440

Performance of the ECMWF and the BoM Ensemble Systems in the Southern Hemisphere

W Bourke[1], R Buizza and
M Naughton[1]

Research Department

May 2004

[1] Bureau of Meteorology Research Centre, GPO Box 1289K,
Melbourne, Victoria, Australia 3001

Submitted for publication to Mon.Wea.Rev.

For additional copies please contact

The Library
ECMWF
Shinfield Park
Reading, Berks RG2 9AX

library@ecmwf.int

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications.html

Abstract

The performance of two Ensemble Prediction Systems, that of the European Centre for Medium-Range Weather Forecasts (EC-EPS) and that of the Australian Bureau of Meteorology (BM-EPS) are compared over the Southern Hemisphere annulus (20-60 degrees South) and over the Australian Region. Ten-day ensemble forecasts for 152 daily cases (from 2 April to 31 August 2001) of 500 hPa geopotential height are examined.

A comprehensive set of verification measures documents the different spread and skill characteristics of the BM-EPS and EC-EPS. Overall, EC-EPS deterministic (i.e. unperturbed control) products and the probabilistic ensemble based products are more skillful than the corresponding BM-EPS products. The utility of the BM-EPS for the Australian region is indicated.

1. Introduction

One of the most recent advances in medium-range numerical weather prediction has been the development of weather prediction systems based on a probabilistic approach, whereby not only a single forecast of the future atmospheric state is provided but an estimate of the probability density function of forecast states is predicted by the application of ensemble/Monte-Carlo strategies. Practically, the prediction of the time evolution of the probability density function of forecast states has been based on a finite number of numerical integrations.

Since December 1992, global ensemble prediction systems have been part of the operational suites at ECMWF (European Centre for Medium-Range Weather Prediction, Reading, UK; see, e.g., Molteni *et al*., 1996) and NCEP (National Centers for Environmental Prediction, USA; see, e.g., Tracton and Kalnay, 1993). Since 1998, an ensemble prediction system has been operational also at CMC (Canadian Meteorological Centre, Montreal, Canada; see, e.g., Houtekamer *et al*., 1996). A recent overview and assessment of the ECMWF, NCEP and CMC systems is given by Buizza *et al* (2003a). Additionally global and regional ensemble systems have been under development at other institutions, such as Meteo-France (J. Nicolau, 2001, personal communication), the UK Meteorological Office (Harrison *et al*., 1999), the Regional Meteorological Service of Emilia-Romagna (Molteni *et al*., 2001, Marsigli *et al*., 2001), KNMI (The Royal Dutch Meteorological Institute; see, e.g., Hersbach *et al*., 2000), and BMRC (Bureau of Meteorology Research Centre, Melbourne, Australia) as described in the present study.

Research in ensemble prediction started at BMRC in 1996. From May 2000 the BMRC global ensemble prediction system has been run daily in research trials. Since July 2001 it has been run in a real-time operational mode by the Bureau of Meteorology (BoM) National Meteorological and Oceanographic Centre (NMOC); it is scheduled to become officially operational in early 2004.

While the three global ensemble systems operational at ECMWF, NCEP and CMC have been designed to provide probabilistic forecasts for both the Northern and the Southern Hemisphere, their performance has been assessed over the northern hemisphere (see, e.g., Buizza, 1997, Toth *et al*., 1997, 1998). The aim of the present study is to provide an assessment of the performance of the ECMWF Ensemble Prediction System (EC-EPS) and the pre-operational BMRC Ensemble Prediction System (BM-EPS) over the southern hemisphere annulus (20-60 degrees south) denoted subsequently by SH and over the Australian region denoted subsequently by AU ; these results focus on the 500 hPa geopotential height field; BM-EPS and EC-EPS daily forecasts from 2 April to 31 August, 2001 (152 cases) are compared.

This study has focused on only the verification at 500hPa in providing an initial evaluation and comparison of these two systems in the southern hemisphere. The recent intercomparison study of the performance of the EPS systems of ECMWF, NCEP and CMC in the northern hemisphere, presented by Buizza *et al* (2003a,) has similarly focused on 500 hPa. Quantitative verification of the EC-EPS for the northern hemisphere has evolved from initial evaluations predominantly in terms of 500 hPa (Molteni *et al*., 1996) to more comprehensive validations of for example quantitative precipitation forecasts for the continental eastern United States (Mullen and Buizza, 2001). A more extensive verification of the BM-EPS system validating quantitative precipitation forecasts in the Australian region is presently being conducted (E. Ebert, 2003 personal communication) to support and encourage wider operational utilization of the full range of the products available from the BM-EPS.

The paper is organized as follows. After a brief description of the two ensemble systems and of the accuracy measures used to assess the ensemble performance (Section 2), deterministic and probabilistic EPS forecasts of 500 hPa geopotential height fields are compared (Sections 3 to 8). Some conclusions are then finally drawn (Section 9).

2. Description of the EC-EPS and BM-EPS

Ensemble systems can differ not only in the way they simulate the effect of initial uncertainties (i.e., errors in the analysis) and the character of model prediction errors, but also through the size of the ensemble and ability to represent the probability distribution function of analysis and forecast states. As far as the simulation of initial uncertainties is concerned, both the EC-EPS and the BM-EPS system have followed the same methodology and define the initial perturbations along the directions of the phase-space of the system characterized by the maximum linear growth of total perturbation energy during a finite time interval. These directions are called the singular vectors of the linear propagator of the model equations (Lorenz, 1965; Farrell 1982, 1989; Buizza and Palmer, 1995). More comprehensive generation of singular vector based perturbations, which are consistent with estimates of analysis error statistics, have been described by Ehrendorfer and Tribbia (1997), Palmer *et al*., (1998), Barkmeijer *et al*., (1999) and Hamill *et al*., (2003); from a practical point of view these are more complex and costly to calculate. Palmer *et al*., (1998) discuss in some detail the properties and appropriateness of the total energy metric for singular vector calculations and suggest that of the simpler metrics (e.g enstrophy, kinetic energy or total energy) total energy is the most appropriate for the predictability problem. An initial evaluation by Barkmeijer *et al*., (1999) of the use of Hessian singular vectors based on static background error covariance, rather than total energy singular vectors, within the EC-EPS system did not improve ensemble prediction performance; the use of a flow dependent covariance in this context, as demonstrated in simulated assimilation studies within an Ensemble Kalman Filter assimilation system by Hamill *et al*., (2003) suggests a very interesting further development of these strategies. The use of singular vectors is different from the methodologies followed at NCEP and at CMC. At NCEP, the initial perturbations are generated using what are described as bred-vectors; the perturbations are generated by comparing pairs of ongoing data-assimilation systems started from randomly perturbed and unperturbed initial states (Toth and Kalnay, 1993). At CMC, each perturbed initial condition is generated by running the CMC data assimilation with perturbed observations and with different model configurations (Houtekamer *et al*., 1996). Further and more recent discussion of the NCEP and CMC systems (and also the EC-EPS system) are included in Buizza *et al* (2003a). See Szunyogh *et al*., (1997), Palmer *et al*., (1998) and Errico and Langland (1999) for detailed discussions of the similarities and differences between singular vectors and bred vectors.

Differing strategies have been developed to simulate the influence of model errors. The EC-EPS simulates the effect of random model errors by introducing random perturbations within the model physical parameterization schemes (Buizza *et al*., 1999). In the approach implemented at CMC, each EPS ensemble member is integrated using a different combination of model parameterization schemes (Houtekamer *et al*., 1996). Within the BM-EPS there is no inclusion of such parameterizations of random model error.

2.1 The EC-EPS

Since 21 October 2001, the EC-EPS has been operational with one unperturbed and 50 perturbed 10-day model integrations with a spectral triangular truncation T_L255 (where subscript L denotes use of linear grid resolution in physical space) and 40 vertical levels (L40). In physical space, the EC-EPS has a grid spacing at mid-latitudes of approximately 80 km. See Simmons and Hollingsworth (2002) and references therein for a recent description of the main characteristics of the ECMWF model and physical parameterization schemes. Prior to March 2001, the operational EC-EPS was running with 1200UTC as starting time; since then ECMWF has been running the EPS system twice a day, at 0000UTC and 1200UTC (the 0000UTC EC-EPS used in this study was not part of the ECMWF operational suite at the time, but was part of an effort to develop an ensemble-based system for severe weather prediction). The results discussed in this paper refer only to the 0000UTC runs of the EC-EPS, as this is the daily starting time used in the initial implementation of the BM-EPS system..

The d^{th} day EC-EPS initial perturbations are generated using the initial-time singular vectors growing between days d and d+2, and the final-time singular vectors growing between day d-2 and d (Barkmeijer *et al*., 1999). In the current configuration, these two sets of singular vectors are linearly combined to define the j^{th} initial perturbations:

$$f_j^{EC}(0) = f_0^{EC}(0) + p_j^{EC}$$
$$p_j^{EC} = \sum_k \alpha_{j,k}^{EC} SV_k^{EC}(d;t=0) + \beta_{j,k}^{EC} SV_k^{EC}(d-2;t=48h)$$

where $f_0^{EC}(0)$ is the initial unperturbed analysis (more precisely, it is the $T_L255L40$ interpolated version of the high-resolution $T_L511L60$ ECMWF analysis) provided by the ECMWF 4D-Variational data assimilation scheme (Mahfouf and Rabier, 2000). The scaling factors $\alpha_{j,k}^{EC}$ and $\beta_{j,k}^{EC}$ are defined by comparing each set of singular vectors with an estimate of the analysis error provided by the ECMWF 4D-Variational data assimilation scheme. Here the individual perturbations are obtained as linear combinations of the component singular vectors through an orthogonal phase space rotation; this rotation, as described in Molteni *et al*. (1996), is designed to maximize the geographical spread of each of the ensemble perturbations, with the magnitude scaled by an estimate of analysis error.

The singular vectors are computed at resolution T42L40 with only simplified physics, namely linearised surface stress, in the tangent forward and adjoint model versions (Buizza, 1994). Two sets of singular vectors, each having maximum growth over a period of 48 hours, are computed, namely:

- northern hemisphere (NH) singular vectors are defined so that they maximize growth in the total energy norm north of 30 degrees North;
- southern hemisphere (SH) singular vectors are defined so that they maximize growth in the total energy norm south of 30 degrees South.

The forecast valid at time-step t for each perturbed member is defined by the numerical integration of the nonlinear perturbed model equations:

$$f_j^{EC}(t) = f_j^{EC}(0) + \int_0^t [A^{EC}(t) + (1 + r_j)P^{EC}(t)]dt$$

where $A^{EC}(t)$ and $P^{EC}(t)$ denote the time t tendencies due to the non-parameterised and the parameterised physical processes respectively. Random model errors due to the parameterised physical processes are simulated by randomly perturbing the tendency due to the parameterised physical processes by the factor r_j, as described in Buizza *et al*. (1999). Currently, the random numbers are defined so that the same random number is used in the vertical, the same random number is used inside a 10-degree box and the random numbers are regenerated every 6 hours. This setting introduces certain spatial and temporal coherent structures in the model-induced perturbations.

2.2 The BM-EPS

The BM-EPS was run in an initial research trial from May 2000 until March 2001 with one unperturbed and 32 perturbed 10 day model integrations on a daily basis with 0000UTC as starting time. The system was configured with a spectral triangular truncation T79 and 19 vertical levels and with perturbations only applied in the southern hemisphere. As of the beginning of April 2001 the BM-EPS trial was upgraded in spectral resolution to T_L119L19 and the use of semi-Lagrangian time integration rather than the previously used Eulerian scheme. In physical space, this BM-EPS has a grid spacing at mid-latitude of approximately 160 km. From 5 July 5 2001 the BM-EPS has run in operational trial mode in the BoM National Meteorological and Oceanographic Centre (NMOC); in December 2001 this operational trial was extended to include twice daily running from 0000UTC and 1200UTC with perturbations in both hemispheres. The results discussed in this paper refer to the 0000UTC runs of the BM-EPS from April 2 to August 31, 2001.

The assimilation systems supporting the EC-EPS and BM-EPS are substantially different with respect to the analysis algorithms used and also have non-trivial differences in observational data usage. The ECMWF assimilation system uses a four-dimensional variational (4D-Var) system (Mahfouf and Rabier, 2000) and together with conventional data utilizes a rather complete range of remotely sensed data including for example much of the data from the then available two NOAA Advanced TOVS (ATOVS) sounders (both Advanced Microwave Sounding Unit (AMSU-A) and High-Resolution Infrared Sounder (HIRS)/3 radiances); additionally marine wind speed estimates from retrievals from two of the Special Sensor Microwave/Imager (SSM/I) on board the Defense Meteorological Satellite Program were in use by ECMWF at this time. The BoM assimilation system utilised three dimensional multivariate statistical interpolation within its data assimilation system (Seaman *et al*., 1995); the analyses relevant to these studies used common conventional data but somewhat less remotely sensed data than that used by ECMWF; the BoM system utilsed only ATOVS (AMSU-A) radiances within a one dimensional variational (1D-Var) retrieval scheme (Harris and Kelly, 2001) from NOAA15 until mid June 2001 and from then onwards additionally used NOAA16 AMSU-A radiances. The BoM assimilation did not access any of the SSM/I data. The more sophisticated assimilation system and the more comprehensive data utilization of the ECMWF system is considered to be providing more accurate initial conditions than available from the BoM system.

Apart from using substantially different data assimilation and model prediction implementations the notable differences between the EC-EPS and the BM-EPS are as summarized in the following discussion.

The d*th* day BM-EPS initial perturbations are based only on the initial time singular vectors growing between days d and d+2; final time evolved singular vectors growing between day d-2 and d are not used in the current BM-EPS approach. The initial unperturbed analysis is provided by the BMRC Multivariate Statistical Interpolation analysis scheme (Seaman *et al*., 1995); this BM-EPS control analysis has a resolution of T_L119L19 and is interpolated from the operational BoM T_L239L29 GASP (Global Assimilation and Prediction System) assimilation analyses.

The BM-EPS singular vectors are computed at resolution T42L19, also with only simplified physics, namely linearised surface stress, in the tangent forward and adjoint model versions. The BM-EPS singular vector decomposition utilises the ARPACK software (Lehoucq *et al*., 1998); this software uses a variant of the Lanczos eigenvector decomposition algorithm, known as the Implicitly Restarted Lanczos Method. The major differences between the BM-EPS and the EC-EPS are summarized in Table 1. The formulation of the BM-EPS prediction model has been described in Bourke *et al*. (1995); the present model implementation differs primarily with respect to its use of semi-Lagrangian time-stepping and the mass-flux convective parameterization (Tiedtke, 1993).

	BM-EPS	EC-EPS
SV horizontal resolution	T42	T42
SV vertical resolution	L19	L40
SV geographical localization	Only Lat < 20S	Lat < 30S and Lat > 30N
Initial perturbations	Only Initial SVs	Initial and evolved SVs
Model horizontal resolution	T_L119	T_L255
Model vertical resolution	L19	L40
Pressure of top level	≈ 10 hPa	≈ 10 hPa
Model error simulation	No	Stochastic simulation
Ensemble size	32 + 1 (control)	50 + 1 (control)
Perturbation Scaling	Constant Scaling	Analysis Error Scaling

Table 1: List of the principal differences between the BM-EPS and the EC-EPS.

Thus, the d*th* day BM-EPS initial perturbations are defined as:

$$f_j^{BM}(0) = f_j^{BM}(0) + p_j^{BM}$$

$$p_j^{BM} = \sum_k \alpha_{j,k}^{BM} SV_k^{BM}(d, t = 0)$$

where $f_0^{BM}(0)$ is the initial unperturbed analysis provided by the BMRC data assimilation scheme (the T_L119L19 interpolated version of the high resolution operational T_L239L29 BoM analysis). Here the individual perturbations are obtained as linear combinations of the component singular vectors through an orthogonal phase space rotation followed by an amplitude scaling factor. The rotation, following the strategy of the EC-EPS and as described in Molteni *et al*. (1996), is designed to maximise geographical spread of each of the ensemble perturbations; the amplitude scaling is defined as a constant factor chosen such that the BM-EPS spread is similar to the error of the BM-EPS control forecast at t=2 days. With the total energy normalisation of the singular vectors set at unity, the scaling factor was set in the present studies at 0.5.

The forecast valid at time-step *t* for each perturbed member is defined by the numerical integration of the nonlinear perturbed model equations:

$$f_j^{BM}(t) = f_j(0) + \int_0^t [A^{BM}(t) + P^{BM}(t)]dt$$

where $A^{BM}(t)$ and $P^{BM}(t)$ denote the time t tendencies due to the non-parameterised and the parameterised physical processes respectively.

2.3 Definitions: ensemble-mean and ensemble spread

The ensemble mean forecast is defined as the average of all the ensemble forecasts, including the control forecast:

$$em^{XX}(t) = \frac{1}{(N_{XX}+1)} \sum_{j=0,N_{XX}} f_j^{XX}(t)$$

where ^{XX} denotes either BM or EC and N_{XX} is the total number of ensemble members.

$$< sp_{CON}^{XX}(t) >= \sqrt{\frac{1}{N_{XX}} \sum_{j=1,N_{XX}} [f_j^{XX}(t) - f_0^{XX}(t)]^2}$$

$$< sp_{EM}^{XX}(t) >= \sqrt{\frac{1}{N_{XX}} \sum_{j=1,N_{XX}} [f_j^{XX}(t) - em^{XX}(t)]^2}$$

Ensemble spread can be defined as the root mean square average distance of the perturbed members from the control forecast or from the ensemble mean forecast:

2.4 Definitions: accuracy measures

The accuracy of deterministic forecasts such as the control, individually perturbed or the ensemble mean forecasts is assessed by computing the average root mean square error or the average anomaly correlation between the forecast and the verifying analysis inside a specified region. The relationship between the ensemble spread around the control forecast and the skill of the control forecast is investigated, as is the spread around the ensemble-mean and the skill of the ensemble-mean forecast.

The accuracy of probabilistic forecasts of the occurrence or non-occurrence of the event "500hPa geopotential height anomaly larger/smaller than defined thresholds" is assessed using the Brier Score and the Brier Skill Score (Brier, 1950), and the area under the Relative Operating Characteristic (ROC-area; Mason, 1982, Stanski *et al*., 1989, Wilks, 1995). The Brier Score measures the mean square error of probability forecasts for a binary event i.e. it measures the squared difference between a forecast probability of an event p and its occurrence o, expressed as 0 or 1 depending on if the event has occurred or not; the Brier Skill Score compares the Brier Score of one ensemble forecast with the Brier Score of a probability forecast given by a climatological distribution. A perfect forecast has a Brier Skill Score of 1, and a forecast less skillful than climatology has a negative Brier Skill Score. The ROC-area is a measure of the capacity of an ensemble system to discriminate between the occurrence and non-occurrence of an event using hit and false alarm rates. A perfect score would have a ROC-area of 1, while a forecast that cannot discriminate between the occurrence and non-occurrence of an event would have a score of 0.5. The Brier Score and Skill score and the ROC-area are measures of the quality of probabilistic forecasts of single categorical events.

Rank histogram statistics are also analysed and the percentage of cases when the analysis falls outside the range of values spanned by the ensemble members is computed (see Talagrand *et al*., 1998 and Buizza *et al*., 1998).

This set of accuracy measures is being used to provide the initial comparison of the accuracy and skill of the two ensemble systems, and to help to understand the reasons for any detected difference. Results are presented for two regions: a region centered on Australia (latitude from 10S to 45S, longitude from 100E to 160E; hereafter AU) and the Southern Hemisphere annulus (latitude from 20S to 60S; hereafter SH). The verification domain of 20S to 60S, namely for the southern hemisphere annulus, has been chosen to reduce the impact on scores of differences in assimilation and prediction for the Antarctic continental area; satellite radiance data are typically rather more difficult to use satisfactorily in the assimilation systems over the ice covered ocean and over high cold terrain. Additionally the prediction of the radiative balance of the cold surface temperatures and the coupling typically to low level temperature inversion over the high continent in wintertime can be very sensitive to specific details of model parameterizations. EC-EPS forecasts have been verified against ECMWF operational analyses as represented by the $T_L255L40$ interpolated version of the high-resolution $T_L511L60$ ECMWF analyses; i.e., equivalent to the control prediction initial condition resolution; the BM-EPS forecasts have been verified against the BoM operational $T_L239L29$ analyses. Both sets of verifying analyses and model predictions of 500 hPa geopotential height, used in this study, have been specified on a regular latitude-longitude grid with 2.5 degree spacing.

3. EC-EPS and BM-EPS initial perturbations

The three fastest growing initial time singular vectors, represented by the stream function shown for the BM-EPS (at sigma level (p/p_surface) = 0.5) and EC-EPS (at sigma level (p/p_surface) ≅ 0.5), are shown in Figures 1(a) to 1(c) and in Figure 1(e) to 1(g) respectively, for the initial date-time of 0000UTC, 15 July 2001; here, although the respective system singular vectors have differing normalisations, the contour count is set to be equivalent. These first three leading singular vectors of each system are seen to have similar patterns, with variance maxima similarly located. It is not expected that the individual singular vectors from each system will have an exact one-to-one correspondence in structure or growth rates. A detailed intercomparison of the equivalence of the independently generated singular vectors has been previously made (A. Doerenbecher, 1998, personal communication); a close correspondence in terms of growth rates and structures was found between the separate EC-EPS and BM-EPS singular vectors, when the vertical level resolution and location (particularly of the lowest model levels) were made close to equivalent in the two systems.

The relationship of these initial time singular vectors to the overall baroclinic instability of the flow is indicated by an evaluation of the Eady Index, which is an estimate of the growth rate of the most unstable Eady Mode (see Hoskins and Valdes, 1990), given by $\sigma \equiv 0.31\frac{f}{N}\frac{\partial u}{\partial z}$ where the static stability, N, and the wind shear, $\frac{\partial u}{\partial z}$ have been estimated from the 300 and 850 hPa potential temperature and wind field data. The Eady Index is shown for the BM-EPS (Figure 1(d)) and for the EC-EPS initial conditions (Figure 1(h)). The geographical location of the first three leading singular vectors is seen to coincide quite well with the high baroclinicity (Eady Index values) in the southern Indian and Atlantic oceans for both sets of singular vectors, although the correspondence is limited and does not extend to the high baroclinicity seen in both

Eady plots over southern South America and the high latitudes of the south-east Pacific. (see below for further comment).

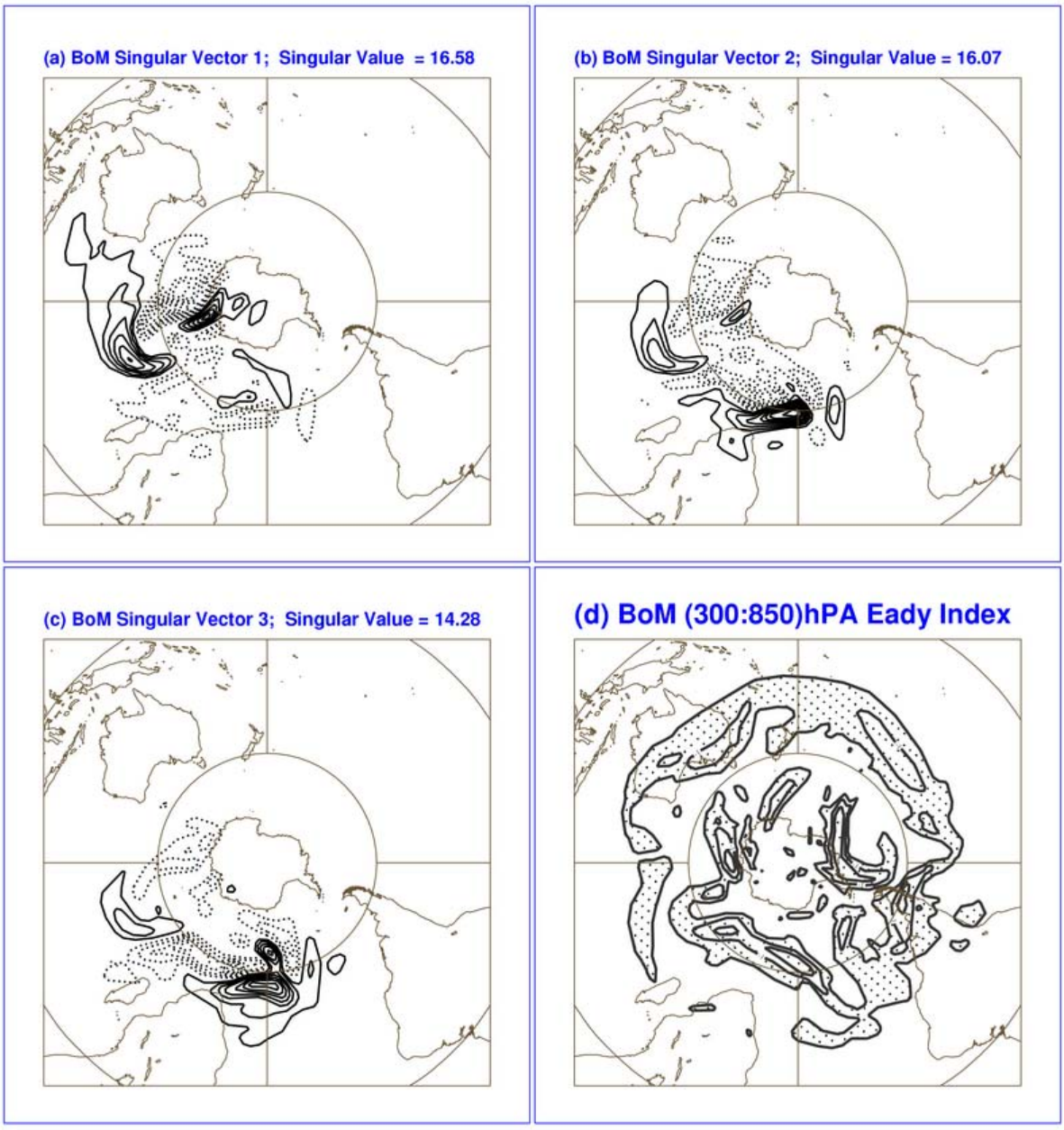

Figure 1. (a-c): Three fastest growing singular vectors for 0000UTC, 15 July 2001 for the BM-EPS, shown in terms of stream function at sigma level (p/p*) = 0.5. Fig.1(d) Eady Index values for 0000UTC, 15 July 2001 for the BM-EPS.
Solid lines indicate positive values; dash lines indicate negative values; the contour count is set to be the same for all singular vector plots.

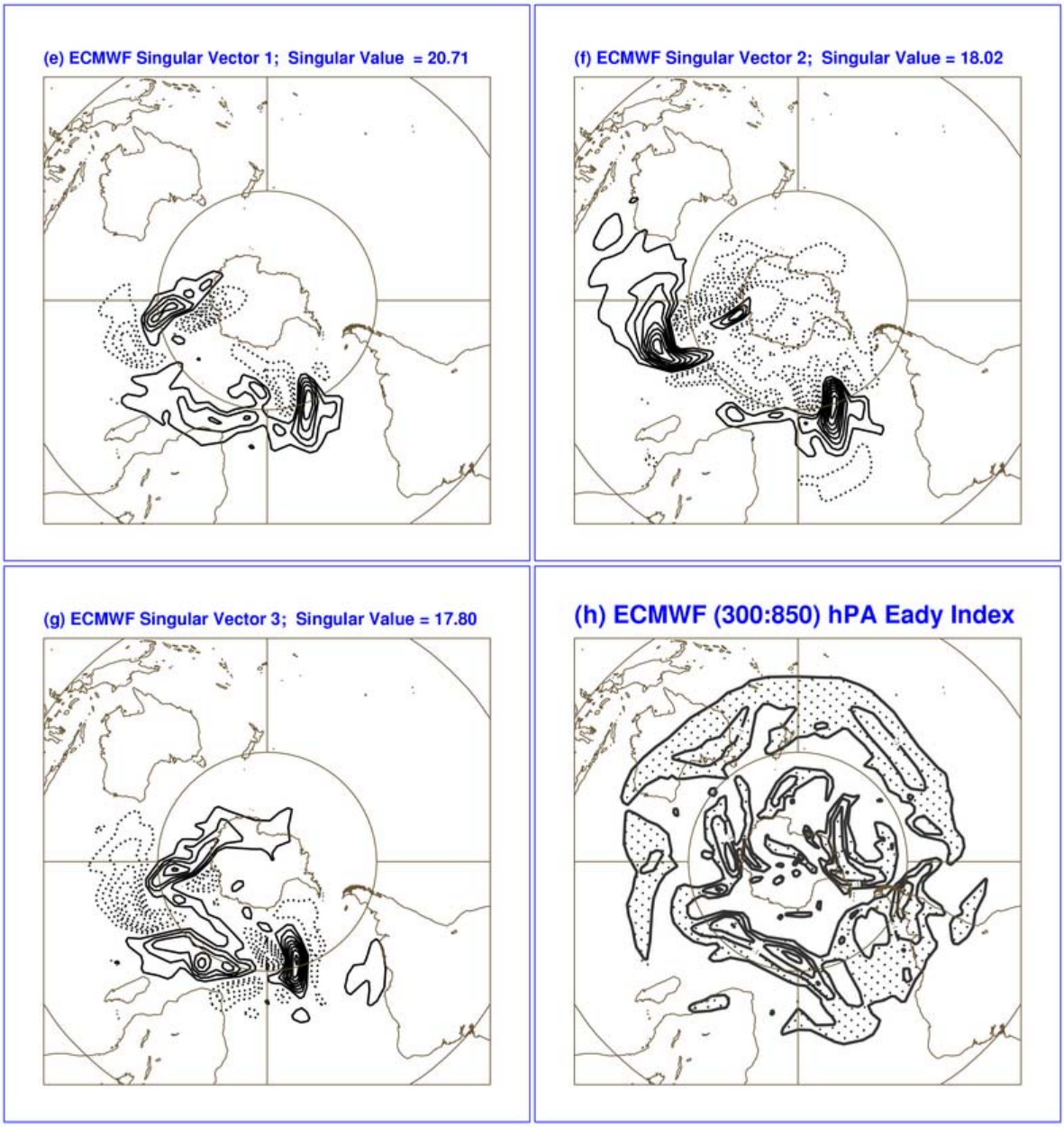

Figure 1(e-g): As in (a-c) but for EC-EPS with sigma level (p/p*) approximately = 0.5. Fig. 1(h): As in Fig 1(d) but for EC-EPS.

Figure 2 shows four BM-EPS (a-d) and four EC-EPS (e-h) initial perturbations of 500 hPa geopotential height at 0000UTC, 15 July 2001; these panels have all been plotted with the same contour interval of 3 m. The comparison of the two sets of figures indicates that the BM-EPS initial perturbations are smaller in both amplitude and scale, with EC-EPS initial perturbations showing local maxima approximately a factor 3 larger (about 50m compared to about 17m) than in the BM-EPS case. The difference in local amplitude is due to three main factors: (a) BM-EPS uses only initial singular vectors while EC-EPS uses both initial and evolved singular vectors; (b) BM-EPS and EC-EPS initial perturbations are scaled differently, with EC-EPS using an analysis error estimate to set the perturbation amplitudes while BM-EPS uses a fixed factor as

indicated in Section 2.2; (c) BM-EPS initial perturbations are defined using 16 singular vectors while the EC-EPS initial perturbations are defined using 25 singular vectors. The BM-EPS initial perturbations have smaller horizontal scales because they are generated using only initial time singular vectors, which are known to have a much smaller scale than final time singular vectors (Buizza and Palmer, 1995). It is useful to note that the perturbations generated from linear combination of the 16 initial- time BM-EPS singular vectors, namely perturbations 1, 3, 5 and 7 and shown in Figures 2(a) to (d), do show structure in the high latitudes of the South American region matching approximately areas of instability suggested by the Eady index of Figure 1(d), and which was not seen in the first three leading singular vectors of Figures 1(a) to (c); (perturbations 2, 4, 6 and 8 match 1, 3, 5 and 7 pairwise but are of opposite sign).

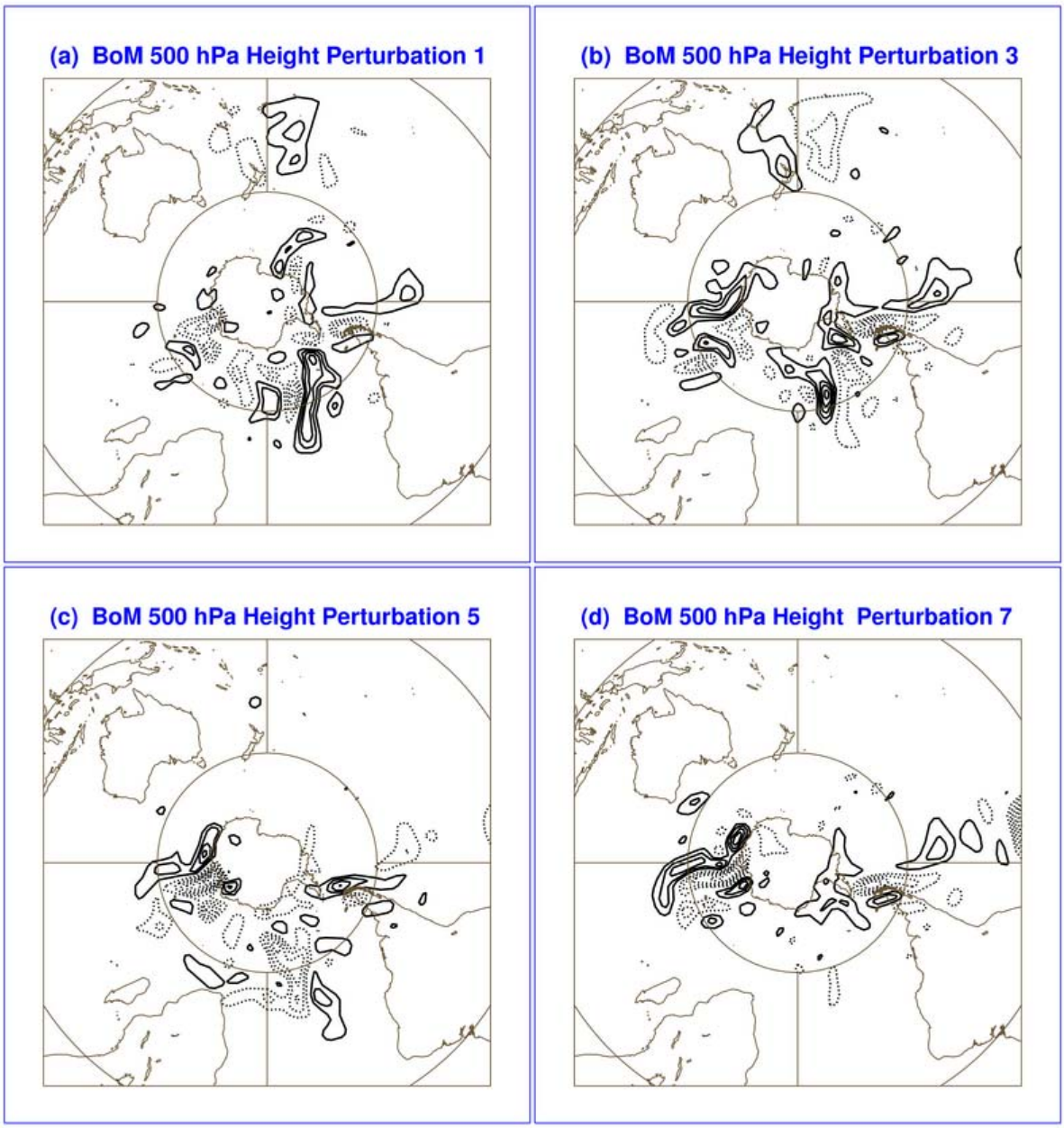

Figure 2(a-d): Four initial perturbations of 500 hPa geopotential height for the BM-EPS for 0000UTC, 15 July 2001: contour interval of 3 m.

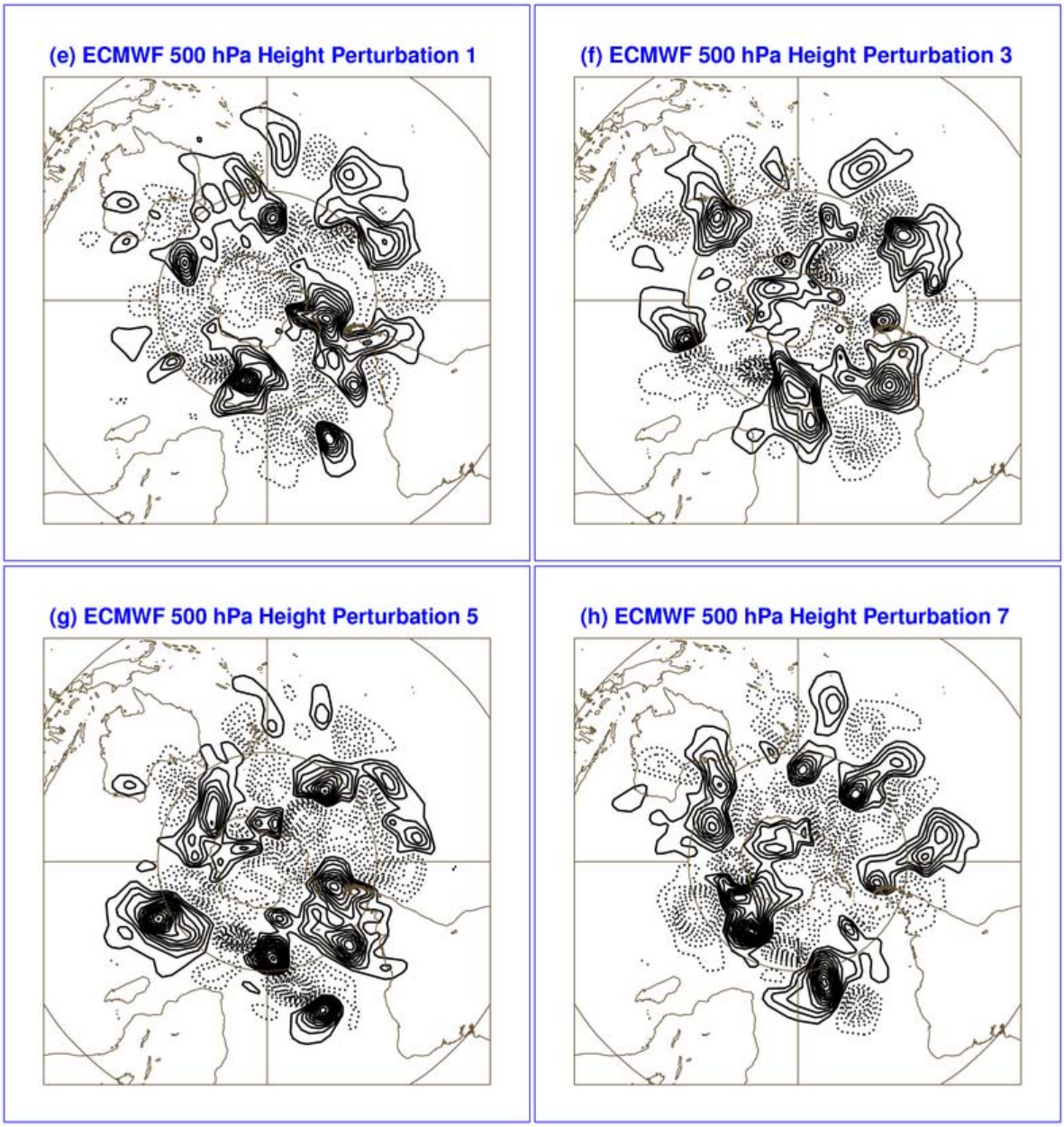

Figure 2 (e-h): As in Fig.2(a-d), but for EC-EPS.

4. BM-EPS and EC-EPS ensemble spread

Figure 3 shows the 152 case probability density function of ensemble spread around the control forecast for the BM-EPS (solid) and the EC-EPS (dash) for the southern hemisphere (lower panels), for the Australian region (upper panels), and at forecast day 3 (left hand panels), day 5 (middle panels) and day 7 (right hand panels). Here each system's spread is calculated in terms of root mean square distance relative to its own control over the southern hemisphere (SH) and over the Australian region (AU); the percentage of cases with ensemble spread inside a specific interval, with intervals defined as multiples of 7m, are represented. The most notable difference between the two systems is the enhanced spread of the EC-EPS system particularly at day 7; at earlier times and particularly for the AU the spread of the BM-EPS is slightly skewed to lower

spread. The results show that the divergence between the members of each ensemble grows at different rates. Figure 4 shows the time evolution of the 152 case ensemble spread for SH and AU (for each system relative to its own control); i.e, the time evolution of the magnitude of the mean spread of the pdf of the ensembles. The differing amplitude of the initial perturbations is evident in Figure 4 at time t=0; the BM-EPS growth rate during the first 2-3 days is greater than that of the EC-EPS. This is to be expected given that the BM-EPS perturbations are defined in terms of initial singular vectors, whereas the EC-EPS perturbations are comprised equally of fast growing initial and more slowly growing evolved singular vectors. The BM-EPS average spread, which is initially about a factor of 3 smaller than the EC-EPS average spread for both SH and AU, is comparable to EC-EPS spread from around day 1.5 to day 4. Subsequently the EC-EPS spread grows at a faster rate than the BM-EPS spread for both regions.

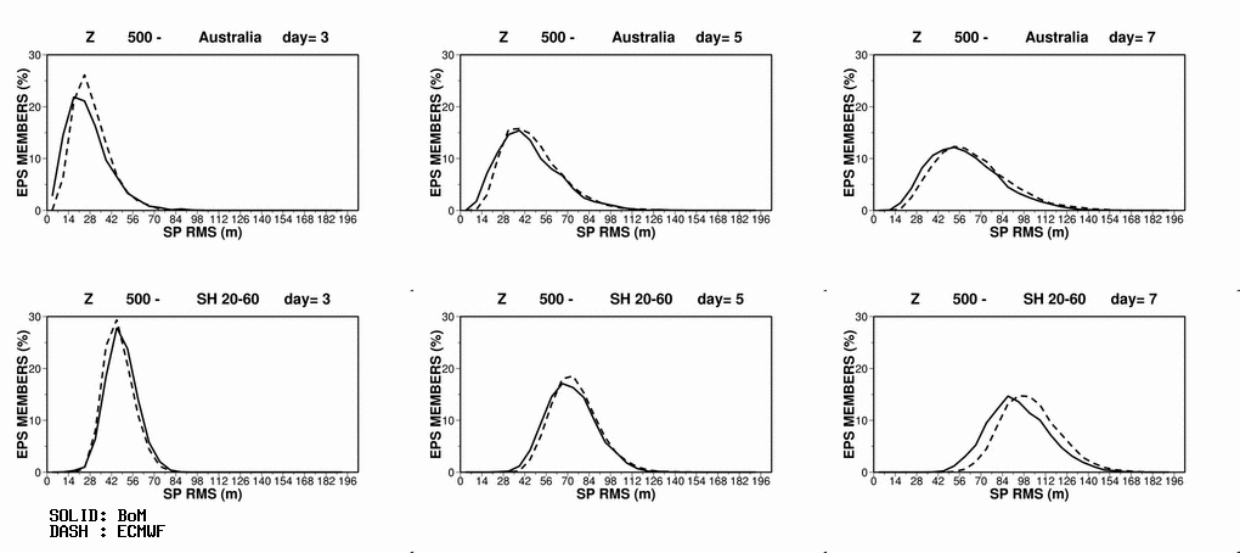

Figure 3 Probability density function of ensemble spread around the control forecast for the BM-EPS (solid) and EC-EPS (dash), for the Australian Region (upper panels) and, for the Southern Hemisphere (lower panels), at forecast day 3 (left panels), day 5 (middle panels) and day 7 (right panels). The probability density function has been constructed using 500 hPa geopotential height predictions for 152 daily cases.

Apart from the fact that the initial perturbations are different, other possible explanations for the different growth rates are that the EC-EPS ensemble members are run at higher resolution, and that the resulting spread benefits from the fact that stochastic model perturbations act as a further source of divergence. Buizza *et al.*(1999) indicated that the impact of stochastic model perturbations on the ensemble spread is rather small, with stochastic physics increasing the ensemble spread by about 3 to 5%. By contrast, Buizza *et al.*(1998) documented an ensemble spread increase when the EC-EPS changed resolution from T63 to T_L159 and from 19 to 31 vertical levels, especially in the 5 to 10 day forecast range. These previous results would suggest that the different growth in ensemble spread in the 5 to 10 day range could be due to the different resolution in the two systems. Another small difference in the present study affecting EC-EPS spread in the southern hemisphere is the additional northern hemisphere perturbations of the EC-EPS; in the medium range and beyond there is inter-hemispheric propagation of perturbations.

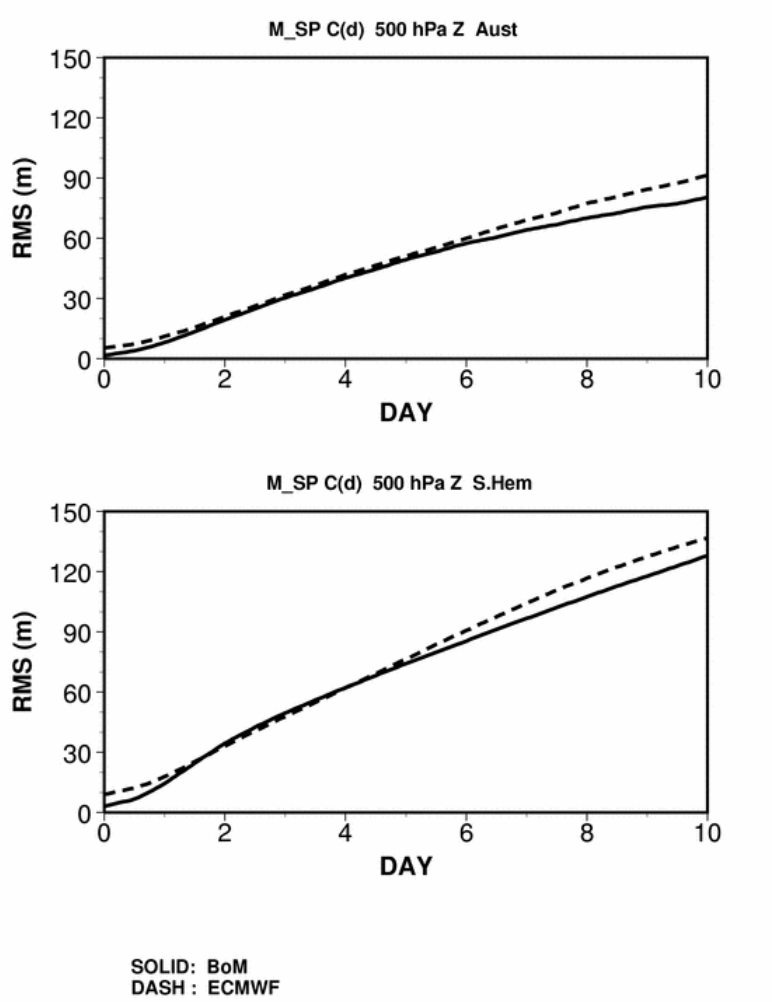

Figure 4: Root mean square average ensemble spread around the control forecast of the BM-EPS (solid) and EC-EPS (dash) for the Australian Region (upper panel), and Southern Hemisphere region (lower panel). [The RMS has been computed over all grid points in each region and over all ensemble members].

The faster growth rate of the BM-EPS for the SH relative to AU in the first few days is consistent with the geographical distribution of the initial singular vectors, which predominantly have variance maxima at latitudes poleward of Australia.

Similar conclusions could have been drawn by considering anomaly correlation instead of root mean square error as a measure of spread (not shown), or by considering the ensemble spread around the ensemble mean (i.e., by considering the ensemble standard deviation, not shown).

5. BM-EPS and EC-EPS Skill

Figure 5 shows the 152 case probability density function of root mean square error of BM-EPS (solid) and the EC-EPS (dash), for the Australian region (upper panels) and for the southern hemisphere (lower panels), at forecast day 3 (left hand panels), day 5 (middle panels) and day 7 (right hand panels). Here each system's root mean square error is calculated in terms of root mean square distance relative to its own analysis over the southern hemisphere (SH) and over the Australian region (AU); the percentages of cases with ensemble root mean square error inside specific intervals, with intervals defined in terms of multiples of 7m, are represented. Results show that for all forecast times the EC-EPS distribution curve is to the left of the BM-

EPS curve, indicating the error of the EC-EPS forecasts is smaller, although this differentiation is less marked for AU.

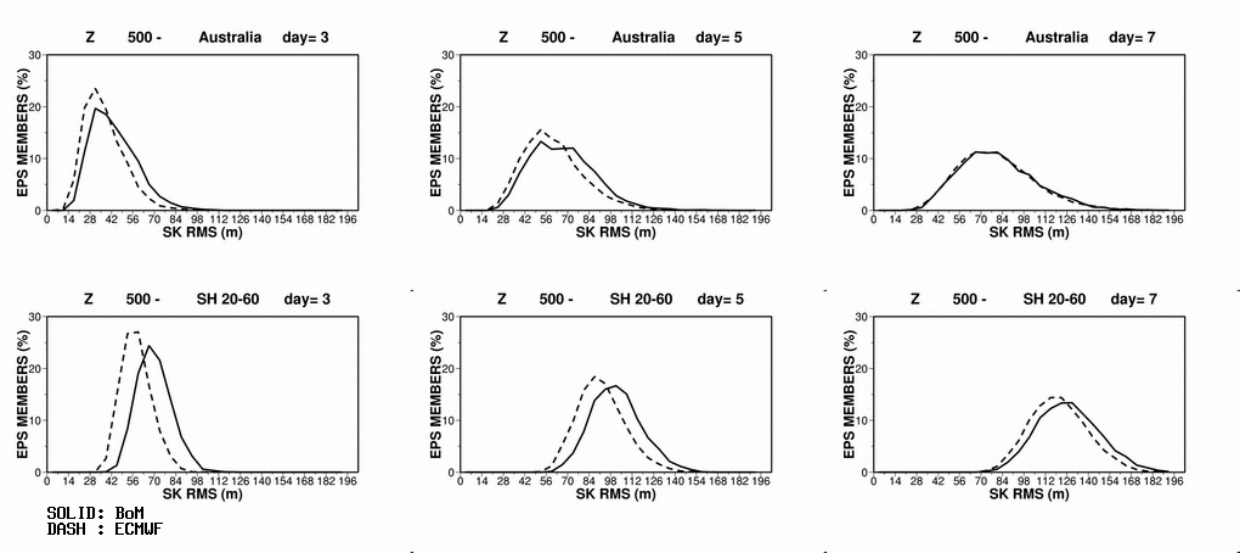

Figure 5: Probability density function of ensemble skill for the BM-EPS (solid) and the EC-EPS (dash), for the Australian Region (upper panels) and, Southern Hemisphere (lower panels), at forecast day 3 (left panels), day 5 (middle panels) and day 7 (right panels). The probability density function has been constructed using 500hPa geopotential height values for 152 cases.

The time evolution of the 152 case root mean square error of the control (Figure 6(a)) and of the root mean square error of the ensemble-mean (Figure 6(b)), of each system, is shown for all forecast times for both AU (upper panels) and SH (lower panels), illustrating further the differences seen in Figure 5. The overall superior skill of the EC-EPS, by a margin of approximately 1 day, is evident for both the control and ensemble mean forecasts. The reduction in 10 day error provided by the ensemble mean is seen in Figures 6(a) and 6(b) to be of the order of 25%.

The time evolution of the mean forecast skill of the ensemble (Figure 6c) is shown for all forecast times for both AU (upper panels) and SH (lower panels) for each system; here the root mean square error of each individual forecast has been averaged over all ensemble members and all 152 cases. The mean errors of the two ensembles (Figure 6(c)) are in closer agreement than the other measures shown in Figures 6 (a) and (b), particularly for AU. This shows that the probability distribution of errors can be more similar than is suggested by the single measures of control or ensemble mean error.

Figure 6(d) shows, for the AU and the SH regions, the percentage of cases from each ensemble where the root mean square prediction error is less than the root mean square error of their respective control predictions; it is seen that for both regions the BM-EPS has a higher percentage of members which perform better than the control. Paradoxically, it is also seen in Figure 6(d) that at time zero there are a number of members in the BM-EPS with lower error than the control; at time zero these errors are of the order of a few metres and very small relative to the magnitude of prediction error at 24 hours and beyond. They are

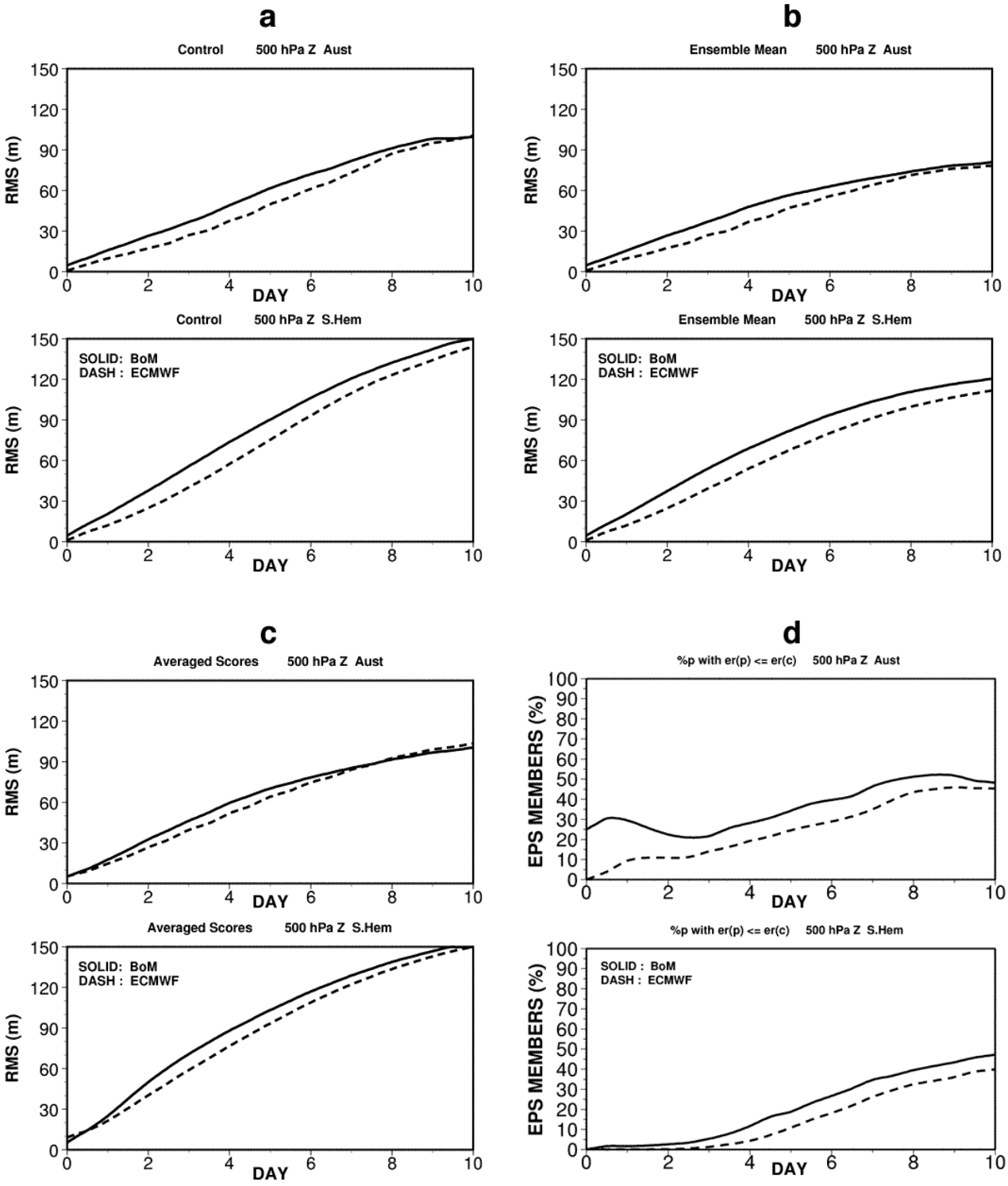

Figure 6 (a) 152 case root mean square error of the control forecast of the BM-EPS (solid) and EC-EPS (dash) over the Australian Region (upper panel) and Southern Hemisphere (lower panel). (b): as (a), but for the ensemble-mean forecast. (c) the root mean square error of each individual forecast averaged over all ensemble members and all 152 cases. (d) as in (a) but the percentage of all members over the 152 cases where the root mean square prediction error is less than root mean square error of the control prediction. Forecast errors have been computed using 500hPa geopotential height data.

attributable to the use of a post-processed $T_L239L29$ operational analyses to verify the BM-EPS initial conditions represented at $T_L119L19$. For the verification of the EC-EPS fields we have used the ECMWF control at t=0 as the verifying analyses, rather than the directly post-processed high resolution $T_L511L60$ ECMWF analyses.

The implication of Figure 6(d) may be that within the BM-EPS system, where the underlying control performance is clearly less satisfactory than that of the EC-EPS, there is clearly scope to enhance prediction performance, and that a relatively greater percentage of perturbations do indeed do this. A further comment on the results seen in Figure 6(d) is that the BM-EPS simpler perturbation strategy may be contributing to these results, although the scope of the present studies does not permit any further clarification of this aspect.

Figure 7 shows the time series of the root mean square error of 5 day forecasts for the SH, of the control and ensemble mean forecasts and the best and worst ensemble members of the BM-EPS (upper panel) and EC-EPS (lower panel); these time series are of the 5 day running mean scores. Additionally the time mean of the 152 case root mean square of the errors shown in Figure 7 are presented in Table 2 together with the standard deviations (in brackets). The results again reflect the fact that the EC-EPS forecasts are in general characterised by smaller error; there is a larger absolute difference between the BM-EPS best member and the BM-EPS control than seen between the corresponding EC-EPS best member and its control, although the reduction in absolute error for the best member is approximately 10% in both the EC-EPS and BM-EPS systems. The equivalent time series for AU are shown in Figure 8; the 152 case time mean of these quantities shown in Figure 8 are also presented in Table 2. Here it is seen for both BM-EPS and EC-EPS systems that the gain of 33% in absolute performance, in the 152 case mean, of the best ensemble member relative to the control for AU, is markedly better than seen for SH.

	SH region		AU region	
	BM-EPS	EC-EPS	BM-EPS	EC-EPS
Control	89.2 (13.9)	74.3 (12.8)	59.3 (17.0)	47.8 (14.4)
Ensemble Mean	81.1 (11.9)	66.9 (9.9)	54.6 (14.8)	45.6 (12.3)
Best Member	77.9 (9.9)	66.8 (7.7)	40.1 (10.6)	31.7 (6.9)
Worst member	127.8 (14.8)	122.1 (14.8)	100.1 (22.0)	101.7 (21.5)

Table 2: Root mean square and standard deviation (in brackets) of 152 cases of day 5 root mean square forecast error for the control, ensemble mean, best ensemble member and worst ensemble member for the BM-EPS and EC-EPS systems.

It is also notable that there is more variability in performance in the BM-EPS system in comparison to the EC-EPS system in both the SH and AU regions as evident in Figures 7 and 8, and as quantified in the standard deviations presented in Table 2. The EC-EPS system is providing a more consistent level of performance, which is interpreted as being primarily associated with the EC-EPS having a more consistent quality in the accuracy of its initial conditions.

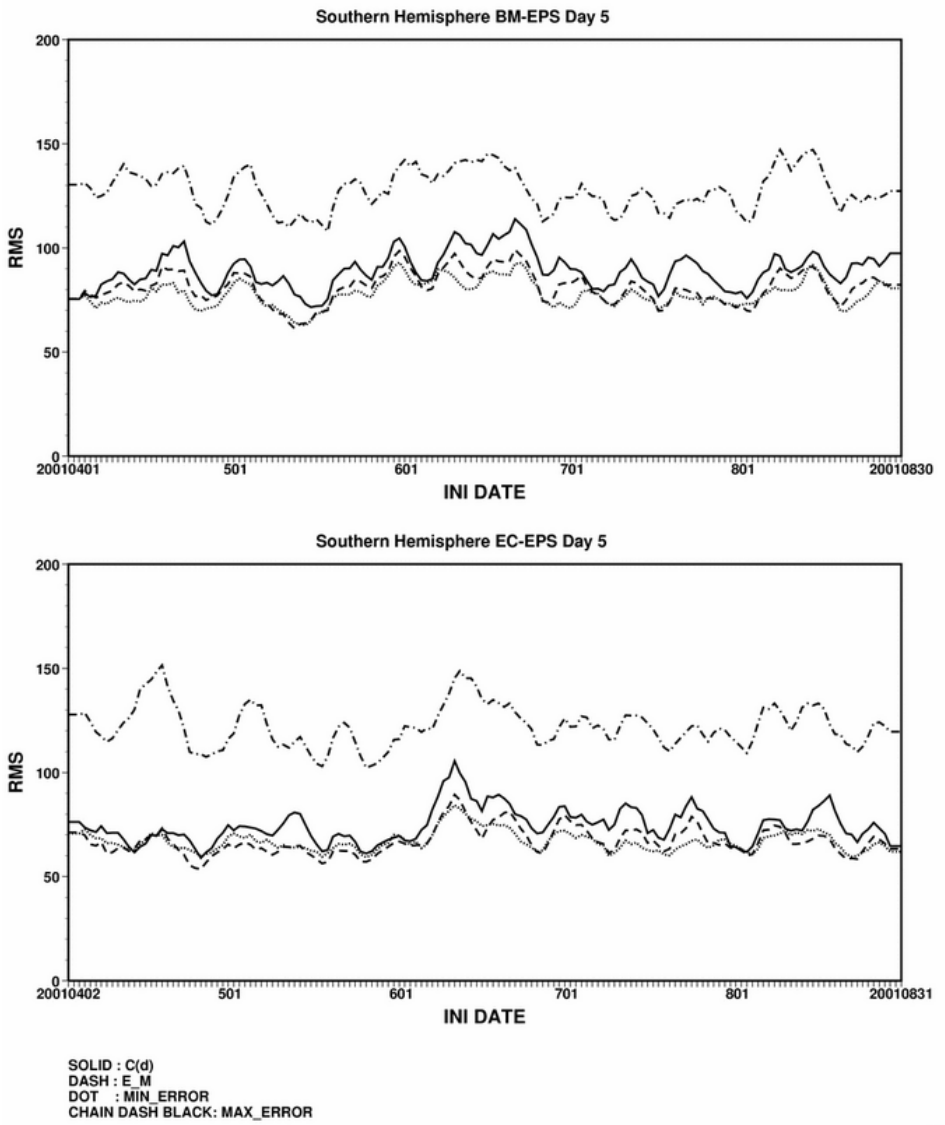

Figure 7: Time series of day-5 root mean square error of the control (solid), ensemble mean (dash), best perturbed member (dotted) and the worst perturbed member (chain-dash), for the BM-EPS (upper) and EC-EPS (lower). Values plotted are 5 day running mean errors and refer to 500 hPa geopotential height forecasts over SH.

A possible explanation of the larger gap in absolute terms between the best member and the control forecast in the BM-EPS compared with the EC-EPS for the SH is the following. There is some evidence that one reason for the poorer skill of the BM-EPS control forecast compared to the EC-EPS control forecast is the fact that the BoM analysis is less accurate; a number of unpublished experiments involving the transplant of ECMWF initial conditions into the BoM prediction system, at $T_L239L29$ resolution, have yielded significant positive impact on prediction performance in the southern hemisphere. Assuming then that the BoM analysis is less accurate, it appears plausible for the potential error reduction induced by the initial perturbations in the BM-EPS to be larger.

It is interesting to compare the difference of the magnitude of the ensemble mean forecast error and the control forecast error. Figure 9 shows for the two systems and for each forecast day the average over all cases of the difference between the root mean square error of the ensemble-mean and the root mean square

error of the control for SH (lower panel) and for AU (upper panel). The ensemble-mean forecast is a smoother field than the control forecast, and it could be expected to have a smaller root mean square error than the control forecast. The difference between the error of the ensemble-mean and the error of the control forecast is a measure of the effectiveness of the filtering induced by the ensemble averaging in defining a single deterministic forecast (namely the ensemble mean) with an error smaller than the control forecast. The error reduction of the ensemble-mean is seen to begin earlier and to be greater for the SH domain than for the AU domain. For the SH, the two curves are very similar, while for AU slightly larger negative values are obtained for the BM-EPS. This indicates that the effect of the filtering of small unpredictable scales by taking the mean of the ensemble is positive for both systems over both regions with the beneficial effect over the AU domain being slightly larger for the BM_EPS.

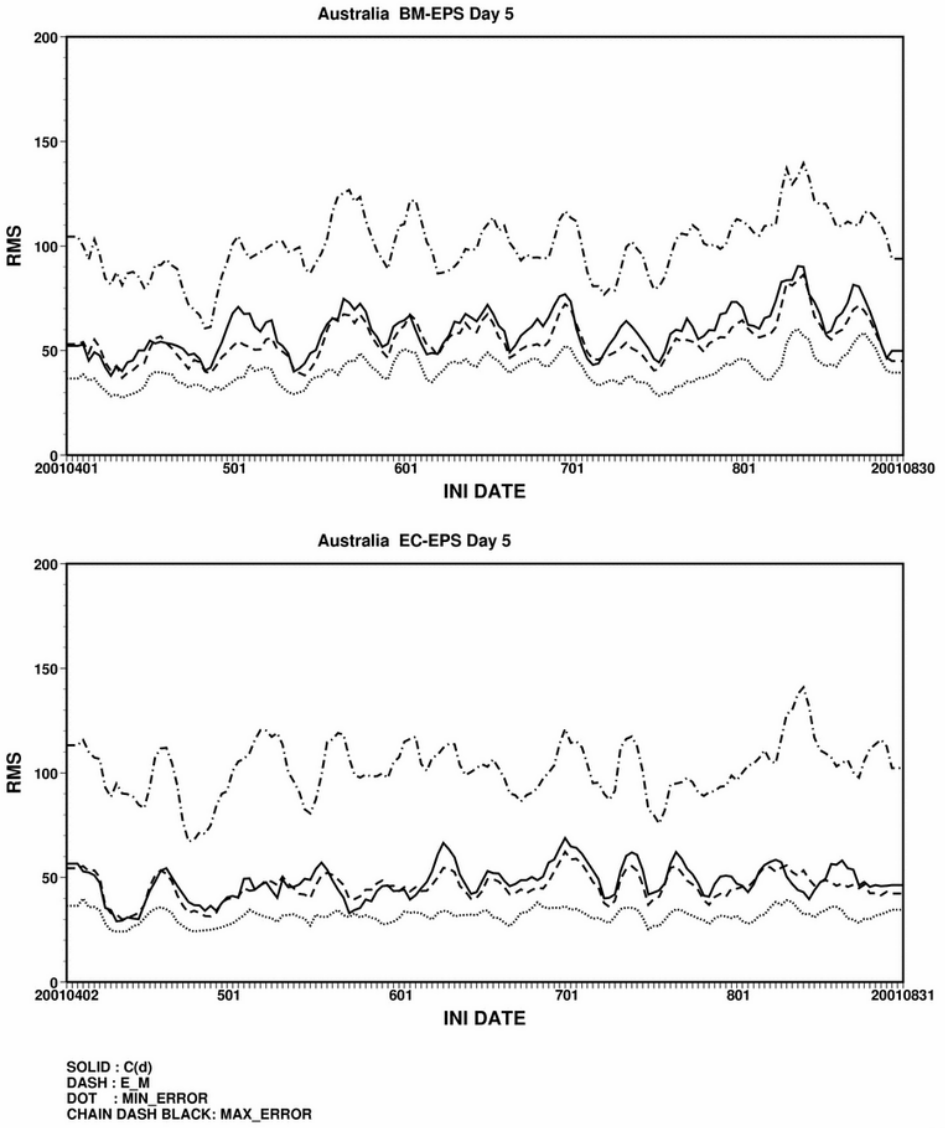

Figure 8: As in Figure 7, but for AU.

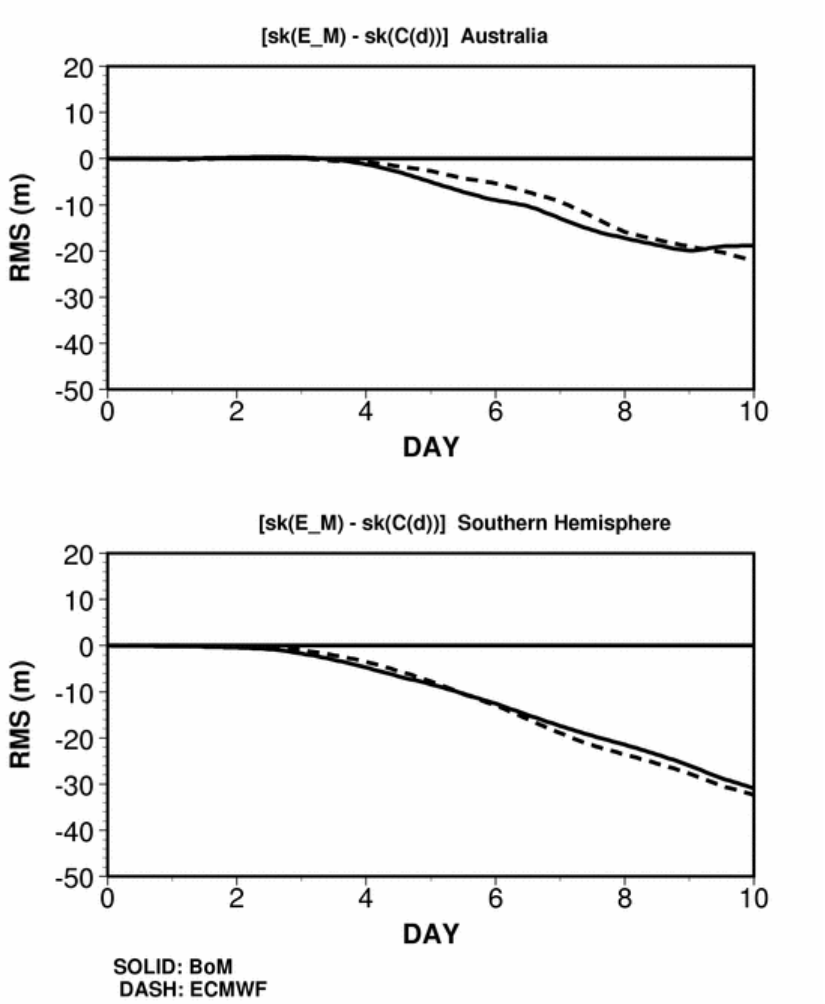

Figure 9: Average of the difference between the root mean square error of the ensemble-mean and the root mean square error of the control forecast of the BM-EPS (solid) and EC-EPS (dash) for AU (upper) and SH (lower). A positive/negative value indicates that the ensemble-mean error is larger/smaller than the control error. The average has been computed using 500hPa geopotential height forecast data over 152 cases.

6. Spread/skill relationships

In earlier studies of a perfect ensemble (Buizza, 1997), the link between ensemble spread and control forecast skill was examined with respect to the following relationships:

- ensemble spread distribution is similar to control error distribution;
- smaller ensemble spread corresponds to better control prediction skill; and
- the verifying analysis is within the range spanned by the ensemble forecasts.

These spread skill relationships have been examined for the EC-EPS and BM-EPS systems in the southern hemisphere in this present study.

Figure 10 shows, for both AU (upper panel) and SH (lower panel), the difference between the ensemble spread of 500 hPa geopotential height prediction around the control forecast and the control error for the two systems. These results indicate that for both the SH and AU region the BM-EPS spread is smaller than the

average error of the control forecast, while the EC-EPS spread is larger than the control error up to forecast day 5 in both regions and then becomes slightly smaller than the control error. (The negative deficit for BM-EPS at time zero is spurious and arises because the verifying analyses are the operational T_L239L29 analyses post-processed to the verifying grid, whereas the control model fields are post-processed from the T_L119L29 resolution. For the verification of the EC-EPS fields we have used the ECMWF control at t=0 as the verifying analyses.)

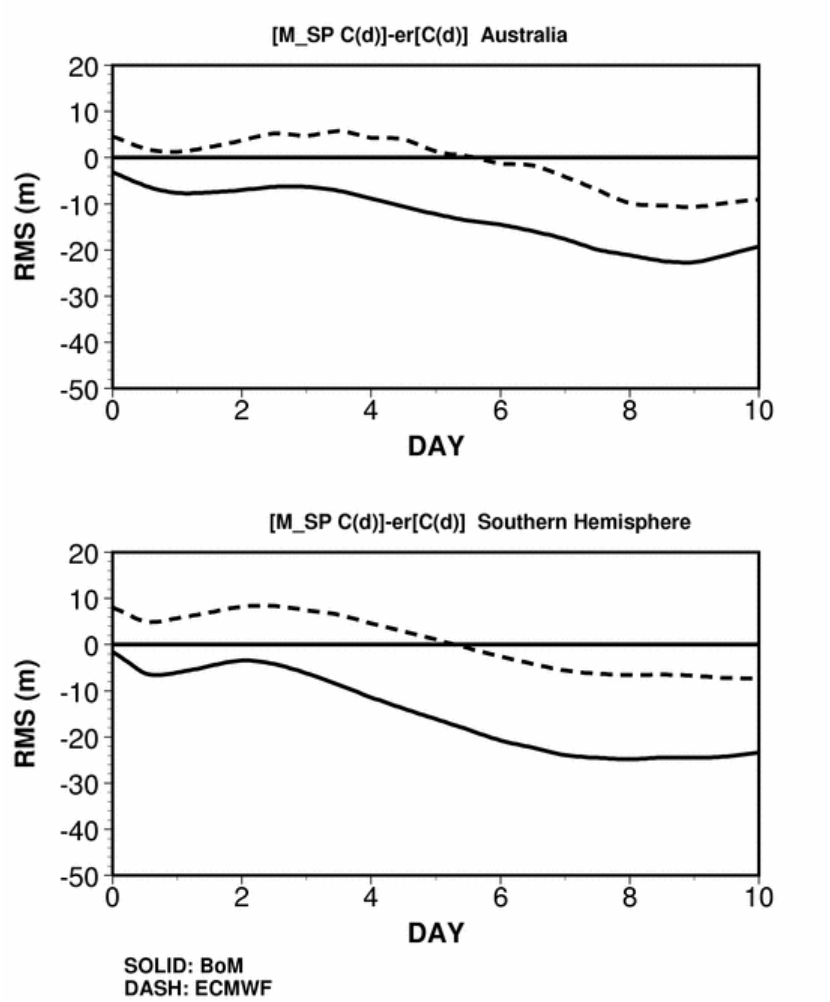

Figure 10: Difference between the average ensemble spread around the control forecast and the error of the control forecast of the BM-EPS (solid) and EC-EPS (dash) for AU (upper) and for SH (lower). A positive/negative value indicates that the ensemble spread is larger/smaller than the control error. The average has been computed using 500hPa geopotential height forecasts for 152 cases.

Table 3(a) summarises the percentage of cases with ensemble spread smaller than control forecast error. An ensemble system with spread comparable to control error on average would have 50% cases with spread lower than the control error. Thus, the ensemble system with the percentage of cases closer to 50% would perform better according to this measure. Table 3(a) indicates that with respect to the equivalence of spread and control error, the EC-EPS system is superior, with the BM-EPS system displaying ensemble spread significantly less than the control error.

Table 3(b) summarises the correlation between the root mean square ensemble spread relative to the control and the control forecast skill. A higher correlation value represents a stronger spread-skill relationship which

is interpreted as better according to this measure. In Table 3(b), the EC-EPS system is superior at later time ranges for SH, whereas the BM-EPS has better performance on this measure for AU.

	SH region		AU region	
	BM-EPS	EC-EPS	BM-EPS	EC-EPS
Fc d+3	78%	16%	70%	25%
Fc d+5	87%	43%	72%	41%
Fc d+7	91%	58%	72%	55%

Table 3(a). Percentage of cases with root mean square spread relative to the control smaller than root mean square error of the control forecast at forecast days 3, 5 and 7 over the SH and AU for the BM-EPS and EC-EPS. Values refer to 500 hPa geopotential height field forecasts.

	SH region		AU region	
	BM-EPS	EC-EPS	BM-EPS	EC-EPS
Fc d+3	25%	22%	26%	22%
Fc d+5	27%	39%	45%	14%
Fc d+7	30%	36%	38%	35%

Table 3(b). Correlation of root mean square spread relative to the control with root mean square error of the control forecast at forecast days 3, 5 and 7 over the SH and AU for the BM-EPS and EC-EPS. Values refer to 500 hPa geopotential height field forecasts.

7. Skill of ensemble probabilistic forecasts of geopotential height anomalies

BM-EPS and EC-EPS ensemble forecasts have been used to generate probability forecasts of events defined in terms of 500 hPa geopotential height anomalies with respect to the monthly mean climate:

$$prob[(f - c) > thr\ ]$$
$$prob[(f - c) < -thr\ ]$$

with threshold thr = 50m.

Figures 11(a) and 11(b) respectively show the area under the Relative Operating Characteristic (ROC-area) curve and the Brier skill score for both the BM-EPS (solid) and EC-EPS (dash) probability forecasts of two events over SH namely: “positive 500 hPa geopotential height anomaly with magnitude larger than 50 m” and “negative 500 hPa geopotential height anomaly with magnitude larger than 50 m”. Details of the evaluation of these scores are given in Buizza and Palmer (1998). Results indicate that both ensemble systems are skillful for the whole 10 day forecast range (positive Brier skill scores and ROC-area larger than 0.5). Comparing one to the other, the EC-EPS shows higher skill scores. Generally speaking, the EC-EPS forecasts valid for day d have skill scores similar to the BM-EPS forecasts valid for day (d-x), with x ranging between 1.5 and 2.0 days.

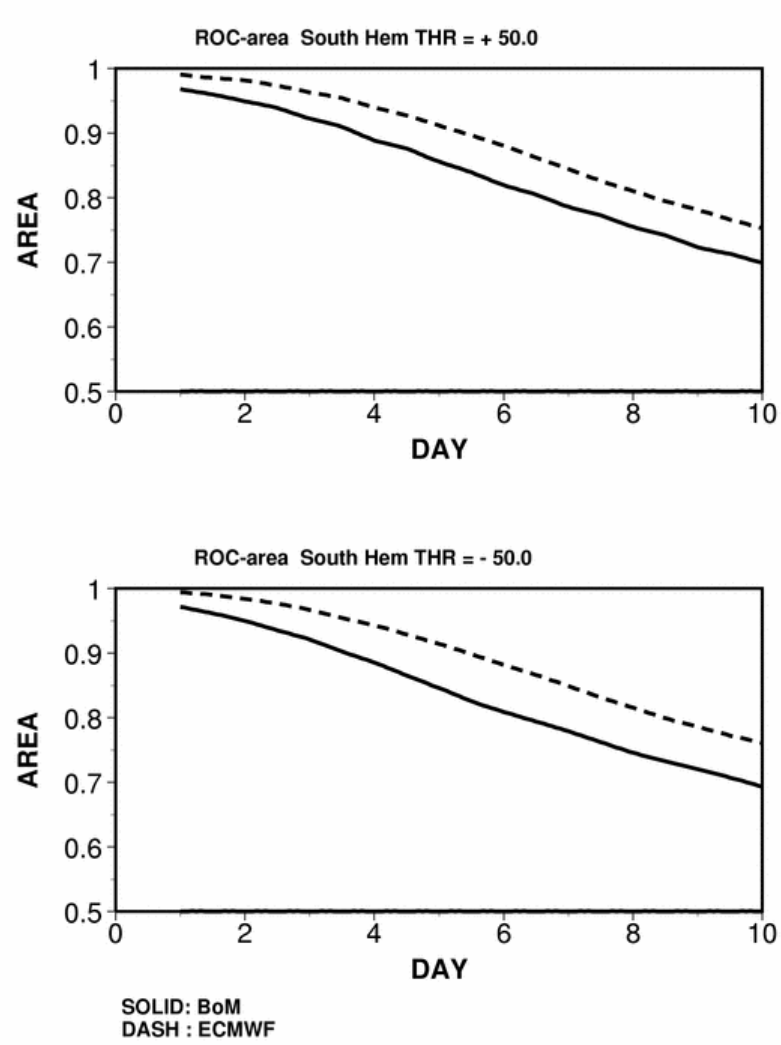

Figure 11(a): ROC-area of the BM-EPS (solid) and EC-EPS (dash) probability forecasts for SH for the event "500 hPa geopotential height positive anomaly with magnitude larger than 50m" (upper panel), and for the event "500 hPa geopotential height negative anomaly with magnitude larger than 50m" (lower panel).

Figures 12(a) and 12(b) are analogous to Figures 11(a) and 11(b) but for AU. Results indicate again that both ensemble systems are skillful for the whole 10 day forecast range, and that the EC-EPS is more skillful than the BM-EPS with x (as defined above) also ranging between 1.5 and 2.0 days. Similar conclusions can be drawn using either the Relative Operating Characteristic (ROC-area) curve or the Brier skill score; both are used here since they measure different ensemble characteristics (e.g., when BSS<0 it is still possible for the ROC area to be greater than >0.5, indicating that the system can still discriminate between the occurrence and non-occurrence of an event.).

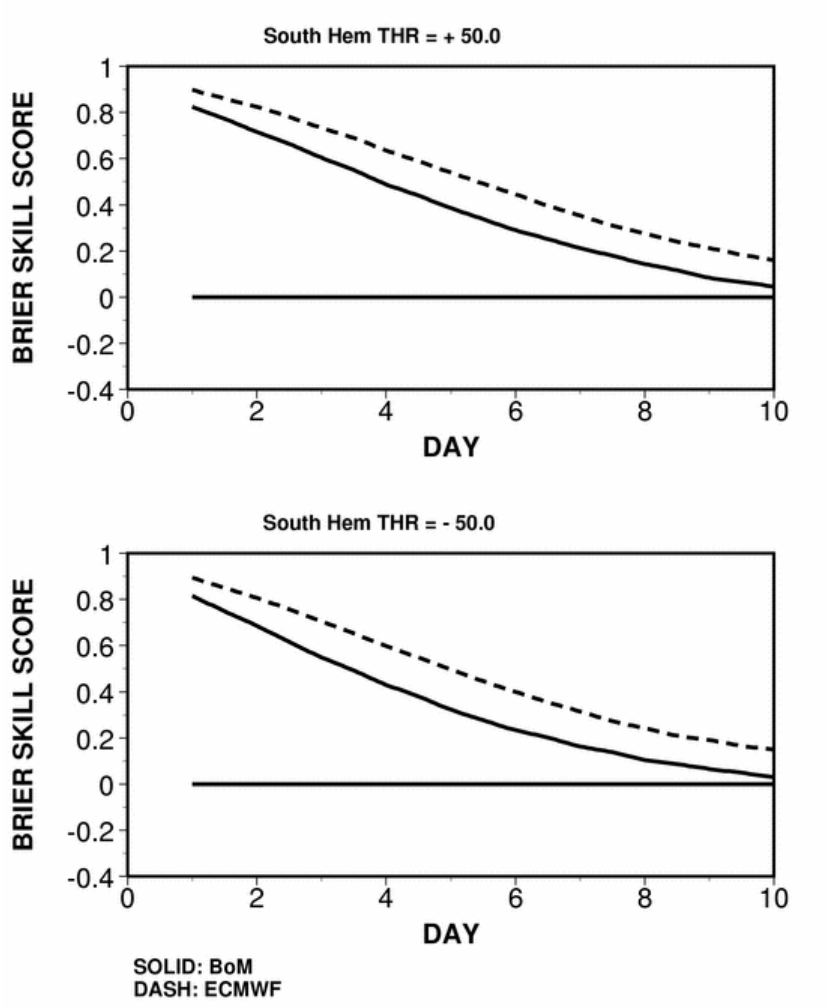

Figure 11(b): As in 11(a), but for Brier skill score instead of the ROC-area.

As seen in the control and ensemble root mean square error of Figures 6(a) and 6(b), the deterministic prediction skill gain of the EC-EPS over the BM-EPS system is of the order of 1 day; however as seen in Figures 11 and 12 in terms of probabilistic verification, the gain of the EC-EPS system is seen to be from 1.5 to 2.0 days. This outcome is consistent with results reported by Buizza *et al*. (2003b), who indicate that when the EC-EPS resolution was increased from T_L159 to TL255 (with 50 members in each ensemble) a larger positive impact was detected on the probabilistic products than on the skill of the EPS control (c.f. resolutions in the present studies of T_L119 and T_L255 for BM-EPS and EC-EPS respectively); these authors state that one of the most striking results from their assessment of increased resolution is that the accuracy of the probabilistic forecasts has been improved more than the accuracy of the deterministic forecasts. Buizza *et al*. (1998) had also found that increased ensemble size (at resolutions of T106) contributes to increased performance of probabilistic scores.

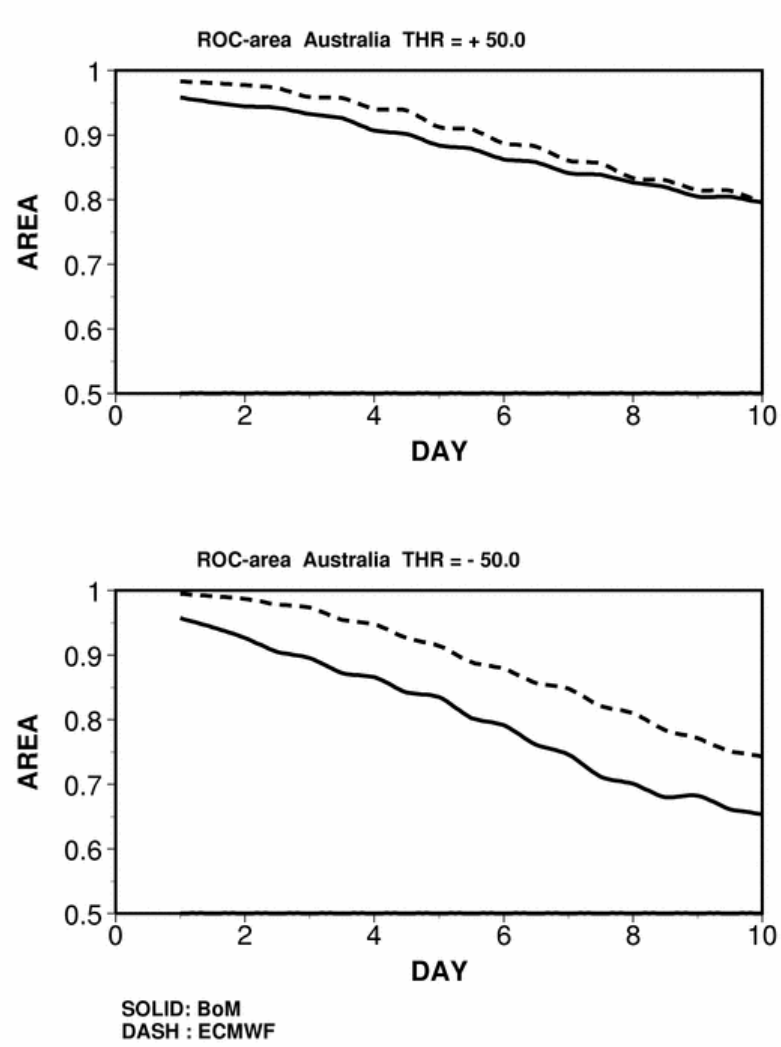

Figure 12(a): As in Figure 11, (a) but for AU.

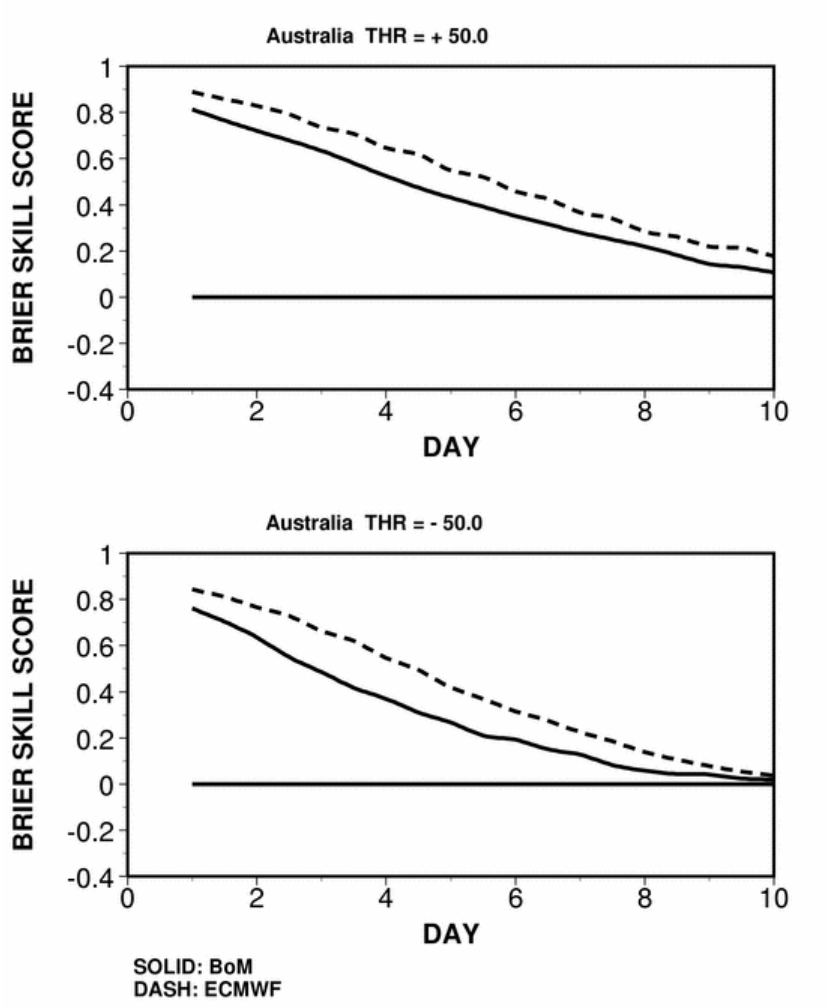

Figure 12(b): As in Figure 11 (b) but for AU

8. Rank histogram statistics

Table 4 shows the percentage of times, over and above the expected value, that the verifying analysis value lies outside the forecast range spanned by the ensemble forecasts, for the BM-EPS and EC-EPS systems. In comparing the probability of the verifying analysis lying outside the forecast range, if the analysis is considered as a random member chosen from the same distribution then the expected percentage of cases where the analysis lies outside the ensemble would be 100*[2/(Nens+1)]; i.e., 5.9% and 3.8% respectively for the BM-EPS and EC-EPS. Except at day 3 for SH, the BM-EPS system is displaying a larger "excess-over-expected" percentage of outliers for both SH and AU, with the difference becoming larger as the forecast duration increases.

For an earlier version of the EC-EPS system (T106) and for the northern hemisphere, Buizza *et al*. (1998) showed similar excess outliers percentages to Table 4. They noted that this is not an unexpected outcome given both the effects of model error and the limitations of the strategies for generating initial perturbations. Furthermore, they showed that increased ensemble size contributes to improved performance with respect to this rank histogram diagnostic; for example, increasing the EC-EPS from 32 to 50 members was found to reduce the outliers by 4% at day 5 in the northern hemisphere. In the present intercomparison the number of

outliers in the BM-EPS 32 member ensemble is 6% larger than that of the 50 member EC-EPS, which is interpreted in part as a further indication of the reduced model error evident in the EC-EPS system.

	SH region		AU region	
	BM-EPS	EC-EPS	BM-EPS	EC-EPS
Fc d+3	22%	25%	36%	32%
Fc d+5	20%	14%	27%	21%
Fc d+7	17%	9%	22%	14%

Table 4 Average excess over expected percentage (over 152 cases) of outliers (i.e., percentage of times the analysis value lies outside the forecast range spanned by the ensemble of forecasts minus the expected value $100*[2/N_{ens}+1)]$ at forecast days 3, 5 and 7 over the SH and AU for the BM-EPS and EC-EPS).

An alternative way of looking at the excess outlier percentage numbers is to scale them by the expected outlier percentages; so for example at day 5 for BM-EPS the excess outliers are 3.4 times as many as the expected number, while for EC-EPS they are 3.6 times. Scaling in this way brings the 32-member and 50-member numbers into closer alignment, both for these BM and EC southern hemisphere results and the EC-EPS northern hemisphere results of Buizza *et al*. (1998).

9. Conclusions

This work has documented the performance in the Southern Hemisphere of the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Australian Bureau of Meteorology (BoM) Ensemble Prediction Systems. The EC-EPS has been providing daily operational probabilistic forecasts since 1992, while the BM-EPS has been run in real-time in BoM schedules since July 2001. Table 1 lists the main differences between the two ensemble systems; notable differences are that the 50 member EC-EPS utilises approximately double the resolution in both the horizontal and vertical in comparison to the 32 member BM-EPS.

Attention has been focused on 500 hPa geopotential height forecasts. The accuracy of deterministic and probabilistic forecasts have been assessed using different accuracy and skill measures over the SH annulus between 20 and 60 degrees South and over Australia (latitude from 10S to 45S, longitude from 100E to 160E). Forecasts for 152 cases, from 2 April to 31 August 2001, have been examined.

Results have indicated that the two systems differ both in ensemble spread characteristics and in ensemble skill. The probability distributions of spread, as seen in Figure 3, show that the most notable difference between the two systems is the enhanced spread of the EC-EPS particularly at day 7; at earlier times, and particularly for the Australian region (AU), the spread of the BM-EPS is slightly skewed to lower spread. Considering the growth rate of ensemble spread as seen in Figure 4, the BM-EPS has initially a smaller spread than the EC-EPS and a subsequent faster growth of the divergence among the different ensemble members up to forecast day 2-3. This initial difference in the ensemble spread is attributed to the fact that the BM-EPS initial perturbations are generated using only initial time singular vectors while the EC-EPS uses a mixture of initial-time and final-time singular vectors. The initial-time singular vectors have a substantially larger growth rate than do the final-time evolved singular vectors (Barkmeijer *et al*., 1999) and thus a more rapid growth of ensemble spread occurs during the first 2-3 days within the BM-EPS system where all the perturbation energy is in these faster growing modes. Both ensemble systems reach a similar value of ensemble spread at around forecast days 2-3, although the BM-EPS initial perturbations are scaled to a 3

times smaller initial amplitude than the EC-EPS initial perturbations. After the first 3 days, there is a time period during which the two ensemble systems have a similar level of spread, to about forecast day 5. After this period, the ensemble spread of the EC-EPS grows at a faster rate. There are several reasons for this differing longer–term behaviour; these include the impact of the higher resolution and larger ensemble size utilised in the EC-EPS system, the impact of the stochastic physics, and the impact of the slower -growing final-time singular vectors used in the EC-EPS initial perturbations. Given the differing strategies of generating perturbations and the differing number and the vertical resolution of the singular vectors, it is perhaps surprising that the spread probability distributions are as similar as has been found in the present study.

Considering the skill of the two ensemble systems, almost all accuracy measures indicate that the EC-EPS performance is better. This skill difference is shown both by deterministic type products (e.g., forecasts given by the control and by the ensemble mean forecasts) and by probabilistic type forecasts (probability of geopotential height anomalies) over both the SH and the Australian Region.

Overall, EC-EPS deterministic products are 1 day more skillful than the corresponding BM-EPS products (see Figure 6(a) and 6(b)). This difference is attributed primarily to the superior quality of the ECMWF analyses, rather than model differences or model resolution. Analysis transplant experiments using ECMWF analyses in the BMRC forecast system (unpublished) have shown results consistent with those of Lalaurette *et al*. (2003) (viz. Figure 18), that the magnitude of 500 hPa deterministic forecast errors in the southern hemisphere can be affected much more by the starting analyses than the model resolution. Within the period of this study, the operational BoM (T239L29) forecast verifications compared with BM-EPS control (T119/L19), in 500 hPa predictions for the SH region for August 2001 (31 cases), is only minimally improved by the increase in resolution; i.e, increase in resolution from that of the BM-EPS to something similar to the EC-EPS does not resolve the poorer performance of the BM-EPS control relative to the EC-EPS control as seen in Figure 6(a). Likewise, the same August 2001 verification for the ECMWF (T511/L60) deterministic predictions compared with EC-EPS control (T255/L40) shows only minimal improvement from the increase in resolution. Model parameterisation differences between the two systems have not been evaluated but are thought to be less important than the impact of differing initial conditions in these present studies.

The EC-EPS probabilistic products are between 1.5 and 2.0 days more skillful than the corresponding BM-EPS products (see Figures 11 and 12). Both larger ensemble size and ensemble resolution of the EC-EPS system are known in other studies of the EC-EPS system to enhance probabilistic skill in the northern hemisphere (Buizza *et al*., 1998, 2003b). The more recent of these two studies, in particular, shows as a function of increased system resolution, a greater relative improvement in probability scores rather than in deterministic scores. It is to be expected that some of the improvements in probabilistic skill identified in Buizza *et al*. (2003b) would be applicable to the BM-EPS in the southern hemisphere with commensurate ensemble size and resolution.

The 33% skill gain, on average, of the best ensemble member relative to the control forecast for the AU region for 5 day prediction for both EPS systems (see Table 2, Figure 8) is markedly larger than seen for SH; this emphasises that the present perturbation strategies do indeed yield ensemble members with marked increase in skill over the control and especially so in the Australian region. It is clearly more unlikely that perturbations over the whole hemispheric domain can realise improved prediction simultaneously over all regions within the hemisphere. The occurrence of a greater percentage of members of the ensemble better

than the control, as seen in the BM-EPS system (see Figure 6(d)), may be due to the poorer control performance, although it may also be a function of the differing perturbation strategies.

In assessing the existence of a correlation between spread and skill (see Table 3(b)), it has been found that the BM-EPS system is showing the strongest correlations for the AU region, particularly at day 5; the EC-EPS system is seen to be superior in this regard at the later time periods for the SH region.

A difference between the two EPS systems that may be impacting positively over the Australian region in the BM-EPS is the use of singular vectors for the latitude range poleward of 20 degrees, rather than poleward of 30 degrees as used in the EC-EPS system.

The present analysis of the BM-EPS system represents the first major diagnosis of its performance since its development and implementation; this comparison with the very well established EC-EPS system is very informative and points to a number of issues within the BM-EPS system that warrant attention. Clearly, enhancing the level of performance of the base control forecast will provide a better framework for the ensemble system. Resolving the more marked shortfall in probabilistic skill, relative to deterministic skill, of the BM-EPS compared with the EC-EPS system, and reducing the percentage of outliers are further immediate issues that need to be considered The implications from EC-EPS studies are that enhanced resolution and ensemble size will alleviate some of these deficiencies.

The results discussed in this paper consider only the respective performance levels of the two EPS systems in terms of 500 hPa geopotential height ensemble prediction; the issues of, for example, the magnitude of spread from the control, and the spread-skill relationships may have different characteristics for other predicted fields.

Acknowledgements

Roberto Buizza would like to thank the Bureau of Meteorology Research Centre for arranging and funding the visit to Melbourne in June 2001. This work is part of the very successful and ongoing collaboration between ECMWF and BMRC. The authors acknowledge the helpful suggestions and clarifications by the two reviewers.

References

Barkmeijer, J., R. Buizza, and T.N. Palmer, 1999: 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. Q. J. R. Meteorol. Soc., 125, 2333-2351.

Bourke, W., T. Hart, P. Steinle, R. Seaman, G. Embery, M. Naughton, L. Rikus, 1995: Evolution of the Bureau of Meteorology's Global Assimilation and Prediction system. Part 2: resolution enhancements and case studies. Aust. Met. Mag., 44, 19-40.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. Mon. Wea. Rev., 78, 1–3.

Buizza, R., 1994: Sensitivity of optimal unstable structures. Q. J. R. Meteorol. Soc., 120, 429-451.

Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF Ensemble Prediction System. Mon. Wea. Rev., 125, 99-119.

Buizza, R. and T.N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. J.Atmos. Sci., 52, 1434-1456.

Buizza, R. and T.N. Palmer, 1998: Impact of ensemble size on ensemble prediction. Mon. Wea. Rev., 126, 2503-2518.

Buizza, R.,T. Petroliagis, T.N. Palmer, J. Barkmeijer, M. Hamrud, A. Hollingsworth, A. Simmons, and N. Wedi, 1998: Impact of model resolution and ensemble size on the performance of an Ensemble Prediction System. Q. J. R. Meteorol. Soc., 124, 1935-1960.

Buizza, R., M. Miller, and T.N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. Q. J. R. Meteorol. Soc., 125, 2887-2908.

Buizza, R., P.L. Houtekamer, Z. Toth, G. Pellerin, M. Wei and Y. Zhu, 2003a: Assessment of the status of global ensemble prediction. Proc. ECMWF Seminar on Predictability of Weather and Climate 9-13 Sept 2002, Reading, United Kingdom, ECMWF, 127-154.

Buizza, R., D.S. Richardson, and T.N. Palmer, 2003b: Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles. Q. J. R. Meteorol. Soc., 129, 1269-1288.

Ehrendorfer, M., and J.J. Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. J. Atmos. Sci., 54, 286-313.

Errico, R.M. and R. Langland, 1999: Notes on the appropriateness of "bred modes" for generating initial perturbations used in ensemble prediction. Tellus, 51A, 431-441.

Farrell, B. F., 1982: The initial growth of disturbances in a baroclinic flow. J. Atmos. Sci., 39, 1663-1686.

Farrell, B. F., 1989: Optimal excitation of baroclinic waves. J. Atmos. Sci., 46, 1193-1206.

Hamill, T.M., C. Synder and J.S. Whitaker, 2003: Ensemble Forecasts and the Properties of Flow-Dependent Analysis-Error Covariance Singular Vectors. Mon. Wea. Rev, 131, 1741–1758.

Harris, B.A., and G. Kelly, 2001: A satellite radiance-bias correction scheme for data assimilation. Q. J. R. Meteorol. Soc., 127, 1453-1468

Harrison, M.S.J., T.N. Palmer, D.S. Richardson, and R. Buizza, 1999: Analysis and model dependencies in medium-range ensembles: two transplant case-studies. Q. J. R. Meteorol. Soc., 125, 2487-2515.

Hersbach, H., R. Mureau, J.D. Opsteegh, and J. Barkmeijer, 2000: A short-range to early-medium-range Ensemble Prediction System for the European Area. Mon. Wea. Rev., 128, 3501 – 3519.

Hoskins, B. J., and P. J. Valdes, 1990: On the existence of storm tracks. J. Atmos. Sci., 47, 1854–1864.

Houtekamer, P. L., L. LeFaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. Mon. Wea. Rev., 124, 1225–1242.

Lalaurette, F., L.Ferranti, A. Ghelli, O. Saetra, and H. Bottger, 2003. Verification statistics and evaluation of ECMWF forecasts in 2001-2002. ECMWF Technical Memorandum 414, Reading, United Kingdom.

Lehoucq, R. B., D. C. Sorenson, and C. Yang, 1998: ARPACK Users Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods. Society for Industrial and Applied Mathematics, 160 pp.

Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. Tellus, 17, 321-333.

Mahfouf, J. F., and F. Rabier, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics. Quart. J. Roy. Meteor. Soc., 126, 1171–1190.

Marsigli, C., A. Montani, F. Nerozzi, T. Paccagnella, S. Tibaldi, F. Molteni, and R. Buizza, 2001: A strategy for high-resolution ensemble prediction. II: Limited-area experiments in four Alpine flood events. Quart. J. Roy. Meteor. Soc., 127, 2095–2116.

Mason, I., 1982: A model for assessment of weather forecasts. Aust. Meteor. Mag., 30, 291–303.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. Quart. J. Roy. Meteor. Soc., 122, 73–119.

Molteni, F., R. Buizza, C. Marsigli, A. Montani, F. Nerozzi, and T. Paccagnella, 2001: A strategy for high-resolution ensemble prediction. I: Definition of representative members and global-model experiments. Quart. J. Roy. Meteorol. Soc., 127, 2069-2094.

Mullen, S.L., and R. Buizza, 2001: Quantitative Precipitation Forecasts over the United States by the ECMWF Ensemble Prediction System. Mon. Wea. Re., 129, 638–663.

Palmer, T. N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics and adaptive observations. J. Atmos. Sci., 55, 633-653.

Seaman, R, W. Bourke, P. Steinle, T. Hart, G. Embery, M. Naughton, and L. Rikus, 1995: Evolution of the Bureau of Meteorology's Global Assimilation and Prediction System. Part 1: analysis and initialisation. Aust. Meteor. Mag., 44, 1-18.

Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. Quart. J. Roy. Meteor. Soc., 128, 647-677

Stanski, H.R., L..J. Wilson, and W.R. Burrows, 1989: Survey of common verification measures in meteorology. World Weather Watch Tech. Rep. 8, WMO, Geneva, 114p.

Szunyogh, I., E. Kalnay, and Z. Toth, 1997: A comparison of Lyapunov and optimal vectors in a low-resolution GCM. Tellus., 49A, 200–227.

Talagrand, O., R. Vautard, and B. Strauss, 1998: Evaluation of probabilistic prediction systems. Proc. ECMWF Seminar on Predictability 20-22 Oct 1997, Reading, United Kingdom, ECMWF, 1–26.

Tiedtke, M., 1993: Representation of clouds in large-scale models. Mon. Wea. Rev., 121, 3040–3061.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. Bull. Am. Meteorolog. Soc., 74, 2317-2330.

Toth, Z., E. Kalnay, S. Tracton, R. Wobus, and J. Irwin, 1997: A synoptic evaluation of the NCEP ensemble. Wea. Forecasting, 12, 140–153.

Toth, Z.,Y. Zhu, T. Marchok, S. Tracton, and E. Kalnay, 1998: Verification of the NCEP global ensemble forecasts. Preprints 12th Conf. On NWP, Phoenix, AZ, Amer. Meteor. Soc., 286-289.

Tracton, M. S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. Wea. Forecasting, 8, 379–398.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, New York, 467pp.