

# Numerical prediction model performance summary October to December 2000

W. Skinner and T. Hart

National Meteorological and Oceanographic Centre, Bureau of Meteorology,  
Australia

(Manuscript received April 2001)

## 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 and Oceanographic Centre (NMOC) Melbourne and from ECMWF (European Centre for Medium-range Weather Forecasts), NCEP (National Centers for Environmental Prediction), UKMO (United Kingdom Meteorological Office) and JMA (Japan Meteorological Agency).

Four models considered from NMOC, Melbourne, are: LAPS\_PT375 (Limited Area Prediction System Point 375); MESO\_LAPS\_PT125 (MESOscale Limited Area Prediction System Point 125) TLAPS\_PT375 (Tropical Limited Area Prediction System Point 375); and GASP (Global Assimulation and Prediction).

Overseas global models included in the comparisons are: ECSP (ECMWF Spectral Assimulation); USAVM (NCEP Spectral Model for Aviation); UKGC (UK 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 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 (see Fig. 5).

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 and easterly points outside the MESO\_LAPS\_PT125 grid. TLAPS\_PT375 and MESO\_LAPS\_PT125 scores are calculated without these points and are therefore not strictly comparable with those from other models.

## Notes on NWP systems

### ECMWF

The resolution of the ECMWF global atmospheric prediction model was increased from  $T_L319$  to  $T_L511$  on 21 November 2000, described as roughly equivalent to a reduction from 60 km to 40 km grid size. The resolution of the data assimilation minimisation used in their 4DVAR analysis was increased from  $T63$  to  $T_L159$  on the same date. (The subscript is used to denote that the grid used for the calculations of non-linear dynamic terms and physics processes is at a lower resolution with the semi-Lagrangian time stepping algorithm than formerly used with the Eulerian scheme.)

The time schedule and types of products disseminated via GTS remain unchanged.

---

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

### LAPS\_PT375

A number of changes have been introduced to the LAPS\_PT375 system which may have led to some noticeable differences in performance.

These include:

- On 16 October 2000, the introduction of a surface moisture analysis scheme whereby screen level dew-point temperatures are directly assimilated.
- The incorporation of new data types including METARs (Aviation Routine Weather Report) on 30 October and additional locally derived cloud-drift winds (hourly IR, water vapour, and high resolution visible winds) on 15 November.
- The implementation of code changes, aimed at improving the running performance during October and November, which affect the data selection in the assimilation.

During a parallel trial, the addition of the extra cloud-drift winds showed a small positive impact at the surface and above. The surface moisture change, and the use of METAR data, should result in improvements in forecasts of rainfall and instability indices.

### USAVM

A package of minor changes was introduced on 1 October.

Observations and analysis:

- new observation error diagnosis;
- rawinsonde radiation correction;
- effects of balloon drift in time and space included.

Forecast model and post processing:

- improved orography;
- reduced grid over polar regions;
- new surface albedo climatology;
- revised single scattering albedo for clouds.

## October to December 2000 intercomparisons

### Local models: (GASP, TLAPS\_PT375, LAPS\_PT375, MESO\_LAPS\_PT125)

Skill scores and root mean square errors for MSLP (Fig. 1(a,b)) and 500 hPa (1(c,d)), show GASP was the best performed model at both levels. LAPS was better than TLAPS only for 500 hPa skill scores and for MSLP root mean square errors for short forecast intervals. However TLAPS does not cover the southern verification points at 50°S and 55°S (Fig. 5).

Individual monthly skill scores for each of the months October, November and December are shown in Fig. 3(a) for MSLP and Fig. 3(b) for 500 hPa. Again GASP has the lowest, or best scores. The relationship between LAPS and TLAPS is the same as in Fig. 1(a) and Fig. 1(c) with TLAPS scoring lower at

MSLP and LAPS at 500 hPa. These figures also include MESO\_LAPS which can be considered a high resolution version of LAPS. Its scores are very much the same as those of LAPS for this period.

The relative superiority of GASP is further demonstrated by the statistics in Table 1 which indicate the relative performance of the models for 24-hour forecasts. For example, GASP in October obtained the best score 25 times and the worst score only four times. There was only one occasion when its score was five points or more worse than the best model on that day. 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. 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. Also the TLAPS\_PT375 southern boundary and MESO\_LAPS\_PT125 eastern boundary fall inside the grid and so do not cover all the verification points. The MESO\_LAPS system provides greater detail than LAPS on small-scale variations and topographic effects and although the skill scores from the two models are very similar, this analysis of best/worst/bad scores shows a significant increase in the 'best' number of occasions attributed to MESO\_LAPS over LAPS. For this period, MESO\_LAPS has provided more reliable forecasts than LAPS. This contrasts to the result from the July-September quarter when the models were very similar.

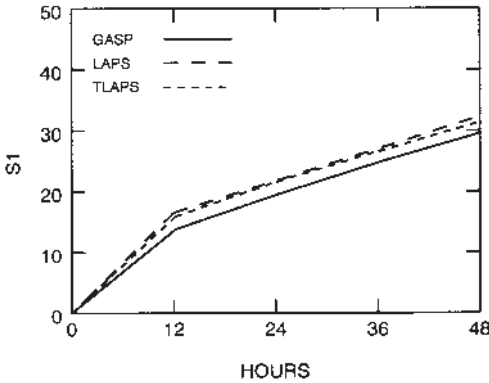
**Table 1. The number of occurrences during October to December when each model's skill scores were (best or equal best)/(worst or equal worst)/(scored five points or more worse than the best) for 24-hour predictions based at 1100 UTC.**

	<i>GASP</i>	<i>TLAPS</i>	<i>LAPS</i>	<i>MESO_LAPS</i>
October (31)	25/ 4/ 1	4/17/ 9	5/ 6/11	5/11/11
November(30)	16/ 3/ 1	9/13/ 5	3/16/14	8/14/11
December(31)	19/ 4/ 0	9/ 9/ 9	2/16/ 8	8/14/ 9
QUARTER	60/11/ 2	22/39/23	10/38/33	21/39/31

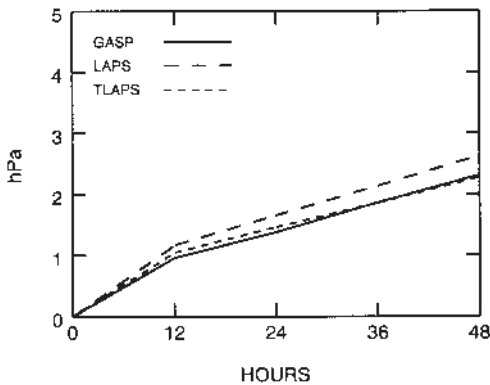
### Synoptic overview for 24 hour MSLP predictions (local models)

Comments on systematic errors in forecast synoptic features are given below and Table 2 provides an overview of performance of the local models. Firstly, however, the worst forecasts are identified for those situations where at least one of the local models scored higher than 30 skill-score points for 24-hour

**Fig. 1(a)** Comparison for LAPS\_PT375/TLAPS\_PT375/GASP from October to December 2000. 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\_PT375/TLAPS\_PT375/GASP from October to December 2000. 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.

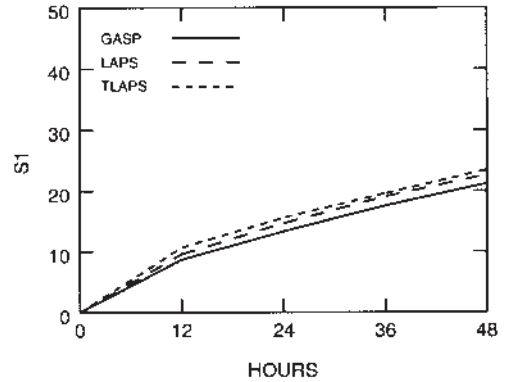


forecasts from 1100 UTC. Models considered are GASP, TLAPS\_PT375, LAPS\_PT375 and MESO\_LAPS and the dates refer to the verifying date of the forecasts.

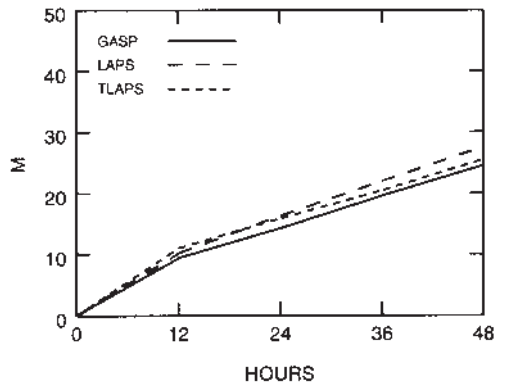
The first of these on 12 October, was a complex trough in the south-eastern Indian Ocean. LAPS and MESO\_LAPS moved the centre too quickly but also developed a secondary and spurious centre to the northwest. GASP's depiction of this system was much more accurate. This trough was also underforecast by LAPS and MESO\_LAPS two days later.

On 25 October, a low index situation saw the worst scores for the month. A cut-off low to the east of Bass Strait was generally underforecast, with the associat-

**Fig. 1(c)** Comparison for LAPS\_PT375/TLAPS\_PT375/GASP from October to December 2000. 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\_PT375/TLAPS\_PT375/GASP from October to December 2000. 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.



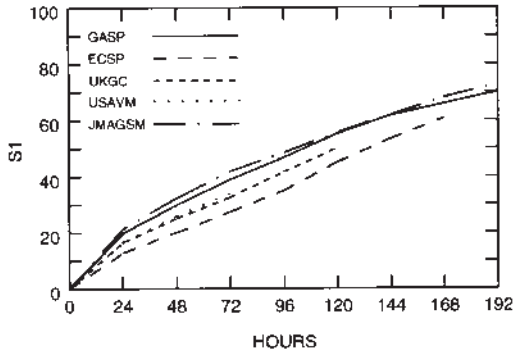
ed cradling high pressure system too strong, and the upstream trough too weak and too slow. All models had high scores and GASP was the most skillful.

Another low index situation occurred on 31 October, this time with the main system centred over the Bight and another easterly low off the Queensland coast. GASP handled it well but all the other models had the Bight low too deep and the secondary low too weak.

On 3 November both LAPS and MESO\_LAPS overestimated the strength of a southerly trough and on 27 November the details of another such trough were poorly forecast by these models.

No models scored 30 or more skill-score points for 24 hour forecasts during December.

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



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

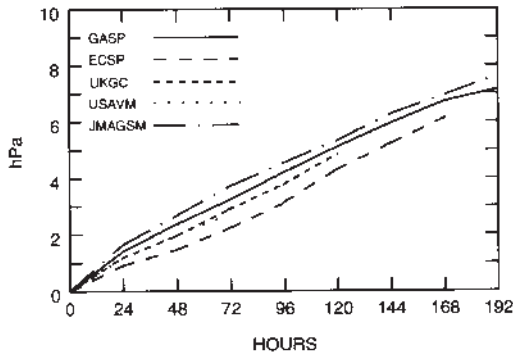
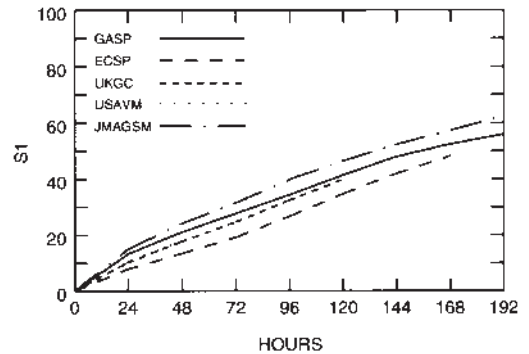


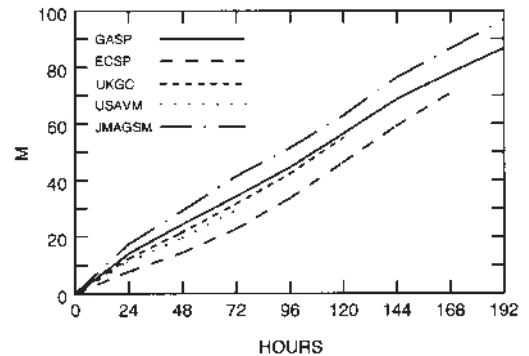
Table 2 provides an assessment of systematic errors for local models similar to that provided for the global models (see discussion below on Table 3). This is for 24-hour forecasts based at 1100 UTC. Since the differences between models at this shorter interval is not so marked as at three days, the number of synoptic categories used was simplified. Forecast errors greater than 4 hPa were used to identify the error region which was restricted to the Australian verification area. This region (shown in Fig. 5) is the same as the one used to produce most of the statistics in this article.

Given the differences in synoptic categories, Tables 2 and 3 are not directly comparable and both suffer the same weaknesses of subjective judgment and model analysis bias. Even allowing for this, some conclusions can be drawn.

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



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



All the local models still show a greater tendency to underestimate the strength of troughs in the westerlies (Category 3b) than to overestimate them (Category 3a) although both error types occurred with all models.

This quarter there was an increase in the number of occasions when each model showed no significant error (Category 1). A tropical cyclone, *Sam*, off the northwest coast from 5 to 10 December was the source of some errors in Category 6 for MESO\_LAPS and LAPS and the west Australian trough also caused some forecast errors as the continent started to heat. It was generally underforecast.

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

The performance of the global models averaged over

Fig. 3(a) Monthly S1 skill-scores of MSLP for LAPS\_PT375/TLAPS\_PT375/GASP/MESO\_LAPS from October to December 2000 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.

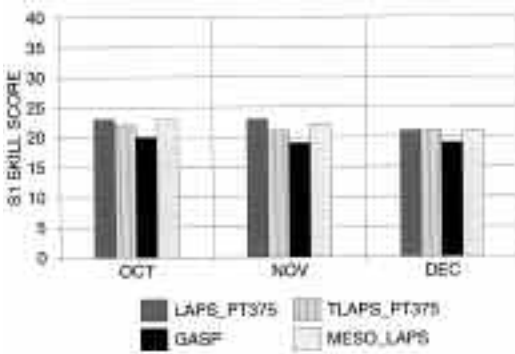
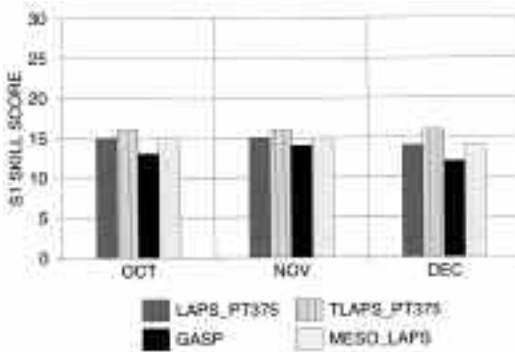


Fig. 3(b) Monthly S1 skill-scores of 500 hPa geopotential height (m) for LAPS\_PT375/TLAPS\_PT375/GASP/MESO\_LAPS from October to December 2000 for base-time 1200 UTC and interval +24 h over the irregular Australian verification grid.



the period for combined 0000 UTC and 1200 UTC base times (Fig. 2) shows ECSP as the most skillful followed by USAVM and UKGC and then GASP and JMAGSM. The two models UKGC and USAVM had very similar results with UKGC scoring slightly better at MSLP and USAVM at 500 hPa. Results for USAVM were available only out to +72 hours.

Figures 3(c) and 3(d) show the monthly breakdown of these results restricted to basetime 1200 UTC and validity +72 hours. The relative performances of ECSP, GASP and JMAGSM are confirmed by this alternative viewpoint, but UKGC and USAVM tend to fluctuate from month to month.

The anomaly correlation (Fig. 4) shows the same order of skill. Again USAVM and UKGC are almost identical when averaged over the period. Using 60 per

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

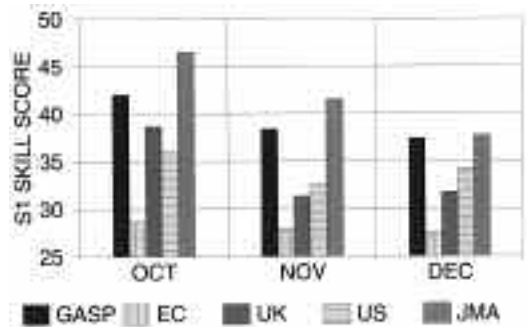
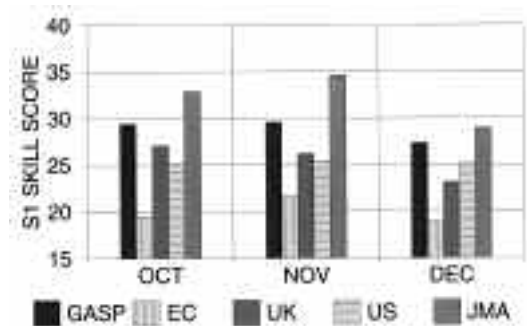


Fig. 3(d) Monthly S1 skill-scores of 500 hPa geopotential height (m) for GASP/EC/UK/US/JMA from October to December 2000 for base-time 1200 UTC and interval +72 h over the irregular Australian verification grid.



cent as the cut-off for useful forecasts, ECSP provides good guidance out to about 150 hours, GASP to about 125 hours and JMAGSM to about 120 hours.

**Synoptic intercomparison of global models**

This section presents a classification of errors into broad synoptic pattern categories with the aim of identifying systematic biases in the model predictions that may assist forecasters in the use of the model output. This study assessed three-day MSLP predictions from the 1200 UTC global model predictions.

For the period October-December 2000 daily forecast error charts (i.e. forecast MSLP minus the verifying analysis field from the same NWP system) were produced for each of the models. For the Australian region (loosely defined as between 10° and 40°S from

**Table 2. Number of occasions during October-December 2000 when errors occurred in each synoptic pattern category in the 24 hour predictions based at 1100 UTC for the local models.**

<i>Error pattern</i>	<i>GASP</i> <i>92 days</i>	<i>TLAPS</i> <i>91 days</i>	<i>LAPS</i> <i>91 days</i>	<i>MESO_LAPS</i> <i>87 days</i>
1. No significant error	44	28	15	18
2 Trough (low) in easterlies				
a. too deep	2	7	5	9
b. too weak	1	5	7	6
c. too fast	4	5	3	5
d. too slow	0	1	2	0
3. Trough (low) in westerlies				
a. too deep	8	9	11	12
b. too weak	18	17	33	26
c. too fast	6	8	11	10
d. too slow	13	11	14	11
4. Ridging				
a. too strong	3	6	6	7
b. too weak	0	4	3	3
c. too fast	0	1	2	2
d. too slow	1	1	3	2
5. WA Trough				
a. too deep	2	0	1	1
b. too weak	0	1	9	11
c. too fast	0	0	0	1
d. too slow	0	0	1	2
6. Tropical Lows				
a. overdeveloped	0	1	0	0
b. underdeveloped or too weak	1	1	4	4
c. mis-located	2	1	2	3

100° to 140°E and between 10° and 45°S from 140° to 160°E), each case of forecast error exceeding 4 hPa in magnitude was assigned subjectively to a synoptic pattern category as listed in Table 3 based on the dominant characteristic of the error. The categorisation is subjective and can clearly be ambiguous. Use of the system's own analysis is another weakness due to the strong influence of the model on the analysis field in data-sparse areas. Table 3 lists the error patterns used in the classification and the number of cases identified in each category for each global NWP system.

### General

There were fewer errors overall than in the previous quarter, especially in the number of errors exceeding 10 hPa in magnitude. This is an expected seasonal trend as the westerly flow and ridging becomes more prevalent over the region. UKMO had 17 cases

(Category 1) where the forecast error was less than 4 hPa over the full area of interest and ECMWF had 13 such cases.

Several general characteristics of the error patterns shown in previous quarters were again evident:

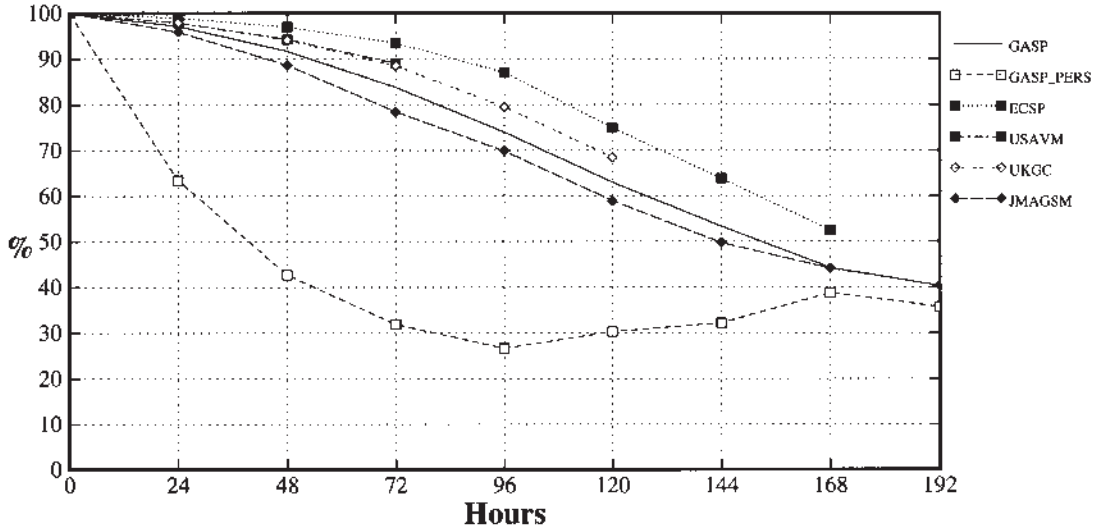
#### **Ridging and high pressure systems (category 7).**

All models again showed a clear tendency to over-predict the extent of high pressure systems and ridging. UKMO again had the most cases but all models had a high number in this category.

**Extratropical lows or troughs in the westerlies (categories 10 to 12).** Apart from ECMWF, the models tended to under-estimate the intensity of these features and to move them too slowly eastwards. The ECMWF predictions did not show a marked bias in either of these characteristics.

**Troughs and lows in easterly flow (categories 3, 4, and 6).** Except for ECMWF the predictions all con-

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



continued to show a slight bias linking such features too much to the westerlies and of moving such systems too rapidly to the east.

With regard to predictions of intensity, in contrast to the previous quarter (but similar to April-June) GASP tended to over-forecast the intensity of such systems. The character of the other model predictions has also varied considerably through the year, making it difficult to draw conclusions about systematic behaviour in intensity predictions.

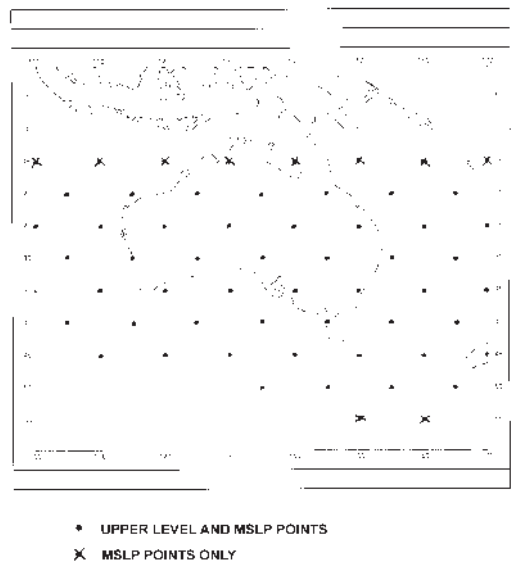
**Tropical systems (category 13).** With the seasonal re-appearance of tropical drepressions and lows some of the traits noted earlier this year re-appeared. In particular, the US predictions were again over-active in the generation of tropical lows and GASP was under-active. In this season JMA was also biased toward the over-development of tropical systems, with UKMO on the under-active side. Most of the mis-location errors were associated with tropical cyclone *Sam*, which is discussed in more detail later.

**Particular features of this season**

**Troughs over Western Australia (category 5).** One category which showed many more errors for all models than the previous three quarters was that of the easterly trough over Western Australia. As noted in previous years the problem is particularly acute for GASP in spring with the sign of the error almost exclusively negative indicating over-development of the trough. The US predictions also showed a bias toward over-predicting this feature, while UKMO clearly showed the opposite character. ECMWF and

JMA had a greater frequency of errors but the sign was more evenly spread. Major forecast errors occasionally resulted such as for GASP and JMA on 26 and 27 October, when spurious synoptic-scale lows were predicted over the south of Western Australia extending over the Bight.

**Fig. 5 Australian verification grid. Upper levels use 48 points while MSLP field verification uses 58 points.**



**Table 3. Number of occasions during October - December 2000 when errors occurred in each synoptic pattern category in the three-day MSLP prediction from each of the global models.**

<i>Error pattern</i>	<i>GASP</i>	<i>ECMWF</i>	<i>JMA</i>	<i>UKMO</i>	<i>USA</i>
1 No significant errors in pattern	1	13	2	17	7
2 Unable to match features	0	0	2	0	0
3 Low in easterlies or cut-off low					
A too linked to westerlies	6	1	4	8	4
B too much in easterlies or cut-off	1	1	1	0	1
4 Trough in easterlies or cut-off low:					
A too deep	20	4	12	6	9
B too weak	12	15	11	17	7
C details of structure	0	2	1	3	1
5 Trough in easterlies over WA:					
A too deep	41	9	8	4	15
B too weak	1	7	11	15	1
6 Trough in easterlies or cut-off:					
A too mobile (i.e. too far east)	10	6	11	7	5
B too slow (i.e. too far west)	5	6	8	5	3
7 Ridging or high pressure system:					
A too strong	15	23	25	33	20
B too weak	3	0	4	0	2
C too mobile	0	0	1	0	0
D too slow	1	0	1	0	0
E latitudinal displacement	1	0	0	0	0
8A Spurious low or marked over-development	8	3	16	2	4
8B Missed development of a low or trough	6	0	8	8	4
9 Retaining a low or trough too long	0	0	0	0	0
10 Trough in westerlies basically correct in structure but:					
A too strong	10	12	6	3	10
B too weak	27	14	18	21	29
C details of structure	1	0	3	4	2
11 Low predicted well in intensity but location poor	0	0	0	0	0
12 Trough in westerlies:					
A too mobile	4	0	2	5	3
B too slow	26	5	20	13	15
13 Tropical lows:					
A over-developed	2	0	10	1	15
B missed or too weak	5	3	0	7	5
C mis-located	7	10	6	5	4

Sensitivity tests in GASP suggest that the problem is related to the surface parametrisations and the definition of soil moisture. It is a data sparse region, subject to seasonally changing surface heat balance terms and complicated by the development of a synoptic-scale heat low which may interact with higher latitude frontal systems.

The synoptic pattern this year was unusual. The

Darwin Tropical Diagnostic Statements for November and December (prepared by the Bureau's Northern Territory Regional Office) noted that the trough over northern Australia was much deeper than normal and displaced southward. Monthly mean MSLP anomalies for November and December were up to 5 hPa below average over central and north-western Australia.

**Tropical cyclone *Sam* and the monsoon trough.**

Only one tropical cyclone (*Sam*) crossed the coast during the quarter. *Sam* crossed the Kimberley coast on 8 December and although its wind strength weakened, it moved eastwards and continued as an identifiable low bringing heavy rain until about 15 December. Its long trajectory was related to the marked and favourably located monsoon trough mentioned above.

There were some reasonable predictions for the development of *Sam* over the sea and as it crossed the coast. The JMA predictions showed the intensity best and gave the most accurate prediction for the coastal crossing. The GASP analyses and predictions were the least successful in capturing the intensity of the system.

Most of the models, except JMA, lost trace of *Sam* as an identity as it moved inland. In some cases it was not present even in the analyses so the predictions did not register an error. The JMA predictions retained a closed low and moved it eastwards at about the correct speed, although the predictions ultimately retained the low too long, leading to over-intense troughs over northeastern Australia from 14 to 16 December.

The stronger than normal monsoon trough was not well captured in general. Some of the errors included in categories for the extent of ridging or strength of lows in the easterlies were due to positive errors over northern Australia. ECMWF showed such errors on 15-16 November, during the track of *Sam* across northern Australia and again on 23-25 December.

**Return of zonal flow in the south.** There were several cases when a period of blocking in eastern longitudes was interrupted by a return to more zonal flow associated with deep lows well south of the Bight. Although in essence the predictions were very good, the timing in all models was a little slow. The clearest case occurred on 10 December but others occurred on 8 and 27 November.

**Problems with cut-off lows.** A low at the head of the Bight on 17 October moved southeastwards over southeastern Australia on 18 October. The predictions were very diverse and all except UKMO had very large errors.

Another case which caused large errors for most of the models was the development of a cut-off low south of Western Australia on 20 December and its subsequent movement. ECMWF modelled the situation very well throughout, but the others underestimated the initial development or had significant position errors.

## References

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

