TECHNICAL MEMORANDUM

280

Current status and future developments of the ECMWF Ensemble Prediction System

R. Buizza, J. Barkmeijer,
T. N. Palmer and D. S. Richardson

Research Department

May 1999

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

ABSTRACT

The two latest changes introduced during 1998 into the Ensemble Prediction System (EPS) operational at the European Centre for Medium-Range Weather Forecasts (ECMWF) are described. The first change, the inclusion of instabilities growing during the data-assimilation period in the generation of the EPS initial perturbations, increased the probability that the analysis lays inside the ensemble forecast range. The second change, the introduction of a simulation of random model errors due to parametrized physical processes, improved in particular the prediction of precipitation. The performance of the ECMWF Ensemble Prediction System from 1 May 1994 to March 1999 is assessed using different statistical measures. Results indicate that the general performance of the EPS has been improving over the years. Finally, ongoing research projects on predictability issues developed either at ECMWF, or at European Research Institutes in collaboration with ECMWF, are discussed.

1. INTRODUCTION

At the European Centre for Medium-Range Weather Forecasts Ensemble Prediction System (ECMWF), the Ensemble Prediction System (EPS) has been producing probabilistic forecasts daily since 1 May 1994, following more than one year of quasi operational trials (from 19 December 1992 until 1 May 1994 the EPS was run only three times a week).

Based on the hypothesis that medium-range forecast errors are predominantly associated with uncertainties in the initial conditions (*Downton and Bell*, 1988; *Richardson*, 1998), the EPS has been realized as a set of multiple integrations of the ECMWF operational model from an ensemble of initial conditions. These initial conditions are created by adding to the operational (unperturbed) analysis perturbations with the fastest energy growth during the early stage of the forecast period, defined by the so-called singular vectors of the tangent forward model (*Buizza & Palmer*, 1995). The choice of using singular vectors was made to reduce sampling errors and to avoid situations in which relatively little dispersion is associated with large forecast errors (*Palmer et al.*, 1997). Since the singular vectors are computed during the early forecast range, they do not sample instabilities growing during the data-assimilation period up to the integration starting time. This fact, or, more generally speaking, the use of a crude representation of the analysis error covariance metric (i.e. the use of the total energy norm, *Palmer et al.*, 1998) to estimate the singular vectors can be seen as a weakness of the EPS.

The EPS configuration implemented in 1992 included 33 integrations with horizontal spectral triangular truncation T63 and 19 vertical levels (T63L19, *Molteni et al.*, 1996). Apart for small changes (see *Buizza*, 1997, for an updated list), the system underwent a major upgrade in December 1996 (*Buizza et al.*, 1998), when the ensemble size was increased from 33 to 51 members, and the resolution was increased from T63L19 (spectral triangular truncation T63 and 19 vertical levels) to T_l159L31 (the subscript 'l' standing for the 'linear grid' option, whereby the computational grid is the same as the standard 'quadratic' grid used for the T106 model version).

Although the hypothesis of the dominant role of initial uncertainties is certainly valid in the early forecast range, model uncertainties do affect forecast accuracy. *Richardson* (1998) and *Harrison et al.* (1999), considering synoptic-scale features, have shown that model errors can become as important as initial condition uncertainties in the medium forecast range. Moreover, due to the uncertainties in the parametrization of physical processes, model errors can be important in the early forecast range for variables such as precipitation and surface

model errors can be important in the early forecast range for variables such as precipitation and surface temperature. An ensemble system which includes both initial and model uncertainties is the one developed at the Atmospheric Environment Service in Canada. Following a system simulation approach to ensemble prediction, *Houtekamer et al.* (1996) developed a procedure where each ensemble member differs both in the initial conditions, which are generated by running independent data assimilation cycles using randomly perturbed observations, and in model characteristics. The absence of simulation of model errors in the ECMWF EPS can be seen as a second weakness of the system.

To overcome, at least partially, these weaknesses, two changes have been introduced into the EPS during 1998:

a) The first, implemented on 26 March 1998, introduced the use of singular vectors growing during the data-assimilation period in the generation of the EPS initial perturbations (*Barkmeijer et al.*, 1998a).
b) The second, implemented on 21 October 1998, introduced a simulation of random model errors related to physical parametrisations (*Buizza et al.*, 1999).

The first purpose of this paper work is to describe these latest changes. The second purpose is to provide the reader with an updated overview of the EPS performance from 1 May 1994 until 2 March 1999. This overview focuses on the prediction of the 500 hPa geopotential height, and is based on various measures of ensemble performance (root-mean-square error of the control and the ensemble mean, correlation between ensemble spread and control error, percentage of outliers, ROC area, Brier Skill Score), which verify first and second order moments of the EPS forecast probability function. The third purpose is to inform the readers about ongoing research on predictability issues at ECMWF, or at European Institutes in collaboration with ECMWF.

The paper is organized as follows. Section 2 gives a brief overview of the EPS in the configuration operational at the time of writing (April 1999) and describes the latest system changes. Section 3 gives an overview of the EPS performance. Section 4 discusses on-going research. Conclusions are drawn in section 5.

2. THE ECMWF ENSEMBLE PREDICTION SYSTEM

2.1. General description

Schematically, consider the forecast model equations

$$\frac{\partial e_j}{\partial t} = A(e_j,t) + P_j'(e_j,t)\,, \tag{1}$$

where e_j is a generic ensemble member, and A and P_j identify non-parametrized and parametrized physical processes. More specifically, P_j identifies the tendency due to parametrized physical processes associated with radiative transfer, turbulent mixing, sub-grid scale orographic drag, moist convection, clouds and surface/soil processes.

Each ensemble member e_j can be seen as the time integration of the model equations (1),

$$e_j(t) = \int_{t=0}^{t} \left[A(e_j,t)+P'(e_j,t)\right]dt\,, \tag{2}$$

starting from the perturbed initial conditions

$$e_j(t=0) \equiv e_0(t=0)+\delta e_j(t=0), \quad (3)$$

where $e_0(t=0)$ is the operational analysis at $t=0$, and the initial perturbations $\delta e_j(t=0)$ are generated using the singular vectors of the linear version of the ECMWF, computed to maximize total energy growth over a 48-hour time interval (*Buizza & Palmer*, 1995). Two sets of singular vectors are currently used, computed independently to maximize total energy in the Northern and Southern Hemisphere extra-tropics. No initial perturbations are currently added in the tropics.

2.2. Initial condition generation

Until March 1998, the EPS initial perturbations were computed to sample instabilities growing in the forecast range, and no account was taken of perturbations that had grown during the data assimilation cycle leading up to the generation of the initial conditions (*Molteni et al.*, 1996). A way to overcome this problem is to use singular vectors growing in the past, and evolved to the current time.

Considering a starting date d, since 26 March 1998 (*Barkmeijer et al.*, 1998a) the initial perturbations in equation (3) have been generated using both the singular vectors growing in the forecast range between day d and day $d+2$ at initial time, and the singular vectors that had grown in the past between day $d-2$ and day d at final time

$$\delta e_j(t=0) \equiv \sum_{i=1}^{25} [\alpha_{j,i}\, v_i^{d,d+2}(t=0) + \beta_{j,i}\, v_i^{d-2,d}(t=48h)], \quad (4)$$

where $v_i^{n,m}(t)$ is the *i-th* singular vector growing between day n and m at time t. The coefficients $\alpha_{j,i}$ and $\beta_{j,i}$, which set the initial amplitude of the ensemble perturbations, are defined by comparing the singular vectors with estimates of analysis errors (*Molteni et al.*, 1996).

The initial perturbations are still scaled at initial time so that, on average over the Northern Hemisphere, the ensemble root-mean-square (rms) spread matches the control rms error at around forecast day 2.

Results, documented in *Barkmeijer et al.* (1998a), indicate that the use of evolved singular vectors improves the ensemble skill and reduces the percentage of times the analysis lays outside the ensemble forecast range, especially in the early forecast range.

2.3. Simulation of random model errors due to parametrized physical processes

The current simulation scheme (*Buizza et al.*, 1999) can be considered as a simple first attempt to simulated random model errors due to parametrized physical processes. It is based on the notion that random errors due to parametrized physical processes are coherent between the different parametrization modules and have a certain coherence on the space and time scales represented by the model. The scheme assumes that the larger the parametrized tendencies, the larger the random error component.

Since implementation on 21 October 1998, each ensemble member e_j can be seen as the time integration of the perturbed model equations

$$e_j(t) = \int_{t=0}^{t} \left[A(e_j,t)+P'_j(e_j,t)\right]dt, \quad (5)$$

where P'_j denotes the perturbed tendency due to parametrized physical processes, defined as

$$P'_j(e_j,t) \equiv <r_j(x,t)>_{D,T} P(e_j,t) , \quad (6)$$

where $<..>_{D,T}$ means that the same random number r_j has been used for all grid points inside a *DxD* degree box and that it has been kept the same for *T* time steps.

The random numbers r_j in equation (6) have been selected uniformly in the interval [0.5,1.5]. In the current formulation, the stochastic term re-scales the tendency for each component of the state vector (i.e. for wind, temperature and specific humidity) by the same factor $<r_j>_{D,T}$ for the entire grid point column. Different random numbers are used for each component. The same random numbers are used inside 10° boxes [i.e. D is set to 10° in equation (6)], and random numbers are generated every 6 hours [i.e. T=6 hours in equation (6)].

The operational implementation of the scheme followed detailed sensitivity tests (since the scheme is based on the randomly determined factors r_j it is also named stochastic physics). Results, documented in *Buizza et al.* (1999), showed that stochastic physics increases the spread of the ensemble and improves the ensemble performance, particularly for the prediction of weather parameters such as precipitation. One measure of accuracy of probability predictions of precipitation is the area under the Relative Operating Characteristic (ROC) curve (*Stanski et al.*, 1989; see Appendix for more details). This measure is positively oriented, with perfect scores having unit ROC area, and with areas smaller than 0.50 indicating no skill. Figure 1 shows that the stochastic physics scheme has a positive impact, especially during summertime.

3. EPS PERFORMANCE: BRIEF OVERVIEW FOR GEOPOTENTIAL HEIGHT AT 500 HPA

A summary of the EPS performance since daily operational implementation (1 May 1994) is given, focusing on the prediction of the 500 hPa geopotential height over the Northern Hemisphere (NH) and Europe.

3.1. Ensemble spread, control error and ensemble-mean error

The ensemble spread with respect to the control forecast (defined as the average distance between the perturbed members and the unperturbed control forecast) is a measure of the ensemble dispersion. A minimum requirement for an ensemble system is that it includes the analysis inside the ensemble forecast range (*Buizza*, 1997). Since the control error measures the distance between the analysis and the control forecast, this requirement implies that the ensemble spread is comparable to the control error. An ensemble spread smaller/larger than the control error is an indication of spread underestimation/overestimation.

Figure 2 shows the difference between the seasonal average ensemble spread and control error for NH and Europe at forecast days 3, 5 and 7. Results indicate that, after the system upgrade of December 1996, the ensemble rms spread is up to 25% smaller than the control rms error over NH (23% over Europe). The comparison of the spread underestimation of summer 1997 and 1998 indicates that the use of evolved singular vectors helped reduce the amount of underestimation.

spread underestimation of summer 1997 and 1998 indicates that the use of evolved singular vectors helped reduce the amount of underestimation.

The first deterministic-type product that can be derived from the EPS forecast data is the ensemble mean, defined as the average of all the 51 EPS members (50 perturbed plus the unperturbed forecast). By construction, the ensemble mean is a smoother field than any of the single members, but with smoothing applied only in regions where the ensemble members differ. As such, the ensemble mean cannot be deduced by filtering all waves with certain scale characteristics.

Figure 3 shows the seasonal average root-mean-square (rms) error of the EPS control forecast, and Figure 4 shows the percentage of rms error reduction of the ensemble mean compared to the control forecast. After December 1996, the ensemble mean has always been more skilful than the control forecast. Note that this was not the case before winter 1995/96. During that period, in fact, to have a reasonable amount of spread after forecast day 2 the ensemble initial perturbations were given a much larger amplitude. Such a very large initial amplitude compensated for the poor activity of the T63L19 model version (*Tibaldi et al.*, 1990) which was one of the causes of a too slow perturbation growth. As a consequence, in the early medium-range all perturbed ensemble members had worse scores than the control, and this was affecting the skill of the ensemble mean.

Ensemble spread can be used as a predictor of the skill of a deterministic forecast. For example, the average spread around the control can be used to predict the skill of the control forecast itself. Ideally, cases with smaller than average ensemble spread should be predictable and should have a smaller than average error. Results (Figure 5) show that in about 15 of the 92 cases from each season, the control error is larger than average (instead of smaller than average) when the ensemble spread is smaller than average. Generally speaking, the correlation coefficient between ensemble spread and control error ranges between 0.11 and 0.73 for the NH and between 0.13 and 0.62 for Europe (Figure 6). It should be pointed out that correlation between ensemble spread and control error remains smaller than 1 also for a perfect ensemble system (*Buizza*, 1997). Thus, one should not expect ensemble rms spread and control rms error to be fully correlated. Both Figures 5 and 6 do not show any improvement after the system upgrade in December 1996.

3.2. Percentage of outliers

Another measure of ensemble performance is the percentage of analysis values lying outside the EPS forecast range (*Strauss & Lanzinger*, 1995; *Buizza*, 1997). Ideally, for an ensemble system with N members which randomly samples the forecast error probability density function, the so-called percentage of outliers should be equal to the reference value $p_{ref}=2/(N+1)$. Figure 7 shows the difference between the percentage of outliers and the reference value [which was 2/34 for the (32+1)*T63L19 system, and is 2/52 for the current (50+1)*T_l159L31 system]. Results indicate that the system upgrade of December 1996 improved the system dramatically, and that the use of evolved singular vectors seems to have had a positive impact.

3.3. Probability forecasts

Figure 8 shows the area under the ROC curve for the probability forecasts of 'positive and negative geopotential anomalies with magnitude greater than one standard deviation', where the mean and the standard deviation are

computed monthly. Results indicate that EPS probability predictions improved substantially after the EPS upgrade of December 1996.

The accuracy of the two probability forecasts of geopotential anomalies has also been assessed using Brier Scores and Brier Skill Scores (*Brier*, 1950), with skill scores defined with respect to climatology (thus, a positive skill score indicate that the EPS forecast is more skilful than a forecast based on climatology). Brier skill scores for NH and Europe (Figure 9) confirm that the EPS probability prediction improved after December 1996. Since then, Brier Skill Scores have been positive in the medium-range, crossing the zero line at about forecast day 10 (not shown).

4. FUTURE DEVELOPMENTS

The following section describes research currently being done at ECMWF in six areas related to ensemble prediction. This list is not intended to be exhaustive, but should give the reader an overview of possible future developments of the ECMWF EPS.

4.1. Use of linear physics in singular vector computation in the tropics

Linearized versions of the most important physical processes has been developed (*Mahfouf*, 1999), and investigation into the behaviour of the linear models in the computation of tropical singular vectors has started. The tropical target area has been chosen because the current EPS lacks perturbations of the initial condition in this area, where moist processes are of key importance. The use of linearized moist processes makes possible the computation of the specific humidity component of the singular vectors (currently, extra tropical singular vectors have no specific humidity component). Note that the inclusion of a non-zero specific humidity component in the initial perturbations required a change in the norm used to define the singular vectors, which has therefore been augmented with a term which takes into account the growth in specific humidity.

A first set of ensembles run with and without tropical singular vectors, both without stochastic perturbations in the physical tendencies, has been compared. This set of results indicates that tropical singular vectors induce considerably more spread in the tropics. As an example, Figure 10 shows the spread of ensembles run without stochastic physics, and with or without tropical singular vectors with initial condition 11 November 1998. The impact of tropical singular vectors on the current EPS configuration with stochastic physics is under investigation.

4.2. Simplified Kalman filter and full quasi-geostrophic Kalman filter

The operational EPS is based on initial perturbations constructed using singular vectors with maximum total energy growth. Total energy singular vectors have no knowledge of analysis error statistics, since they identify unstable atmospheric regions regardless of the presence of analysis errors in such regions. As a consequence, the current system is not able to discriminate between regions with large (e.g. the oceans) or small (e.g. Europe or North America) analysis errors, and to sample more the former than the latter regions. Generally speaking, it would be desirable to use information about analysis error characteristics in the singular vector computation.

One way of improving upon this is to use in the singular vector computation statistics generated by the data assimilation system. At ECMWF, work is in progress to use the Hessian of the cost function of the 3-dimensional

(or 4-dimensional) variational assimilation system (3D/4D-Var) to define singular vectors (*Barkmeijer et al.*, 1998b). These so-called Hessian singular vectors are constrained at initial time by analysis error statistics but still produce fast perturbation growth during the first few days of the forecast.

Singular vector properties, such as total energy spectrum and vertical profile of total energy, change substantially when the Hessian of the 3D/4D-Var cost function rather than the total energy is used in their computation. At initial time, Hessian singular vectors have larger scale than total energy singular vectors (Figure 11a,b). Considering the vertical profile of total energy, while total energy singular vectors show an upward energy propagation (Fig. 11c), Hessian singular vectors concentrate at jet level both at the initial and the final times (Fig. 11d).

Earlier experimentation with perturbations derived from Hessian singular vectors showed that the performance of this EPS configuration was comparable to that using total energy based perturbations. Introduction of a simplified Kalman filter (*Fisher*, 1998) may improve upon this. The simplified Kalman filter makes it possible to have flow dependent background error statistics. Experimentation is under way to study the impact of this modified background error covariance on the 4D-Var analysis procedure and the EPS using Hessian singular vectors.

Parallel to this, a full quasi-geostrophic Kalman filter has been developed (*Ehrendorfer & Bouttier*, 1998). It will be a valuable tool in finding an optimal configuration of the simplified Kalman filter.

4.3. Increase of horizontal resolution

The major improvement of the EPS performance came in December 1996, mainly because of the resolution increase from T63L19 to T_l159L31 (*Buizza et al.*, 1998). The EPS performance improvement was particularly large because the T63L19 model version had statistical properties different from the higher resolution versions, which were closer to reality (e.g. global transient eddy kinetic energy, *Tibaldi et al.*, 1990). Moreover, the fact that before December 1996 the EPS (T63L19 resolution) and the ECMWF high-resolution deterministic forecast (T213L31 resolution) had very different statistical properties, made their combined use sub-optimal.

Since December 1996 the situation has improved, because the EPS T_l159L31 model and the high-resolution T_l319L31 deterministic model have rather similar statistical properties. Any future resolution increase in the ECMWF deterministic model could re-introduce undesirable differences between the EPS and the high-resolution deterministic forecast if not coupled with a proportional increase of the EPS resolution. It is thus of interest to investigate the impact of an increase in EPS resolution on the forecast quality.

Preliminary results of the comparison of ensembles run at T_l159L31 and T_l319L31 during four cases of intense precipitation in the European region suggest that such a resolution increase has a very positive impact on the prediction of weather parameters such as precipitation, although it has a small impact on the accuracy of the prediction of the synoptic-scale flow (e.g. geopotential height at 500 hPa). As an example, Figure 12 shows the low and high resolution precipitation prediction for one of these four cases, with starting date of 12Z on 23 September 1993. Generally speaking, the control and some members of the T_l319L31 ensemble correctly predict

intense precipitation over Northern Italy where more than 150 mm/24 h was observed (Figure 12e). This has a strong positive impact on the probability prediction of the event 'precipitation in excess of 50 mm over 24 hours' (Figure 12c,d). Similar indications can be drawn from the other four cases (not shown).

4.4. Limited area ensembles nested in the EPS

Limited area ensemble systems have been operational for some years. For example, short-range ensemble forecasting based on nested limited area models has been operational at the National Centers for Environmental Prediction (NCEP) since 1995 (*Tracton & Du*, 1998). The current NCEP configuration has 15 members and uses boundary and initial conditions from the NCEP global ensemble system (*Toth & Kalnay*, 1993) and from independent analyses. Work is in progress to set up a new system which would run twice a day (at 00Z and 12Z), with initial perturbations defined with regional breeding and based on two models, the ETA and the regional spectral model (*Z. Toth*, 1998, personal communication).

In collaboration with ECMWF, three groups have been investigating the usefulness of a short-range ensemble system focused to produce forecasts for the European region in the forecast range from day 1 to days 3-4:

a) A group at the Royal Netherlands Meteorological Institute (KNMI, *Hersbach et al.*, 1999) has been testing the possibility of running a targeted version of the current EPS, with perturbations generated using singular vectors with maximum energy at final time inside a region covering Europe and the Eastern Atlantic ocean only. Results from this so-called targeted EPS based on 51 ensembles (summer, winter and autumn cases) indicate that it can outperform the operational EPS, especially for the prediction of rare, extreme events.

b) A group at Oslo University, Norway (*I.-L. Frogner & T. Iversen*, 1998, personal communication) has been testing the possibility of running an ensemble of limited area model integrations with initial and boundary conditions defined by the EPS, or by the targeted EPS described above. This research project includes the comparison of the EPS and of the targeted EPS over the European region, to define which of the two systems is more skilful. Preliminary results based on 20 winter and 15 summer cases indicate that the probabilistic prediction for large-scale precipitation and 2m temperature of the targeted EPS is more skilful.

c) A group involving the Regional Meteorological Service of Emilia Romagna, ERSA/ARPA, and the Centro di Calcolo Inter-Universitario dell'Italia Nord-Orientale (CINECA) has been studying the possibility of running a small number of integrations of a limited area model nested within members of the 50 global members of the EPS (*Marsigli C., Molteni, F., Montani, A., Nerozzi, F., Paccagnella, T., Tibaldi, S.*, 1998, personal communication). Preliminary tests during four cases of intense precipitation over Europe have indicated that an ensemble based on only five members can be used successfully to generate an ensemble of high-resolution limited area forecasts.

4.5. Multi-analysis multi-model ensemble

Ensemble experiments using the UK Meteorological Office (UKMO) and the ECMWF analysis and forecasting systems have shown that to maximize the information available to forecasters, both models and both analyses should be used in the generation of ensemble prediction (*Harrison et al.*, 1999). To confirm statistically these results, a version of the UKMO model has been implemented at ECMWF to generate a multi-analysis multi-model ensemble system using the UKMO and the ECMWF analysis to generate the initial conditions. One of the purposes of this research project is to compare ensemble configurations with 27 ECMWF members, 27 UKMO members, and a mix of 27 ECMWF and 27 Met-Office members with the 51 member ECMWF EPS (*R. Barnes*, 1998, personal communication).

4.6. Ensemble from a consensus analysis

Parallel to the multi-model multi-analysis study, work is in progress at ECMWF to investigate whether a so-called consensus analysis, defined as the average of analyses produced by different weather centres, is a better estimate of the atmospheric initial state than the ECMWF analysis (*D. S. Richardson*, 1998, personal communication). The operational EPS configuration has been run from the consensus analysis, average of the ECMWF, UKMO (UK Meteorological Office), Météo-France, NCEP (National Centers for Environmental Prediction, Washington) and DWD (Deutscher WetterDienst, Offenbach) analyses. The same perturbations as used in the operational EPS have been added to the consensus analysis to create the 50 perturbed initial conditions.

Preliminary results based on nine cases selected in April and May 1998 show that the skill of the control forecast is improved if the consensus analysis is used instead of the ECMWF analysis as the unperturbed initial condition. Results also indicate that the difference between the spread in the two systems is rather small, while the ensemble-mean forecast of the system started from the consensus analysis is more skilful.

5. Conclusions

The two latest EPS changes implemented in 1998 have been described. The first one introduced the use of evolved singular vectors, which sample the instability of the atmospheric flow during the data-assimilation period, into the generation of the EPS initial perturbations (*Barkmeijer et al.*, 1998a). The second one introduced a stochastic simulation of random model errors due to the parametrization of physical processes (*Buizza et al.*, 1999). Both changes have improved the system performance, especially the probability that the analysis is included in the EPS forecast range and the prediction of precipitation probabilities.

The performance of the ECMWF Ensemble Prediction System from 1 May 1994 to 2 March 1999 has been assessed using different statistical measures. Results confirm that the major system upgrade implemented in December 1996, when the ensemble resolution was increased from T63L19 to T_l159L31 and the ensemble size was enlarged from 33 to 51 members, improved the EPS skill substantially. It has increased the skill of the ensemble mean, the probability that the analysis is included in the EPS forecast range (percentage of outliers), and the skill of probabilistic forecasts of geopotential anomalies measured by the ROC area and the Brier Skill Score.

Finally, ongoing research on predictability issues developed either at or in collaboration with ECMWF has been discussed. Research is under way to use more linearized physical processes (e.g. moist processes) in the singular vector computation, especially in the tropical region, and to use of the Hessian of the 3D/4D-Var cost function (instead of the total energy) as a metric in the singular vector computation. Other experimentation has started to verify whether a further increase in the EPS resolution would improve the EPS prediction, especially of extreme and rare events such as floods. On a regional scale, some research groups are starting to nest limited area models within the EPS perturbed members, to provide short-range ensemble systems run with very high resolution. On a global scale, multi-model multi-analysis ensemble systems are being experimented with, and compared to the EPS and to an ensemble system starting from a so-called consensus analysis, defined as the average of analyses produced by different Meteorological Centres.

APPENDIX: RELATIVE OPERATING CHARACTERISTIC

Following *Stanski et al* (1989), consider a two category contingency table:

	Forecast=YES	Forecast=NO	Total Observed
Observed=YES	X	Y	X+Y
Observed=NO	Z	W	Z+W
Total forecast	X+Z	Y+W	

where X can be referred to as the hits and the Z as the false alarms. Define the hit rate (i.e. the percentage of correct forecast) as *X/(X+Y)* and the false alarm rate (i.e. the percentage of forecasts of the event given that the event did not occur) as *Z/(Z+W)*. If these two rates are plotted against one each other on a graph, a single point results.

Similar contingency tables can be constructed for probabilistic forecasts. Consider a forecast distribution stratified according to observation into ten 10% wide categories:

Probability range	Observed=NO	Observed=YES
$(j-1)*10\% \le prob < j*10\%$	a_j	b_j

with *j=1,.. 10*, with the last category *j=10* including also *prob=100%*. For a given probability threshold *prob=j*10%*, the entries of this table can be summed to produce the four entries of a two by two contingency table, the hit and false alarm rates calculated, and a point plotted on a graph. Specifically, the four entries of a two by two contingency table for probability threshold *prob=j*10%* are:

$$X_j = \sum_{i=j+1}^{10} b_i \quad Y_j = \sum_{i=1}^{j} b_i$$
$$Z_j = \sum_{i=j+1}^{10} a_i \quad W_j = \sum_{i=1}^{j} a_i$$

If the process is repeated for all probability thresholds from *0%* to *100%*, the result is a smooth curve called the Relative Operating Characteristic (ROC).

One convenient measure associated with a ROC curve is the area under the curve, which decreases from 1 toward 0 as more false alarm rates occur. A value of 0.5 is considered as the lower bound for a useful forecast, since a system with such a ROC-area cannot discriminate between occurrence and non-occurrence of the event.

REFERENCES

Barkmeijer, J., Buizza, R., & Palmer, T. N. (1998a). 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, in press.

Barkmeijer, J., Van Gijzen, M., & F. Bouttier (1998b). Singular vectors and estimates of the analysis-error covariance metric. *Q. J. R. Meteorol. Soc.*, **124**, 1695-1713.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1-3.

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., & Palmer, T. N. (1995). The singular-vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 1434-1456.

Buizza, R., Petroliagis, T., Palmer, T. N., Barkmeijer, J., Hamrud, M., 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., Miller, M., & Palmer, T. N. (1999). Stochastic simulation of model uncertainties. *Q. J. R. Meteorol. Soc.*, in press. (Also published as *ECMWF Tech.Memo.* No.279.)

Downton, R A, and Bell, R S, 1988. The impact of analysis differences on a medium-range forecast. *Meteorol. Mag.*, **117**, 279-285.

Ehrendorfer, M., & Bouttier, F. (1998). An explicit low-resolution Kalman filter: Implementation and preliminary experimentation. ECMWF Technical Memorandum No. 259, ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.

Fisher, M. (1998). Development of a simplified Kalman filter. ECMWF Technical Memorandum No. 260, ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.

Harrison, M. S. J., Palmer, T. N., Richardson, D. S., & Buizza, R. (1999). Analysis and model dependencies in medium-range ensembles: two transplant case studies. *Q. J. R. Meteorol. Soc.*, in press.

Hershbach, H., Mureau, R., Opsteegh, J. D., & Barkmeijer, J (1999). A short to early-medium range ensemble prediction system for the European area. *Mon. Wea. Rev.*, submitted.

Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H., & Mitchell, H. (1996). A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225-1242.

Mahfouf, J.-F. (1999). Influence of physical processes on the tangent-linear approximation. *Tellus*, **51A**, 147-166.

Molteni, F., Buizza, R., Palmer, T. N., & Petroliagis, T. (1996). The ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.*, **122**, 73-119.

Palmer, T. N., Barkmeijer, J., Buizza, R., & Petroliagis, T. (1997). The ECMWF Ensemble Prediction System. *Meteorol. Appl.* **4**, 301-304.

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

Richardson, D. S. (1998). The relative effect of model and analysis differences on ECMWF and UKMO operational forecasts. *Proceedings of the ECMWF Workshop on Predictability*, ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.

Stanski, H. R., Wilson, L. J., & Burrows, W. R. (1989). Survey of common verification methods in meteorology. Research Report n 89-5, Atmospheric Environment service, Forecast research division, 4905 Dufferin Street, Downsview, Ontario, Canada M3H 5T4, pp 114 (also published as WMO Technical report n. 8).

Strauss, B., & Lanzinger, A. (1995). Validation of the ECMWF EPS. *Proc. ECMWF Seminar on Predictability*, Vol. 2, ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom, 157-166.

Tibaldi, S., Palmer, T. N., Brancović, Č., & Cubash, U. (1990). Extended-range predictions wić ECMWF models: influence of horizontal resolution on systematic error and forecast skill. *Q. J. R. Meteorol. Soc.*, **116**, 835-866.

Toth, Z., & Kalnay, E. (1993). Ensemble forecasting at NMC: the generation of initial perturbations. *Bull. Am. Meteorol. Soc.*, **74**, 2317-2330.

Tracton, M. S. & Du, J. (1998). Short Range Ensemble Forecasting (SREF) at NCEP/EMC. Preprints, *12th Conference on NWP, Phoenix, AZ, Amer. Met. Soc.*, 269-272.

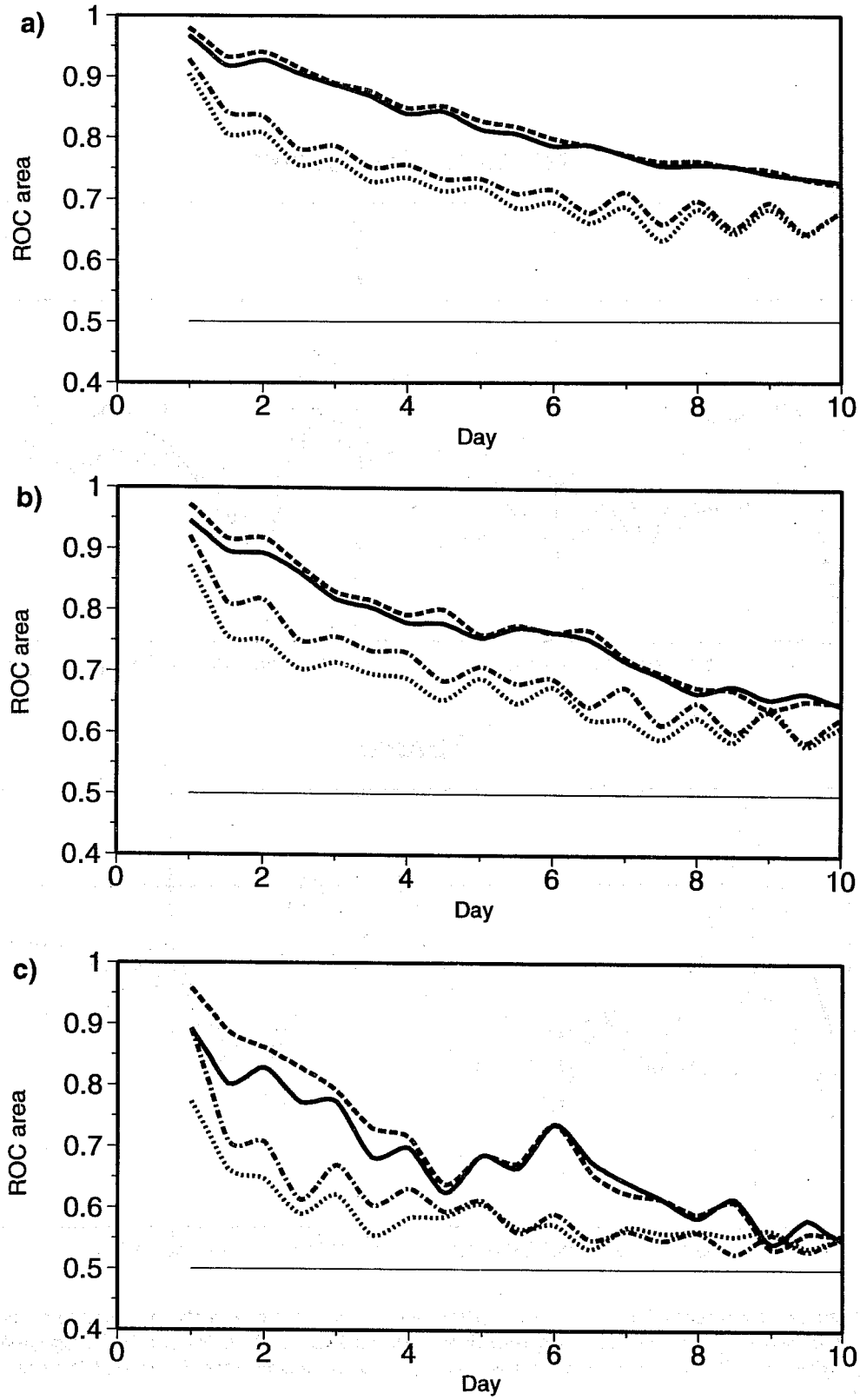

Fig. 1 (a) area under the Relative Operating Characteristic curve relative to the event "12h accumulated precipitation greater than 5 mm", (b) "12h accumulated precipitation greater than 10 mm", and (c) "12h accumulated precipitation greater than 20 mm", for the reference (solid) and the stochastically perturbed system (dash) for the week 16-22-December 1997, and for the reference EPS (dot) and the stochastically perturbed system (chain dash) for the week 29 June-5 July 1997.

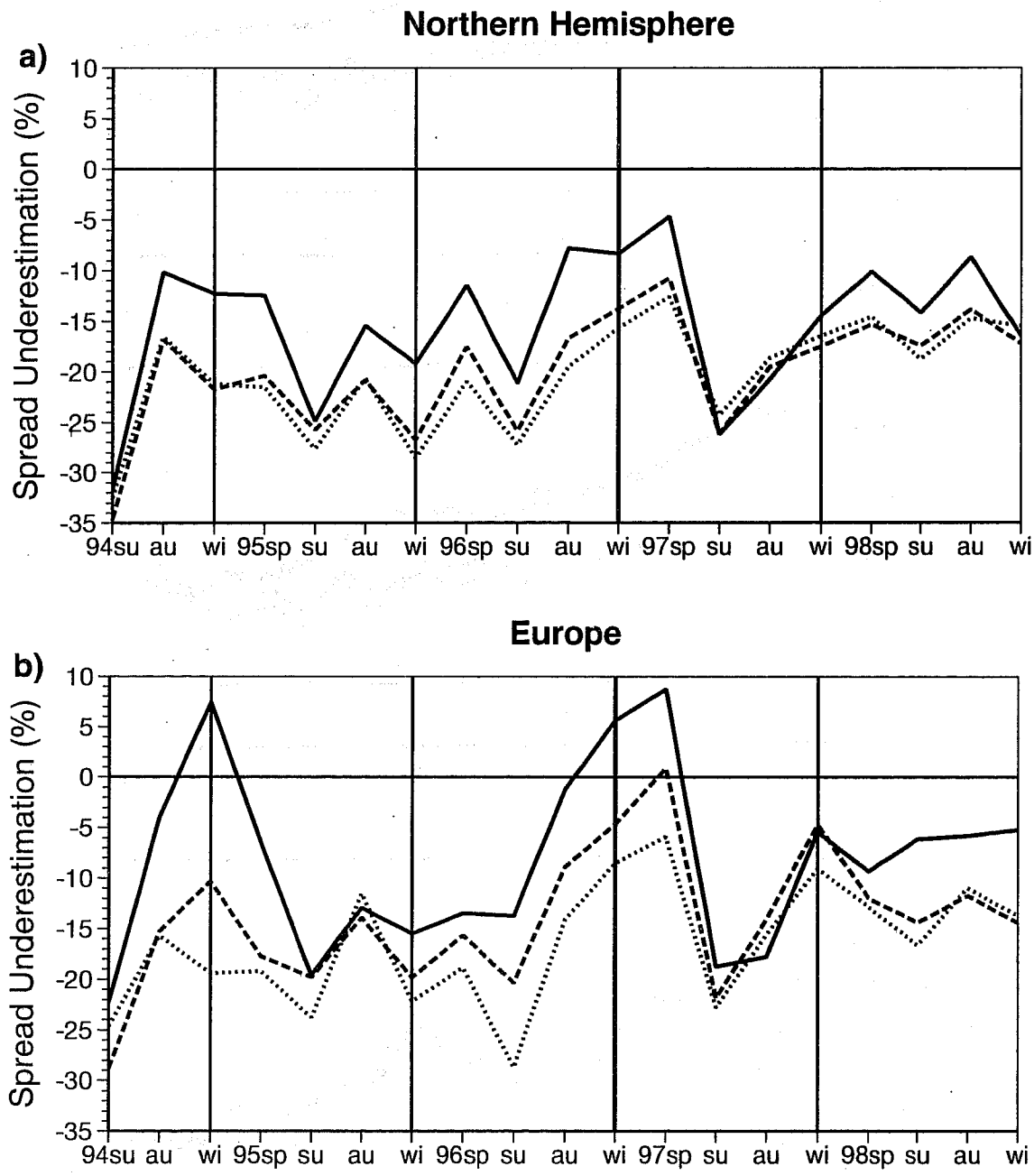

Fig. 2 Ensemble spread underestimation, defined as the difference between the ensemble rms spread *spr* and the control rms error errc divided by the control error, *(spr-errc)/errc,* computed over the (a) NH and (b) Europe, at forecast day 3 (solid), 5 (dash) and 7 (dot). Vertical lines mark values for winter seasons (bold identifies the system upgrade of December 1996).

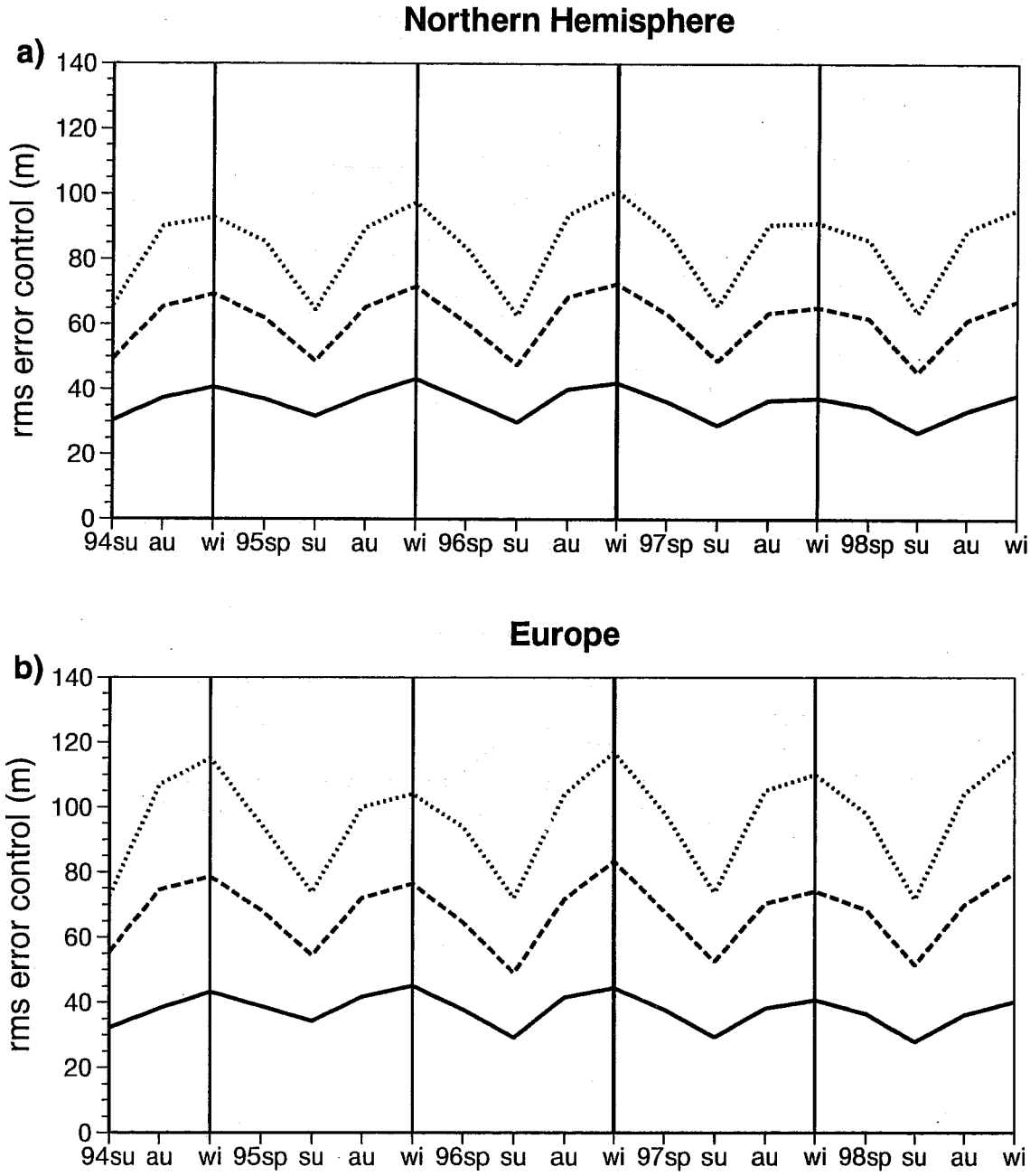

Fig. 3 **Control rms error computed over the (a) NH and (b) Europe, at forecast day 3 (solid), 5 (dash) and 7 (dot). Vertical lines mark values for winter seasons (bold identifies the system upgrade of December 1996).**

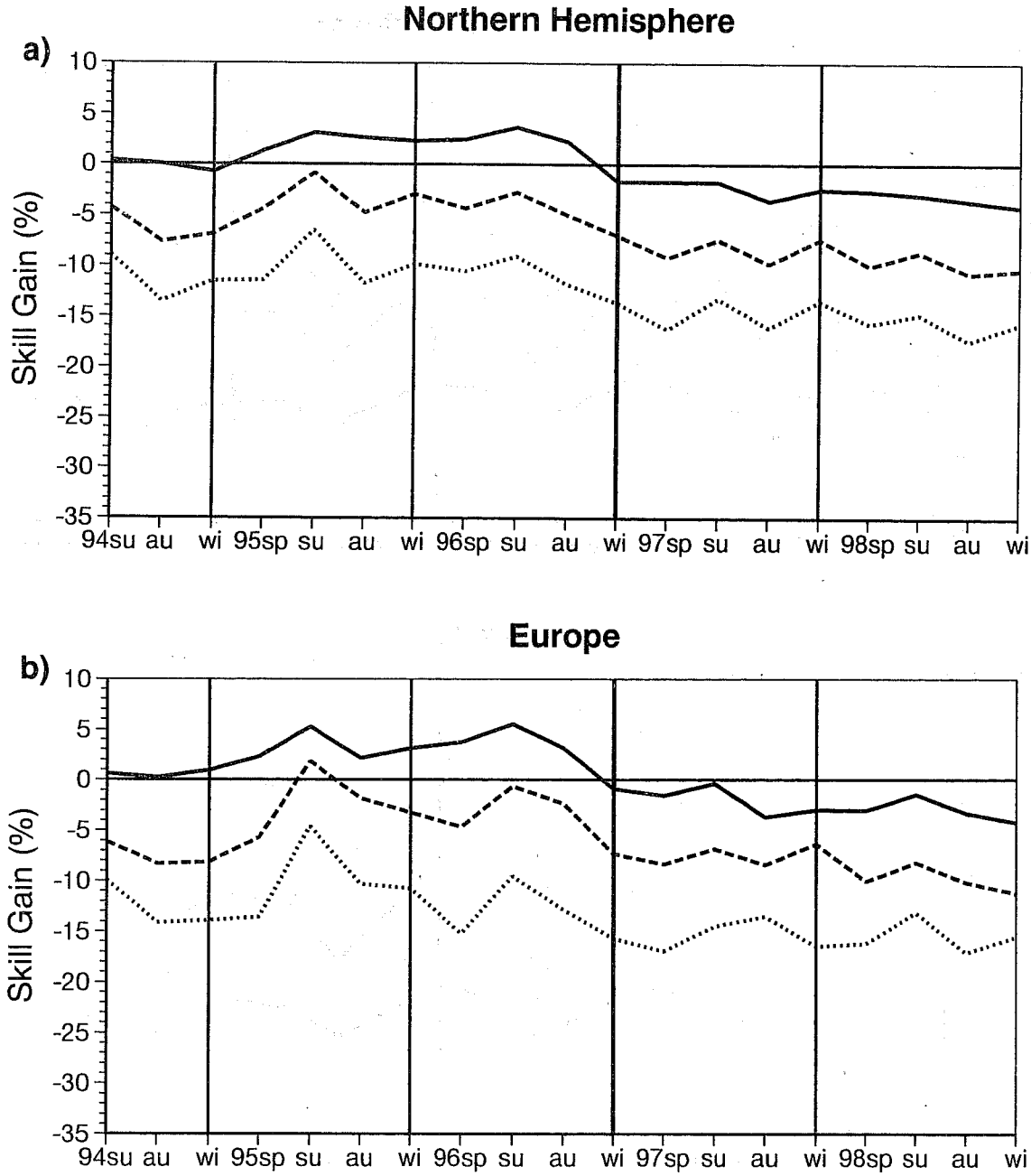

Fig. 4 Percentage of ensemble-mean skill gain, defined as the difference between the rms error of the ensemble mean *errm* and of the control error *errc* divided by the control error, *(errm-errc)/errc,* computed over the (a) NH and (b) Europe, at forecast day 3 (solid), 5 (dash) and 7 (dot). Vertical lines mark values for winter seasons (bold identifies the system upgrade of December 1996).

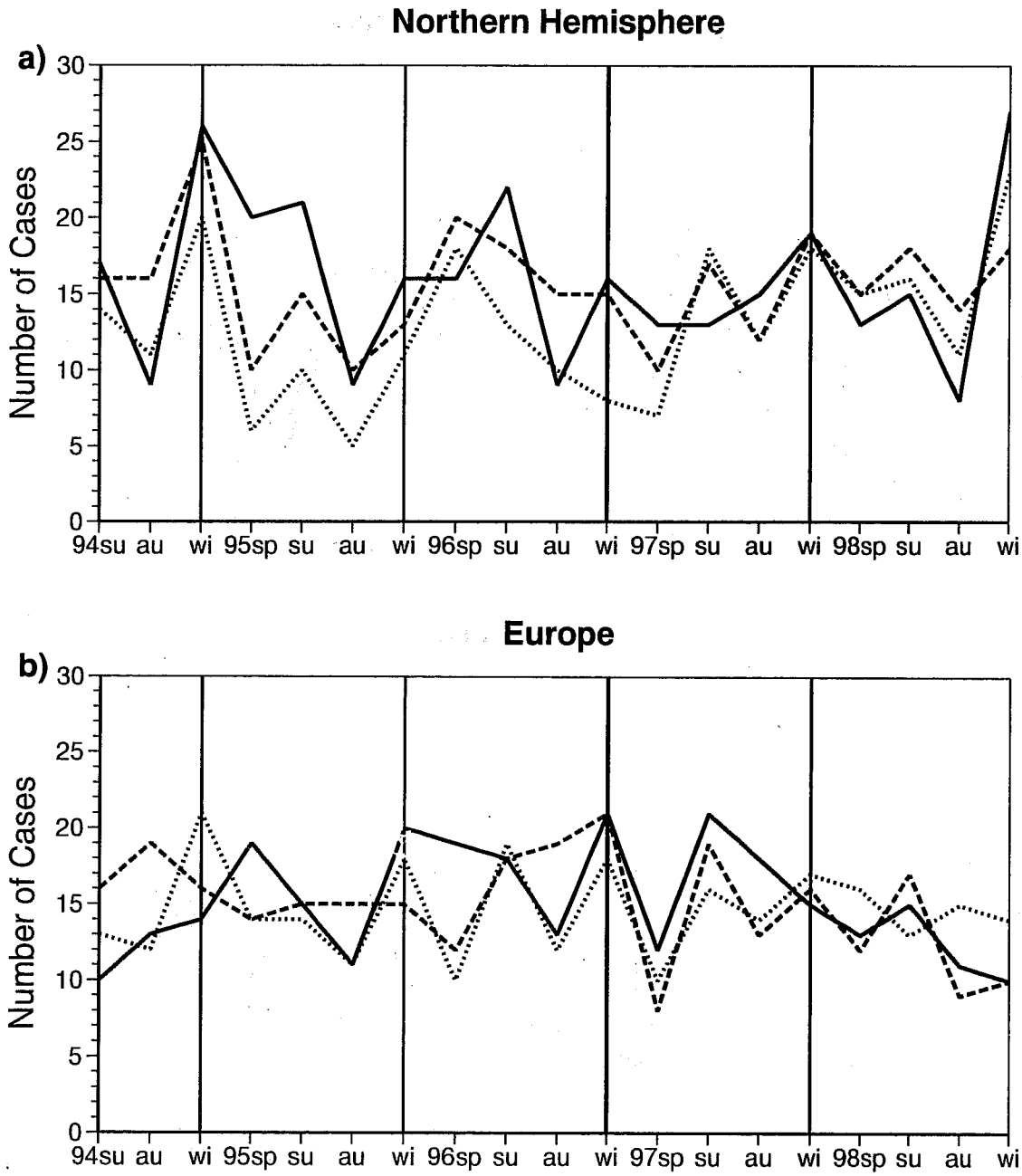

Fig. 5 **Number of cases with smaller-than-average ensemble rms spread and larger-than-average control rms error, computed over the (a) NH and (b) Europe, at forecast day 3 (solid), 5 (dash) and 7 (dot). Vertical lines mark values for winter seasons (bold identifies the system upgrade of December 1996).**

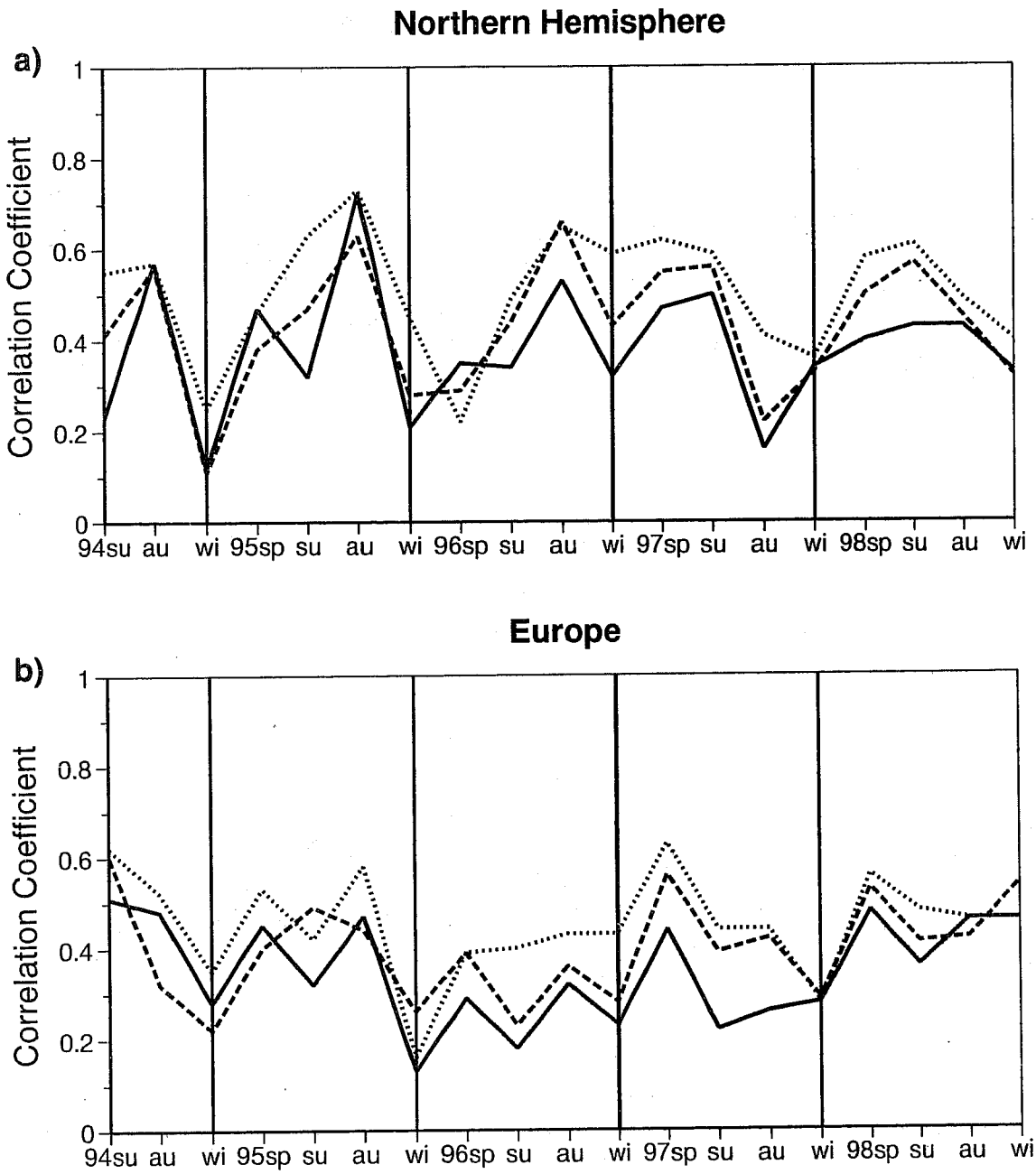

Fig. 6 Correlation coefficient between ensemble rms spread and the control rms error, computed over the (a) NH and (b) Europe, at forecast day 3 (solid), 5 (dash) and 7 (dot). Vertical lines mark values for winter seasons (bold identifies the system upgrade of December 1996).

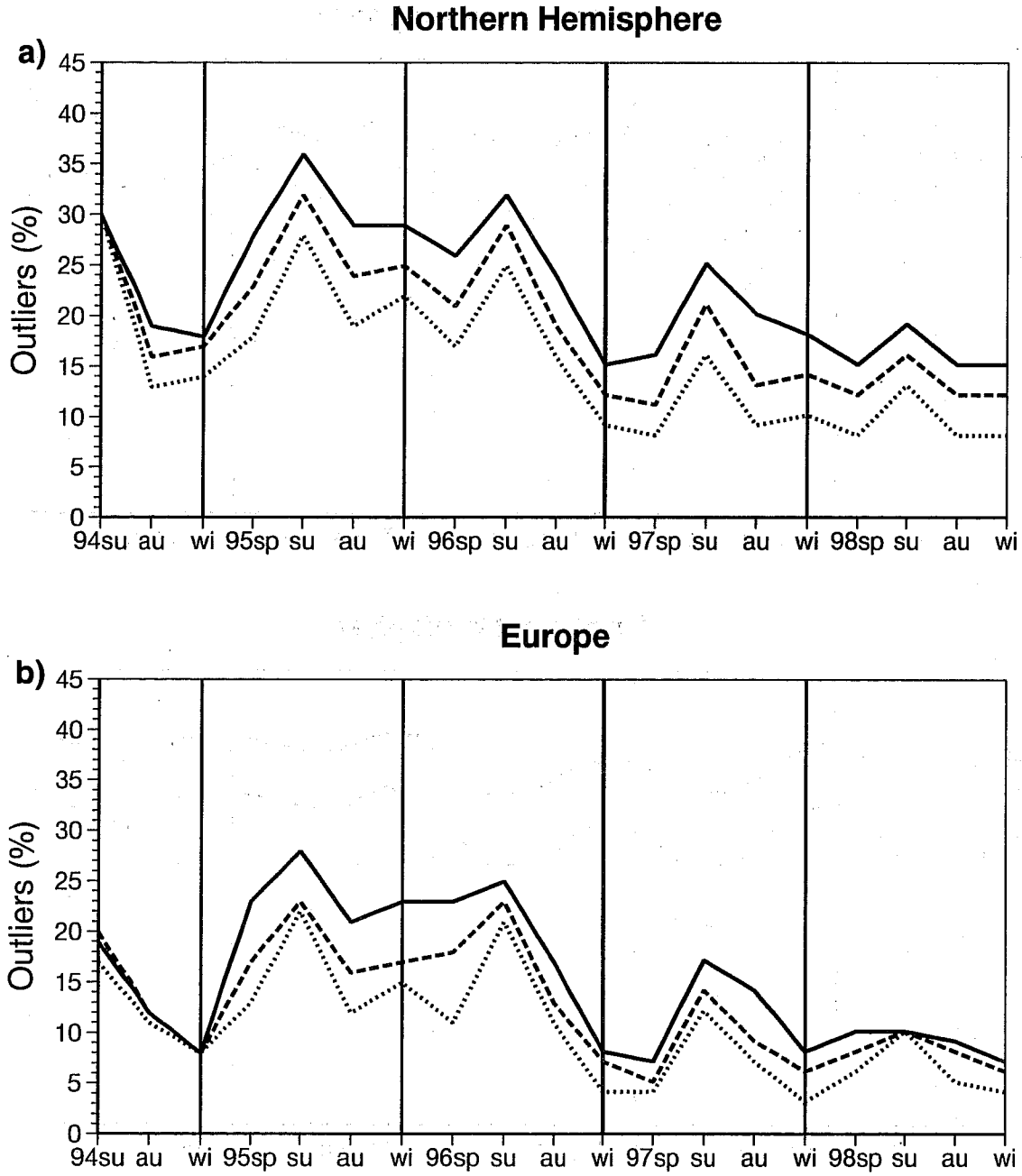

Fig. 7 Percentage of outliers, defined as the percentage of cases when the analysis is not included in the ensemble forecast range, minus reference (see text for details) computed over the (a) NH and (b) Europe, at forecast day 3 (solid), 5 (dash) and 7 (dot). Vertical lines mark values for winter seasons (bold identifies the system upgrade of December 1996).

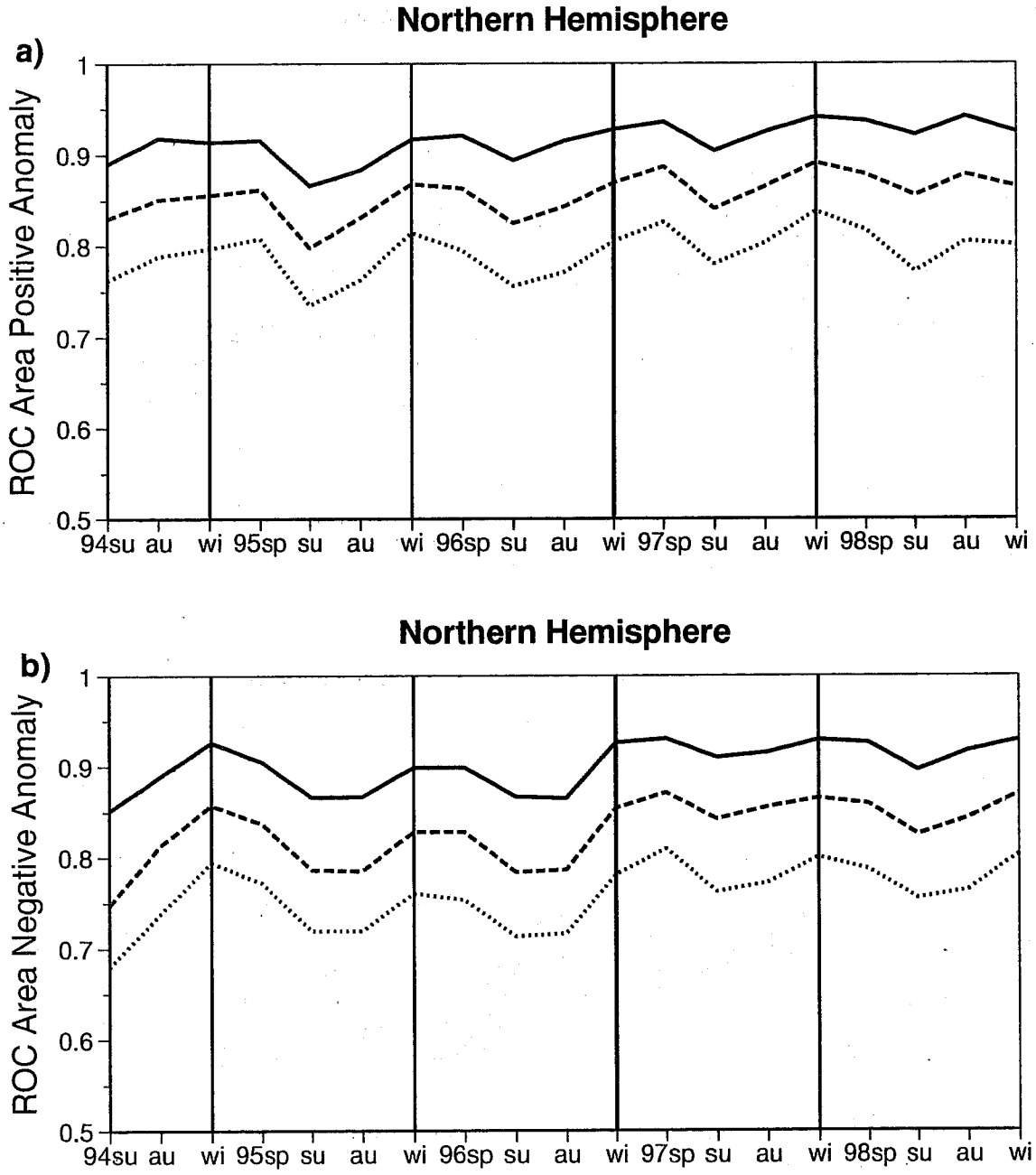

Fig. 8 (a) Area under the ROC curve for the probability prediction of 'geopotential height positive anomaly with magnitude greater than one standard deviation' over the NH, at forecast day 3 (solid), 5 (dash) and 7 (dot). (b) As (a) but for the event 'geopotential height negative anomaly with magnitude greater than one standard deviation'.

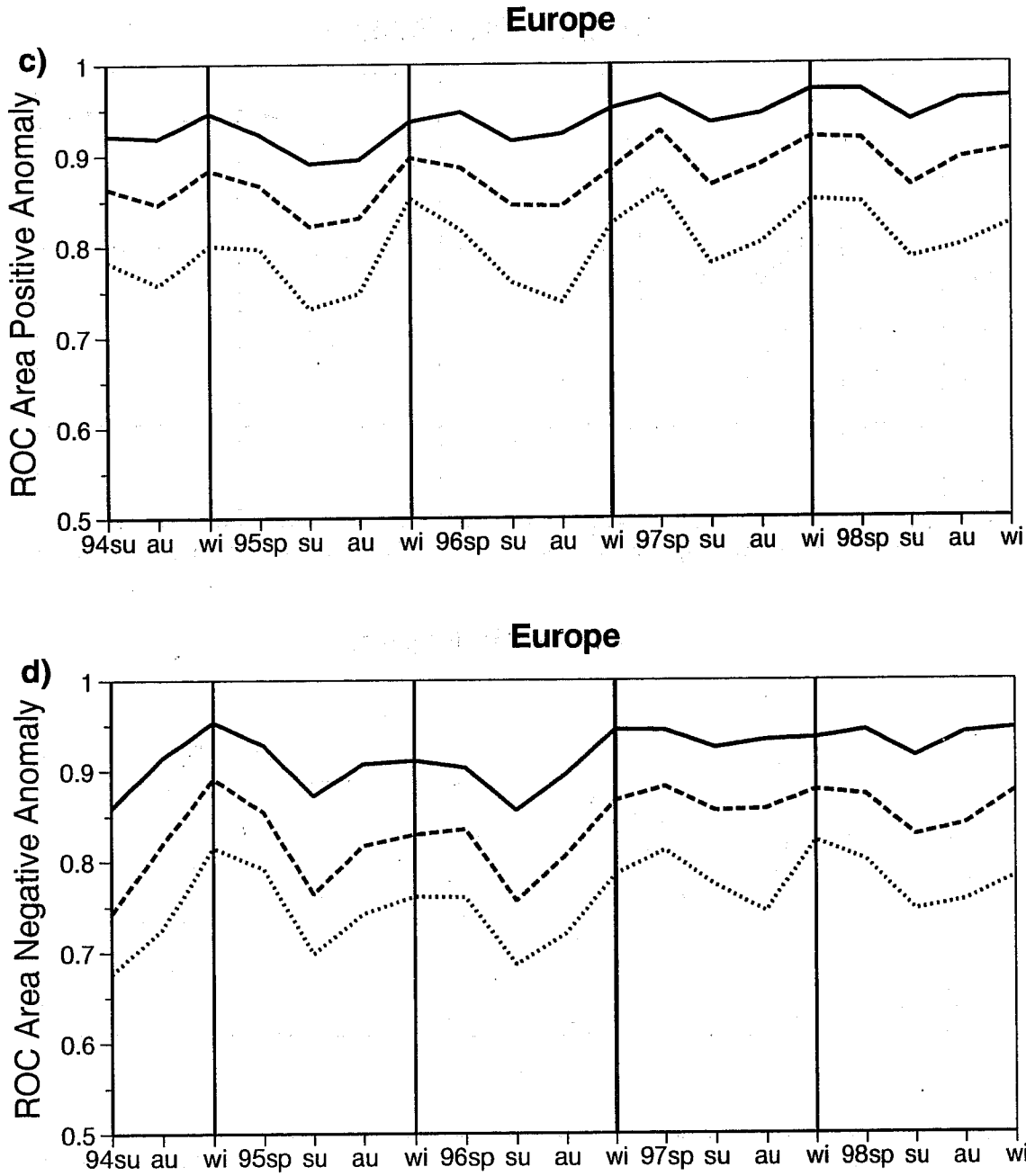

Fig. 8 (cont) (c), (d) As (a), (b) but for Europe. Vertical lines mark values for winter seasons (bold identifies the system upgrade of December 1996).

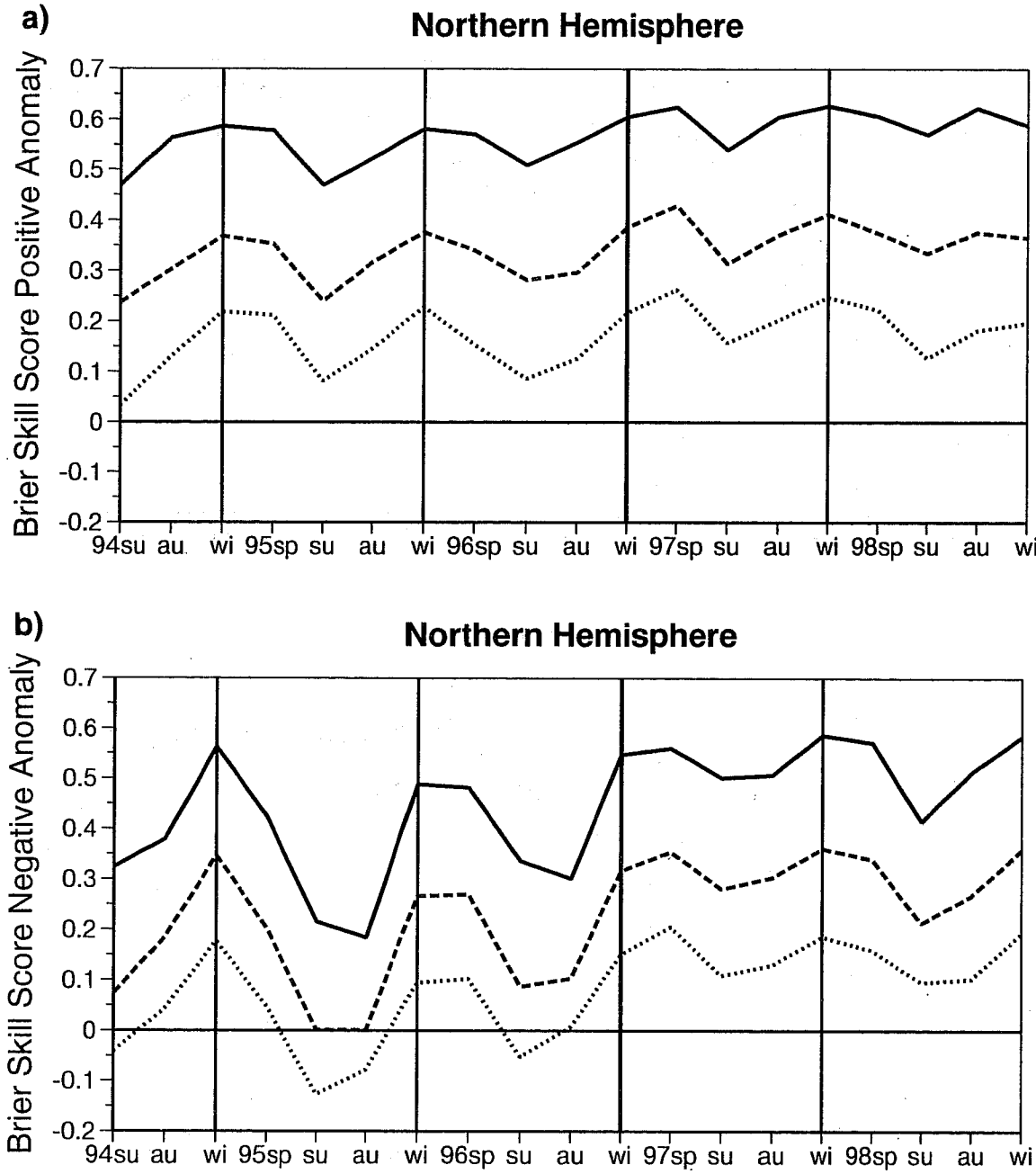

Fig. 9 (a) Brier skill Score for the probability prediction of 'geopotential height positive anomaly with magnitude greater than one standard deviation' over the NH, at forecast day 3 (solid), 5 (dash) and 7 (dot). (b) As (a) but for the event 'geopotential height negative anomaly with magnitude greater than one standard deviation'.

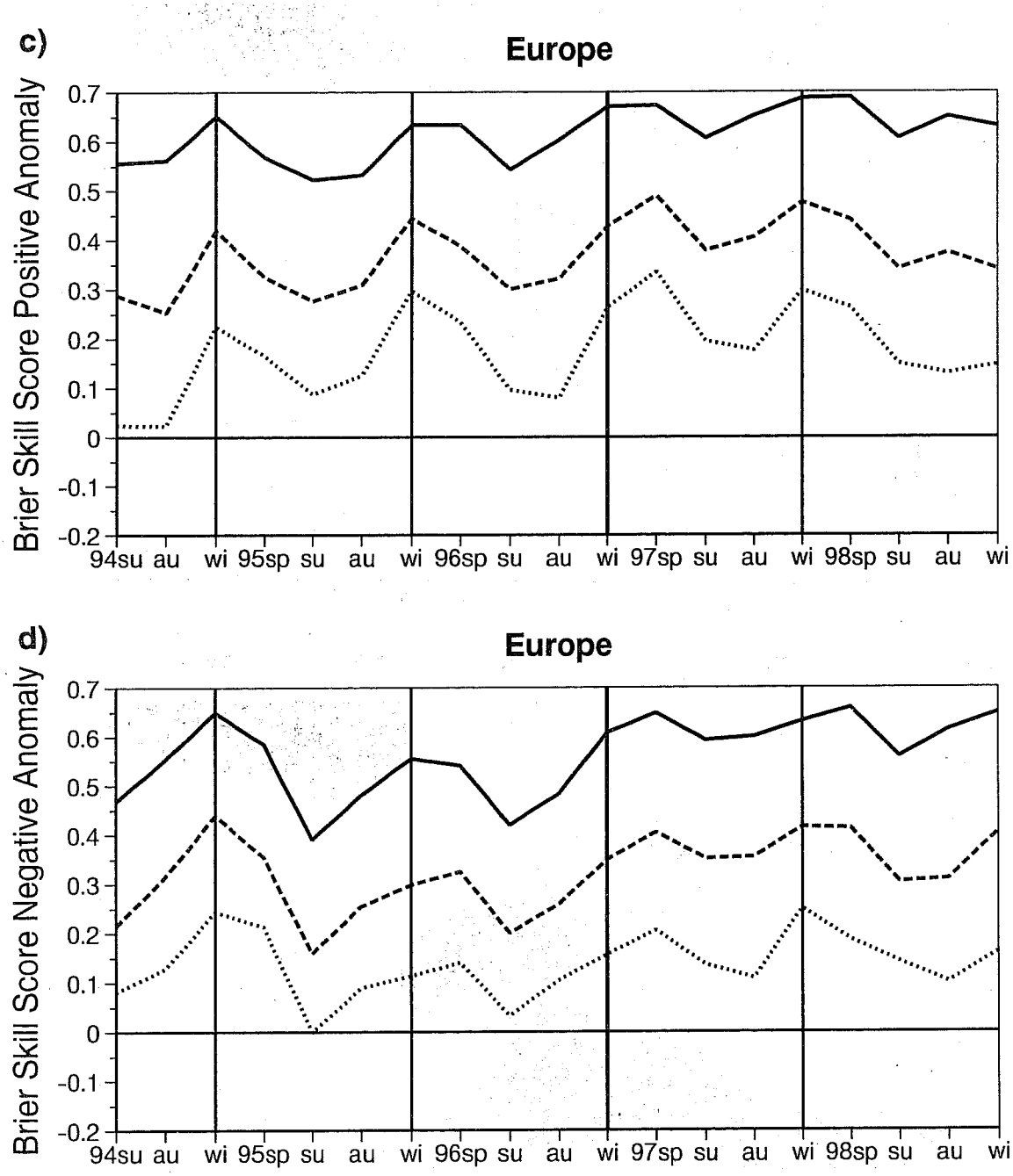

Fig. 9 (cont) (c), (d) As (a), (b) but for Europe. Vertical lines mark values for winter seasons (bold identifies the system upgrade of December 1996).

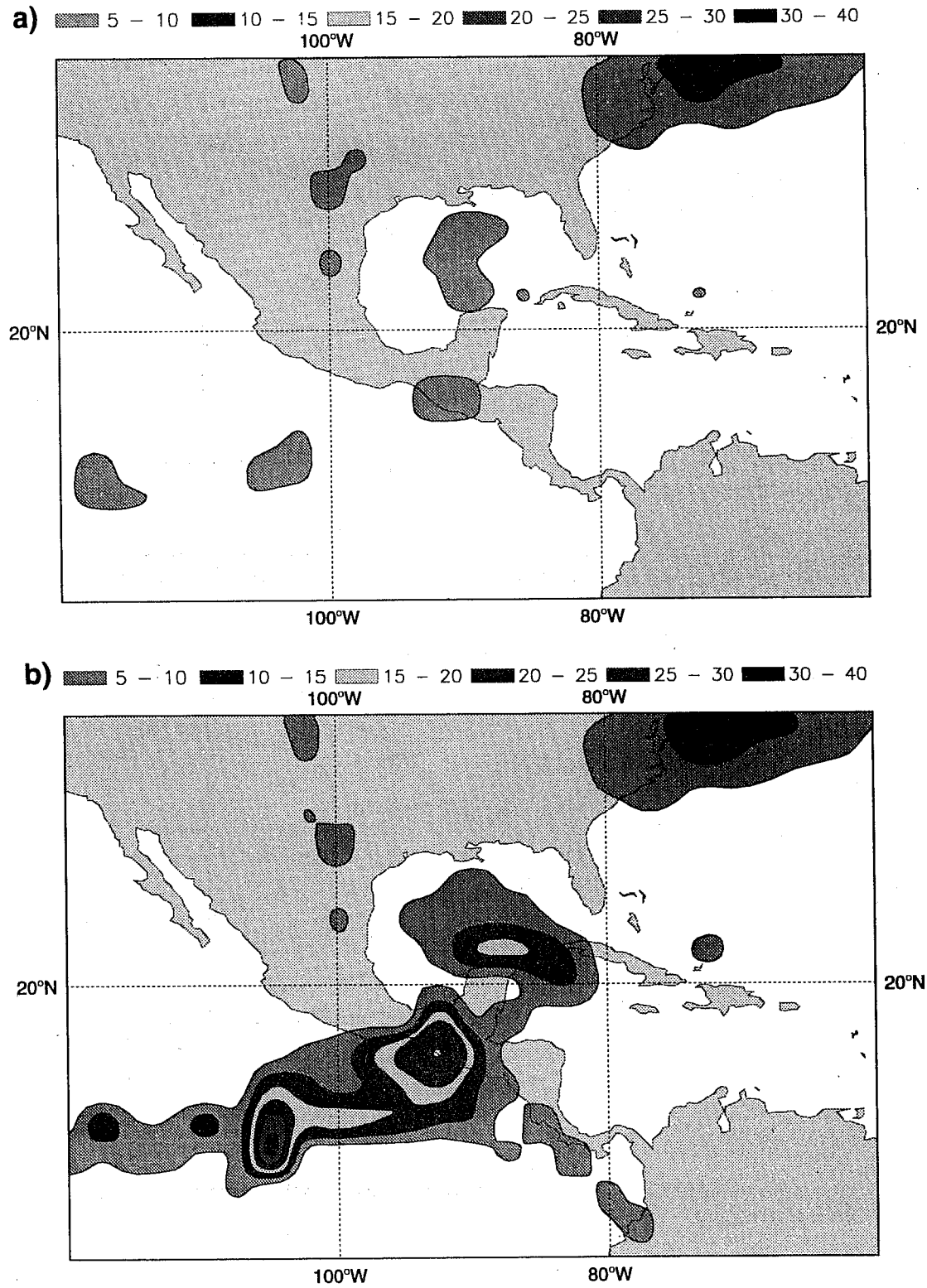

Fig. 10 Spread in terms of 24-h accumulated rain at forecast day 3 of ensembles started on 11 November 1998, (a) without and (b) with tropical singular vectors, over Central America. Contour isolines every 5 mm.

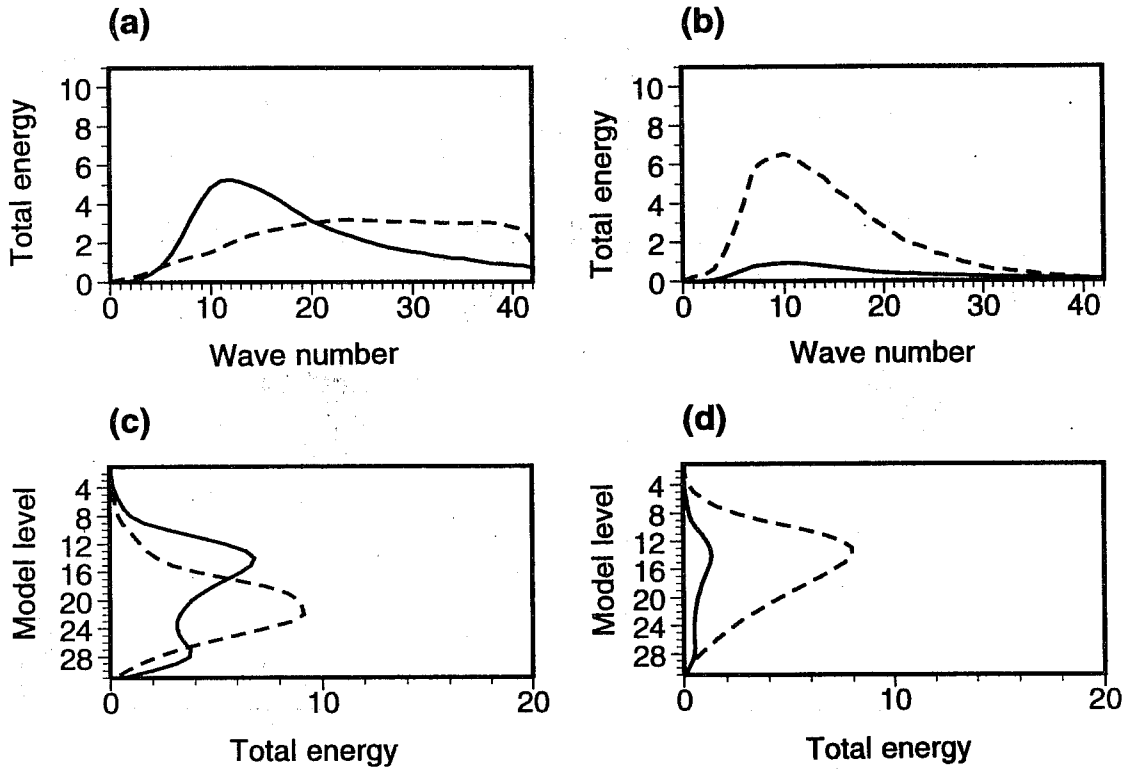

Fig. 11 Total energy spectrum of (a) total energy and (b) Hessian singular vectors, at initial (dash, values multiplied by 100) and final (solid) time. (c), (d) As (a), (b) but for the vertical profile of total energy.

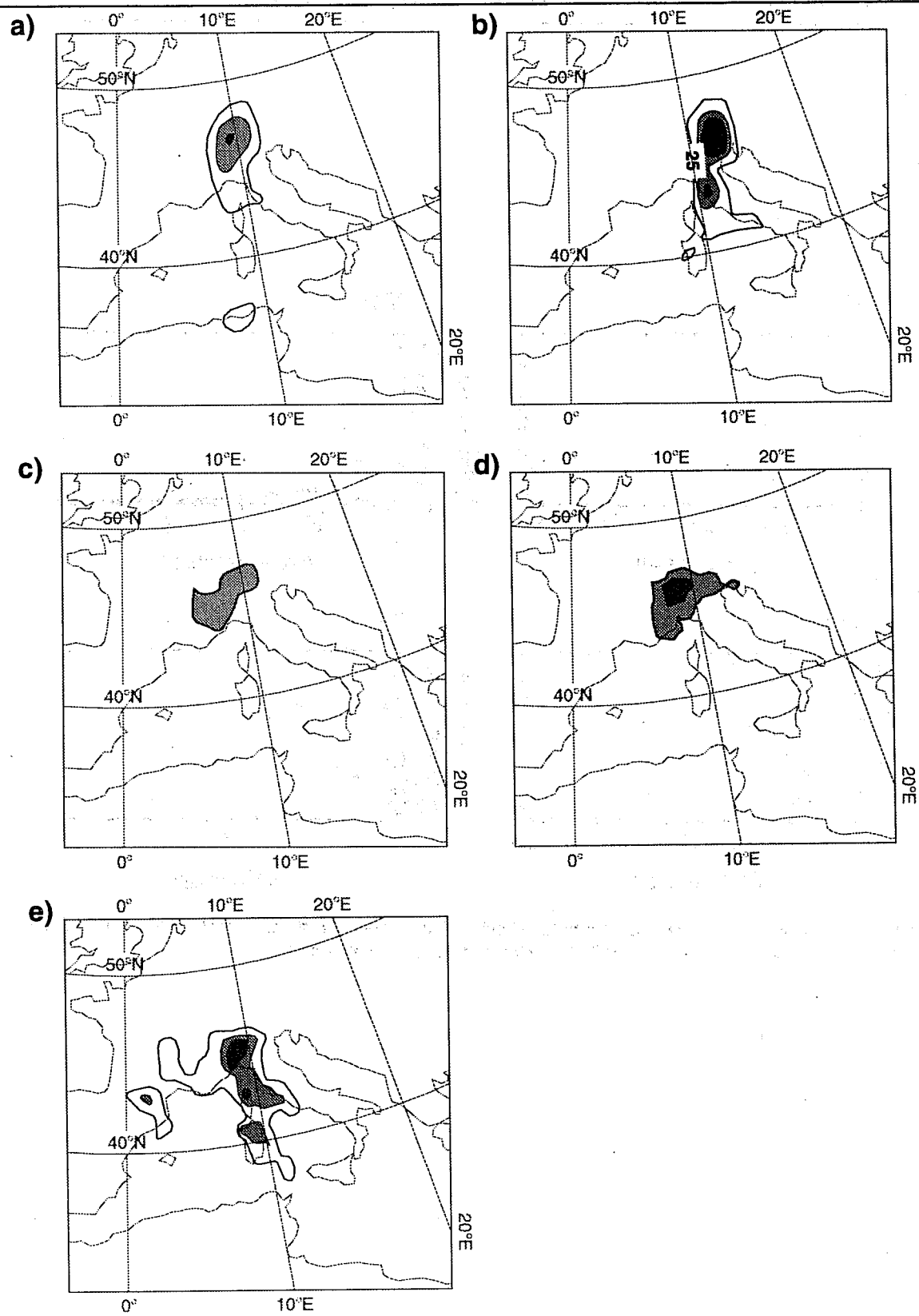

Fig. 12 24-h accumulated precipitation predicted between forecast at t+96h and t+120h from (a) the TI159 and (b) the TI319 control forecasts, and probability prediction of more than 50 mm of rain between forecast at t+96h and t+120h from (c) the TI159 and (d) the TI319 ensembles, for the forecast started at 12Z of 19 September 1993. (e) Verification, defined as the 00 24h accumulated precipitation predicted by the T213 model started at 12Z of 23 September 1993. Contouring: (a), (b), (e) isoline for 25 mm, grey shading for 25mm<prec<50mm and black shading for 50mm<prec; (c), (d) grey shading for 10%<prob<20%, and black shading for 20%<prob<30%.