Aust. Met. Mag. 44 (1995) 275–286

Verification of rainfall forecasts from the Australian Bureau of Meteorology's Global Assimilation and Prognosis (GASP) system

David Noone
School of Earth Sciences, University of Melbourne, Australia
and
Harvey Stern
Victorian Regional Office, Bureau of Meteorology, Australia

(Manuscript received February 1995; revised May 1995)

The Australian Bureau of Meteorology (BOM) implemented a high resolution version of its Global Assimilation and Prognosis System (GASP) on 22 March 1994. A dataset comprising rainfall forecasts for three Australian cities from the first available 100 days of GASP quantitative precipitation forecasts was constructed and verified against observed rainfall. Several measures of skill were used to evaluate GASP predictions. These were comparisons with persistence, climatology, 'model output statistics', 'perfect prog' and BOM Regional Forecast Centre (RFC) predictions.

GASP generally provided useful guidance up to four days in advance. However, it poorly forecast rain events associated with stream weather and post-frontal cold air, while precipitation accompanying fronts was indicated with considerable reliability.

Introduction

The Australian Bureau of Meteorology (BOM) implemented a high resolution version of its Global Assimilation and Prognosis System (GASP) on 22 March 1994 (National Meteorological Centre (NMC) 1994). The pre-existing GASP model had a horizontal resolution of wave number 31 and 19 levels in the vertical (R31L19). The new GASP increased horizontal resolution to wave number 53 (R53L19) (corresponding to a grid resolution of approximately 150 to 200 km in Australian latitudes), as well as improving the analysis scheme to a multivariate system, upgrading the radiation scheme and improved quality control of observations. This high resolution version of the GASP model was a project undertaken at the Bureau of Meteorology Research Centre (BMRC).

Corresponding author address: Mr D. Noone, School of Earth Sciences, University of Melbourne, Parkville, Vic. 3052, Australia.

In testing the model out to seven days rather than the previous five, BMRC considered that, due to the upgrading, the forecasts for days six and seven would be of use and the GASP model was implemented operationally out to seven days. Reconstruction of the code for greater efficiency, and utilisation of the multiprocessing capabilities of the Cray Y-MP computer on which GASP was run, resulted in a run time of around 75 minutes.

One feature of the R53L19 version of GASP is the quantitative precipitation forecast (QPF) field. Verification of the precipitation forecasts is important for two reasons. Firstly, the limitations of the model can be identified and taken into consideration by the forecaster. Secondly, weaknesses in the model in certain situations can be highlighted and could be addressed in a future upgrading.

The fields produced by numerical weather prediction (NWP) models, such as the BOM's GASP system, are interpreted by forecasters in terms of local weather. The purpose of this paper is to evaluate the skill of the GASP's QPFs at a number of locations during the model's first 100 days of operational life.

Specifically, the questions addressed are: are the GASP QPFs of sufficient accuracy to provide useful guidance? if so, for how many days in advance? and, which synoptic patterns are associated with the better (or poorer) forecasts? By 'sufficient accuracy' we mean better than no-skill measure such as persistence, and of comparable accuracy to pre-existing objective methods.

The dataset examined is of a short length, but this study provides an initial insight into the reliability of the GASP QPFs.

Verification method

GASP dataset

QPFs were obtained from the GASP prognoses via manual interpolation of the forecast charts.

QPFs for GASP's first 100 days were verified. The study period was set from 29 March, seven days after first run, to obtain a complete sequence of forecasts for each day, to 8 July 1994. Output was not produced on two of the days, due to computer problems.

Three Australian stations were selected for rainfall verification: Melbourne RFC; Perth Airport; and Brisbane Airport. In each instance, the topography in the vicinity of the respective site is such that anomalous localised orographic influences are negligible, the rainfall occurring likely to be representative of a broad area around the site. These three cities were used because they represented three different climate regimes — temperate uniform rainfall (Melbourne), subtropical winter rainfall (Perth) and subtropical summer rainfall (Brisbane). The airports were used in the latter two cases because these were the main observing sites for those cities. The charts of the GASP output issued daily by NMC were interpolated to obtain QPFs for up to seven days before the target day. The charts used were the products issued from the 2000 UTC run of GASP valid for a 24-hour period ending 1200 UTC the following day, and so on for the seven-day period. Only one station in each city was used due to lack of availability of hourly information. This means rainfall variation across a city is, unfortunately, not taken into account.

Comparative dataset

For the GASP forecast dataset, a number of verification statistics were calculated. Each statistic defined was also calculated for the datasets of the different forecast models and procedures. To establish a base against which skill could be established, two 'no-skill' QPFs were used, namely, the persistence and climatology QPFs. They are both considered to display 'no skill' because they can be derived automatically and without reference to sophisticated reasoning.

The persistence QPF is the rainfall in the 24-hour period prior to the forecast period, repeated for each of the seven days. The climatology QPF is taken to be the median 24-hour rainfall at the time of year under consideration, repeated for each of the seven days. From rainday records for each of the three cities, for the same days on the calendar as the forecast dataset, there is, on average, a less than fifty per cent probability of rain. This leads to a climatology QPF forecast for each day of no rain. An alternative to this could have been the 'seasonal average'. However, the climatology forecast 'no rain' has been used (rather than the seasonal average) because the most 'typical' day does, indeed, receive 'no rain' and much of the verification statistics presented relate to the rain/no rain occurrence issue.

GASP would be required to outperform the two 'no-skill' QPFs in order for usefulness to be demonstrated.

For further comparison, verification statistics for the official forecasts issued by the RFCs in each city were also calculated. The RFC forecasts were issued at 1000 UTC for the eastern cities and 1200 UTC for Perth (2000 local time) for a 24-hour period beginning 2300 UTC for the eastern cities and 0100 UTC (0900 local time) the following day. Hence, a different set of observed rainfall figures from the GASP dataset had to be used for verification.

Other specific guidance for rainfall prediction has been available in RFCs since the early 1980s. For example, Model Output Statistics (MOS) (Glahn and Lowry 1972) produces estimates of probability of precipitation occurrence and amount, and has been applied in the Australian region (Tapp et al. 1986). These MOS forecasts were issued at around 1400 EST based on 0000 UTC analysis fields from the Regional Assimilation and Prognosis system (RASP) (Leslie et al. 1985). This method was seen to produce a standard forecast to which GASP could be compared.

As computer systems at the BOM were being upgraded in late June, the MOS code was not converted, thus production of MOS forecasts ceased on 24 June. Combined with individual gaps in the records, MOS forecasts were only performed for 71 of the 100 days. The rainfall forecast was valid for 24 hours starting 0000 UTC the next day.

Stern (1994) describes the application of parameter enveloping to a 'perfect prog' (PP) system of forecast guidance which diagnoses the output of an NWP model in terms of local weather.

The process of parameter enveloping, in its simplest form, is defined by Stern (1994) as follows:

(a) suppose one is required to derive a regression equation to forecast a value for predictand F using predictors P and Q;

(b) linear regression, if applied to the data, leads to an equation of the form

$$F = a + bP + cQ,$$

a, b and c being constants;

(c) in order to establish the spectrum of prediction equations that might describe the variation in the value of the constants with respect to the westerly and northerly components of the surface flow, u and v, a new regression equation is derived that includes as predictors all the terms in the product $(1+u+v)*(1+P+Q)$, so leading to a prediction equation of the form

$$F=a+bP+cQ+du+ev+fuP+guQ+hvP+ivQ$$

where d,...i are constants, so providing 'envelopes' defining how the surface flow affects the manner in which the other predictors affect the value of the predictand. For example, the partial derivative of F with respect to P is

$$b+fu+hv$$

Stern's (1994) example addressed maximum temperature forecasting at Melbourne, but also referred to its extension to the forecasting of rainfall without documenting its performance. This PP system is applied here in order to provide further evidence of the effectiveness of GASP at describing precipitation events at Melbourne.

Measures of skill

Several measures of skill were employed to evaluate the QPFs. Such scores include Ratio tests (proportion correct), Hanssen and Kuipers (1965) score (Allen and Le Marshall 1994), the root mean square (rms) error and the correlation coefficient.

Mason (1979, 1989) highlights the manner in which many scores (with the exception of the ratio test) are dependent on the sample frequency, and this has to be borne in mind when considering the results which follow. Schaefer (1990), however, does justify their use provided they are applied in the same environment.

Categorical forecasts can be evaluated by use of a contingency table that shows the frequency of rain events forecast in categories compared to the observed category (Murphy and Winkler 1987). However, it is convenient to collapse such a table into a single score for comparison of forecasts. Many different scores have been defined for a 2×2 rain/no-rain contingency table (Table 1) in terms of mathematical manipulation of the elements, A, B, C and D.

For example, the ratio test is defined as the ratio of the total number of correctly forecast events and non-events to the total number of forecasts. That is

$$R=(A+D)/N$$

A perfect forecast system would yield R=1 and for a system that was always wrong R=0 (notwithstanding the fact that a system that is always wrong is just as skilful as one that is always right). The Hanssen and Kuipers (1965) score, defined as

$$V = (AD-BC)/((A+B)(C+D))$$

A, B, C and D as illustrated in Table 1

where V varies from −1 for total inaccuracy to 1 for perfect forecasts, was also used to provide a measure of skill at forecasting the occurrence or non-occurrence of rain. Woodcock (1976) showed the Hanssen and Kuipers score to provide an unbiased and acceptable measure of goodness of a forecast. However, the ratio tests and this score do not take into account the amount. On the other hand, rms error of the forecast rainfall category, defined at Table 2, and the correlation coefficient are measures that take into account the amount.

The correlation coefficient can also be used as such a measure. The value ranges from 1 for a perfect forecast to −1 for completely wrong and 0 for an essentially random distribution. Although Woodcock showed this to be an unacceptable measure for forecast evaluation in terms of scientific or administrative purposes, it can still be used as a comparison guide in an initial stage or to map how the forecast behaves over time. The correlation coefficient should only be used with the knowledge that it has limitations, as Woodcock outlines. The rms error can also be used as a comparative measure of the forecasts. It is, however, not conventionally used for ranking the skill of rainfall forecasts due to the skewed distribution of the rainfall categories recorded.

Table 1. 2×2 rain/no-rain contingency table defining the elements A, B, C and D for calculation of verification discriminants based on elements where N is the total number of observations.

Observed rain	*Forecast rain*		
	yes	*no*	*Total*
yes	A	B	
no	C	D	
Total			N

Table 2. Lower bound of rainfall categories defined by 24-hour rainfall ranges as used by the Bureau of Meteorology.

Category	0	1	2	3	4	5	6
Lower limit (mm)	0	0.2	2.6	5.6	10.6	20.6	40.6

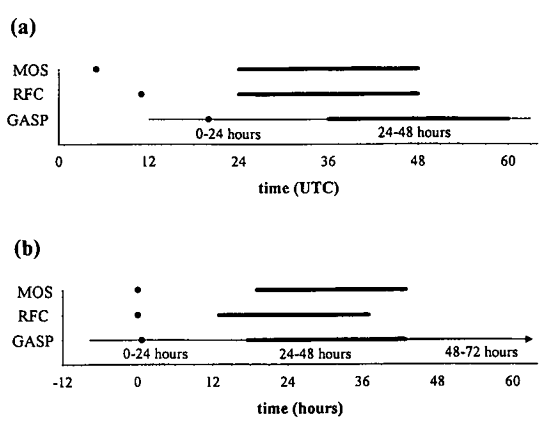
A feature of the GASP forecast was that rain amounts are predicted for seven days. To see how GASP performs over this time-scale, the various measures of accuracy can be calculated for each day of the forecasts period and plotted as a time series and compared with the persistence and climatology forecasts.

There was a difference in the lead and validity times of GASP/PP, MOS and RFC forecasts. The difference between the three systems is shown schematically in Fig. 1(a). If the time of issue for the different forecasts is defined as time zero, the overlap of the different validity times of the 24-hour period was sufficient for comparison to be made. This overlap is seen in Fig. 1(b).

Method of analysis

The RFC forecasts were issued as categories, rather than amounts, as defined in Table 2. The GASP, persistence, climatological and MOS forecasts were therefore also converted into these categories before the analysis took place for consistency and to normalise the effects of the skewed nature of the rainfall distribution.

Fig. 1 (a) Times (UTC) of issue and validity of GASP, MOS and RFC forecasts. (b) Time in hours of validity of GASP/PP, MOS and RFC forecasts after time of issue. Dots represent the time of issue and lines represent the 24-hour period for which the forecast is valid. Note that the GASP forecasts shown are the 0 to 24, 24 to 48 and in (b) 48 to 72-hour forecasts.



One of the simplest quantitative methods of verification is direct comparison of the number of rain days observed and forecast. This can be extended to comparison of the frequency of individual categories forecast and observed.

Insight into the reliability of GASP could be further studied by looking at individual significant or heavy rain events. These events were defined by the occurrence of at least 5.6 mm (category 3) rainfall being recorded. In particular the events were examined to categorise the weather into one of two types: frontal-related precipitation; and cold stream weather or outbreak-related precipitation.

Results

Quantitative analysis

Figure 1 shows that while the lead and validity times for GASP/PP, MOS and RFC forecasts are different, the time from issue to validity period is similar. For consistency, our analysis focuses on the 24 to 48-hour GASP forecast for comparison.

Data from the three cities were treated individually (because, as mentioned earlier, they do have somewhat different climates), as well as combined (given their locations fall within a narrow (10 deg.) latitude range). Although combining data masks specific information on each of the cities, common features are seen more readily.

Firstly, the percentage of rain days forecast by GASP was compared to that observed and to climatology. Table 3 presents these values which are plotted at Fig. 2. Figure 2 shows how well GASP estimates the frequency of rain days. For Brisbane, GASP significantly overestimated the number of rain days, but not for the other capitals. However, in the case of Brisbane, in particular, many of the forecasts were of category 1. It is feasible that although no rain was recorded at the station used, some light showers could have been present in the surrounding areas that would have rendered the forecast correct. This may be because Brisbane is a subtropical location exposed to the southeast trades and has more pre-

Table 3. Percentage of rain days predicted by GASP, expected from climatology and observed in the 100-day period studied.

	Observed	*GASP*	*Climatology*
Melbourne	41	53	43
Brisbane	31	52	39
Perth	39	35	43
Three cities	37	47	42

Fig. 2 Comparative plot of the percentage of rain days predicted by GASP, expected from climatology and observed in the 100-day period studied.

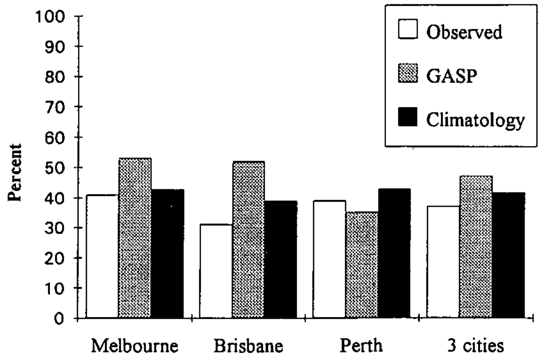
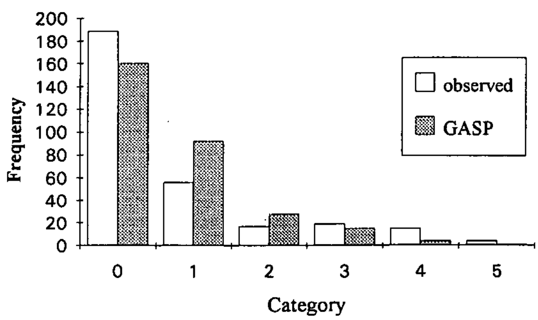


Fig. 3 Distribution or frequency spectrum of categorical GASP forecasts and rainfall observations for the 300 forecasts from the combined three-cities dataset.



cipitation of a convective nature than the other stations. Most importantly, GASP did not produce results that have unusually biased partitioning of the rain/no-rain forecast.

This comparison provides no indication of whether GASP is forecasting the rainfall amount accurately. A simple comparison that achieves this is the use of the rainfall category spectrum. Here the frequency of forecast categories from GASP are compared with the frequency of observed categories. Figure 3, which presents this information, shows a consistency between the two frequency distributions. It is noted that in this combined three-city dataset the deficit in no-rain events (when compared with forecasts by GASP) is made up for by an excess in category 1. This could be because while there may have been no rain at the verifying locality, light rain could have fallen nearby. The higher category frequencies tend to fall off exponentially. The number of observations was not sufficient to allow comment on higher categories.

What the two comparative analyses show is that GASP meets the necessary condition for the climatology to be reproduced. One weakness of this method is that, although the approximate correct number of a particular category is being forecast, the corresponding observation may be completely different. Categorical contingency tables can be constructed for the three-city 24 to 48-hour GASP forecast as in Table 4 to allow greater analysis of the forecast distribution. In particular, the relative frequency of a specific category forecast with another category being observed can be shown. The GASP forecasts can be compared to MOS (Table 5) and RFC (Table 6) forecasts. Due to the availability of data, comparison of the tables becomes easier upon scaling the 213 MOS forecasts, the 304 RFC forecasts, and the 300 forecasts as recorded for GASP.* The variations in the observed percentages were due to scaling the fraction of the total number of days in each dataset that a particular category was observed. This was because the percentage depends closely on the individual days included in the dataset. The different percentages of observations were also due to the differences in the time of day for which the GASP forecasts were valid compared to the forecasts of MOS and RFC. The differences were such that direct comparison of the tables was still justified.

In these tables, perfect forecasts are seen as values across the diagonal. From these tables, the accuracy of the forecasts can be evaluated. It was seen that although MOS tended to predict no-rain events well, the other category forecasts were not well-predicted. On the other hand, GASP seemed to predict categories with similar skill to the RFC forecasts. Of particular note was the large number of category 1 and 2 GASP forecasts for days which ended up with no rain, while the RFC forecasts were more representative of the observed categories.

From collapsing these tables into the 2×2 case as in Table 2, the rain/no-rain ratio score, categorical ratio score with an error of 0, 1 and 2 categories and the Hanssen and Kuipers score for rain/no-rain forecasts can be calculated. The values calculated effectively summarise these contingency tables and are tabulated in Table 7.

It is noted that the ratio test with a tolerance of two categories can be considered as a measure of bad forecasts by considering the cases not included. That is, the complement of this score will show the percentage of cases with an error of three or more categories.

*The reason for the different numbers of forecasts is that the trial period ended when there were 100 GASP forecasts. Because GASP did not run on two occasions within the period, there is a potential 306 (102x3) city predictions. However, two Brisbane forecasts were unavailable, leaving 304 city predictions.

Table 4. Relative frequencies of the three cities 24 to 48-hour GASP forecasts and corresponding recorded rainfall based on 300 observations.

Observed category	*Forecast rainfall category*						
	0	*1*	*2*	*3*	*4*	*5*	*Total*
0	43.7	16.0	2.3	0.7	0.3	0	63.0
1	7.3	7.0	2.3	1.3	0.7	0	18.7
2	1.3	2.3	1.7	0.3	0	0	5.7
3	0.7	2.3	1.3	2.0	0	0	6.3
4	0.3	2.0	1.7	0.7	0	0.3	5.0
5	0	1.0	0	0	0.3	0	1.3
Total	53.3	30.7	9.3	5.0	1.3	0.3	100.0

Table 5. Relative frequencies of the three cities 24 to 48-hour MOS forecasts and corresponding recorded rainfall based on 213 observations.

Observed category	*Forecast rainfall category*						
	0	*1*	*2*	*3*	*4*	*5*	*Total*
0	51.6	0.5	5.6	3.3	0.9	0.5	62.4
1	10.3	0.9	7.0	1.4	0	0	19.7
2	3.3	0.9	0.9	3.8	0.9	0	9.9
3	0.9	0.5	0.9	1.4	0	0	3.8
4	0.5	0	0.9	0.9	0.5	0	2.8
5	0	0	0.9	0.5	0	0	1.4
Total	66.7	2.8	16.4	11.3	2.4	0.5	100.0

Table 6. Relative frequencies of the three cities 24 to 48-hour RFC forecasts and corresponding recorded rainfall based on 304 observations.

Observed category	*Forecast rainfall category*						
	0	*1*	*2*	*3*	*4*	*5*	*Total*
0	46.4	10.2	2.0	0.3	0	0	58.9
1	8.6	9.5	1.6	0.7	0	0	20.4
2	2.0	4.9	2.6	1.0	0.3	0	10.9
3	0.7	2.0	0.3	1.3	0.3	0	4.6
4	1.0	0.3	1.0	0.7	0	0.3	3.3
5	0	0.7	0	0.7	0	0.7	2.0
Total	58.6	27.6	7.6	4.6	0.7	1.0	100.0

From Table 7 it is seen that in all cases GASP is better than persistence except for the categorical ratio score with 0 tolerance in Brisbane. In Melbourne, GASP is superior to MOS, while this relation is reversed in Brisbane (during the short period under consideration Melbourne appears to have had a higher frequency of widespread rain events that were well-replicated by GASP). In both Perth and the three cities combined, GASP and MOS are generally of similar value with some measures favouring one over the other depending on the instance.

To indicate the significance of the results presented, use was made of the variance of the Hanssen and Kuipers score as defined by Woodcock (1976). In each city over the seven-day forecast period, the variance became larger. For Melbourne, the variance was found to be 0.008 (95% confidence interval ± 0.18) for the 0 to 24-hour forecast and 0.011 (95% confidence interval ± 0.21) for the seven-day forecast. Similarly for the other two cities, the variance for Brisbane went from 0.010 (95% confidence interval ± 0.18) to 0.011 (95% confidence interval ± 0.21) and for Perth from 0.006 (95% confidence interval ± 0.20) to 0.010 (95% confidence interval ± 0.21) over the 7 days. Looking at the combined data for the three cities, this value reduced to

Table 7. Summary of verification scores for forecasts. For each city, GASP and Persistence results are calculated from 100 days, while the RFC forecasts and MOS are based on 102 and 71 days respectively. The number of days used in calculations of the three-city set is the sum of the number of days of each component city's forecasts.

Forecast type	*Rain/fine ratio test (%)*	*Categorical ratio test with number of categories tolerance (%)*			*Hanssen and Kuipers score*	*Root mean square error*
		0	*1*	*2*		
Melbourne						
GASP (24–48)	74	55	83	96	0.51	1.11
Persistence	53	46	75	86	0.02	1.46
MOS	67	40	78	99	0.33	0.95
RFC	74	60	89	98	0.47	0.98
PP	73	55	81	97	0.42	1.08
Brisbane						
GASP	61	50	88	91	0.27	1.18
Persistence	59	52	73	86	0.05	1.73
MOS	78	65	85	93	0.37	1.08
RFC	67	59	86	95	0.25	1.14
Perth						
GASP	78	58	85	95	0.52	1.12
Persistence	76	54	76	85	0.50	1.61
MOS	79	63	76	88	0.52	1.11
RFC	85	63	90	98	0.71	0.91
Three Cities						
GASP	71	54	85	94	0.42	1.17
Persistence	63	51	75	86	0.20	1.61
MOS	75	56	80	93	0.43	1.26
RFC	75	61	88	97	0.49	1.00

around 0.003 (95% confidence interval ± 0.11) for the entire forecast period, the smaller confidence interval due mostly to the larger sample size. As variance is a measure of spread of the data, a variance of this size suggests that the values obtained from the small sample studied would be a satisfactory reflection of a larger dataset, were it available.

Although comparing 24-hour forecasts was a good basis for the verification, an aspect of the GASP forecasts that was also to be examined was how the forecasts perform in terms of the seven-day period for which they were valid. Time series were produced as in Fig. 4 and show how, as expected, the forecast quality was reduced the earlier it was made. In Fig. 4, the measures plotted with time are the Hanssen and Kuipers score, correlation coefficient and rms error. As well as persistence, climatology forecasts were available for comparison for this period, but were excluded in the first two series plots. This was because the Hanssen and Kuipers score of the climatology forecast described is zero by definition and in the case of the correlation coefficient, climatology does not improve the comparison.

Summary

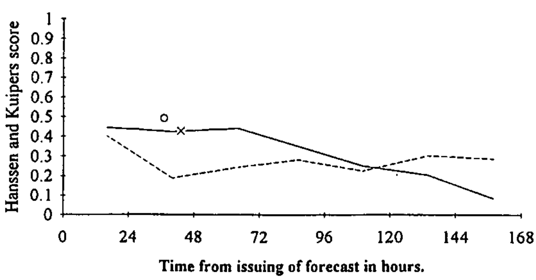
From these diagrams, it is evident that GASP was sufficiently superior to persistence, as expected if the model was to be of any use. There seems to be a point at about the five-day mark where GASP reduces to essentially a no-skill situation. From examination of Fig. 4(a), the MOS forecast seems to have a similar usefulness to GASP. A significance test (Seaman 1992) was performed and showed ($\alpha=0.05$) that indeed for the three cities, GASP held no significant advantage over MOS. The regression equations used by MOS were based on the old version of the RASP model so it would be expected that the high resolution GASP out-perform MOS. This is observed in Fig. 4(b) and Fig. 4(c) although statistical significance for correlation coefficient and rms error is not presented. It was further revealed that the RFC forecasts were still generally superior. This result was not surprising as the RFC forecasters have access to all guidance, including numerical, historical and observational, on which to base their prediction.

If the cities are again considered individually, other notable aspects of the forecast arise. The series of only the Hanssen and Kuipers scores is presented in Fig. 5, as it is the best for such verification procedures, as Woodcock (1976) shows.

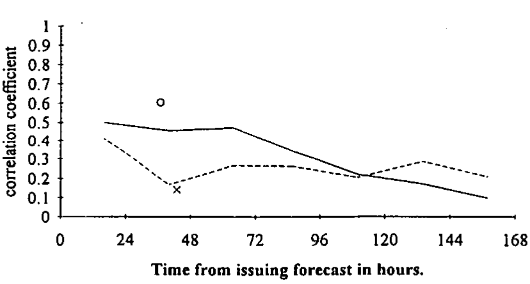
One feature both Fig. 5(a) and 5(b) show is that the best forecast is issued for periods ending the 48 and 72-hour periods — not the most immediate period. This may indicate a weakness in the model's initialisation scheme or, alternatively, it could be caused by the model needing time to

Fig. 4 Time sequences of verification scores for the combined three-city dataset. The data are based on 300 days of data for GASP and persistence, 213 days for MOS and 304 days for the RFC forecast. Points are plotted as times at the end of the 24-hour forecast period based on the issue time at time=0. The solid line represents GASP, the dashed line represents persistence and for further comparison, MOS and RFC forecasts are shown as a cross and a circle respectively. In (c), the climatological forecast is also included as a horizontal line. (a) The Hanssen and Kuipers score series. (b) The correlation coefficient series. (c) The root mean square error (in categories) series.

(a)



(b)



(c)

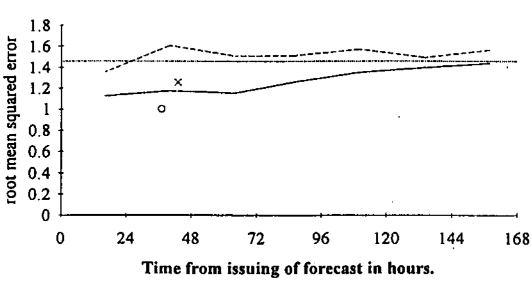
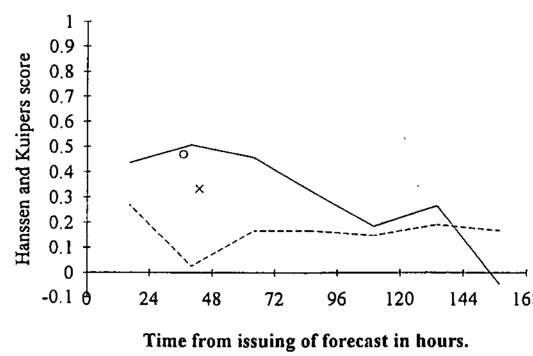
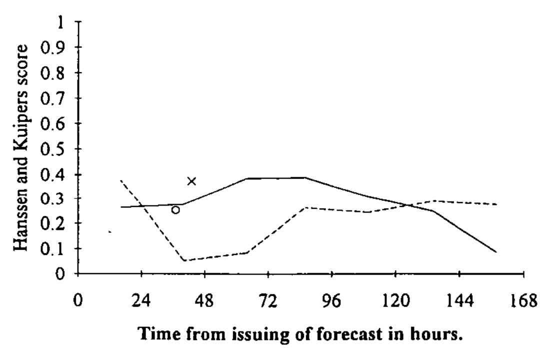


Fig. 5 Hanssen and Kuipers scores as a time series for Melbourne, Brisbane and Perth. The scores are based on 100 days of data for GASP and persistence, 71 days for MOS and 102 days for the RFC forecast. Points are plotted as times at the end of the 24-hour forecast period based on the issue time as time=0. The solid line represents GASP, the dashed line represents persistence, MOS is shown as a cross and RFC is shown as a circle. (a) Hanssen and Kuipers score series for Melbourne. (b) Hanssen and Kuipers score series for Brisbane. (c) Hanssen and Kuipers score series for Perth.

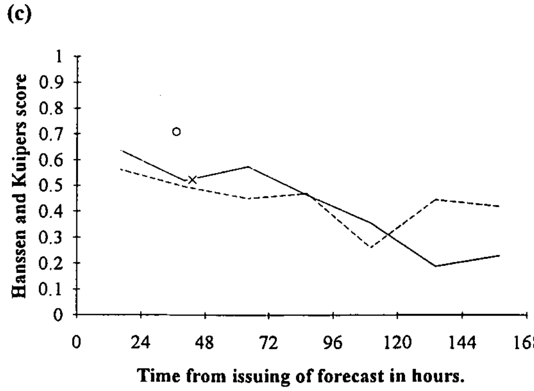
(a)



(b)



(c)



develop or 'spin-up' vertical motions within the dynamics before performing at its most effective. The latter is more probable in respect to the modelling of tropical storms — during the period under consideration a number of minor systems of tropical or sub-tropical origin 'grazed' south-east Queensland, and it is interesting that this feature is more prominent in the most tropical city, Brisbane, where the persistence forecast outperforms GASP for the first 24 hours.

In Melbourne the guidance offered by GASP is comparable to the forecast produced by the RFC for the 100-day time span examined and superior to the MOS guidance. Significance testing showed again that at about the fifth day GASP was no better than persistence.

In Perth, GASP was comparable to MOS and significantly better than persistence for the period of three days ending 24, 48 and 72 hours and again significantly better on day five.

It is noticed that the highest scores were for Perth. This was mostly due to the type of weather recorded in this city for the 100-day period. The first 50 days of the period studied had little or no rain while there were frequent rain events and showers in the second 50 days. This meant that the persistence forecast was quite accurate and that the rain was usually more predictable, as is evident from the generally higher scores for this city. Brisbane had the lowest scores of the three cities, probably due to a higher frequency of convective precipitation. This was true for GASP, MOS, RFC and persistence. This consistency suggests greater difficulty in issuing accurate rainfall forecasts in this region. The important feature, however, is that the GASP forecasts were still considerably more skilful than persistence. Looking at the dataset shows mostly dry conditions interspersed with occasional rain events. This lack of definable, systematic pattern in rainfall events leads to a lower score for the persistence forecast. Forecasting problems arise from the position of Brisbane in a region of a predominant high pressure ridge with large spacings between isobars. Thus the GASP forecasts also score low.

Fig. 6(a) GASP analysis of synoptic situation on 10 May 1994. Note the cold air ridge to the south of Melbourne with south to southeasterly flow. Solid lines represent isobars and dashed lines represent contours of equal thickness which can be interpreted as approximate isotherms.

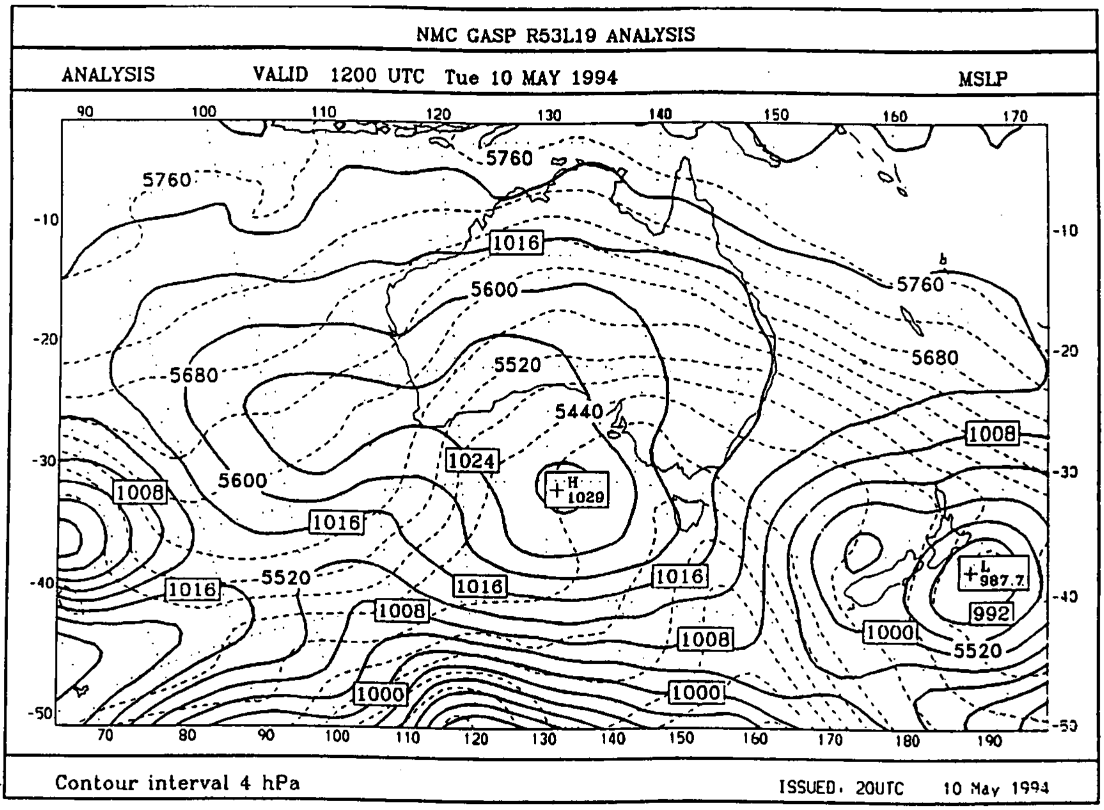
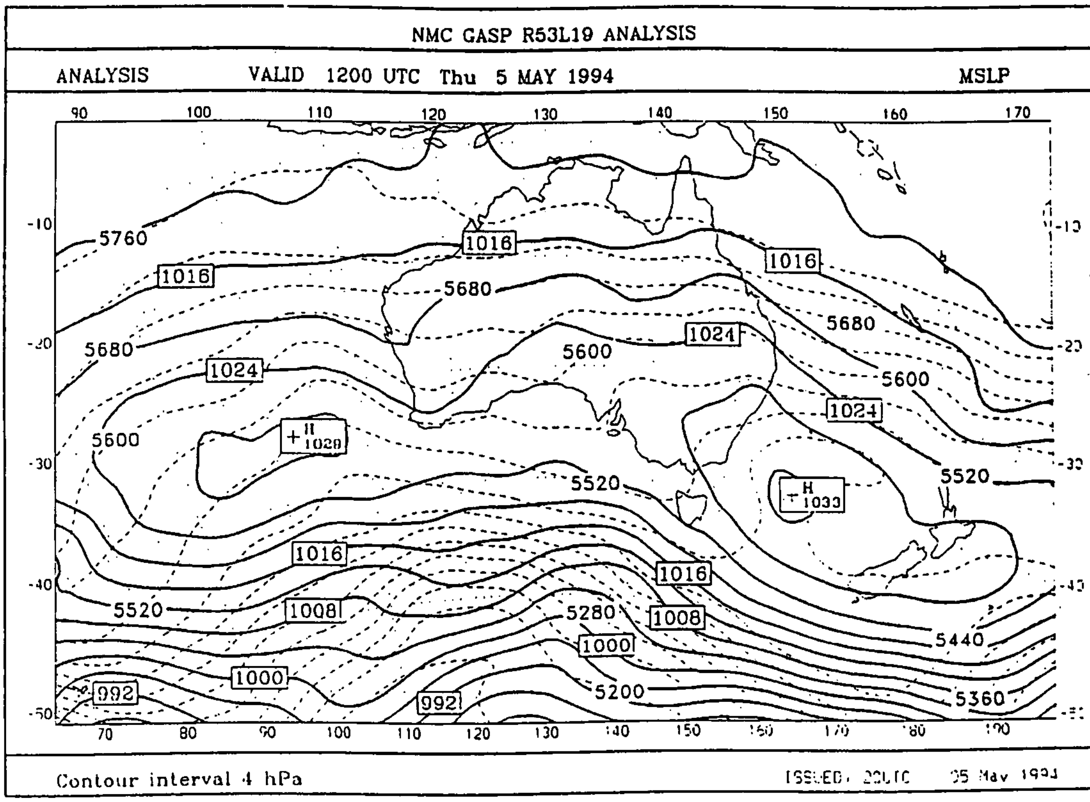


Fig. 6(b) GASP analysis of synoptic situation on 5 May 1994. Note the cold air ridge to the southeast of Brisbane with southeasterly flow. Solid lines represent isobars and dashed lines represent contours of equal thickness which can be interpreted as approximate isotherms.



Qualitative analysis

The synoptic patterns observed in each of the three cities are quite different and therefore must be examined separately.

Melbourne In Melbourne, there were eight days with more than 5 mm during the period. Six of these events were frontal related; the other two were caused by a cold SSE stream as in Fig. 6(a). This stream became unstable on its northward passage because the ocean surface is warmer than the southern origin air, and heated it from below. The resulting convective cells caused the formation of the characteristic cumulus cloud and showery precipitation associated with such stream weather.

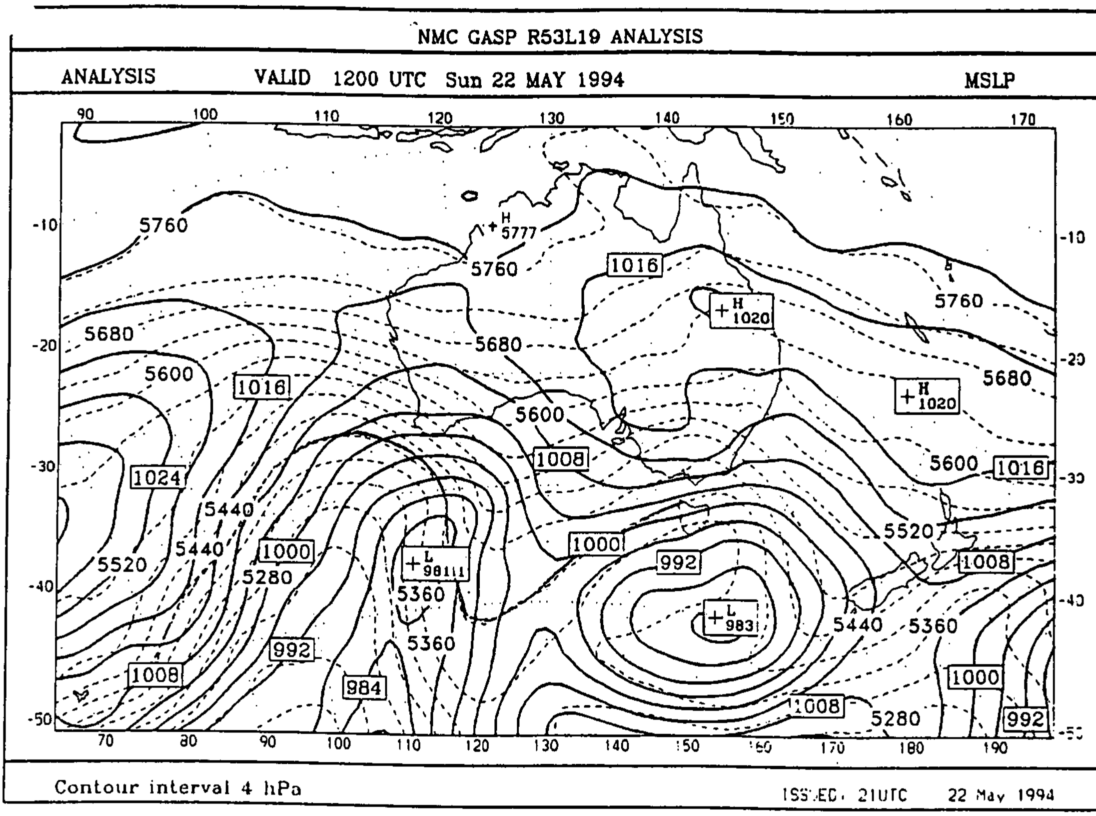
One of the stream events was predicted with great accuracy up to four days before the event while the other was overlooked, seemingly due to the non-forecast of the cold Tasman Sea air. In one three-day period for which GASP predicted weak frontal-related precipitation, rainfall was not observed.

GASP QPFs indicated little more than minimum precipitation with the fronts. In particular, GASP QPFs were often of about 3 or 4 mm while 8 to 10 mm were observed. Frontal precipitation was better predicted on the 24 to 48 and 48 to 72-hour forecasts.

It is noted from examination of the mean sea-level pressure (MSL) charts that fronts are generally well timed and positioned by GASP.

Brisbane The lower latitude of Brisbane allowed most of the frontal weather seen at Melbourne to pass to the south. There were nine days with more than 5 mm of rain. In most cases, there was cool air to the southeast with a southeasterly flow often associated with a high pressure system centred to the east of Victoria as in Fig. 6(b). In only one instance was the rain in Brisbane caused by a front. With each high centre that passed over the eastern coast, rain fell in Brisbane, but the amount of rain was not always above 5 mm. In only two instances of the 0 to 24-hour forecast was the stream weather predicted. Again, better forecasts

Fig. 6(c) **GASP analysis of synoptic situation on 22 May 1994. Notice the trough approaching Perth. Solid lines represent isobars and dashed lines represent contours of equal thickness which can be interpreted as approximate isotherms.**



were from the 24 to 48-hour period. There was a lower frequency of these rain events in the second half of the 100-day period (seven of the nine) due to the northward migration of the high pressure belt as winter approaches.

Perth The first 52 days saw only three days of rain in Perth, whereas during the remainder of the period it rained on most days, with 20 days recording a fall in excess of 5 mm. All were due to frontal-type activity (for example, see Fig. 6(c)), either raining directly at the front or due to the instability of the cold air pool left in the front's wake. Eight of the 20 were predicted well and in general these eight were the first days of the consecutive days on which there was rain. GASP did not predict the showers in the post-frontal cold air.

Summary

The accuracy of the GASP QPFs has been compared with a number of other products. RFC QPFs were superior, but, when compared with MOS and PP estimates, the measures used variously suggested superiority and inferiority and no clear conclusion could be drawn. A particular difficulty applied to the MOS guidance because it was discontinued about two-thirds of the way through the trial.

In general, GASP predicted frontal rain better than the stream showers. However, because precipitation was depicted on the QPF charts by patches that could easily miss a city by only a few hundred kilometres, sometimes no rain was indicated by GASP for a city while a patch of rain was shown as passing nearby. This latter problem reduced the apparent skill with which GASP predicts rain.

Conclusion

A preliminary examination of the first 100 days of the GASP QPFs demonstrates useful skill out to four days. Although on most measures the RFC forecasts were superior to those of GASP, this could be a result of the RFC having the advantage of the GASP estimates at the time of preparation of their forecasts. GASP, PP and MOS QPFs were of similar skill. The least skill was displayed by climatology and persistence.

Although GASP's useful skill is limited to four days in advance, the seven-day forecast period is a useful feature as it enables the forecaster to monitor the development of rain systems as the target day approaches.

By looking at specific events, it was evident that GASP could predict frontal precipitation with reasonable skill while stream weather and cold post-frontal air mass precipitation was poorly forecast. This suggests that instability or convection processes might not be resolved or simulated well enough in the GASP model.

Rainfall can be a relatively local phenomenon and often occurs on the subgrid-scale. Local topography, for example, has a large influence on the spatial distribution of precipitation, particularly in stream weather. The GASP would need to resolve such scales in order to better account for such influences.

This study represents only the second time that there has been an objective verification of numerical rainfall forecasts reported in Australia. The first such verification was reported by Mills and Logan (1994), but that study examined those of the operational RASP model over a 24-hour period. Mills and Logan (1994) faced the same limitation as the present investigation, namely that due to the verification sample being small. This limitation led to them concluding in their paper that 'a longer period of statistics encompassing a larger sample of heavy rainfall events will provide a more robust assessment'. We now conclude the present paper similarly, urging that objective verification of numerical rainfall forecasts in Australia should become routine, in order that trends in prediction performance may be more effectively monitored.

Acknowledgments

Special thanks go to Noel Davidson and Graham Mills of the Bureau of Meteorology Research Centre, to Noel Fitt of the Victorian Regional Office, and to Ian Simmonds of Melbourne University's School of Earth Science for their help and advice.

References

Allen, G. and Le Marshall, J.F. 1994. An evaluation of neural networks and discriminant analysis methods for application in operational forecasting. *Aust. Met. Mag., 43*, 17–28.

Glahn, H.R. and Lowry, A.D. 1972. The use of model output statistics in objective weather prediction. *Jnl appl. Met., 11*, 1203–11.

Hanssen, A.W. and Kuipers, W.H.A. 1965. On the relationship between frequency of rain and various meteorological parameters. *Koninklijke Nederlands Meteorlogisch Institut, Meded. Verhand., 81*, 2–15.

Leslie, L.M., Mills, G.A., Logan, L.W., Gauntlett, D.J., Kelly, G.A., Manton, M.J., McGregor, J.L. and Sardie, J.M. 1985. A high resolution primitive equations NWP model for operations and research. *Aust. Met. Mag., 33*, 11–35.

Mason, I. 1979. On reducing probability forecasts to yes/no forecasts. *Mon. Weath. Rev., 107*, 207–11.

Mason, I. 1989. Dependence of the critical success index on sample climate and threshold probability. *Aust. Met. Mag., 37*, 75–81.

Mills, G.A. and Logan, L.W. 1994. Upgrades to the Australian limited area assimilation system. *Aust. Met. Mag., 43*, 167–80.

Murphy, A.H. and Winkler, R.L. 1987. A general framework for forecast verification. *Mon. Weath. Rev., 115*, 1330–8.

NMC 1994. Operational implementation of the R53L19 Version of the Global Model (GASP). *NMC Operations Bulletin, No. 16*, 15 March. Bureau of Meteorology, Australia.

Schaefer, J.T. 1990. The critical success index as an indicator of warning skill. *Weath. forecasting, 5*, 570–5.

Seaman, R.S. 1992. Serial correlation considerations when assessing differences in prediction skill. *Aust. Met. Mag., 40*, 227–37.

Stern, H. 1994. Enveloping predictor parameters in a regression equation – an alternative to model output statistics (MOS). *Abstracts. 10th Conf. Num. Weath. Pred.*, Amer. Met. Soc.

Tapp, R.G., Woodcock, F. and Mills, G.A. 1986. The application of model output statistics to precipitation prediction in Australia. *Mon. Weath. Rev., 114*, 50–61.

Woodcock, F. 1976. The evaluation of yes/no forecasts for scientific and administrative purposes. *Mon. Weath. Rev., 104*, 1209–14.