TECHNICAL MEMORANDUM

279

Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System

R. Buizza, M. Miller and T.N. Palmer

Research Department

April 1999

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

ABSTRACT

A stochastic representation of random error associated with parametrized physical processes ("stochastic physics") is described, and its impact in the European Centre for Medium-Range Weather Forecasts Ensemble Prediction System (ECMWF EPS) is discussed. Model random errors associated with physical parametrizations are simulated by multiplying the total parametrized tendencies by a random number r sampled from a uniform distribution between 0.5 and 1.5. A number of diagnostics are described and a choice of parameters is made. It is shown how the scheme increases the spread of the ensemble, and improves the skill of the probabilistic prediction of weather parameters such as precipitation. A choice of stochastic parameters is made for operational implementation. The scheme was implemented successfully in the operational ECMWF EPS on 21 October 1998.

1. INTRODUCTION

Routine real-time execution of the European Centre for Medium-Range Weather Forecasts Ensemble Prediction System (ECMWF EPS) began in 1992 with a 31-member T63L19 configuration (spectral triangular truncation T63 and 19 vertical levels, *Palmer et al*, 1993, *Molteni et al*, 1996). A major upgrade to a 51-member T_L159L31 system (spectral triangular truncation T_L159 with linear grid) took place in 1996 (*Buizza et al*, 1998). From its inception, the EPS has been based on the premise that medium-range forecast errors are predominantly associated with uncertainties in initial conditions. As such, the EPS is based on multiple integrations of the ECMWF operational model (albeit at lower resolution) from an ensemble of initial conditions, created by adding perturbations to the operational analysis.

The philosophy of basing the EPS on perturbed initial conditions, or in other words on a "perfect forecast model assumption", is consistent with results from *Downton and Bell* (1988), and, more recently, from *Richardson* (1998). In these studies, substantial forecast differences between the ECMWF and the UKMO (United Kingdom Meteorological Office) operational models could mostly be traced to differences between the two operational analyses, rather than between the two forecast models. A similar "perfect model" strategy is followed at the US National Centers for Environmental Prediction (NCEP), using initial perturbations generated using bred-vectors (*Tracton and Kalnay*, 1993; *Toth and Kalnay*, 1993).

On the other hand, recent results from *Harrison et al* (1999) indicate that the impact of model uncertainties on forecast error cannot be ignored, and that in some respects an ensemble system based on two (or more) models may be superior to an ensemble based on just one.

There is further evidence that uncertainties in model formulation may be a significant factor in accounting for forecast errors in the medium range. Specifically, although the root-mean-square (rms) spread of the ECMWF EPS agrees well with the error of the unperturbed control forecast at around day 2, the spread is notably smaller than the control rms error later in the medium range.

Whilst the use of perturbations with larger initial amplitude would reduce spread underestimation in the late forecast range, it would also deteriorate performance in the early range, when all perturbed forecast would be unrealistically far from the control forecast compared with the analysis.

An approach to the representation of model uncertainties by inflating the divergence between perturbed and unperturbed forecasts during the integration of the ensemble forecast has been suggested by *Toth and Kalnay* (1995). Their preliminary results suggested that this approach could lead to a better agreement between ensemble spread and control error, and to a more skilful ensemble mean.

Houtekamer et al (1996) first included model uncertainties in ensemble prediction. Following a system simulation approach to ensemble prediction, they developed a procedure where each ensemble member differs both in the initial conditions, and in subgridscale parameters. In this approach, each ensemble member is integrated using different parametrizations of horizontal diffusion, convection, radiation, gravity wave drag, and with different orography.

There are certainly good grounds for believing that there is a significant source of random error associated with the parametrized physical processes. For example, consider a grid point over the tropical warm pool area during a period of organised deep convection. By definition, the actual contributions to the tendencies due to parametrized physical processes are often associated with organised mesoscale convective systems whose spatial extent may be comparable with the model resolution. In such a case, the notion of a quasi-equilibrium ensemble of sub-grid-scale processes, upon which all current parametrizations schemes are based, cannot be a fully-appropriate concept for representing the actual parametrized heating (e.g. *Palmer*, 1997). For example, even if the parametrized heating fields agree on average (i.e. over many time steps) at the chosen grid point, there must inevitably be some standard deviation in the time-step by time-step difference between observed and modelled heating.

In this paper a simple stochastic scheme for perturbing the parametrized tendency is developed, and its impact on the performance of the EPS is studied. The scheme is based on the notion that the sort of random error in parametrized forcing, discussed above, will be coherent between the different parametrization modules, and will have a certain coherence on the space and time scales associated, for example, with organized convection schemes. Moreover, the scheme assumes that the larger the parametrized tendencies, the larger the random error component will be. The notion of coherence between modules allows the stochastic perturbation to be based on the total tendency from all parametrized processes, rather than on the parametrized tendencies from each of the individual modules. In this case the present scheme differs conceptually from that of *Houtekamer et al* (1996). The notion of space-time coherence assumes that organised systems have some intrinsic space and time-scales which may span more than one model time step and more than one model grid point. Making the stochastic uncertainty proportional to the tendency is based on the concept that organisation (away from the notion of a quasi-equilibrium ensemble of sub-grid processes) is likely to be stronger, the stronger is the parametrized contribution. A certain space-time correlation is introduced in order to have tendency perturbations with the same spatial and time scales as observed organisation.

Full details of the scheme are described in section 2; the impact of the scheme on numerical integrations is investigated in section 3. The impact of stochastic forcing with different spatio-temporal correlation and amplitude on the ECMWF EPS is analyzed in section 4. Further results from tests leading to the operational implementation of the scheme in the ECMWF EPS are discussed in section 5. Conclusions are summarised in section 6.

2. STOCHASTIC FORCING SIMULATION

2.1 The original ECMWF Ensemble Prediction System

The original ECMWF EPS (*Palmer et al*, 1993, *Molteni et al*, 1996, *Buizza et al*, 1998) can be described schematically as follows. Each ensemble member e_j (with j=0 identifying the control) can be seen as the time integration

$$e_j(t) = \int_{t=0}^{t} [A(e_j;t)+P(e_j;t)]\ dt \tag{1}$$

of the model equations

$$\frac{\partial e_j}{\partial t} = A(e_j;t)+P(e_j;t)\ , \tag{2}$$

(where A and P identify the non-parametrized and the parametrized processes) starting from initial conditions

$$e_j(t=0) \equiv e_0(t=0)+\delta e_j(t=0)\ , \tag{3}$$

where $e_0(t=0)$ is the operational analysis at $t=0$. The initial perturbations $e_j(t=0) \equiv e_0(t=0)+\delta e_j(t=0)$ are generated using the singular vectors of the linear version of the ECMWF, computed to maximize the total energy norm over a 48-hour time interval (*Buizza and Palmer*, 1995), and scaled to have an amplitude comparable to analysis error estimates.

2.2 The new ECMWF Ensemble Prediction System

Compared with the original, the new system includes a scheme to simulate model uncertainties deriving from parametrized physical processes. Schematically, each ensemble member e_j can be seen as the time integration

$$e_j(t) = \int_{t=0}^{t} [A(e_j,t)+P'(e_j,t)]\ dt \tag{4}$$

of the perturbed model equations

$$\frac{\partial e_j}{\partial t} = A(e_j;t)+P'(e_j;t)\ , \tag{5}$$

starting again from the perturbed initial conditions as in Eq. (3). For each grid point $x=(\lambda,\phi,\sigma)$ (identified by its latitude, longitude and vertical hybrid coordinate), the perturbed parametrized tendency (of each state vector component) is defined as

$$P'_j(e_j;t) \equiv [<r_j(\lambda,\phi;t)>_{D,T}]\ P_j(e_j;t)\ , \tag{6}$$

where $<..>_{D,T}$ means that the same random number r_j has been used for all grid points inside a DxD degree box and over T time steps. Random numbers have been sampled uniformly from three different intervals for so-called high, medium and low amplitude stochastic forcing configurations,

- high (H): $r_j \in [0,2.00]$, (7a)

- medium (M): $r_j \in [0.5,1.50]$, (7b)

- low (L): $r_j \in [0.75,1.25]$. (7c)

Note that the stochastic term re-scales the parametrized tendency for any component of the state vector (i.e. for wind, temperature and specific humidity) for the entire grid point column by the same factor $<r_j>_{D,T}$, and different measures of spatio-temporal autocorrelation through the D and T parameters. More sophisticated formulations will be tested in due course. Specifically, configurations have been tested with $D=1°$ *deg* (low, L), $D=5°$ *deg* (medium, M) and $D=10°$ *deg* (high, H), and with $T=45$ *min* (low, L), $T=3$ *h* (medium, M) and $T=12$ *h* (high, H). Note that the current EPS spectral resolution T_l159L31 corresponds to about 1° degree resolution in physical space, and thus in experiments with $D=1°$ *deg* each grid point uses a different random number. Analogously, with an integration time step of 2700 seconds (which was the time step used at the time of this work in the T_l159L31 non-linear integrations), with $T=45$ *min* random numbers are re-sampled every time step.

A configuration (e.g. L-M-H) is specified in order of amplitude, spatial D and temporal T autocorrelation parameters.

2.3 Experimental set-up

Single deterministic integrations and ensembles have been run for five case studies (starting dates 96.10.23, 96.12.19, 97.05.14, 97.07.02 and 97.07.31) in many but not all 27 possible configurations. Considering configuration M-M-M as starting point, sensitivity experiments have been run with both stronger (M-H-M, M-M-H, H-L-L and H-L-M) and weaker (M-M-L, M-L-M, M-L-L, L-H-H, L-H-M, L-M-H, and L-L-H) forcing. First, these configurations have been tested performing single deterministic runs, and comparing the score of perturbed and unperturbed forecasts. Such single deterministic runs indicated that, for example, the stochastic forcing induced by configurations H-M-L or H-M-M is too strong, while the stochastic forcing induced by configurations L-M-M, L-H-L, L-L-L is ineffective.

Results from single deterministic forecasts determined the configurations to be tested in the EPS, with ensemble performed at T_l159L31 resolution, with 50 perturbed members, and with the ECMWF IFS model cycle CY16R1.

Ensembles have been run with stochastic forcing only (ST ensembles), with initial perturbations only (IC ensembles), and with stochastic forcing and initial condition perturbations (ICST ensembles). Both the IC and the ICST ensemble initial perturbations have been defined using T42L31 singular vectors, optimized over a 48-hour time interval to have maximum total energy norm in the Northern Hemisphere (NH) extra-tropics, and set to have local initial amplitude comparable to analysis error estimates provided by the ECMWF Data Assimilation procedure. These settings are as in the ECMWF EPS operational at the time of this work (autumn 1997 and winter 1997-98, *Buizza et al*, 1998).

3. IMPACT OF STOCHASTIC FORCING IN NUMERICAL INTEGRATIONS

The section describes results from a set of sensitivity studies have been performed on deterministic numerical integrations. Note that, because of the absence of moist processes in the forward and adjoint tangent models used here, ensemble initial perturbations are optimized for the extra-tropics. Thus, the impact of stochastic forcing in the tropics is analyzed only for single deterministic forecasts. Since the relative role of the parametrized terms is much larger in the deep tropics, the perturbation technique is correspondingly tested more carefully. Attention is restricted to the reference configuration IC and to the perturbed configuration ICST M-M-M.

3.1 Precipitation comparison with TOGA COARE data

Lin and Johnson (1996) compared rainfall rates from moisture budget and from ECMWF forecasts with rates from rainfall estimations from satellite data (SSM/I data, from the DMSP Special Sensor Microwave/Imager, and GPI data, from the GOES Precipitation Index). Focusing on the TOGA COARE intensive flux area located over the western Pacific warm pool (150°E-160°E;10°S-0°) and over a time period from 1 November 1992 to 28 February 1993. It was shown (see Fig. 8 in *Lin and Johnson*, 1996) that forecasts and rainfall estimations differ substantially, with peaks of more than 20 mm/day. The differences between the ECMWF forecasts and the different estimates can be considered as an upper limit to acceptable rainfall differences induced by stochastic forcing.

Deterministic 10-day forecasts with starting dates 1 to 28 December 1992, without and with M-M-M stochastic forcing have been run, and their rainfall predictions in the TOGA COARE intensive flux area have been compared. Figure 1 shows daily values of precipitation forecast accumulated between the 48 and the 72 hour forecasts and observation estimates from GPI retrievals (as in *Lin and Johnson*, 1996). Results indicate that 24-hour rainfall forecast differences are smaller than the differences between each of the forecasts and the GPI estimates. As such, M-M-M forcing does not appear to be excessive.

3.2 Mean impact of stochastic forcing on forecast precipitation

Since precipitation is sensitive to changes in parametrized tendencies, one way to assess the impact of stochastic forcing is to examine monthly average precipitation fields for the deterministic forecasts run for December 1992 without and with the M-M-M stochastic forcing. Indeed, the monthly mean precipitation accumulated between forecast days 6-7 of the two experiments are very similar (Fig. 2a-b), with differences mostly concentrated in the tropics (Fig. 2c). Similar differences characterize mean precipitation accumulated between forecast days 1-2 (not shown). A study of the impact of the stochastic physics scheme on the climatology of the model is in progress and will be reported elsewhere.

3.3 Parametrized tendency diagnostics

The impact of the M-M-M stochastic forcing on the parametrized tendencies has been investigated on single deterministic forecast with starting date 96.10.23 (Fig. 3).

Ten 10° square regions characterized by very different weather conditions have been defined, and average unperturbed and M-M-M stochastically perturbed tendencies have been compared for the first 48 hour forecast. Results for three of these regions will be discussed here: the first region is located under the Pacific anticyclone (140°W-130°W;30°N-40°N), the second region is located in the Rockies and characterized by intense precipitation (130°W-120°W;40°N-50°N), and the third region is located in the Sahara desert (10°W-0°W;20°N-30°N). All results relate to model level 23 (around 700 hPa).

Figure 4 shows the time variation of normalized wind, temperature and specific humidity tendencies (each tendency has been normalized by its time average value), and the time variation of the stochastic scaling factor $<r_j>_{D,T}$ for the first region (Pacific, dry, see Fig. 3). Note that only at the end of the forecast period are both the wind tendencies (Fig. 4a-b) and the stochastic scaling factors (Fig. 4e) large, and the impact on the tendencies becomes visible. Compared to the first region, the tendencies for the second region (Rockies, wet, see Fig. 3) are more variable (only the tendency for specific humidity is shown in Fig. 5a), essentially because of moist processes (note that the average value for specific humidity tendency is about 50 times larger than that for region 1). The third region (Sahara desert, dry, see Fig. 3) is characterized by small wind tendencies, by a strong daily cycle in the temperature tendency (Fig. 5c), and by small specific humidity tendencies (not shown). Similar results are obtained for the other 7 regions. Generally speaking, results show that the character of the tendency time variation does not change between the unperturbed and the perturbed cases.

4. IMPACT OF STOCHASTIC FORCING IN ENSEMBLE PREDICTIONS

Sensitivity studies of the impact of different stochastic forcing configurations on the ECMWF EPS are discussed in this section.

4.1 rms spread and rms error

Table 1a summarizes the impact of stochastic forcing on the ensemble spread; Table 1b and 1c on mean rms error of the perturbed members and on rms error of the ensemble mean.

Table 1a shows that, for both ST and ICST ensembles, configuration M-M-H is characterized by the largest spread, followed closely by configuration H-L-M, and then by configurations L-M-H, M-H-M and M-M-M. Table 1a shows that a stochastic forcing with small amplitude unchanged for 12 hours (L-M-H) can be as effective as a forcing with double amplitude unchanged for 3 hours (M-M-M). Moreover, Table 1a suggests that configurations in which the random forcing is changed every time step have a small impact on ensemble.

Focusing on forecast error at day 7, Table 1b shows that there are six ICST configurations for which the average rms error of the perturbed forecasts is smaller than the reference value given by configuration IC. Table 1c shows that configurations M-M-H and M-L-M have the most skilful ensemble mean.

In summary (Table 1d), configuration M-M-M is the only one with increased ensemble spread, decreased average rms error of the perturbed forecasts, and decreased ensemble mean rms error. Four other configurations (H-L-M, M-M-H, L-H-H and L-M-H) have larger ensemble spread and a positive impact on at least one of the skill scores.

The comparison of the skill of the ensemble mean for the full forecast range for these 5 configurations indicates that configurations H-L-M and M-M-H have a negative impact on the skill of the ensemble mean up to forecast day 4. Hereafter, attention will focus on configurations M-M-M, L-H-H and L-M-H, which have similar spread throughout the whole forecast range.

It is interesting to note that these results are comparable to the findings of *Houtekamer et al* (1996), who show (their Table 2) that by perturbing the model the spread at forecast day 7 is increased by about 5.6%.

4.2 Other measures of ensemble skill

Considering the results discussed in the previous section, configurations M-M-M, L-H-H and L-M-H are compared with the reference IC EPS configuration using other measures of ensemble skill.

Significant differences among the four configurations were detected when two grid-point measures of ensemble skill were used, the percentage of analysis values lying outside the ensemble forecast range (percentage of outliers) and the rms error of the so-called grid-point best ensemble member (*Buizza and Palmer*, 1998). Table 2 shows that, according to these two measures, the stochastically perturbed ensembles perform better using these measures.

Comparison of precipitation predictions by the ensembles are made using probabilistic measures such as the area under a Relative Operating Characteristic (ROC) curve and the Brier skill score (*Stanski et al* 1989, *Palmer et al.* 1993, *Molteni et al.* 1996, *Zhu et al* 1996, *Toth et al* 1998, *Buizza at al* 1999). Given a forecast of the occurrence/non-occurrence of a specific event, the ROC curve is a graph of the hit and false alarm rates, and the area under a ROC curve is a measure of skill. A ROC area of 0.5 is considered as the lower bound for a skilful forecast, since a system with such a ROC area cannot discriminate between occurrence and non-occurrence of the event.

The comparison of the area under a ROC curve for the prediction of the events "12-hour accumulated precipitation greater than 2, 10, 20 and 30 mm/day" (Fig. 6) indicate that ensembles run with stochastic forcing are more skilful than the reference IC configuration. Defining an area of 0.7 as the limit for useful prediction (*Buizza at al* 1999), Fig. 6d shows that about 18 hours of predictability are gained for the largest precipitation threshold. This positive impact is related to the increased ensemble spread induced by stochastic forcing, which increases the hit rates of forecasts of precipitation events without a proportional increase in false-alarm rates. A similar although smaller positive impact can be detected in the Brier skill score for prediction amounts up to 20 mm/day (at most, predictability increases by about 12 hours for the 10 mm/day threshold, not shown).

4.3 Synoptic evaluation for 500 hPa geopotential height

The structure of the forecast divergence induced by stochastic forcing is analyzed in this section. Since differences between the three stochastically perturbed ensembles M-M-M, L-H-H and L-M-H are small, attention is restricted to configuration M-M-M and to the reference ensemble IC.

Table 3 lists the rms error over Europe (20°W-45°E;30°N-75°N) of the best IC and ICST M-M-M ensemble member and of the control forecast, for the 500 hPa geopotential height at forecast day 7 for the 97.07.31 case. The error of the IC 8th member (Fig. 7a) is notably smaller than the control error (Fig. 7c), because its divergence from the control (Fig. 7e) brings it closer to the verifying analysis (Fig. 7d). Furthermore, the error of the ICST M-M-M 8th member (Fig. 7b) is smaller than that of the equivalent IC 8th member over England and Russia because the stochastic forcing increases the geopotential height in this region (Fig. 7f).
A more complete picture of the difference between the IC and ICST 8th members at forecast day 7 in terms of temperature is given in Fig. 8 for the whole Atlantic-European region, and in Fig. 9 for two vertical cross sections, one from the Labrador Peninsula to Sardinia (Fig. 9a-b), and one from Newfoundland to the North Cape (Fig. 9c-d). Generally speaking, these figures confirm that (within the current set-up) initial perturbations lead to similar scale but larger forecast divergence than stochastic forcing. In some regions, e.g. around Iceland (Fig. 9a-b), stochastic forcing reinforces the divergence induced by the initial perturbation, while in other regions, as around (40°W,57°N) (Fig. 9a-b), stochastic forcing produces a divergence which is opposite to the one given by the initial perturbation. Further, there are regions, as in the Norwegian sea (Fig. 8a-b), where the stochastic forcing is active but the initial perturbation has no effect on forecast divergence.

Similar conclusions could have been drawn from other case studies.

5. FURTHER RESULTS

Following the results reported in section 4, further tests have been performed using a more recent version of the ECMWF model (cycle CY18R6) for one summer and one winter week. By contrast to the results presented so far, these new experiments have been performed using NH initial time and evolved singular vectors, to take into consideration the latest EPS changes introduced on 25 March 1998 (*Barkmeijer et al*, 1999).

By contrast with the experiments discussed so far, a check has been introduced to avoid supersaturation leading to possible inconsistencies in the cloud scheme (*C. Jakob*, 1998, personal communication). For any grid point and any time step, if the updated specific humidity (after stochastic tendency perturbation) is greater than the saturation value, then the temperature and specific humidity tendencies are not perturbed. As might be anticipated, the introduction of this check reduces the dispersion induced by the stochastic physics. It has been found that a spread similar to the one induced by configuration M-M-M without the check is therefore given by a configuration with amplitude randomly selected between 0.5 and 1.5 (as in configuration M-M-M discussed above), but with higher spatio-temporal correlations; more precisely, by using the same random number for all grid points inside a 10°x10° and updated every 6 hours.

Results from a reference set of experiments and a stochastically-perturbed set with the super-saturation check are summarized. To reduce computer costs, only NH singular vectors were used, as for the ensembles discussed in

section 4. Results relate to one NH summer (29 June to 5 July 1997) and one NH winter (16 to 22 December 1997) period.

Considering 500 hPa geopotential height over the NH, the stochastically perturbed ensemble has a slightly larger spread especially in the NH summer period, and is characterized by a slightly larger average rms error of the perturbed members (not shown). No detectable impact is shown in the rms error of the ensemble-mean (not shown).

Figure 10 shows, for the NH, the impact of the stochastic scheme on probability precipitation prediction measured by the area under the ROC curve for the events "12 hour cumulated precipitation greater than 5, 10 and 20 mm". The positive impact of the scheme is more evident during the NH summer period for the higher thresholds. Similar results are obtained using the Brier Skill Score (not shown).

Since Southern Hemisphere (SH) singular vectors have not been used in these tests, the impact of the operational implementation of stochastic forcing on the performance of the EPS for the SH cannot be estimated quantitatively, although it is expected to be qualitatively similar to the impact over the NH. For completeness some results for the Tropics (30°S-30°N) are briefly discussed.

Figure 11 shows, for 850 hPa Temperature, the impact of the stochastic forcing on the rms spread over the NH and the Tropics (30°S-30°N). The impact on skill measures such as the rms error of the ensemble-mean or the average rms error of the perturbed forecast is very small (not shown). Finally, Fig. 12 shows the impact of stochastic forcing on the area under the ROC curve for the prediction of the event "850 hPa temperature warmer than average". The impact is small but positive over the NH, while it is positive over the Tropics. Similar conclusions can be drawn from Brier Skill Scores (not shown).

6. CONCLUSIONS

A simple approach to representing unavoidable random errors associated with the parametrization of subgridscale physical processes is described. The approach is based on adding a stochastic perturbation to the tendency due to the parametrized physical processes, with amplitude proportional to the total parametrized tendency itself. In its current formulation, there are three basic parameters governing overall amplitude, and spatio-temporal correlation. A number of diagnostics have been described and applied to single deterministic and ensemble integrations.

Results from a set of deterministic integrations suggested a number of possible choices of parameters for the stochastic scheme. These choices have been tested in the ECMWF Ensemble Prediction System. Ensemble results have shown that stochastic physics increases the spread of the ensemble and improves the performance of the ensemble, particularly for the probabilistic prediction of precipitation.

New experiments were performed for a week in summer and a week in winter, using a more recent version of the ECMWF model, and with initial conditions defined using both initial and evolved singular vectors. These new experiments were made with a check introduced to avoid super-saturation, with amplitude r_j uniformly sampled

in the interval *[0.5;1.5]*, spatial correlation length $D=10°$ and temporal correlation length $T=6$ *h*. Results confirmed the positive impact of the stochastic scheme, and defined this latter configuration, with the super-saturation check, as the operational scheme. Operational implementation took place on 21 October 1998.

This "tuning" process indicates that the parameters in this stochastic scheme are poorly known. Comparison with either observations or cloud-resolving models is required to determine these parameters with more confidence.

It is possible that this parametrization will be useful for longer-timescale ensemble predictions, e.g. on the seasonal or climate timescale. The impact of this stochastic scheme on climate timescales will be reported separately.

Acknowledgements

The authors thank Lars Isaksen for his role in revising the software simulating the stochastic forcing, Pedro Viterbo for providing diagnostic software to analyse tendencies, and David Gregory for diagnosing 120 day runs. The advise of Christian Jakob on implementing a super-saturation check on the perturbed humidity field is acknowledged. Thanks are also due to Eugenia Kalnay and Zoltan Toth for their careful revision and comments which help improve an earlier version of the manuscript.

References

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

Buizza, R, and Palmer, T N, 1995. The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 9, 1434-1456.

Buizza, R, and Palmer, T N, 1998. Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 9, 2503-2518.

Buizza, R, Petroliagis, T, Palmer, T N, Barkmeijer, J, Hamrud, M, Hollingsworth, A, Simmons, A, and Wedi, N, 1998. Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **124**, 550, 1935-1960.

Buizza, R, Hollingsworth, A, Lalaurette, F, and Ghelli, A, 1999. Probabilistic predictions of precipitation using the ECMWF Ensemble Prediction System. *Weather and Forecasting*, **14**., 2, 168-189.

Courtier, P, Freyder, C, Geleyn, J F, Rabier, F, and Rochas, M, 1991. The Arpege project at Météo France. Proceedings of the ECMWF Seminar on *Numerical methods in atmospheric models*, 9-13 September 1991, Vol. 2, ECMWF, Shinfield Park, Reading RG2 9AX, UK, pp 324.

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

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

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

Lin, X, and Johnson, R H, 1996. Heating, moistening, and rainfall over the Western Pacific Warm Pool during TOGA COARE. *J. Atmos. Sci.*, **53**, 22, 3367-3383.

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

Palmer, T. N., 1997: On parametrizing scales that are only somewhat smaller than the smallest resolved scales, with application to convection and orography. Proceedings of the ECMWF Workshop on *New insights and approaches to convective parametrization*, 4-7 November 1996, ECMWF, Shinfield Park, Reading RG2 9AX, UK, 328-337.

Palmer, T N, Molteni, F, Mureau, R, and Buizza, R, 1993. Ensemble prediction. ECMWF Seminar proceedings *'Validation of models over Europe: Vol. 1'*, ECMWF, Shinfield Park, Reading, RG2-9AX, UK, 21-66.

Richardson, D, 1998. The relative effect of model and analysis differences on ECMWF and UKMO operational forecasts. Proceedings of the ECMWF Workshop on *Predictability*, 20-22 October 1997, ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Stanski, H R, Wilson, L J, and Burrows, W R, 1989. Survey of common verification methods in meteorology. *World Weather Watch Tech. rep. 8*, WMO, Geneva, pp 114.

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

Toth, Z, and Kalnay, E, 1995. Ensemble forecasting with imperfect models. WMO CAS/JSC WGNE report on *Research activities in atmospheric and oceanic modelling*, WMO/TD-No. 665, 6.30.

Toth, Z, Zhu, Y, Marchok, T, Tracton, S, and Kalnay, E, 1998. Verification of the NCEP global ensemble forecasts. *AMS 12-th Conference on Numerical Weather Prediction*, 11-16 January, Phoenix, Arizona, 286-289.

Tracton, M S, and Kalnay, E, 1993. Operational ensemble prediction at the National Meteorological Center. *Weather and Forecasting*, **8**, 379-398.

Zhu, Y, Iyengar, G, Toth, Z, Tracton, S, and Marchok, T, 1996. Objective evaluation of the NCEP global ensemble forecasting system. *AMS 11-th Conference on Numerical Weather Prediction*, August 19-23, Norfolk, Virginia, J79-J82.

(a)

	<spr>		
Exp	fc d+3 ST - ICST	fc d+5 ST - ICST	fc d+7 ST - ICST
H-L-M	15.4 - 31.5	30.5 - 55.7	49.3 - 74.6
H-L-L	6.6 - 28.7	15.6 - 50.8	29.2 - 69.3
M-H-M	9.7 - 29.1	22.5 - 52.3	38.2 - 70.8
M-M-H *	17.1 - 32.4	33.8 - 57.3	53.0 - 76.3
M-M-M	8.7 - 29.2	20.9 - 52.1	36.7 - 70.4
M-M-L	11.3 - 28.6	21.8 - 50.6	34.4 - 67.9
M-L-M	6.2 - 28.7	15.4 - 51.2	29.6 - 69.4
M-L-L	4.0 - 28.4	10.8 - 50.9	22.4 - 68.7
L-H-H	8.6 - 29.1	20.5 - 51.7	36.5 - 70.2
L-H-M	5.2 - 28.5	13.5 - 50.9	25.9 - 68.6
L-M-H	7.5 - 29.1	18.4 - 52.5	33.8 - 71.1
L-L-H	4.3 - 28.5	11.7 - 51.0	23.5 - 68.8
IC	28.3	50.2	67.8
Error of control forecast	34.2	60.6	88.9

(b)

	$<err_{pf}>$		
Exp	fc d+3 ST - ICST	fc d+5 ST - ICST	fc d+7 ST - ICST
H-L-M	35.8 - 44.4	64.0 - 73.6	91.7 - 93.0
H-L-L	34.1 - 43.1	59.7 - 71.1	88.5 - 92.3
M-H-M	35.0 - 43.4	62.3 - 72.5	89.9 - 93.4
M-M-H	35.9 - 44.7	65.4 - 75.2	91.7 - 94.4
M-M-M	34.5 - 43.4	61.8 - 72.1	89.4 - 92.6
M-M-L	35.7 - 43.4	62.5 - 71.8	90.8 - 92.5
M-L-M*	34.2 - 43.2	60.7 - 71.9	88.8 - 92.2
M-L-L*	34.1 - 43.2	60.6 - 71.8	89.1 - 92.2
L-H-H	34.7 - 43.4	62.2 - 72.2	89.8 - 92.9
L-H-M	34.4 - 43.3	61.3 - 72.0	89.5 - 93.1
L-M-H	34.3 - 43.4	61.8 - 72.9	89.3 - 93.1
L-L-H	34.2 - 43.2	61.0 - 71.9	89.2 - 92.7
IC	43.4	71.9	92.7
Error of control forecast	34.2	60.6	88.9

Table 1. (a) Average (among the 5 case studies) of the ensemble rms spread (m) and error of control forecast; ICST spread values closest to the control rms error are highlighted in bold. (b) Mean rms error of the perturbed forecasts (m) and error of control forecast; ICST smallest error values are highlighted in bold. (c) Difference between the rms error of the control of configuration IC and the rms error of the ensemble mean (m); ICST largest values are highlighted in bold. In each table, a * near the configuration identifier marks the ICST configuration with the best result at forecast day 7. (d) Comparison of the results of all ICST configurations with the reference IC configuration at forecast d+7 (from Tables 1a-c): the first column reports the change induced on the ensemble spread (*ds*), the second column the change induced on the mean error of the perturbed forecasts (de_p), and the third column the change induced in the error of the ensemble-mean forecast (de_m). All results refer to the 500 hPa geopotential height over the NH extra-tropics.

(c)

	$\langle err^{IC}_{con} \rangle - \langle err_{mean} \rangle$		
Exp	fc d+3 ST - ICST	fc d+5 ST - ICST	fc d+7 ST - ICST
H-L-M	-0.3 - -0.4	0.0 - 2.6	3.0 - 14.6
H-L-L	0.4 - 0.4	2.0 - **4.4**	2.9 - 14.5
M-H-M	0.3 - 0.4	1.2 - 3.7	4.2 - 14.0
M-M-H *	0.8 - 0.4	0.8 - 2.5	6.2 - **15.0**
M-M-M	0.6 - **0.5**	1.4 - 4.0	4.4 - 14.7
M-M-L	0.1 - 0.1	1.0 - 3.8	2.5 - 14.1
M-L-M *	0.3 - 0.3	1.0 - 3.8	2.7 - **15.0**
M-L-L	0.2 - 0.2	0.6 - 3.8	1.2 - 14.8
L-H-H	0.4 - **0.5**	1.1 - 4.0	4.2 - 14.6
L-H-M	0.1 - 0.2	0.4 - 3.8	2.0 - 13.8
L-M-H	0.5 - **0.5**	0.9 - 3.4	3.8 - 14.9
L-L-H	0.2 - 0.3	0.3 - 4.0	1.6 - 14.6
IC	0.1	3.7	14.2

(d)

Experiment	ds=($\langle spr^{ICST} \rangle$-$\langle spr^{IC} \rangle$) / $\langle spr^{IC} \rangle$	de_{pf}=($\langle err^{ICST}_{pf} \rangle$-$\langle err^{IC}_{pf} \rangle$) / $\langle err^{IC}_{pf} \rangle$	de_m=($\langle err^{IC}_{con} \rangle$-$\langle err_{mean} \rangle$) / $\langle err_{mean} \rangle$
> H-L-M	$10 \le ds$		$-3.5 \le de_{mean} \le 0$
H-L-L		$de_{pf} \le 0$	$-3.5 \le de_{mean} \le 0$
M-H-M	$3.5 \le ds \le 10$		
> M-M-H	$10 \le ds$ *	$de_{pf} \le 0$	$de_{mean} \le -3.5$ *
> M-M-M *	$3.5 \le ds \le 10$	$de_{pf} \le 0$ *	$de_{mean} \le -3.5$
M-M-L		$de_{pf} \le 0$ *	
M-L-M		$de_{pf} \le 0$	$de_{mean} \le -3.5$ *
M-L-L			$de_{mean} \le -3.5$
> L-H-H	$3.5 \le ds \le 10$		$-3.5 \le de_{mean} \le 0$
L-H-M			
> L-M-H	$3.5 \le ds \le 10$		$de_{mean} \le -3.5$
L-L-H		$de_{pf} \le 0$	$-3.5 \le de_{mean} \le 0$

Table 1. (a) Average (among the 5 case studies) of the ensemble rms spread (m) and error of control forecast; ICST spread values closest to the control rms error are highlighted in bold. (b) Mean rms error of the perturbed forecasts (m) and error of control forecast; ICST smallest error values are highlighted in bold. (c) Difference between the rms error of the control of configuration IC and the rms error of the ensemble mean (m); ICST largest values are highlighted in bold. In each table, a * near the configuration identifier marks the ICST configuration with the best result at forecast day 7. (d) Comparison of the results of all ICST configurations with the reference IC configuration at forecast d+7 (from Tables 1a-c): the first column reports the change induced on the ensemble spread (*ds*), the second column the change induced on the mean error of the perturbed forecasts (de_{pf}), and the third column the change induced in the error of the ensemble-mean forecast (de_m). All results refer to the 500 hPa geopotential height over the NH extra-tropics.

Configuration	fc day+3 outlier (%)	fc day+3 rms (m)	fc day +5 outlier (%)	fc day +5 rms (m)	fc day +7 outlier (%)	fc day +7 rms (m)
IC	24.2	10.4	21.5	12.3	16.1	13.6
ICST M-M-M	17.5	9.4	17.3	11.2	13.2	12.3
ICST L-H-H	17.4	9.4	16.4	11.2	11.8	13.2
ICST L-M-H	18.5	9.4	17.3	11.3	11.8	13.0

Table 2. Percentage of analysis values lying outside the ensemble forecast range (outlier) and rms error of the grid-point best forecast, for the reference ensemble IC and for configurations M-M-M, L-H-H and L-M-H, at forecast day 3, 5 and 7 for the NH 500 hPa geopotential height.

	rms error (m)	
Configuration (97.07.31)	IC	ICST M-M-M
member n. 41 (best IC)	46.3	48.7
member n. 8 (best ICST M-M-M)	51.3	**43.4**
IC control	57.4	
T213	61.3	

Table 3. Root-mean-square error over Europe of ensemble members 8 (best ICST M-M-M member) and 41 (best IC member), of the control and the T213L31 hight resolution forecast, for the 97.07.31 case, for the 500 hPa geopotential height day 7 forecast.

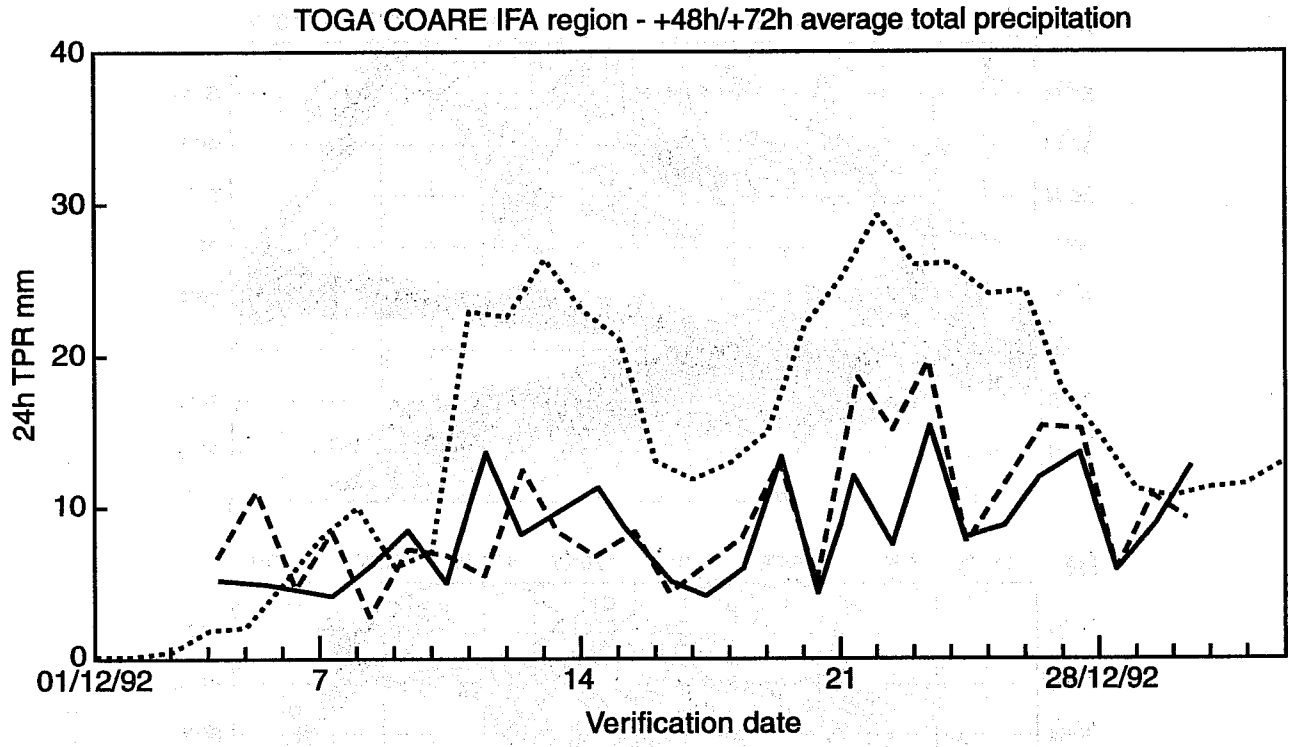

Figure 1 Comparison of the daily-mean rainfall rates (mm/d) in the TOGA COARE IFA region (0°–10°S;150°E-160°E) predicted from the unperturbed (solid) and M-M-M stochastically perturbed (dash) integrations for December 1992, and estimated from GPI retrievals (dotted, Lin and Johnson, 1996).

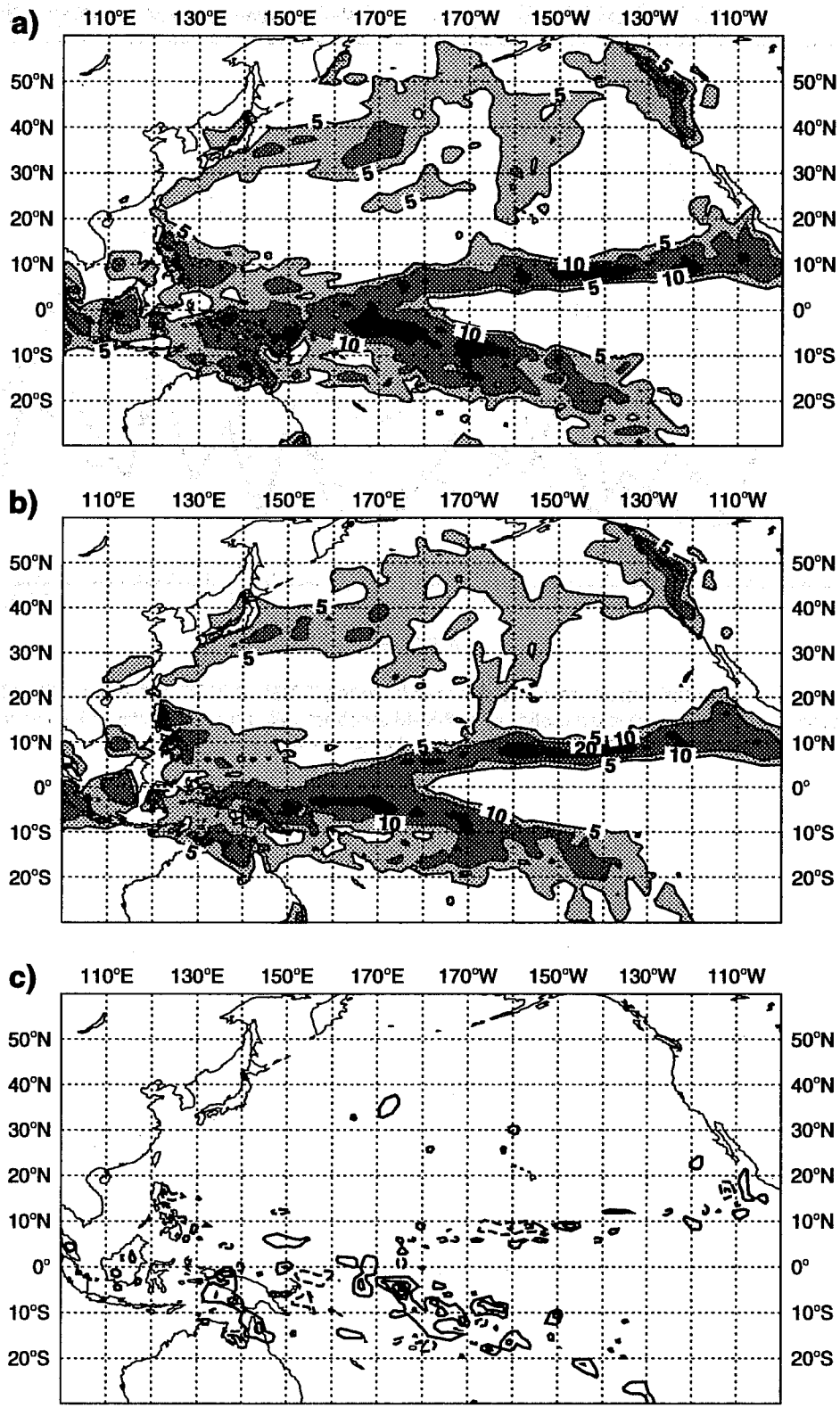

Figure 2 December 1992 average precipitation accumulated between forecast day 6-7 for (a) the unperturbed and (b) the M-M-M stochastically perturbed integrations; (c): difference between (b) and (a). Shading in panels (a-b) for 5, 10, 20 and 40 mm/day, and 5 mm/day contour interval in panel (c).

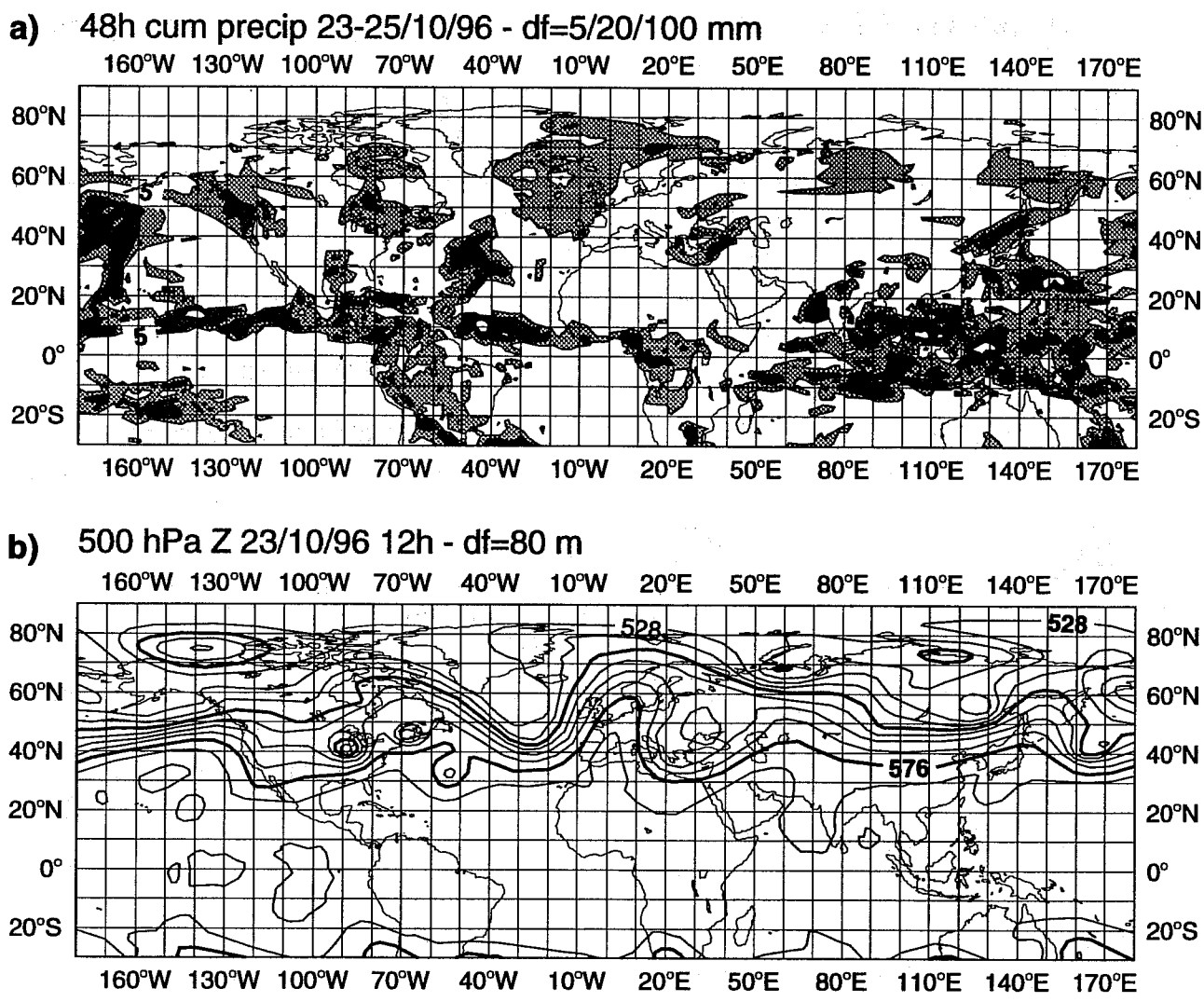

Figure 3 (a) 48-h accumulated precipitation between 12GMT of 23 and 25 March 1996; (b): 500 hPa geopotential height at 12 GMT of 23 March 1996. Contour isolines for 5, 20 and 100 mm/day for precipitation, and contour interval 8 dam for geopotential height.

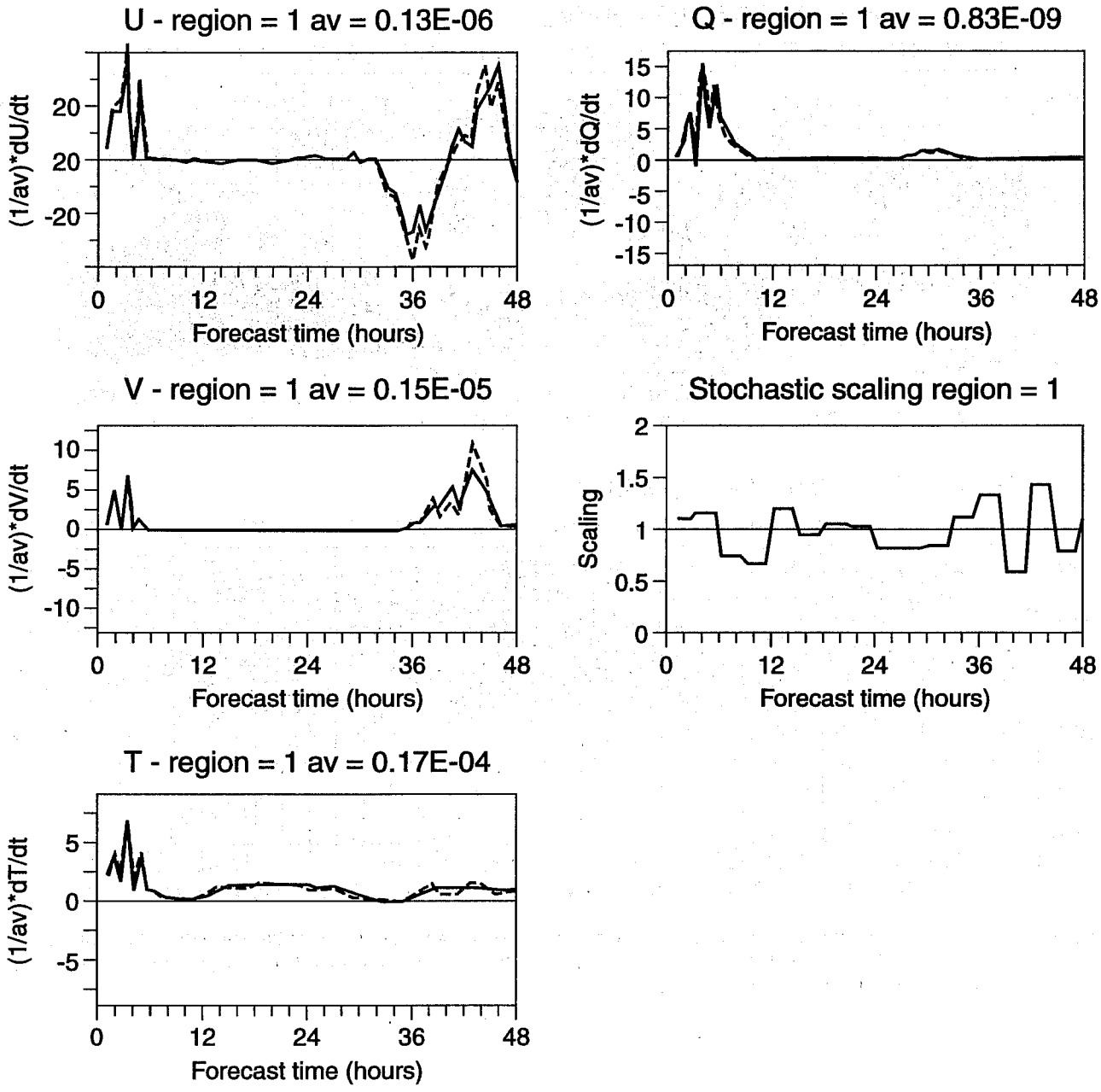

Figure 4 Pacific region (140°W-130°W, 30°N-40°N): average value of unperturbed (solid) and perturbed (dash) tendencies for (a) U and (b) V wind components, (c) temperature, and (d) specific humidity during the first 48 hour forecasts normalized by their respective average value. (c) stochastic scaling factors <rj>D,T. Note that tendencies have been normalized by their time average value av, and that the average value av is reported in the title of each panel.

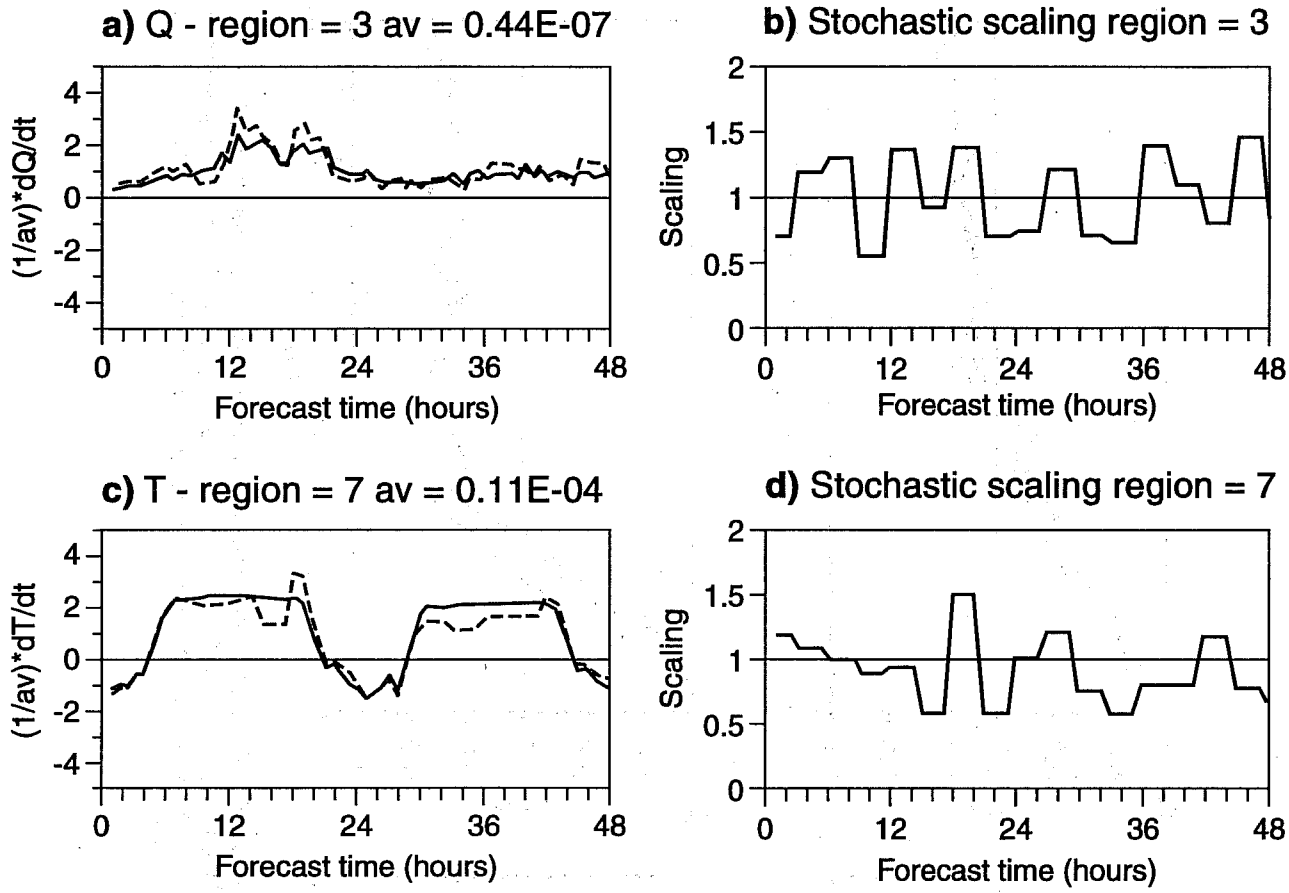

Figure 5 (a-b) Rockies region (110°W-100°W,40°N-50°N): (a) average value of unperturbed (solid) and perturbed (dash) tendencies for specific humidity, and (b) stochastic scaling factors <rj>D,T. (c-d) Sahara region (10°W-0°,20°N-30°N): (c) average value of unperturbed (solid) and perturbed (dash) tendencies for temperature and (d) stochastic scaling factors <rj>D,T.

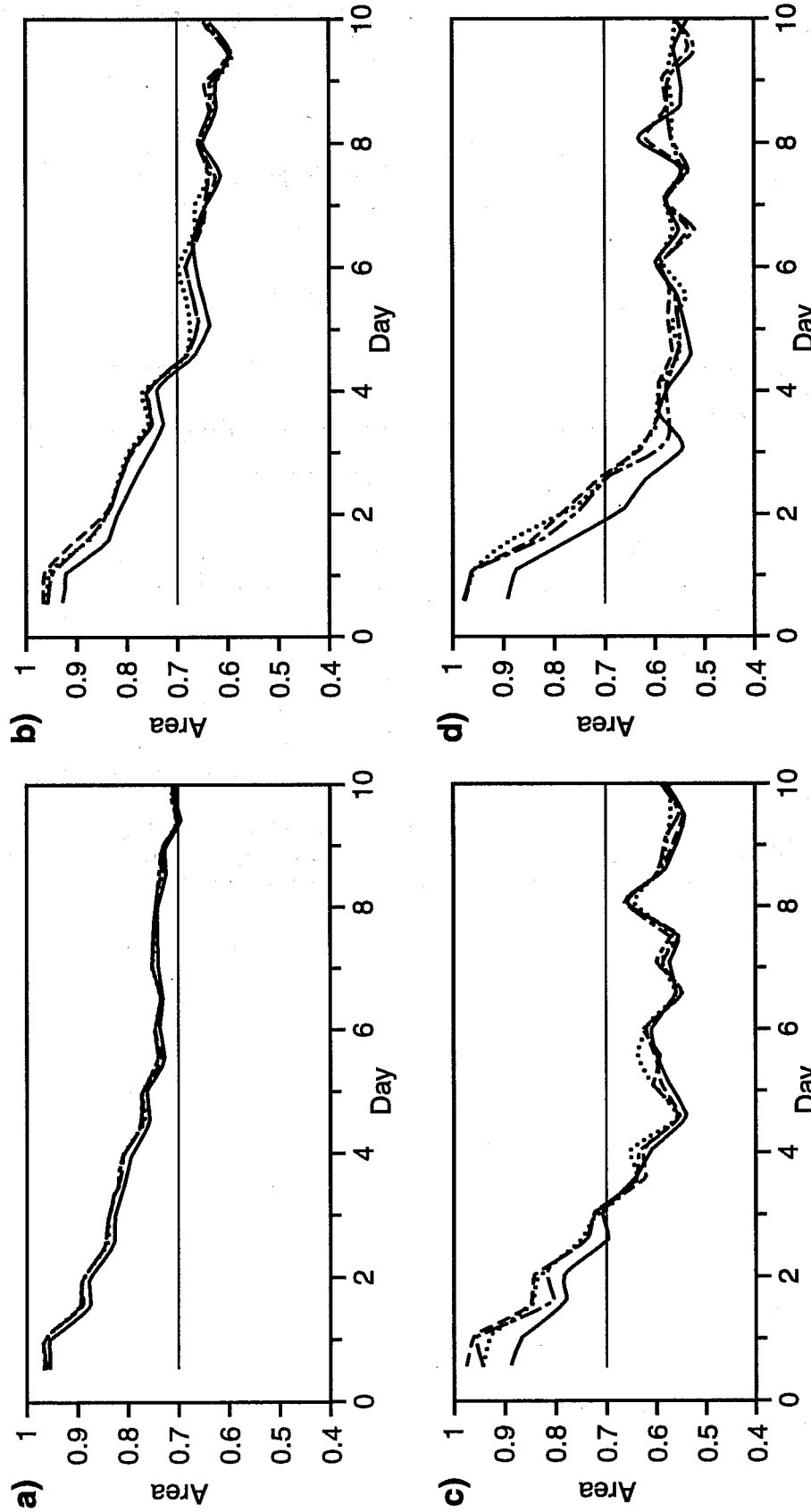

Figure 6 (a) Area under the average ROC curve for the prediction of the event "12-h accumulated precipitation larger than 2mm/12h" for the NH extra-tropics, for configurations IC (solid), ICST M-M-M (dash), ICST L-H-H (dot) and ICST L-M-H (chain dash). (b): as (a) but for the event "12-h accumulated precipitation larger than 10mm/12h". (c): as (a) but for the event "12-h accumulated precipitation larger than 20mm/12h". (d): as (a) but for the event "12-h accumulated precipitation larger than 30mm/12h".

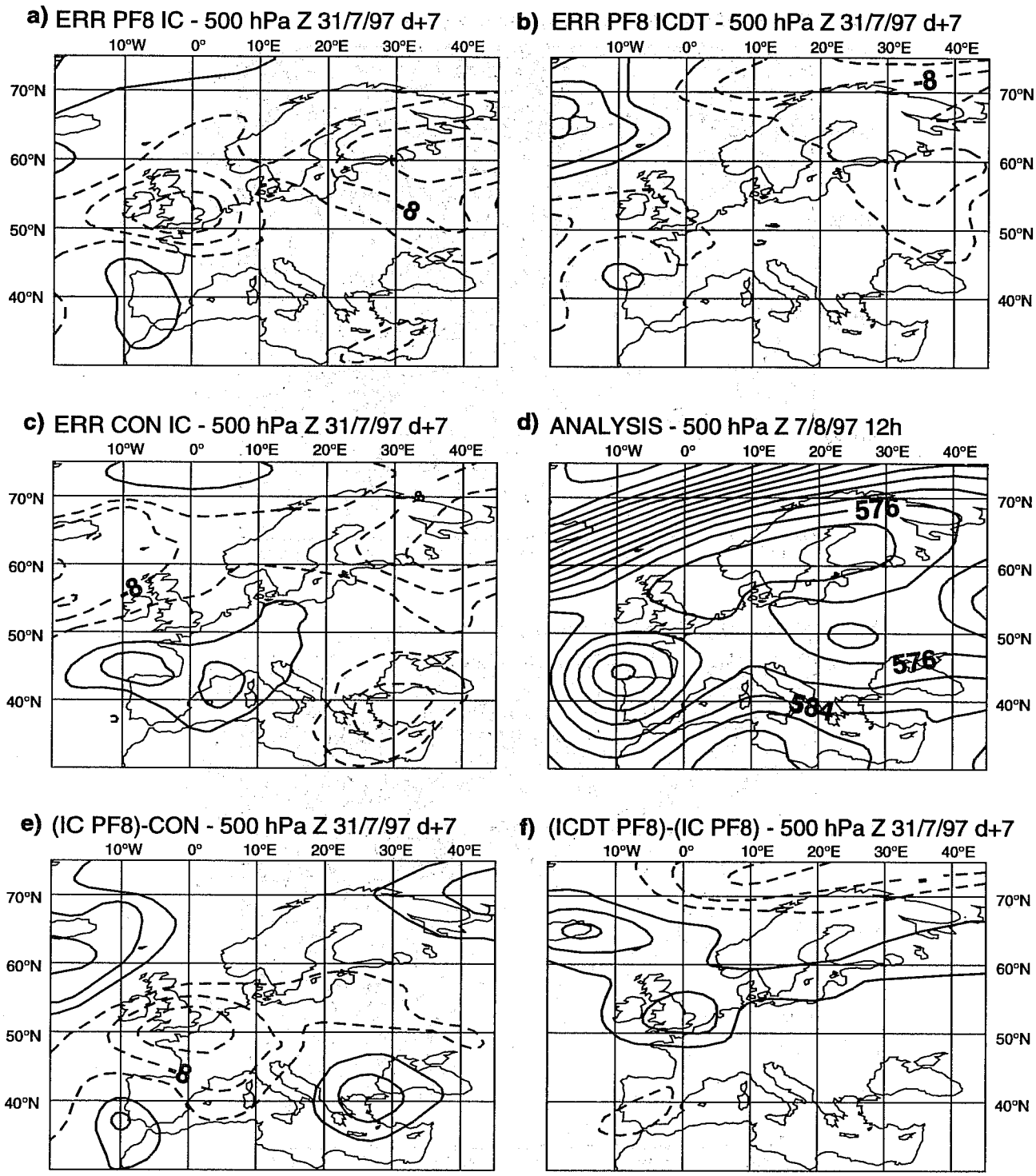

Figure 7 Forecast error at day 7 of the 8th member of configurations (a) IC and (b) ICST M-M-M, and of (c) the control forecast. (d): verifying analysis. (e): difference between the 8th member of configuration IC and the control, and (f) difference between the 8th member of configuration ICST M-M-M and of configuration IC. Results refer to the 500 hPa geopotential height field for the 97.07.31 case study. Contour interval 4 dam.

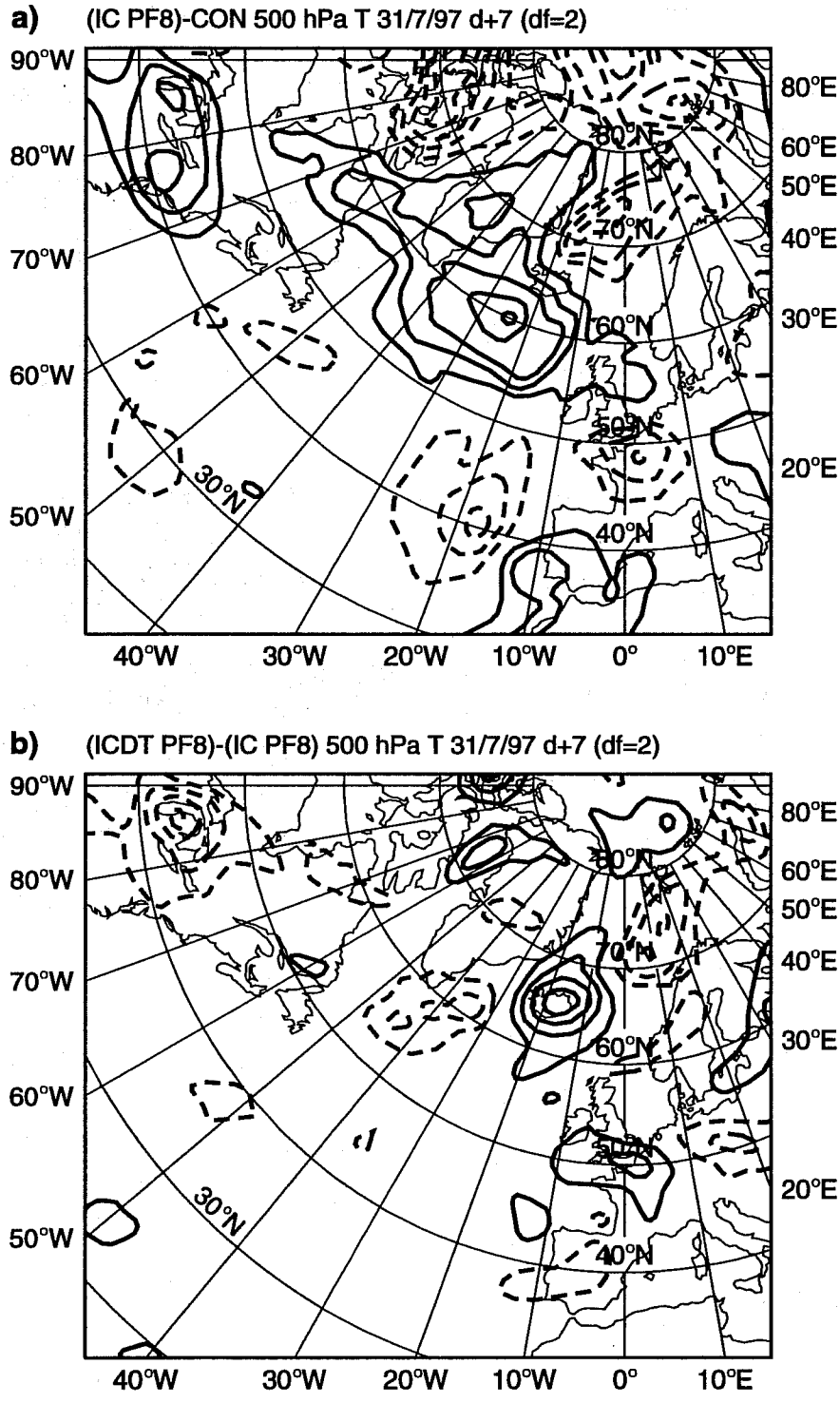

Figure 8 500 hPa temperature day 7 forecast (a) difference between the 8th member of configuration IC and the control, and (b) difference at forecast day 7 between the 8th member of configuration ICST M-M-M and configuration IC, for the 97.07.31 case. Contour interval 2 degrees.

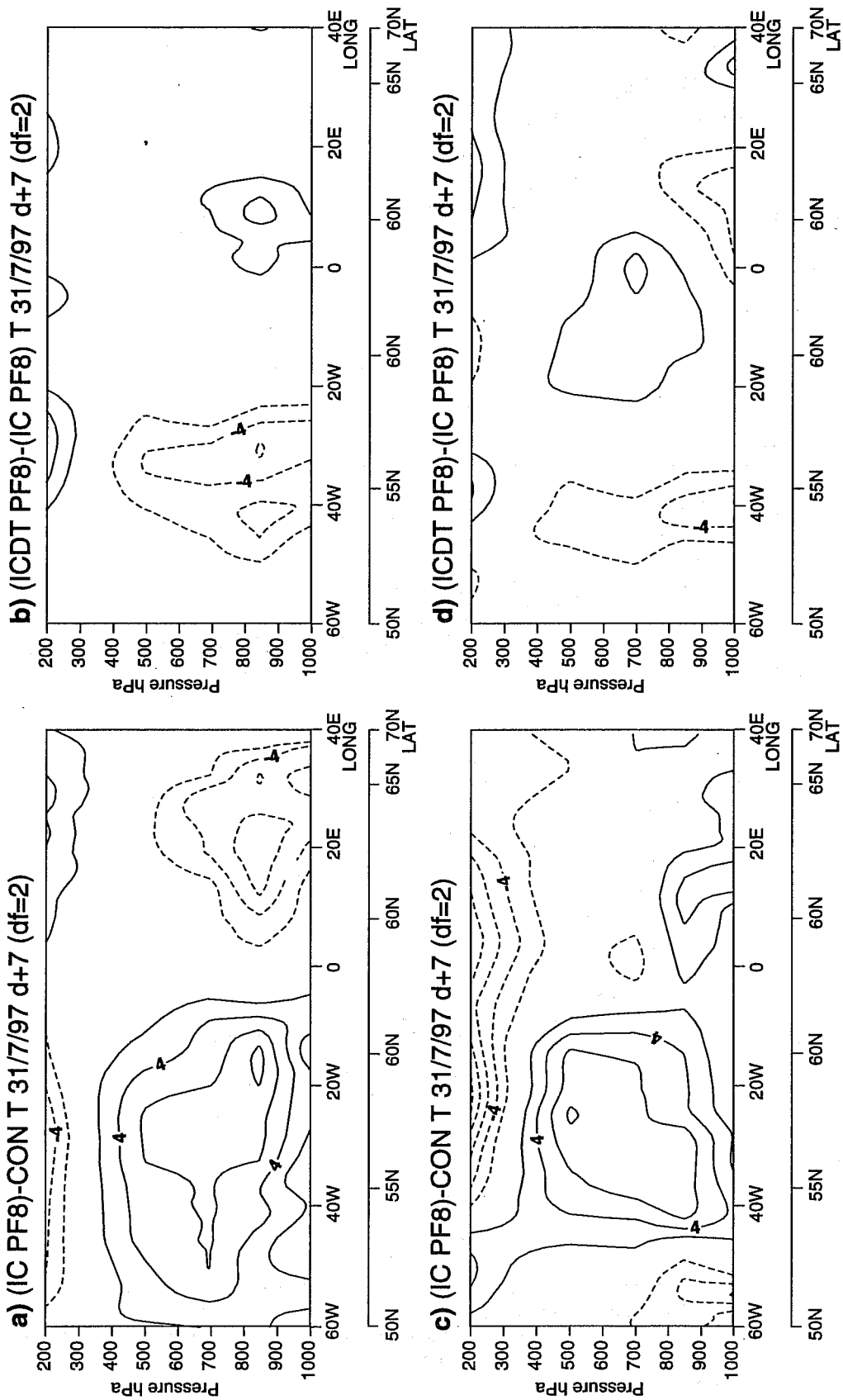

Figure 9 Temperature vertical cross section from the Labrador Peninsula to Sardinia at forecast day of (a) difference between the 8th member of configuration IC and the control, and (b) difference at forecast day 7 between the 8th member of configuration ICST M-M-M and configuration IC. (c-d): as (a-b) but for a vertical cross section from Newfoundland to the North Cape. Contour interval 2 degrees.

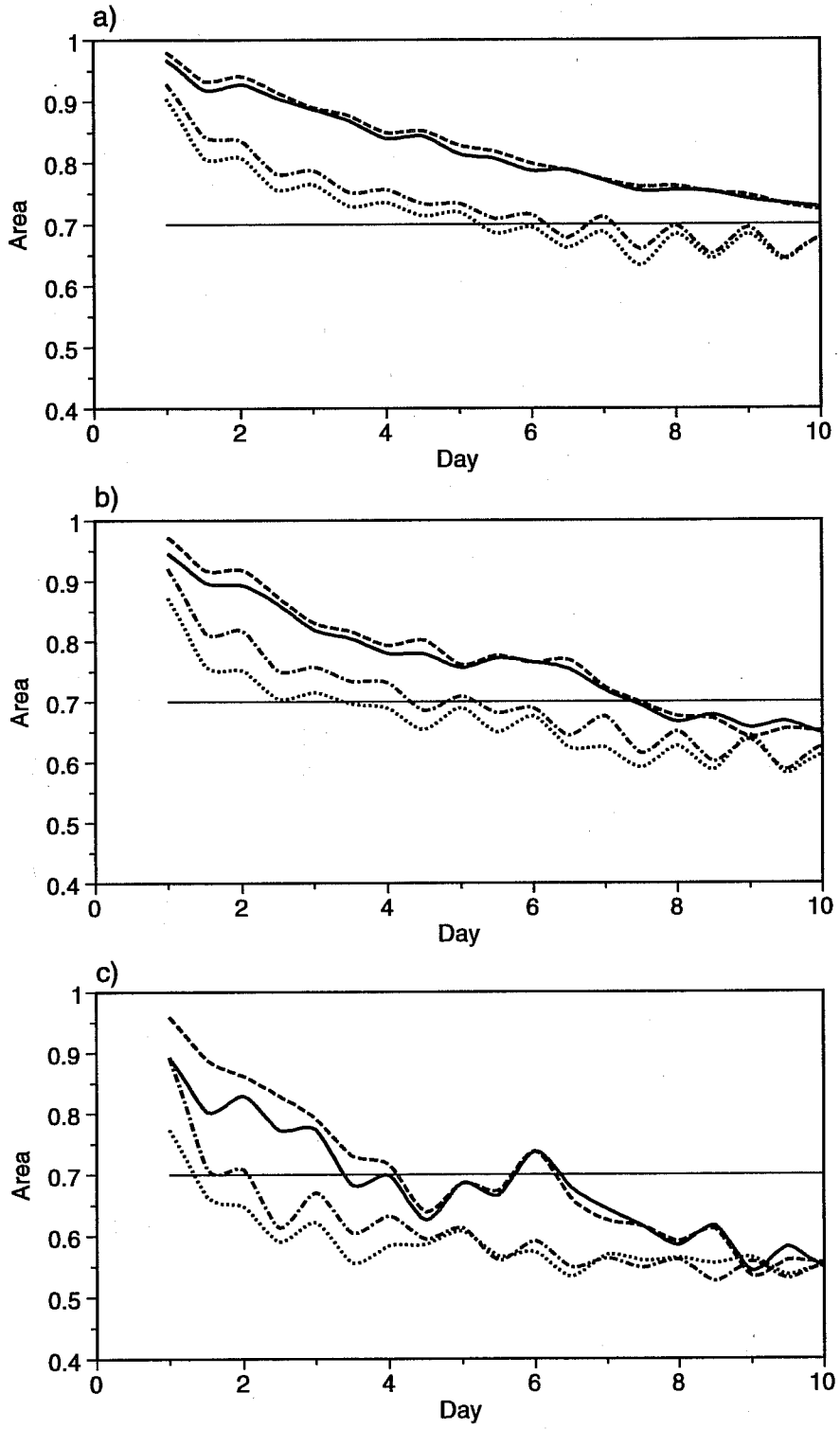

Figure 10 (a) area under the ROC curve relative to the event the event "12-h accumulated precipitation larger than 5 mm/12h", (b) "12-h accumulated precipitation larger than 10 mm/12h" and (c) "12-h accumulated precipitation larger than 20 mm/12h", for the reference (solid) and stochastically perturbed (dash) ensembles for the week 16-22 December 1997, and for the reference (dot) and stochastically perturbed (chain-dash) ensembles for the week 29 June-5 July 1997, over NH.

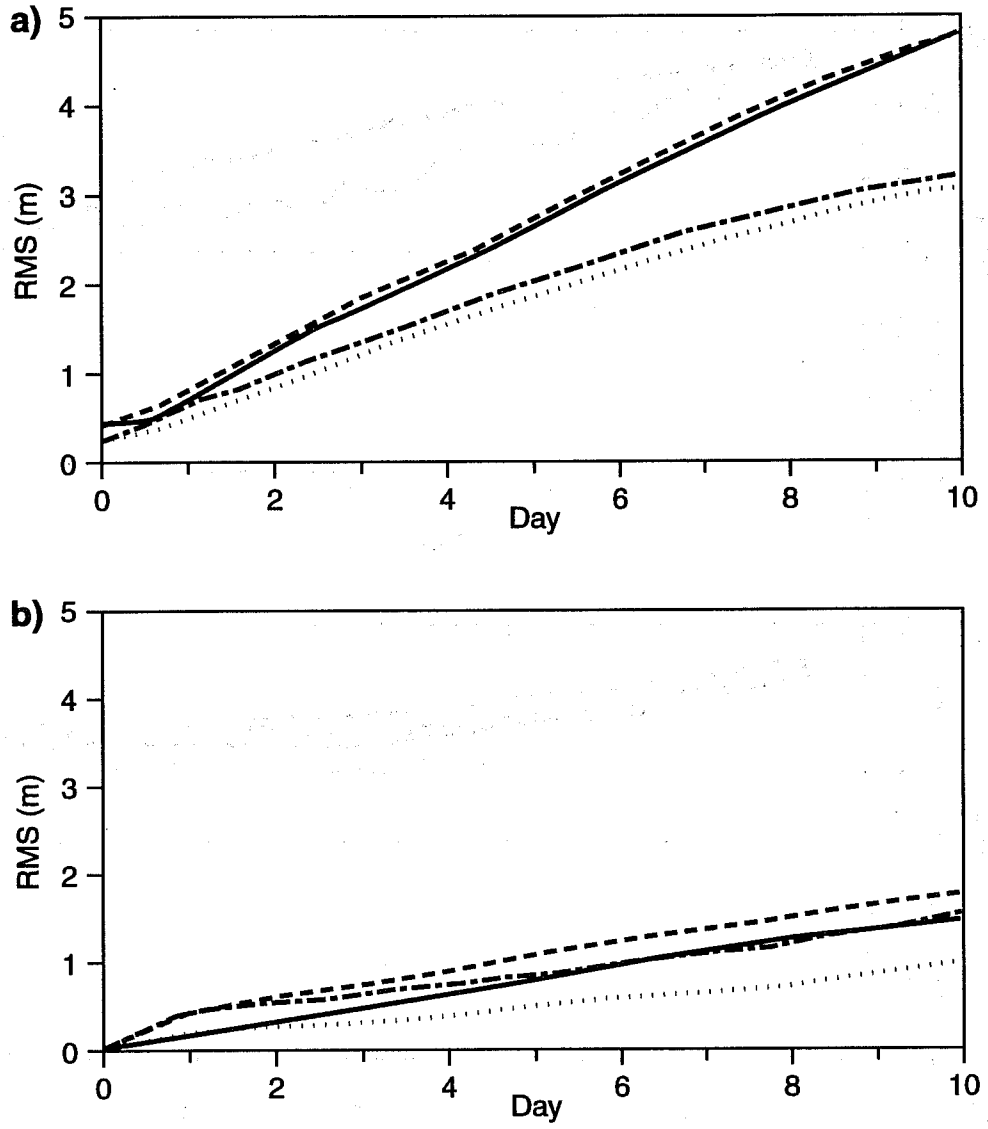

Figure 11 (a) average rms spread for the 850 hPa temperature for the reference (solid) and stochastically perturbed (dash) ensembles for the week 16-22 December 1997, and for the reference (dot) and stochastically perturbed (chain-dash) ensembles for the week 29 June-5 July 1997, over the NH. (b): as (a) but for the tropics.

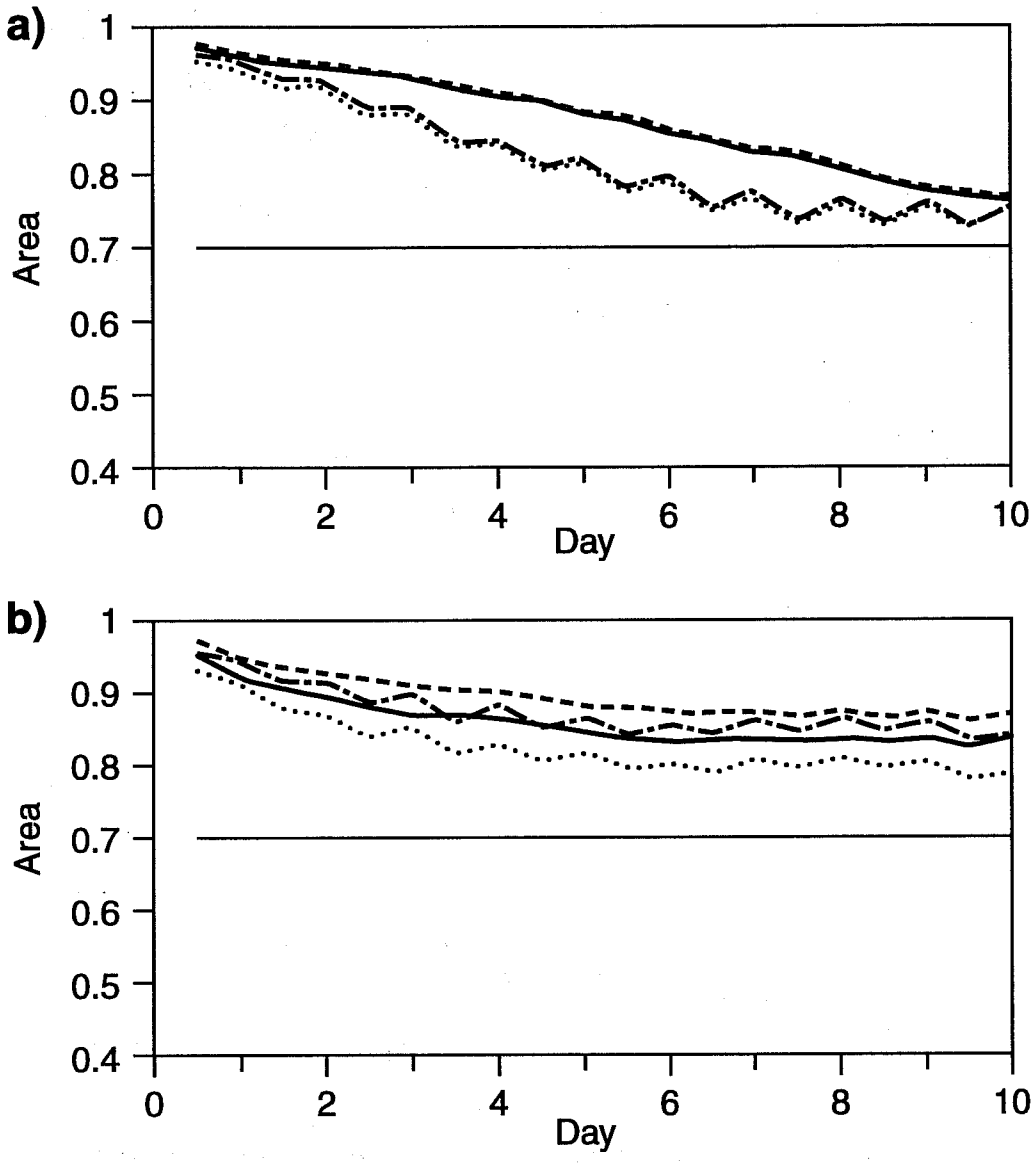

Figure 12 (a) area under the ROC curve for the prediction of the event "850 hPa temperature warmer than climatology" for the reference (solid) and stochastically perturbed (dash) ensembles for the week 16-22 December 1997, and for the reference (dot) and stochastically perturbed (chain-dash) ensembles for the week 29 June-5 July 1997, over the NH. (b): as (a) but for the tropics.