC /1200UTC and intervals +12, +24, +36, +48 h over the irregular Australian verification grid.**



ridge too slowly but TLAPS was the worst with a score 17 points above its monthly average. The subsequent movement was well forecast.

Yet another deepening easterly trough, this time southwest of Western Australia on 29 February, resulted in high scores for all models. GASP was best, while TLAPS missed it altogether. All models captured the deepening and southeastward movement on 1 March.

Tropical cyclone *Norman* formed in the northeastern Indian Ocean and moved west during 1 to 4 March before passing out of the verification area. All models provided good general guidance although localised errors in position occurred.

Tropical cyclone *Steve* formed over the Coral Sea, then passed over the Queensland coast on 27 February and weakened. It then moved northwesterly as a tropical low, reintensified and crossed into the Northern Territory on 1 March. It subsequently moved west and southwest as a tropical low until being re-classisified as a TC after crossing the coast near Broome on 5 March. It tracked southwest and south, and re-crossed the coast near Exmouth on 9 March, then moved southeast as an easterly low to be eventually captured by the westerlies on 13 March.

Table 1. Showing (best or equal best)/(worst or equal worst)/(scores ≥ 5 points worse than the best) no of occurrences for each model during the period January to March. For example, GASP in March had the best score 18 times and the worst score only 8 times. There were no occasions when its score was 5 points or more worse than the best model that day.

	GASP	*TLAPS_PT375*	*LAPS_PT375*
January	13/9/5	18/8/6	5/19/11
February	17/4/1	10/10/6	4/16/7
March	18/8/0	9/10/6	6/19/5
Quarter	48/21/6	37/28/18	15/54/23

The models handled this complex movement well although some errors occurred due to slight position errors of a small intense system. Later in the period, the models' predictions of the intensity of the extra-tropical low were variable.

On 12 March, a low over the south island of New Zealand was badly under-forecast by all models with LAPS havinging errors of 16 hPa .

Tropical cyclone *Olga* off the northwest coast, was correctly forecast to move southwestward during 17 to 19 March. Its central pressure tended to be under-estimated.

On 17 March, LAPS badly overestimated the depth of a low southwest of Tasmania whereas GASP produced a good forecast.

On 23 March, a trough from the eastern Indian Ocean moved rapidly southeast and was replaced by a ridge moving across Tasmania and a low in the Southern Ocean near 160°E. All models moved this ridge system too slowly and weakened the trough so that there were major errors associated with both features.

Many errors during the three-month period involved underestimation of the strength of troughs or lows. This is consistent with the behavior of GASP documented in Table 2 for +120 hours. In particular, categories 4B, 5B, 8B and 13B all have significant representation in the +24 hour predictions.

Table 1 shows the (best or equal best)/(worst or equal worst)/(scores ≥ 5 points lower than the best) no of occurrences for each model during the period January to March 2000. GASP was the most reliable model overall with the highest number of days of 'best' scores, the lowest number of days with 'worst' scores and also the lowest number of 'bad' forecasts. The only exception was in January when TLAPS_PT375 performed better. It should be noted that these comparisons of the operational products are affected by two constraints. GASP is run later than the other regional models and therefore has the best data coverage, and TLAPS_PT375's southern boundary does not include the most southerly verification points.

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

Global model performance, averaged over January to March at MSLP (Fig. 2(a)), again shows the clear superiority of ECSP followed by UKGC, USAVM and JMAGSM. It is only after 144 hours that GASP edges ahead of the Japanese model. The monthly averages at 72 hours (Fig. 3(c)) repeat the pattern. Low scores were evident in February, particularly for ECSP, but all models showed an improvement over February 1999 by at least five skill points. This is in step with a lower score for +72 hour persistence as well, implying that February 2000 was a relatively easy month to predict at MSLP.

At 500 hPa (Fig. 2(c)), GASP outperformed JMAGSM after +72 hours. During January and March (Fig. 3(d)), GASP scored better than JMAGSM at +72 hours, but was poor in February when other models scored well.

Root mean square errors (Fig. 2(b), 2(d)) also show GASP with an edge over JMAGSM, particularly at 500 hPa (Fig. 2(d)) when averaged over the whole three-month period.

The anomaly correlation (Fig. 4) of these five models shows similar relative performances. ECSP has the best correlation, then UKGC and USAVM. GASP and JMAGSM are very close out to +96 hours while GASP is slightly ahead in the intervals +96 to +168 hours. If 60 per cent is the cut-off correlation for identifying a useful prognosis, then ECSP is useful to about 150 hours while GASP and JMAGSM fall below 60 per cent at about 130 hours.

The historical series (Fig. 3(h)) also shows February 2000 500 hPa skill-scores as trending up (i.e. less skill) for GASP and down or level for the other models.

Synoptic intercomparison of global models

What are the systematic errors in the prediction of synoptic features by the global NWP models? Errors have been classified 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. We have concentrated on five-day MSLP predictions from 1200 UTC global model output 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).

Table 2. Number of occasions during January-March 2000 when errors 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	0	1	0	0
2 Unable to match features	6	3	7	4
3 Low in easterlies or cut-off low				
A too linked to westerlies	8	6	4	7
B too much in easterlies or cut-off	3	1	1	3
4 Trough in easterlies or cut-off low:				
A too deep	18	19	23	24
B too weak	31	21	15	21
C details of structure	2	2	0	0
5 Trough in easterlies over WA:				
A too deep	15	2	7	7
B too weak	10	11	9	13
6 Trough in easterlies or cut-off:				
A too mobile (i.e. too far east)	3	5	1	7
B too slow (i.e. too far west)	4	7	8	10
7 Ridging or high pressure system:				
A too strong	20	30	30	26
B too weak	12	8	10	8
C too mobile	1	0	1	1
D too slow	4	0	5	4
E latitudinal displacement	2	1	0	1
8 A Spurious low or marked over-development	10	6	17	15
B Missed development of a low or trough	19	15	24	19
9 Retaining a low or trough too long	2	-	-	-
10 Trough in westerlies basically correct in structure but:				
A too strong	15	11	12	15
B too weak	6	9	5	7
C details of structure	0	2	0	0
11 Low predicted well in intensity but location poor	3	8	5	4
12 Trough in westerlies:				
A too mobile	6	3	5	7
B too slow	14	11	11	9
13 Tropical lows:				
A over-developed	4	19	12	18
B missed or too weak	19	8	4	7
C mis-located	6	7	6	8

Daily forecast error charts (i.e. five-day forecast MSLP minus the verifying analysis field from the same NWP system) were produced for each of the models. Each case of forecast error exceeding 4 hPa in magnitude was assigned subjectively to a synoptic pattern category as listed in Table 2, based on the dominant characteristic of the error. A glance at the categories indicates that this process is obviously not unambiguous (e.g. some judgement is involved in differentiating between a category of ridging being too strong (7A) or a trough being too weak (10B)). In some cases during this summer period the general flow was quite weak and the patterns difficult to categorise. Use of the system's own analysis is another

Fig. 2(a) **Comparison for GASP/ECSP/USAVM/UKGC/JMAGSM from January to March 2000. S1 skill-scores of MSLP for combined base-times 0000UTC/1200UTC and intervals +24 h to +192 h over the irregular Australian verification**

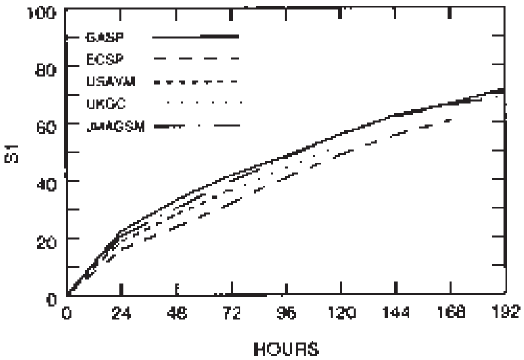


Fig. 2(b) **Comparison for GASP/ECSP/USAVM/UKGC/JMAGSM from January to March 2000. Root mean square errors of MSLP for combined base-times 0000UTC/1200UTC and intervals +24 h to +192 h over the irregular Australian verification grid.**

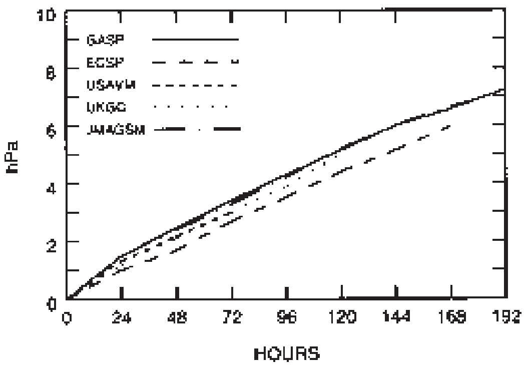


Fig. 2(c) **Comparison for GASP/ECSP/USAVM/UKGC/JMAGSM from January to March 2000. S1 skill-scores of 500 hPa geopotential height (m) for combined base-times 0000UTC/1200UTC and intervals +24 h to +192 h over the irregular Australian verification grid.**

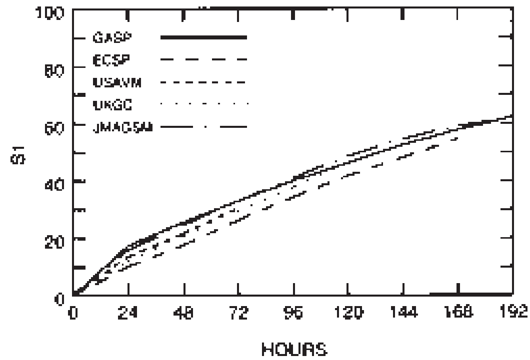
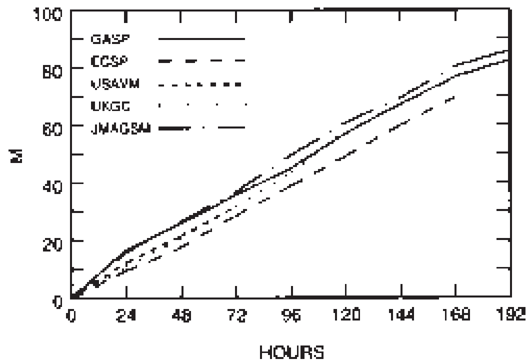


Fig. 2(d) **Comparison for GASP/ECSP/USAVM/UKGC/JMAGSM from January to March 2000. Root mean square errors of 500 hPa geopotential height (m) for combined base-times 0000UTC/1200UTC and intervals +24 h to +192 h over the irregular Australian verification**



weakness due to the strong influence of the model on the analysis field and consequent differences among the analyses. However, at the five-day forecast range this problem is not major, except in cases of tropical cyclones which are poorly represented in some systems. 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 January-March period was characterised by protracted strong monsoon flow over northern and central Australia, bringing a persistent easterly regime for much of the period. As in the previous quarter, outside the tropics, the five-day predictions generally showed a reasonable correspondence between the predicted and observed synoptic features. However, there was only one case (the ECMWF prediction for 2 January) where the forecast error was less than 4 hPa over the full area of interest. This is considerably fewer than for the previous quarter. On the other hand, there were cases for each model where it was difficult to match the observed and predicted features such as the situation of 17 January, common to all the models. The synoptic situation in this case was a cut-off low over the Bight. Two other cases of poor forecasts, common to all the predictions except ECMWF, were 11 January, which involved the interaction of a low in

Fig. 3(a) Monthly S1 skill-scores of MSLP for LAPS_PT375/TLAPS_PT375/GASP from January to March 2000 for base-time 1200UTC and interval +24 h over the irregular Australian verification grid.

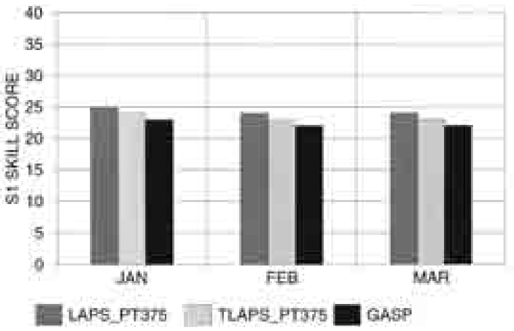


Fig. 3(b) Monthly S1 skill-scores of 500 hPa geopotential height (m) for LAPS_PT375/ TLAPS_PT375/ GASP from January to March 2000 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.

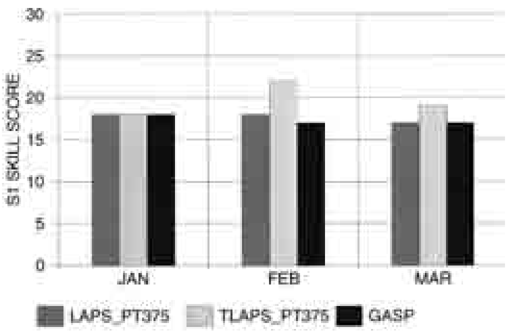


Fig. 3(c) Monthly S1 skill-scores of MSLP for GASP/EC/UK/US/JMA from January to March 2000 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.

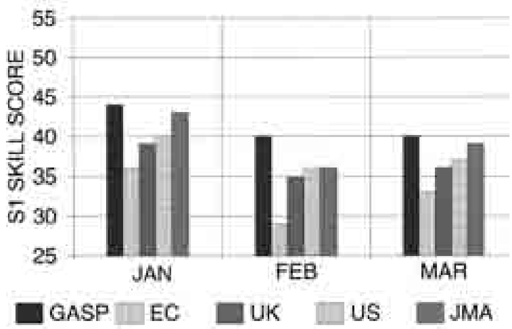


Fig. 3(d) Monthly S1 skill-scores of 500 hPa geopotential height (m) for GASP/EC/UK/US/JMA from January to March 2000 for base-time 1200UTC and interval +72 h over the irregular Australian verification grid.

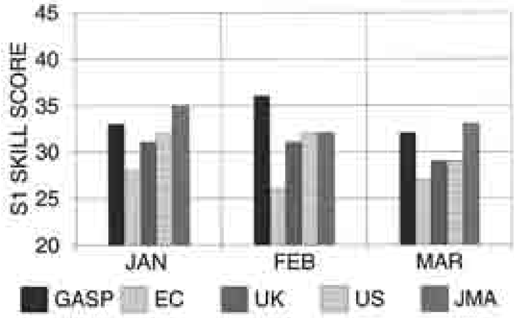


Fig. 3(e) Monthly S1 skill-scores of MSLP for LAPS/LAPS_PT375/TLAPS/TLAPS_PT375 from January 1997 to March 2000 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.

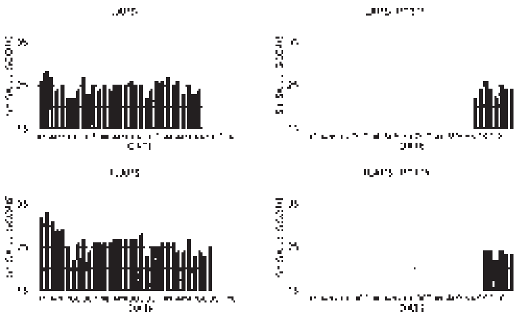


Fig. 3(f) Monthly S1 skill-scores of 500 hPa geopotential height for LAPS/LAPS_PT375/TLAPS/ TLAPS_PT375 from January 1997 to March 2000 for base-time 1200 UTC and interval +24 h over the irregular Australian verification

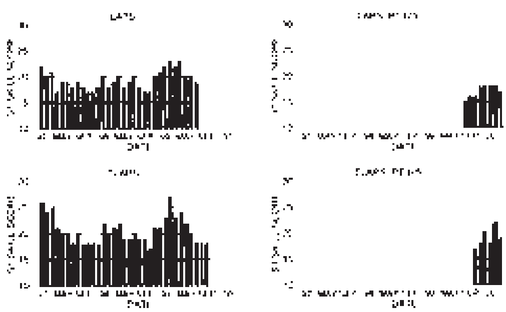


Fig. 3(g) Monthly S1 skill-scores of MSLP for GASP/EC/UK/US from July 1997 to March 2000 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.

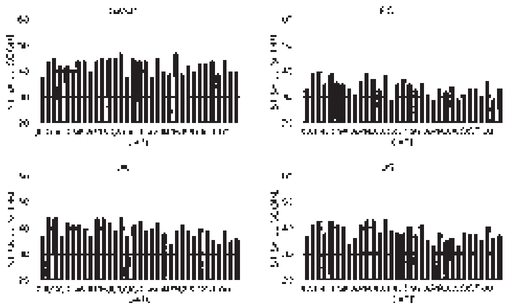


Fig. 3(h) Monthly S1 skill-scores of 500 hPa geopotential height for GASP/EC/UK/US from July 1997 to March 2000 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.

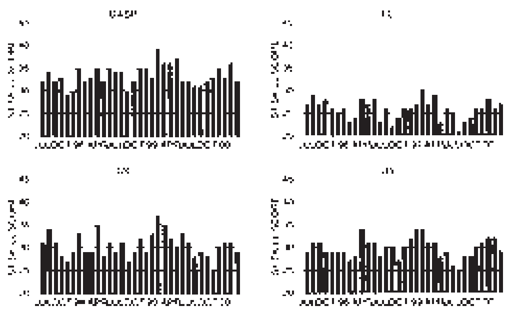


Fig. 3(i) Monthly S1 skill-scores of MSLP for GASP/EC/UK/JMA from July 1997 to March 2000 for base-time 1200 UTC and interval +120 h over the irregular Australian verification grid.

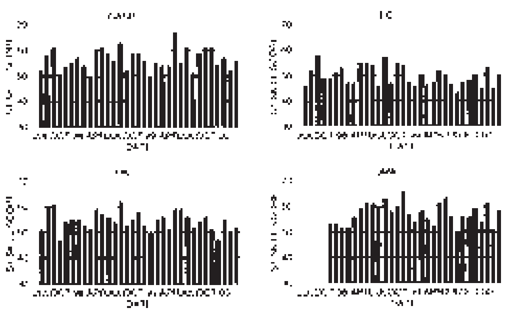
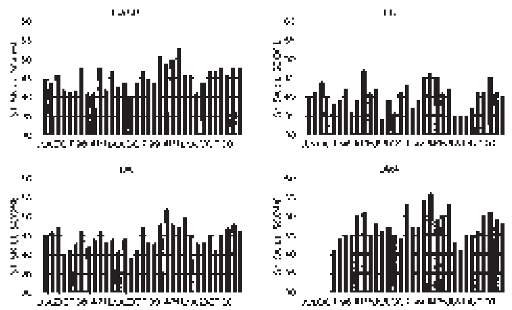


Fig. 3(j) Monthly S1 skill-scores of 500 hPa geopotential height for GASP/EC/UK/JMA from July 1997 to March 2000 for base-time 1200 UTC and interval +120 h over the irregular Australian verification grid.



the easterlies over Western Australia with a higher latitude trough, and 31 March, a case of rapid ridging over the Bight.

From a forecasting point of view the common errors (noted below and in Table 2) in failing to predict the occurrence or at least in significantly under-predicting developments in troughs or lows is worrying. Perhaps in a larger ensemble there would be a member which did capture the development more realistically.

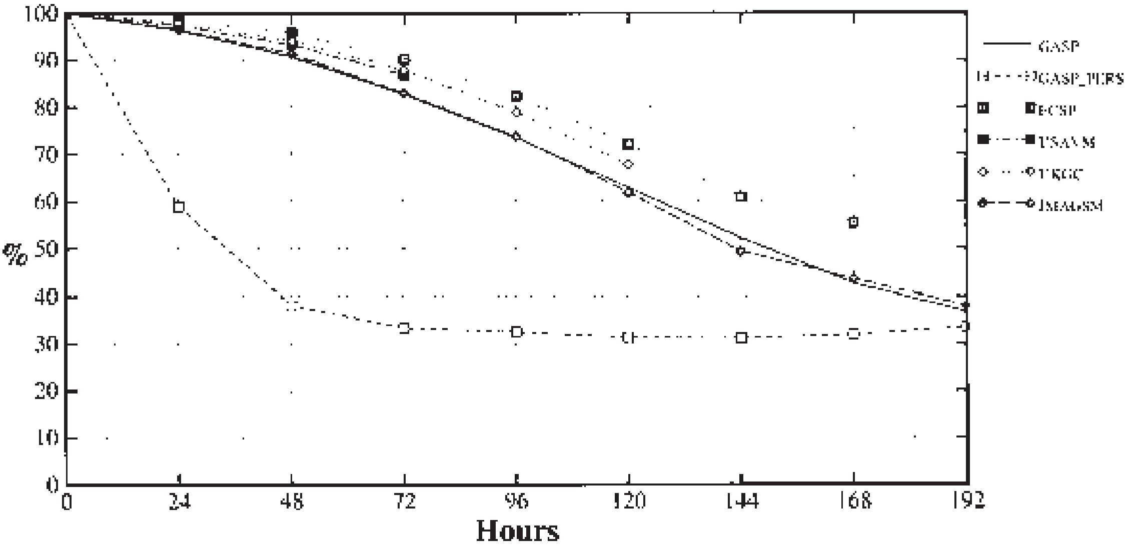
The following are some comments on particular synoptic features:

(a) Troughs and lows in easterly flow: Cut-offs and easterly dips were common during the January-March period associated with the protracted monsoon trough. There were many errors in predicted intensity, but these were fairly evenly spread between being too strong or too weak (in contrast to the previous quarter). GASP showed a bias toward under-prediction. All again showed the characteristic of linking such features too much or too early to the westerlies.

  For the particular case of troughs over Western Australia, GASP continued its bias of over-prediction. The frequency of over-prediction appears less than in the previous quarter but the 4 hPa threshold was applied more consistently this time. Conversely, UKMO again and also ECMWF showed a bias to under-prediction of the strength of easterly troughs over Western Australia.

(b) Extratropical lows or troughs in the westerlies: For low pressure systems or major troughs there were of the order of 10 to 20 cases of both spurious and under-forecast developments, with the latter predominating in contrast to October-December

Fig. 4 Anomaly Correlation of MSLP for GASP/ECSP/USAVM/UKGC/JMAGSM from January to March 2000 for combined base-times 0000 UTC/1200 UTC over the irregular Australian verification grid.



1999. ECMWF in particular showed a change from a slight preponderance of over-development to a clear majority of under-development.

In two cases (9 and 28 January) the over-development was common to all four models. The case of 9 January was another interaction of an easterly trough over Western Australia with a higher latitude trough. The predictions for 28 January all showed some form of mobile mid-latitude trough well south of the Bight when the analyses showed a strong high over the region. In two other cases (10 January and 23 March) the problem occurred in all the predictions except ECMWF.

There were more cases of missed or under-development of a low or trough than in the October-December 1999 period. This deficiency was identified in at least one of the models on 31 different dates during January-March. As in the previous three month period almost half of these occasions were common to at least three of the models.

The cases common to all four models were:

- 4 January - all the models failed to predict the development of a small intense low just east of Bass Strait. The forecast error patterns for each model were very similar and reached about 20 hPa in magnitude.
- 19-21 January - a deep low well south of the Bight on 19 January which moved steadily eastwards.
- 14-15 February - a low not predicted over the Tasman Sea with errors again up to 20 hPa.
- 18 February - a low in the easterlies over the Bight was under-predicted.
- 12 March - underestimates of a low over the Bight which had developed from TC *Steve*. (Given the tropical origin the forecasts were in fact remarkably good for the five-day forecast range).

The cases of error in three models were:

- 6 January - a continuation of the problems for 4 January.
- 12 February - a poorly forecast trough in the westerlies south of 40°S.
- 20 February - the same low as 18 February but now further east.
- 29 February - a cut-off low off southwest Western Australia.

For mobile westerly troughs where the structure and mobility were acceptable, the intensity was over-estimated (contrast to previous quarter) by a factor of 2 to 3. The predictions showed a strong tendency to be too slow.

(c) Ridging and high pressure systems. All models again over-predicted the strength of high pressure systems, and the extent of ridging, particularly in March.

(d) Tropical systems. As in the previous quarter GASP continued to show a clear propensity to be less active in the tropics, in the sense of generating fewer discrete lows, than the other models. The other models showed the opposite characteristic with a clear bias toward over-prediction.

For systems which could be classed as tropical cyclones there were some excellent forecasts of genesis and movement 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.

References

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