Stochastic simulation of model uncertainties in the ECMWF Ensemble Prediction System

R Buizza, M Miller and T N Palmer

European Centre for Medium-Range Weather Forecasts

ABSTRACT

A parametrization of random error associated with representation of diabatic processes ("stochastic physics") is described, and its impact in the European Centre for Medium-Range Weather Forecasts Ensemble Prediction System is discussed.

The scheme is based on the notion that random errors in diabatic forcing are coherent between the different parametrization modules, and have a certain coherence on the space and time scales represented by the model. Moreover, the scheme assumes that the larger the parametrized tendencies, the larger the random error component is.

A number of diagnostics are described and a choice of parameters is made. The primary application of the scheme is for the European Centre for Medium-Range Weather Forecasts Ensemble Prediction System. It is shown how the scheme increases the spread of the ensemble, and improves the skill of the probabilistic prediction of weather parameters as precipitation.

A choice of stochastic parameters is made for possible operational implementation.

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 (*Palmer et al*, 1993; *Molteni et al*, 1996). A major upgrade to a 51-member TL159L31 system 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

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 models. A similar "perfect model" strategy is followed at the US National Centers for Environmental Prediction (NCEP), using initial perturbations generated using so-called bred-vectors (*Tracton and Kalnay*, 1993; *Toth and Kalnay*, 1993).

On the other hand, recent results from *Harrison et al* (1998) 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 error in the medium range. Specifically, although the spread of the ECMWF EPS agrees well with the error of the unperturbed control forecast at, say, day 2 or 3, the spread is notably smaller than the control rms error later in the medium range.

Whilst using perturbations with larger initial amplitude would reduce the spread underestimation in the late forecast range, it would also deteriorate the performance in the early range, since all perturbed forecast would be at an unrealistic distance from the control forecast than the analysis. [The old EPS system, which used larger initial perturbations to compensate from the slower perturbation growth after forecast day 3 (*Buizza et al*, 1998), suffered from this problem.]

Houtekamer et al (1996) first included model uncertainties in ensemble prediction. 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. In fact, each ensemble member is integrated using different parametrization 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 diabatic tendencies. For example, consider a grid point over the warm pool area during a period of organised deep convection. By definition, the actual diabatic terms are 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 diabatic heating. For example, even if the parametrized and actual diabatic 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 diabatic 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 diabatic forcing, discussed above, will be coherent between the different parametrization modules, and will have a certain coherence on the space and time scales represented by the model. 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 full diabatic tendency, rather than on the diabatic tendencies from all the individual modules. 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.

Full details of the scheme are described in section 2, and the impact of the scheme on the EPS is investigated in section 3. A further detailed analysis of unperturbed and perturbed diabatic tendencies is presented in section 4. Conclusions are drawn in section 5.

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

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

of the model equations

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

(where A and P identify the adiabatic and the diabatic processes) starting from the initial conditions

$$e_j(t=0) \equiv e_0(t=0)+\delta e_j(t=0)\ . \qquad (3)$$

$e_0(t=0)$ 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 the total energy norm over a 48-hour time interval (*Buizza and Palmer*, 1995).

2.2 The new ECMWF Ensemble Prediction System

Compared to the original one, the new system includes a scheme to simulate model uncertainties deriving from the diabatic terms. Schematically, each ensemble member e_j can be seen as the time integration

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

of the perturbed model equations

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

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 diabatic

tendency (of each state vector component) is defined as

$$\boldsymbol{P}'_j(\boldsymbol{e}_j;t) \equiv \langle r_j(\lambda,\phi;t)\rangle_{D,T}\ \boldsymbol{P}_j(\boldsymbol{e}_j;t) \quad , \tag{6}$$

where $\langle..\rangle_{D,T}$ 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. Random numbers have been selected uniformly from three different intervals for the so-called high, medium and low amplitude stochastic forcing configurations (Table 1),

$$r_j \in [0,2] \quad , \tag{7a}$$

$$r_j \in [0.5,1.5] \quad , \tag{7b}$$

$$r_j \in [0.75,1.25] \quad . \tag{7c}$$

Note that, in the current formulation, the stochastic term re-scales the diabatic 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 $\langle r_j\rangle_{D,T}$, and different degrees of randomness can be applied in the horizontal space coordinates and in time by setting the D and T parameters. More sophisticated formulations will be tested in due course.

Different combinations of the stochastic forcing amplitude, and spatial and temporal parameters have been tested, with each parameter set to either a low, a medium or a high value. Table 1 lists them, named according to the (in alphabetic order) amplitude, space and time parameter (configuration H-M-L, for example, has high amplitude as in Eq. (7a), medium box size $D=5°$ *deg*, and low time interval $T=3$ *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), $T=0.75$ *h* indicates that the random numbers are re-generated every time step.

A detailed diagnostic on perturbed and unperturbed tendencies will be presented in Section 3.

Configuration	amplitude	box size D (deg)	time interval T (h)
H-M-M	[0;2]	5	3
H-L-M	[0;2]	1	3
H-L-L	[0;2]	1	0.75
M-H-M	[0.5;1.5]	10	3
M-M-H	[0.5;1.5]	5	12
M-M-M	[0.5;1.5]	5	3
M-M-L	[0.5;1.5]	5	0.75
M-L-M	[0.5;1.5]	1	3
M-L-L	[0.5;1.5]	1	0.75
L-H-H	[0.75;1.25]	10	12
L-H-M	[0.75;1.25]	10	3
L-M-H	[0.75;1.25]	5	12
L-L-H	[0.75;1.25]	1	12

Table 1. Experiment list.

2.3 Experimental set-up

Ensembles in all configurations listed in Table 1 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). All ensemble have been performed at T_l159L31 resolution and with 50 perturbed members (with the ECMWF IFS model cycle CY16R1).

For each date, ensembles have been run with diabatic stochastic forcing only (DT ensembles), with initial perturbations only (IC ensembles), and with stochastic forcing and initial condition perturbations (ICDT ensembles). The ICDT 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 extra-tropics, and have been 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

Because of the lack of moist processes in the forward and adjoint tangent models used here, ensemble initial perturbations are optimized for the extra-tropics (work is in progress to add moist physical processes in the tangent model, to allow for singular vector computation in the tropical region, but results are not available yet). Thus, instead of ensemble predictions, single deterministic forecasts will be compared to investigate the impact of stochastic forcing on the tropical belt. Since the relative role of the diabatic terms is much larger in the deep tropics, the perturbation technique is correspondingly tested more severely. Attention is restricted to the reference configuration IC and to the average perturbed configuration ICDT M-M-M.

Firstly, the impact of stochastic forcing on precipitation in the western Pacific warm pool is compared with observations collected during the TOGA COARE experiment (*Lin and Johnson*, 1996). Secondly, the model climatology of configurations IC and ICDT M-M-M are compared.

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, they showed (see Fig. 1, from *Lin and Johnson*, 1996) that not only forecasts and rainfall estimations differ substantially, but also that SSM/I and GPI retrievals can differ by several mm/day, with peaks of more than 20 mm/day. These amounts 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 the M-M-M type of diabatic stochastic forcing have been run, and their rainfall predictions in the TOGA COARE intensive flux area have been compared. Figure 2 shows the differences between the daily values of precipitation forecast accumulated between the 48 and the 72 hour forecasts. Results indicate that 24-hour rainfall forecast differences are smaller than the differences in *Lin and Johnson* (1996) at both forecast ranges, with maximum values of about 10mm/day.

3.2 Impact of stochastic forcing on model climatology

Since precipitation is one of the weather parameters more sensitive to changes in the diabatic tendencies, one way to assess the impact of stochastic forcing on model climatology is to examine monthly average precipitation fields for the deterministic forecasts run for December 1992 without and with the M-M-M diabatic stochastic forcing. Indeed, the monthly mean precipitation accumulated between forecast days 6-7 of the two experiments are very similar (Fig. 3a-b), with differences mostly concentrated in the tropics (Fig. 3c). Similar differences characterize mean precipitation accumulated between forecast days 1-2 (not shown).

A further verification of the small impact of the stochastic perturbations is obtained from longer simulations. Figures 4a-b show the total rainfall for the unperturbed and perturbed runs averaged over three months (JJA). These mean precipitation maps represent very similar mean circulations, with some evidence of spatially smoother fields in the perturbed case.

3.3 Diabatic tendency diagnostics

The impact of the M-M-M stochastic forcing on the diabatic tendencies is studied, to further

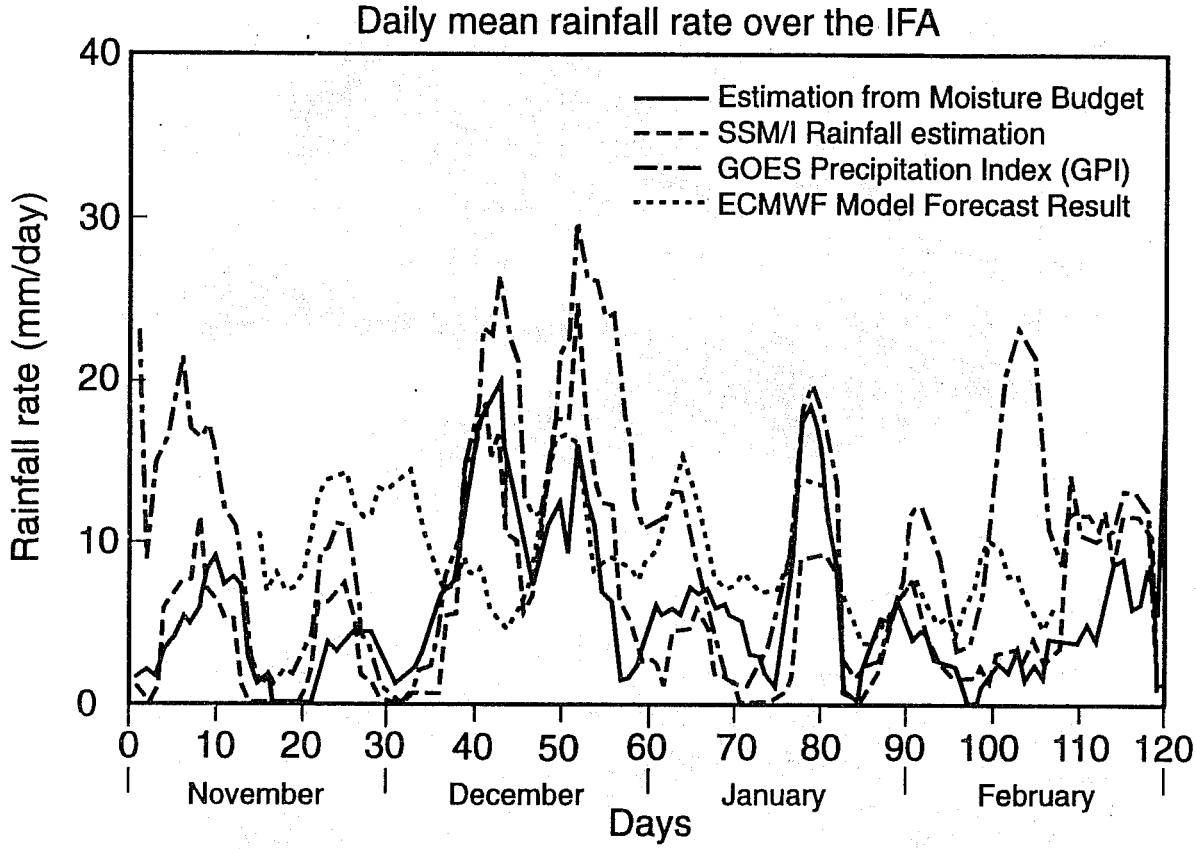

Fig. 1 Comparison of daily-mean rainfall rates (mm/d) over the IFA region (0°- 10°S;150°E-160°E) from moisture budget, SSM/I retrievals, GPI, and ECMWF model forecast results (5-day running mean) for the TOGA COARE period (1 November 1992 to 28 February 1993, from Lin and Johnson, 1996).

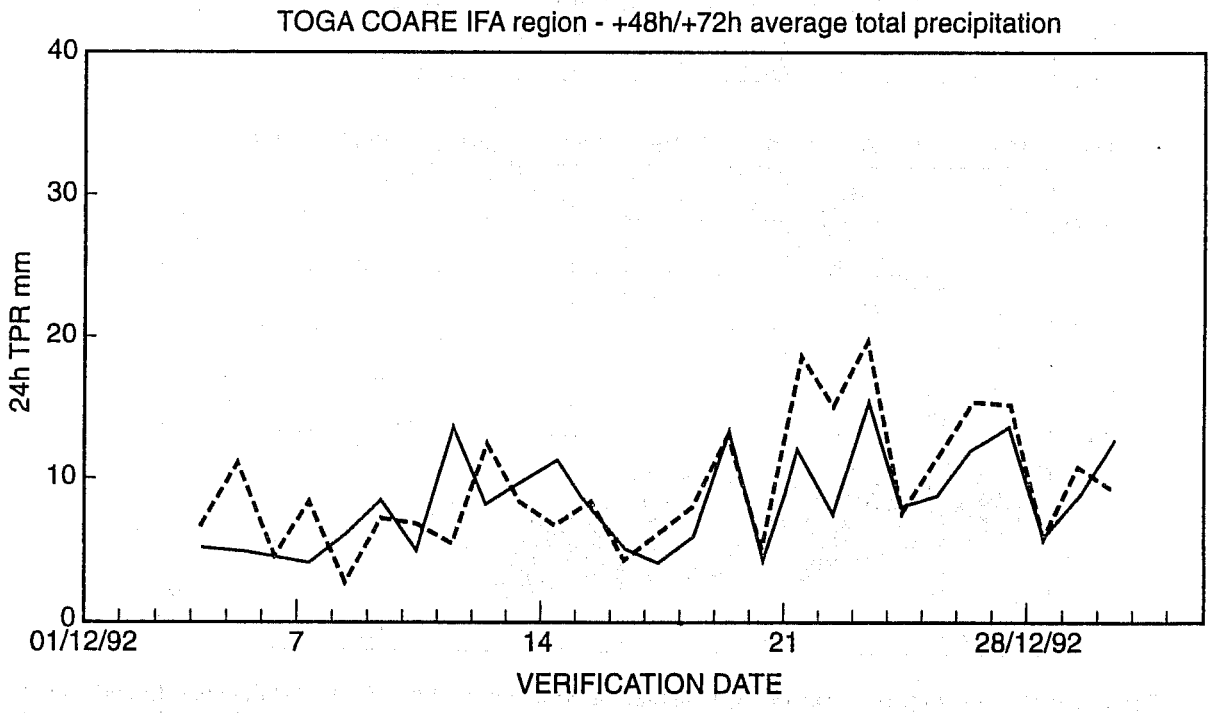

Fig. 2 Comparison of the daily-mean rainfall rates (mm/d) over the region (0°- 10°S;150°E-160°E) from the unperturbed (solid) and M-M-M stochastically perturbed (dash) integrations for December 1992.

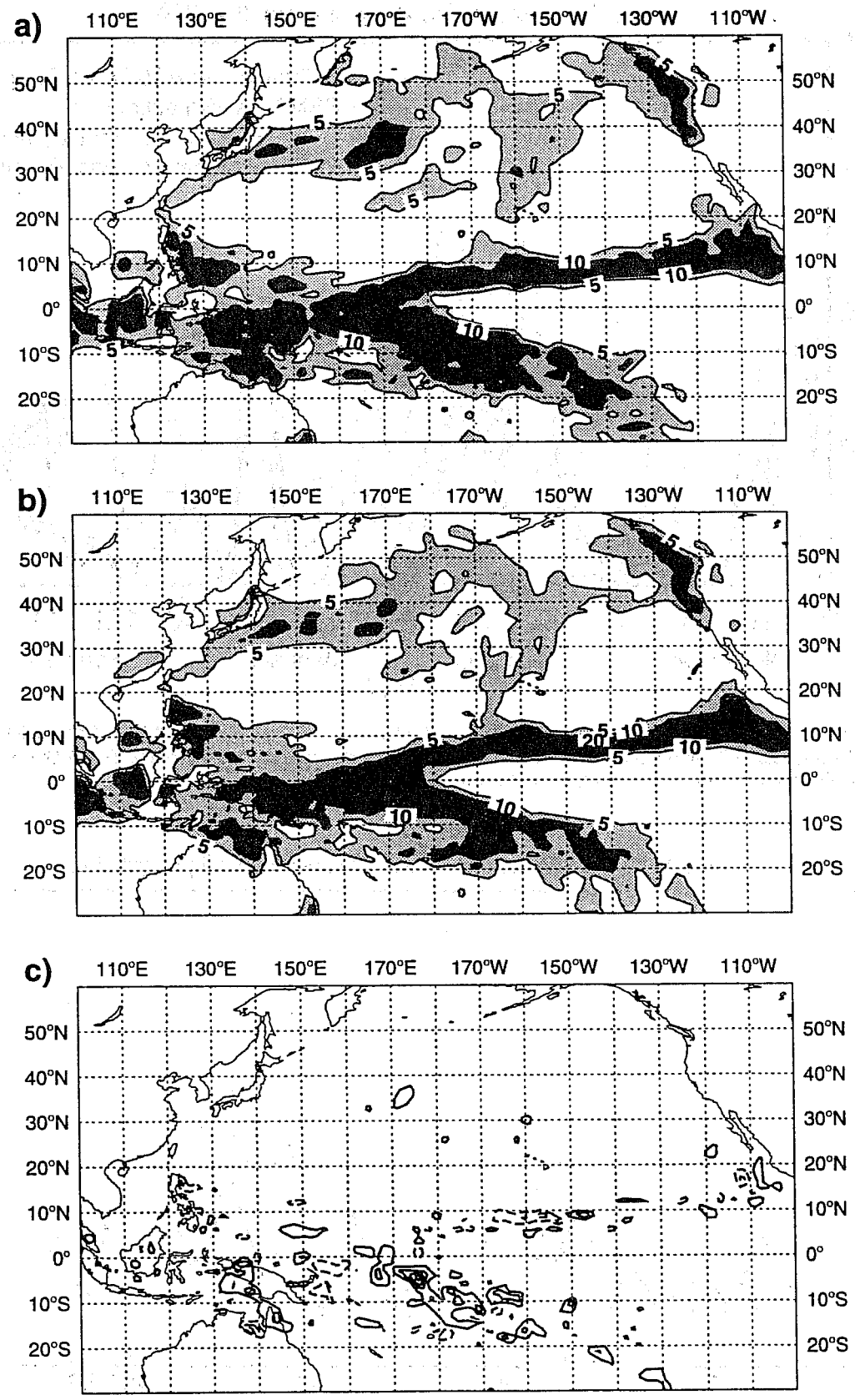

Fig. 3 December 1992 average precipitation accumulated between forecast day 6-7 for (a) the unperturbed and (b) the M-M-M stochastically perturbed integrations; (c): difference between (b) and (a). Shading in panels (a-b) for 5, 10, 20 and 40 mm/day, and 5 mm/day contour interval in panel (c).

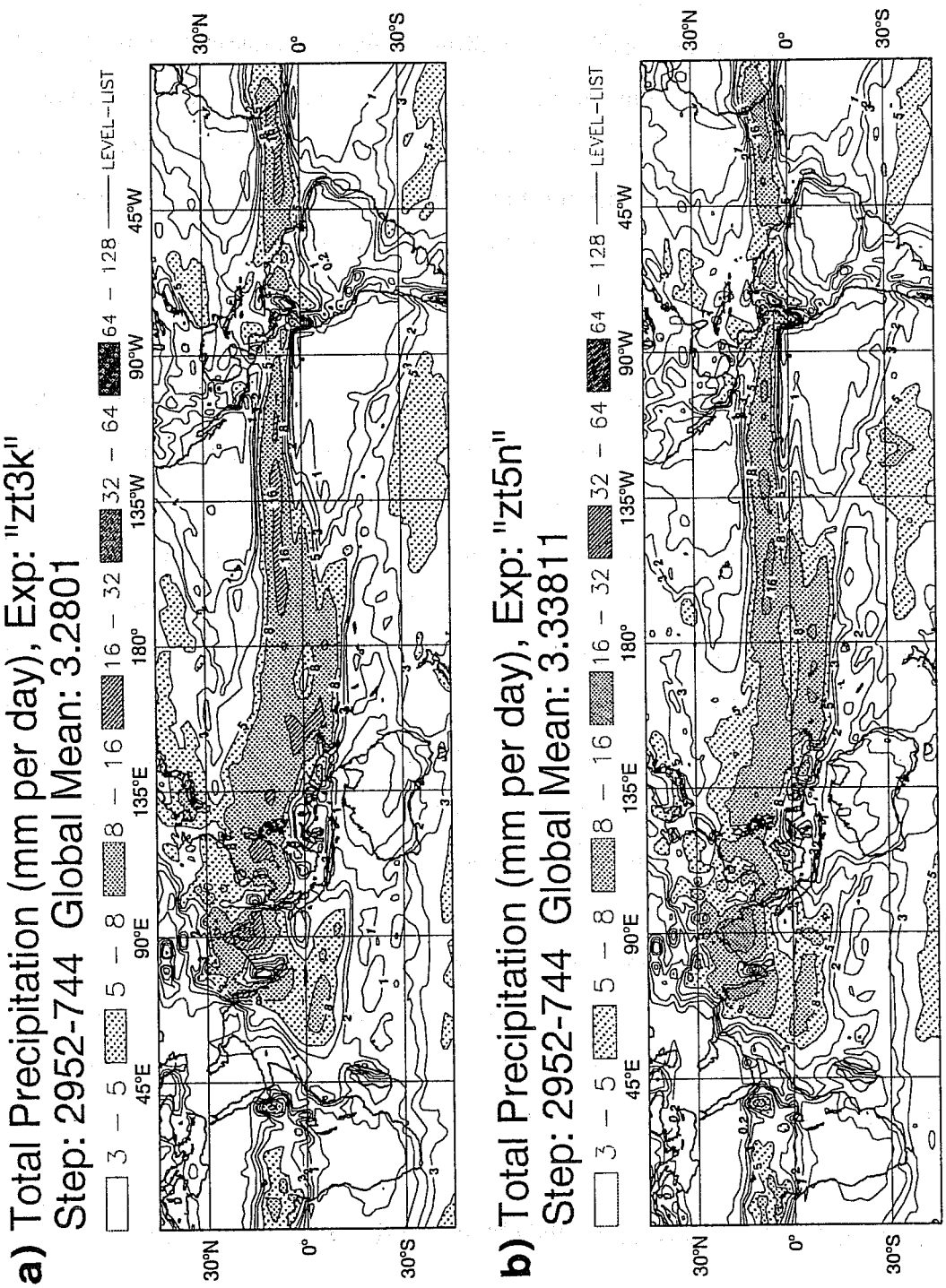

Fig. 4 Total precipitation for the (a) unperturbed and (b) DT M-M-M perturbed long runs.

verify that tendencies are modified in a reasonable and meteorological way. This investigation concentrates on the forecast with starting date 96.10.23 (the same analysis and conclusions could have been drawn by selecting any other case).

Ten 10° square regions characterized by very different weather conditions have been defined (Table 2 and Fig. 5), and average unperturbed and M-M-M stochastically perturbed tendencies have been compared for the first 48 hour forecast. Results for only three of these regions will be discussed: region 1 located under the Pacific high, region 3 located in the Rockies and characterized by intense precipitation, and region 7 located in the Sahara desert. All results relate to model level 23 (around 700 hPa).

Figure 6 shows the time variation of normalized wind, temperature and specific humidity tendencies (each tendency have been normalized by its time average value), and the time variation of the stochastic scaling factor $<r_j>_{D,T}$ for region 1 (Pacific high, dry). Note that only at the end of the forecast period both the wind tendencies (Fig. 6a-b) and the stochastic scaling factors (Fig. 6e) are large, and the impact on the tendencies is visible. Compared to region 1, region 3 (Rockies, wet, Fig. 7) tendencies are characterized by a stronger time variability, essentially because of moist processes (note that the average value for specific humidity tendency is about 50 times larger than that for region 1). Region 7 (Sahara desert, Fig. 8) is characterized by small wind tendencies (Fig. 8a-b), by a very strong daily cycle in the temperature tendency (Fig. 8c), and by small specific humidity tendencies (Fig. 8d).

Generally speaking, results show that the same type of tendency time variation can be seen in the unperturbed and the perturbed cases. Indeed, one could hardly distinguish between unperturbed and perturbed tendencies for all cases apart from the tendency for temperature over the Sahara desert (Fig. 8c), which one expect to be constant during day time.

4. IMPACT OF STOCHASTIC FORCING IN ENSEMBLE PREDICTIONS

4.1 rms spread and rms error

Table 3a summarizes the impact of stochastic forcing on the ensemble spread, and Table 3b and

	Region	Coordinate
>	1	140°W-130°W ; 30°N - 40°N
	2	140°W-130°W ; 10°N - 20°N
>	3	130°W-120°W ; 40°N - 50°N
	4	110°W-100°W ; 30°N - 40°N
	5	50°W- 40°W ; 30°N - 40°N
	6	50°W- 40°W ; 50°N - 60°N
>	7	10°W- 0°W ; 20°N - 30°N
>	8	10°W- 0°W ; 50°N - 60°N
	9	140°E -150°E ; 10°N - 20°N
	10	150°E -160°E ; 0° - 10°S

Table2. Coordinate of the 10° square regions for which diabatic tendencies of unperturbed and DT M-M-M perturbed integrations have been compared. Pointers identify regions for which results are shown in Figs. 8-11.

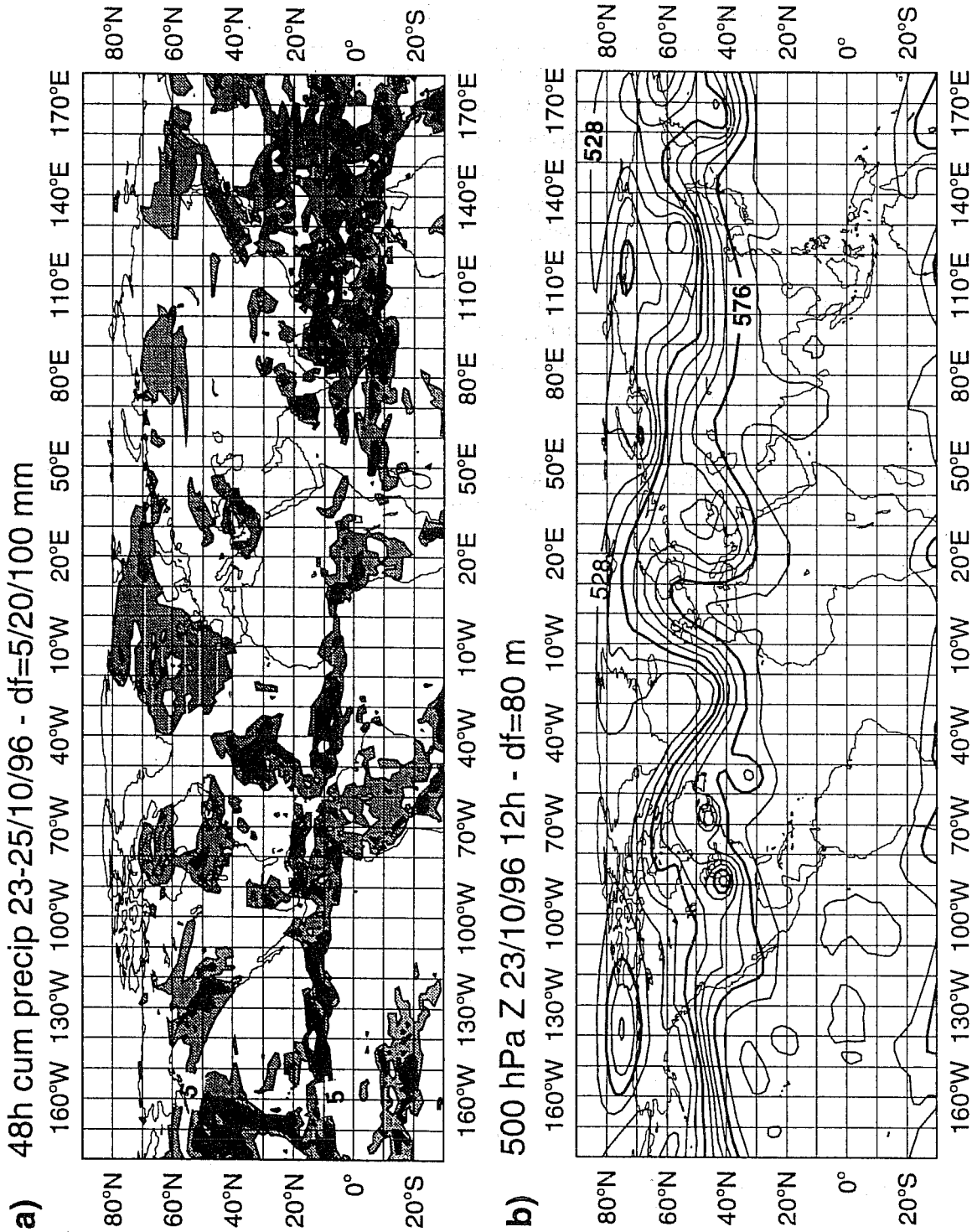

Fig. 5 (a) 48-h accumulated precipitation between 12GMT of 23 and 25 March 1996, and location of the ten 10° square regions where diabatic tendencies have been analyzed (see text). (b): 500 hPa geopotential height at 12 GMT of 23 March 1996. Contour isolines for 5, 20 and 100 mm/day for precipitation, and contour interval 8 dam for geopotential height.

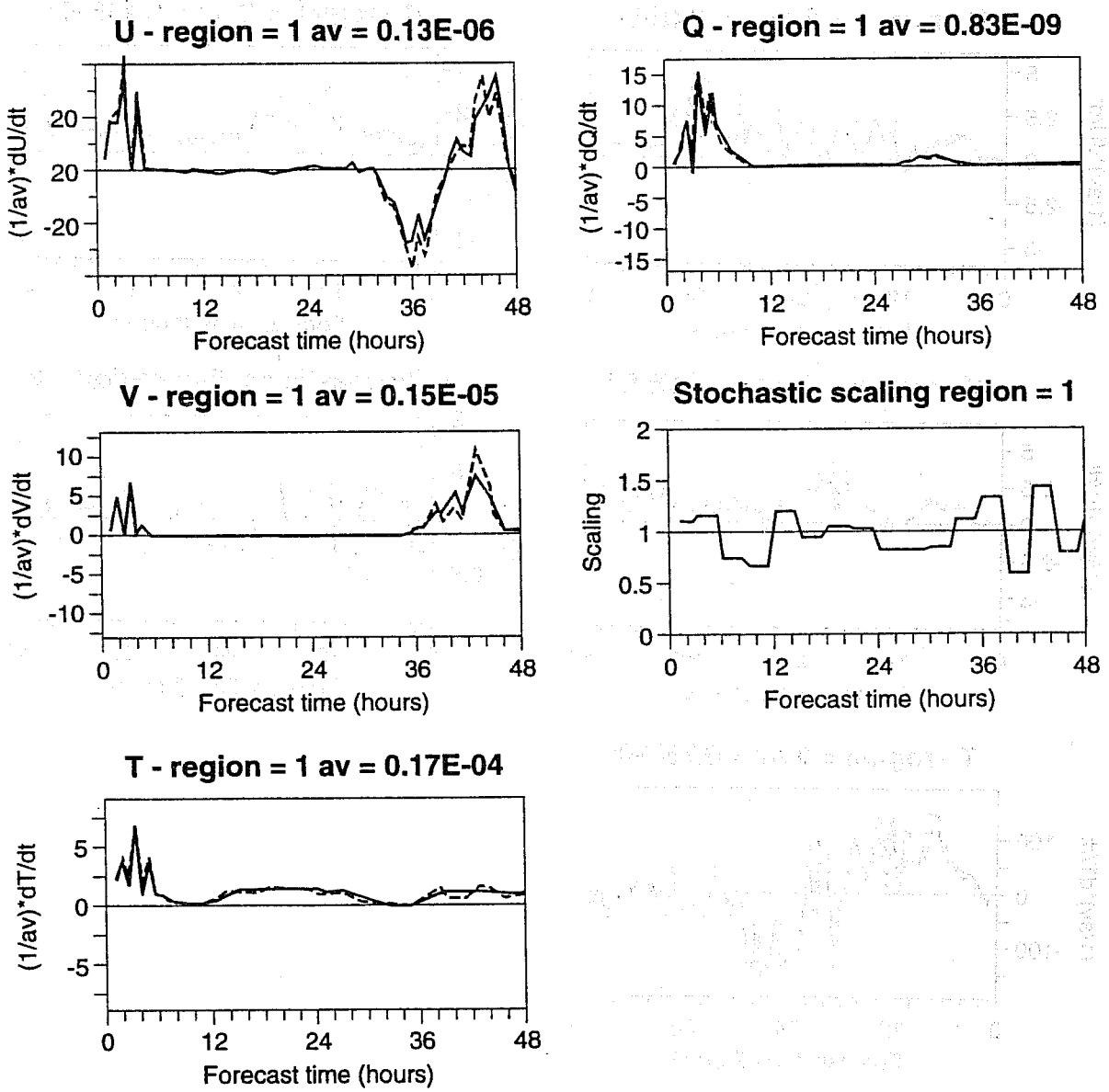

Fig. 6 Region 1 (Pacific high, 140°W-130°W, 30°N-40°N) average value of unperturbed (solid) and perturbed (dash) diabatic tendencies (a) for U and (b) for V wind components, (c) for temperature, and (d) specific humidity during the first 48 hour forecasts normalized by their respective average value. (c) stochastic scaling factors $\langle r_j \rangle_{D,T}$. 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.

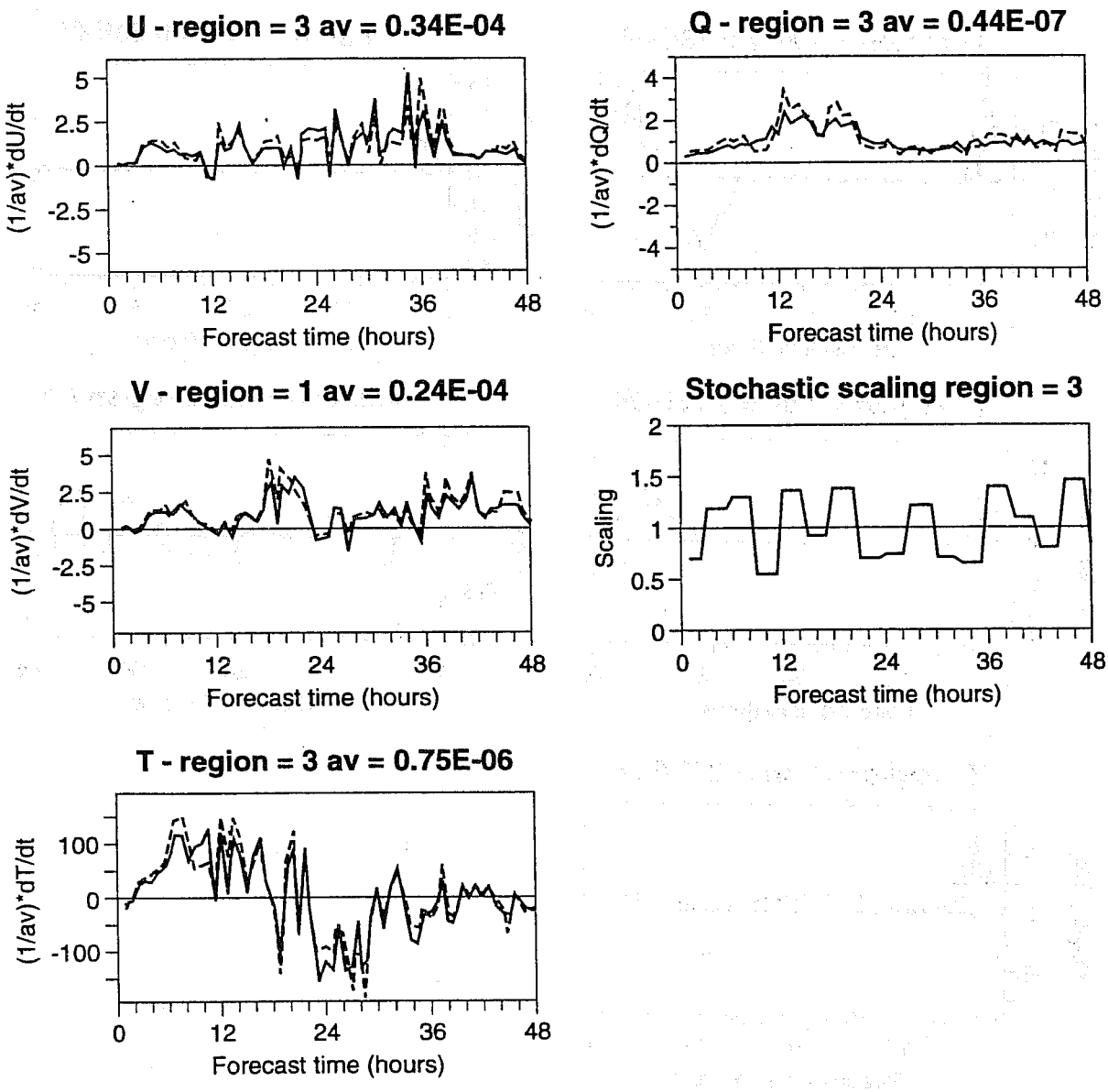

Fig. 7 As Fig. 8 but for region 3 (Rockies, 110°W-100°W, 40°N-50°N).

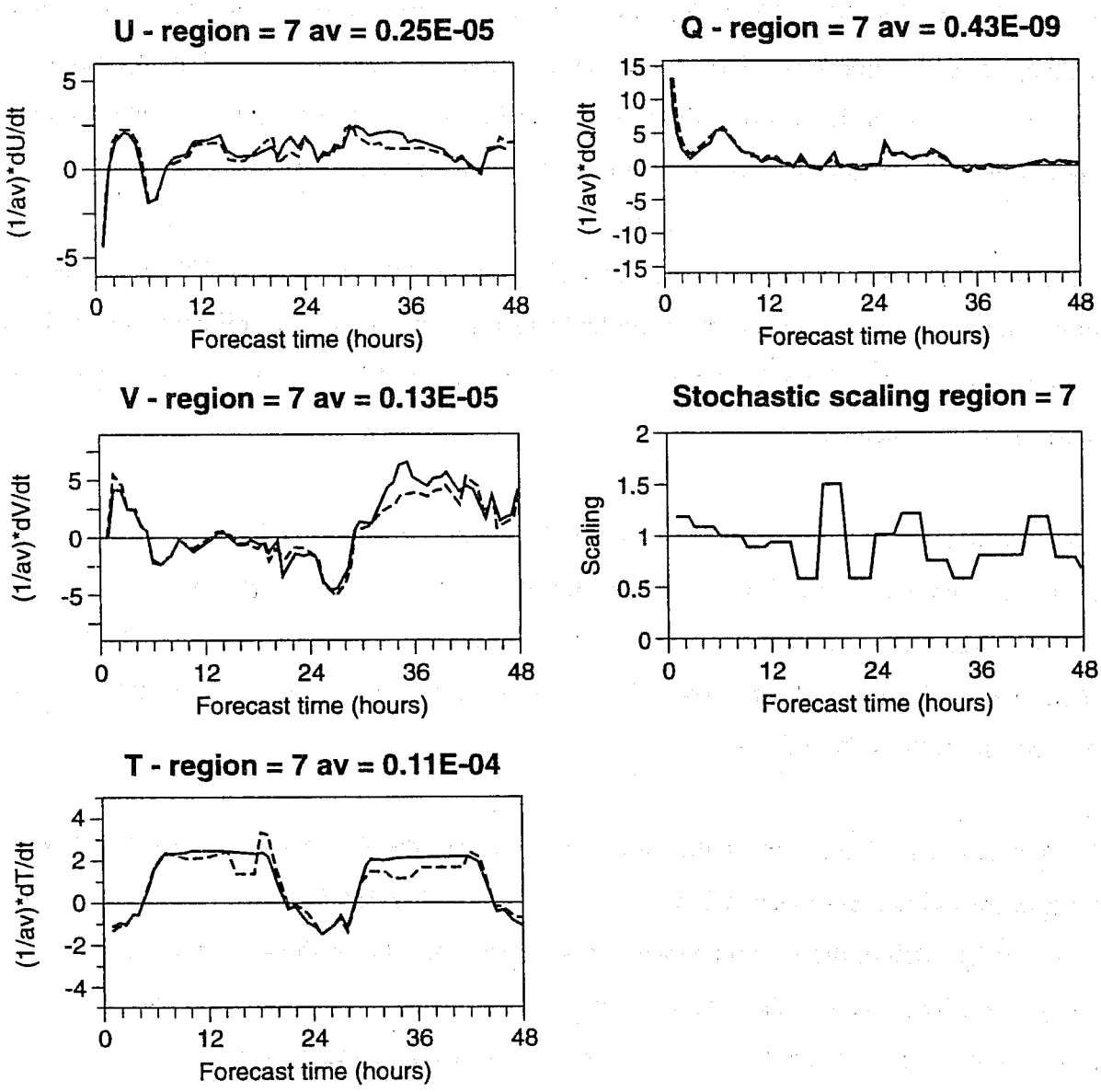

Fig. 8 As Fig. 8 but for region 7 (Sahara desert, 10°W-0°, 20°N-30°N).

3c on mean rms error of the perturbed members and on rms error of the ensemble mean.

Table 3a shows that, for both DT and ICDT 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 3a 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 3a suggests that configurations in which the random forcing is changed every time step induce a small ensemble spread increase.

Focusing on the forecast error at day 7, Table 3b shows that there are six ICDT configurations for which the average rms error of the perturbed forecast is smaller than the reference value given by configuration IC. Table 3c shows that configurations M-M-H and M-L-M have the most skilful ensemble mean scores.

Table 3d summarizes these results, and shows that configuration M-M-M is the only one with an increased ensemble spread, with a decreased average rms error of the perturbed forecasts, and with a decreased ensemble mean rms error, and shows that for four other configurations (H-L-M, M-M-H, L-H-H and L-M-H) the ensemble spread increases and there is a positive impact on at least on one of the skill indicators.

The comparison of the skill of the ensemble mean for the full forecast range for these 5 configurations indicates that configurations H-L-M (solid) and M-M-H (dash) have a negative impact on the skill of the ensemble mean up to forecast day 4 (not shown), and thus should be disregarded. The analysis of the ensemble spread shows that configurations M-M-M, L-H-H and L-M-H have a very similar level of spread during the whole forecast range.

It is interesting to contrast the results in Table 3a with the results in Table 4 of *Houtekamer et al* (1996), which shows that perturbing the model increased the day 7 spread by about 5.6%.

4.2 Other measures of ensemble skill

The results presented so far have restricted the range of possible configurations to M-M-M, L-H-H and L-M-H. These three configurations were compared with the reference IC ensemble.

	<*spr*>			
Exp	fc d+3 DT - ICDT	fc d+5 DT - ICDT	fc d+7 DT - ICDT	% spr increase d+7 ICDT
H-L-M	15.4 - 31.5	30.5 - 55.7	49.3 - 74.6	$10 \le ds$
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	$3.5 \le ds \le 10$
M-M-H *	17.1 - **32.4**	33.8 - **57.3**	53.0 - **76.3**	$10 \le ds$
M-M-M	8.7 - 29.2	20.9 - 52.1	36.7 - 70.4	$3.5 \le ds \le 10$
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	$3.5 \le ds \le 10$
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	$3.5 \le ds \le 10$
L-L-H	4.3 - 28.5	11.7 - 51.0	23.5 - 68.8	
IC	28.3	50.2	67.8	
control	34.2	60.6	88.9	

Table 3a.

	<*err_{pf}*>			
Exp	fc d+3 DT - ICDT	fc d+5 DT - ICDT	fc d+7 DT - ICDT	$derr_{pf}$ d+7 ICDT
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	$derr_{pf} \le 0$
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	$derr_{pf} \le 0$
M-M-L	35.7 - 43.4	62.5 - 71.8	90.8 - 92.5	$derr_{pf} \le 0$
M-L-M *	34.2 - 43.2	60.7 - 71.9	88.8 - **92.2**	$derr_{pf} \le 0$
M-L-L *	34.1 - 43.2	60.6 - 71.8	89.1 - **92.2**	$derr_{pf} \le 0$
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	$derr_{pf} \le 0$
IC	43.3	71.9	92.7	
control	34.2	60.6	88.9	

Table 3b.

	$\langle err^{IC}_{con} \rangle - \langle err_{mean} \rangle$			
Exp	fc d+3 DT - ICDT	fc d+5 DT - ICDT	fc d+7 DT - ICDT	$derr_{mean}$ d+7 (%) ICDT
H-L-M	-0.3 - -0.4	0.0 - 2.6	3.0 - 14.6	$-3.5 \le derr_{mean} \le 0$
H-L-L	0.4 - 0.4	2.0 - **4.4**	2.9 - 14.5	$-3.5 \le derr_{mean} \le 0$
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**	$derr_{mean} \le -3.5$
M-M-M	0.6 - **0.5**	1.4 - 4.0	4.4 - 14.7	$derr_{mean} \le -3.5$
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**	$derr_{mean} \le -3.5$
M-L-L	0.2 - 0.2	0.6 - 3.8	1.2 - 14.8	$derr_{mean} \le -3.5$
L-H-H	0.4 - **0.5**	1.1 - 4.0	4.2 - 14.6	$-3.5 \le derr_{mean} \le 0$
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	$derr_{mean} \le -3.5$
L-L-H	0.2 - 0.3	0.3 - 4.0	1.6 - 14.6	$-3.5 \le derr_{mean} \le 0$
IC	0.1	3.7	14.2	

Table 3c.

Exp	spr increase (%) (Table 3a)	$\langle err_{pf} \rangle$ (Table 3b)	$\langle err^{IC}_{con} \rangle - \langle err_{mean} \rangle$ (%) (Table 3c)
> H-L-M	$10 \le ds$		$-3.5 \le derr_{mean} \le 0$
H-L-L		$derr_{pf} \le 0$	$-3.5 \le derr_{mean} \le 0$
M-H-M	$3.5 \le ds \le 10$		
> M-M-H	$10 \le ds$ *		$derr_{mean} \le -3.5$ *
> M-M-M *	$3.5 \le ds \le 10$	$derr_{pf} \le 0$	$derr_{mean} \le -3.5$
M-M-L		$derr_{pf} \le 0$ *	
M-L-M		$derr_{pf} \le 0$ *	$derr_{mean} \le -3.5$ *
M-L-L		$derr_{pf} \le 0$	$derr_{mean} \le -3.5$
> L-H-H	$3.5 \le ds \le 10$		$-3.5 \le derr_{mean} \le 0$
L-H-M			
> L-M-H	$3.5 \le ds \le 10$		$derr_{mean} \le -3.5$
L-L-H		$derr_{pf} \le 0$	$-3.5 \le derr_{mean} \le 0$

Table 3d.

Table 3. List of the average among the 5 case studies of (a) the rms ensemble spread (m), (b) the mean rms error of the perturbed forecasts (m), (c) the difference between the rms error of the control of configuration IC and the rms error of the ensemble mean (m), and (d) summary of the results of Tables 3a-c. In each Table, the best result at each forecast time is in bold, and a * near the configuration identifier marks the ICDT configuration with the best result at forecast day 7. Results refer to the 500 hPa geopotential height over the NH extra-tropics.

Significant differences between 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 4 shows that, according to these two measures, the stochastically perturbed ensembles, due to the larger grid-point ensemble spread induced by the perturbed tendencies, have better skill scores.

Let us now consider precipitation prediction, and compare ensemble performances using probabilistic measures as the Ranked Probability Skill Score (RPSS), the area under a Relative Operating Characteristic (ROC) curve and the Brier skill score (*Stanski et al*, 1989, *Buizza and Hollingsworth*, 1998). The comparison of the RPSS (computed using consistency as standard forecast) for 12-hour precipitation predictions indicate that ensembles run with stochastic forcing are more skilful than the reference IC configuration (Fig. 9). Indeed, Fig. 9 shows that predictability increases by almost 2-days in the stochastically perturbed ensembles.

The positive impact of stochastic forcing is confirmed by 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. 10). If we consider an area of 0.7 as the limit for an useful prediction (*Buizza and Hollingsworth*, 1998), Fig. 10d shows that about 18 hours of predictability are gained for the largest precipitation threshold. 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).

3.3 Synoptic evaluation for 500 hPa geopotential height

In this Section we analyze the structure of the forecast divergence induced by stochastic forcing, and we compare single forecasts run with and without stochastic forcing. Since differences between the three stochastically perturbed ensembles M-M-M, L-H-H and L-M-H are small, we will restrict our attention to configuration M-M-M and to the reference ensemble IC.

Table 5 lists the rms error over Europe (20°W-45°E;30°N-75°N) of the best IC and ICDT 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. 11a) is smaller than the control

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

Table 4.

Table 4. 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.

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

Table 5.

Table 5. Root-mean-square error over Europe of ensemble members 8 (best ICDT 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.

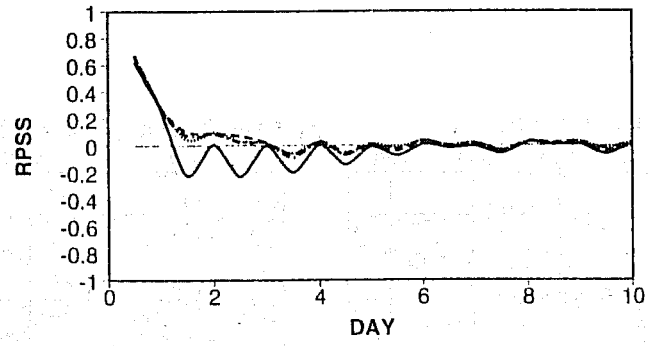

Fig. 9 Ranked Probability Skill Score for precipitation prediction over the NH extra- tropics for configurations IC (solid), ICDT M-M-M (dash), ICDT L-H-H (dot) and ICDT L-M-H (chain dash).

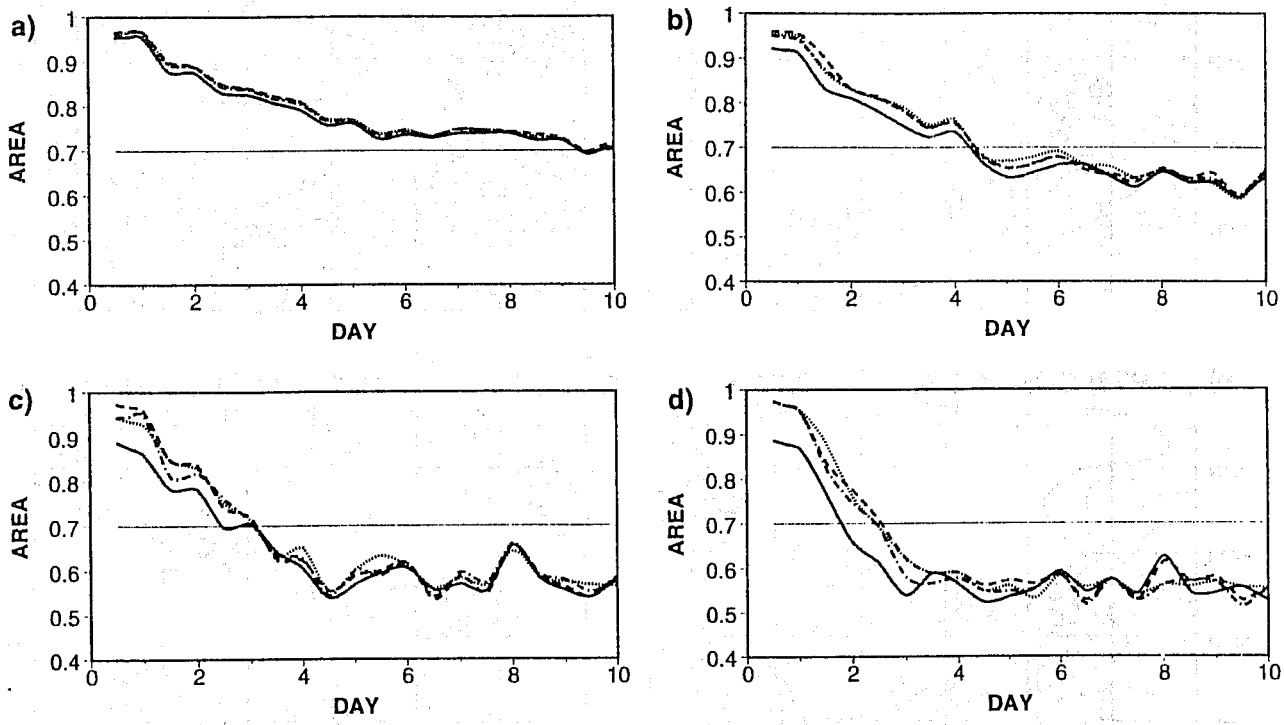

Fig. 10 (a) Area under the average ROC curve for the prediction of the event "12-h accumulated precipitation larger than 2mm/12h" for the NH extra-tropics, for configurations IC (solid), ICDT M-M-M (dash), ICDT L-H-H (dot) and ICDT L-M-H (chain dash). (b): as (a) but for the event "12-h accumulated precipitation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

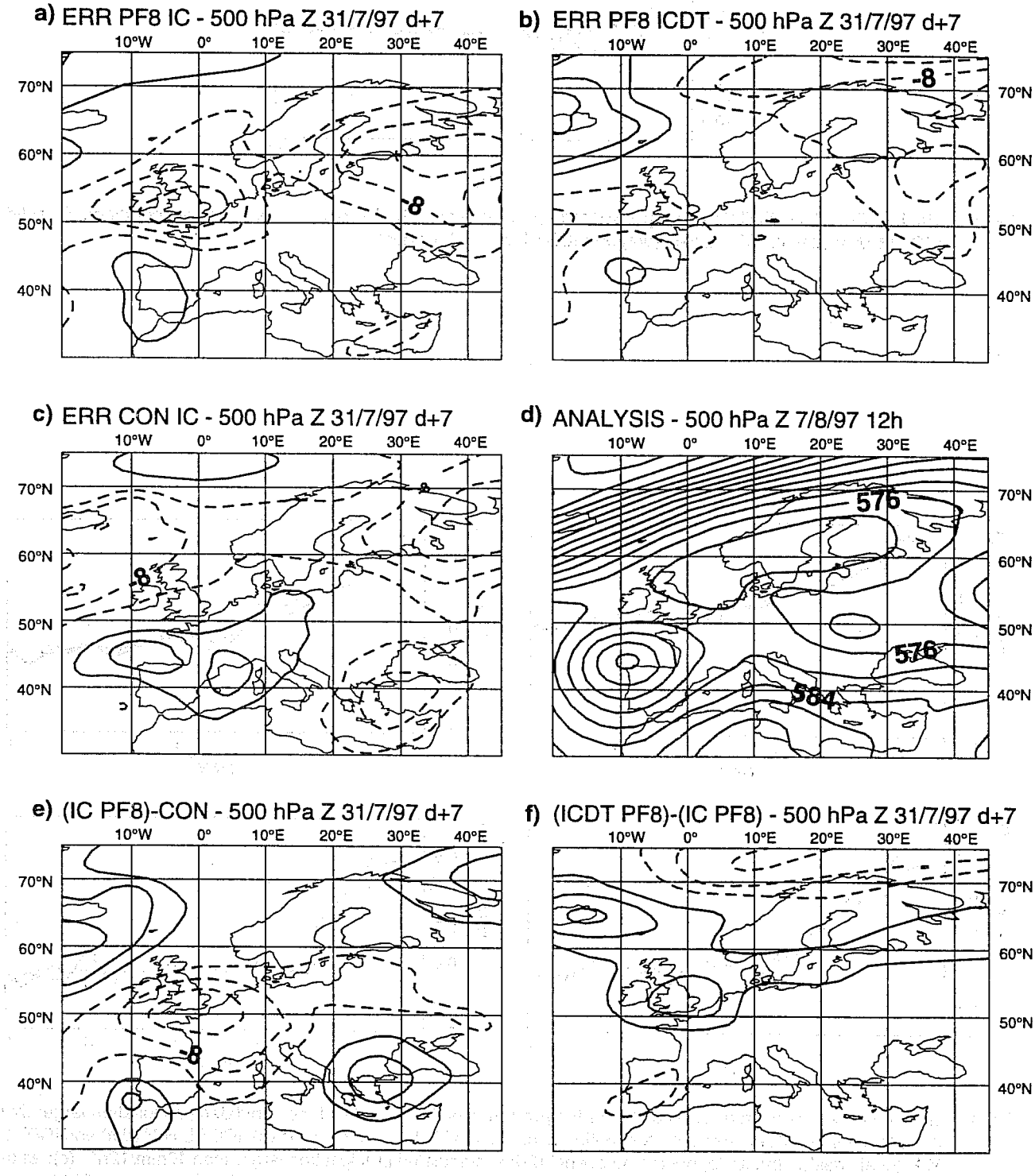

Fig. 11 Forecast error at day 7 of the 8th member of configurations (a) IC and (b) ICDT M-M-M, and of (c) the control forecast. (d): verifying analysis. (e): difference between the 8th member of configuration IC and the control, and (f) difference between the 8th member of configuration ICDT M-M-M and of configuration IC. Results refer to the 500 hPa geopotential height field for the 97.07.31 case study. Contour interval 4 dam.

error (Fig. 11c), because its divergence from the control (Fig. 11e) induced by its initial perturbation brings it closer to the verifying analysis (Fig. 11d). Furthermore, the error of the ICDT M-M-M 8th member (Fig. 11b) 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. 11f).

A more complete picture of the difference between the IC and ICDT 8th members at forecast day 7 in terms of temperature is given in Fig. 12 for the whole Atlantic-European region, and in Fig. 13 for two vertical cross sections, one from the Labrador Peninsula to Sardinia (Fig. 13a-b), and one from Newfoundland to the North Cape (Fig. 13c-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, as round Iceland (Fig. 13a-b), stochastic forcing reinforces the divergence induced by the initial perturbation, while in other regions, as around (40°W,57°N) (Fig. 13a-b), stochastic forcing produces a divergence which is opposite to the one given by the initial perturbations. Further, there are regions, as in the Norwegian sea (Fig. 12a-b), where the stochastic forcing is active but initial perturbations have no effect on forecast divergence.

Considering the average rms ensemble spread, the current formulation stochastic forcing induces an average divergence in areas where initial conditions have already produced ensemble spread (not shown).

Similar considerations could have been drawn considering other case studies.

5. CONCLUSIONS

A parametrization of random errors in the representation of diabatic forcing is described. The parametrization is based on a stochastic perturbation to the parametrized diabatic tendencies whose amplitude is proportional to the total diabatic tendency. In its current formulation, there are three basic parameters governing overall amplitude, and spatio-temporal correlation.

A number of diagnostics have been described and a choice of parameters is made. The primary

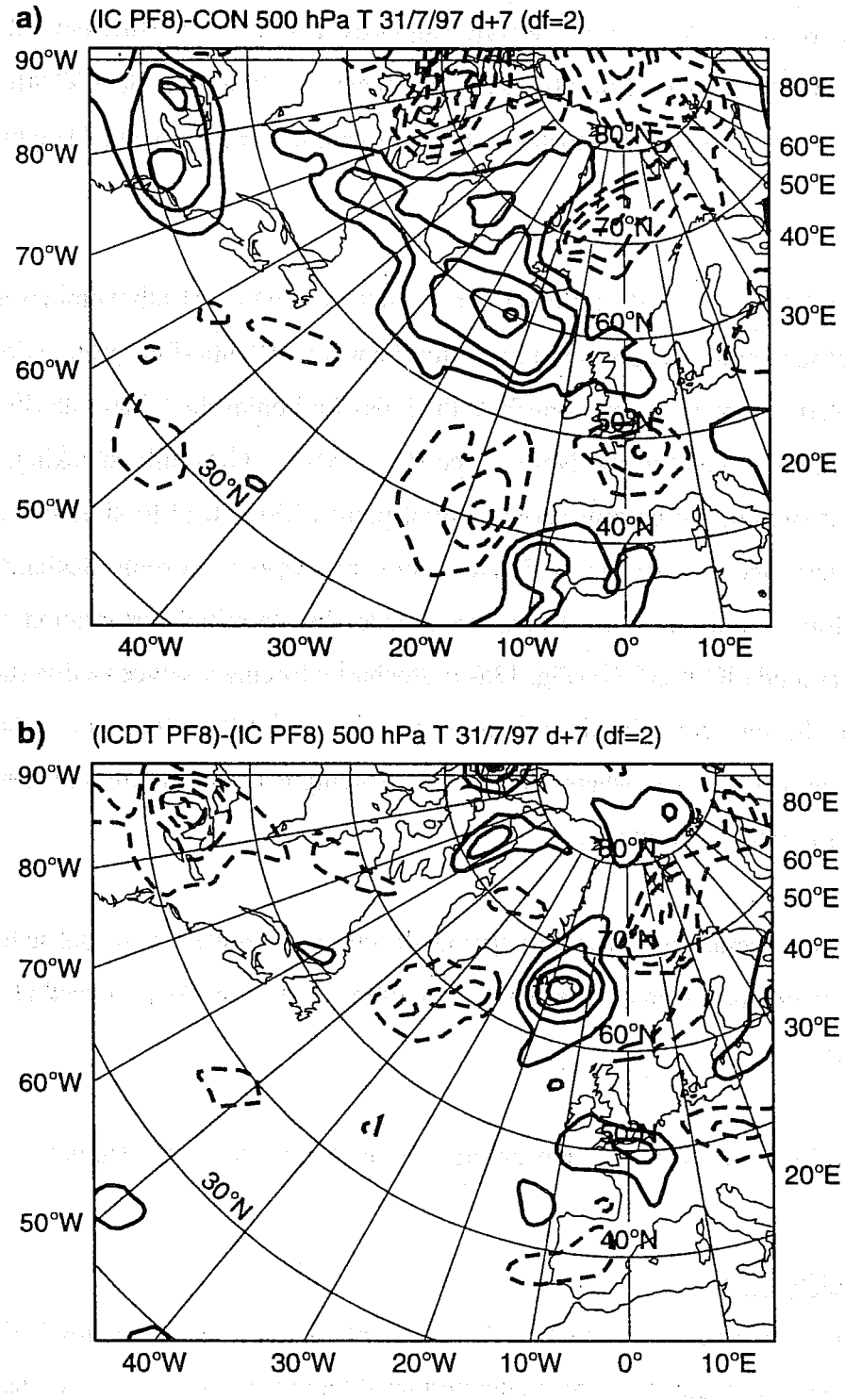

Fig. 12 500 hPa temperature day 7 forecast (a) difference between the 8th member of configuration IC and the control, and (b) difference at forecast day 7 between the 8th member of configuration ICDT M-M-M and configuration IC, for the 97.07.31 case. Contour interval 2 degrees.

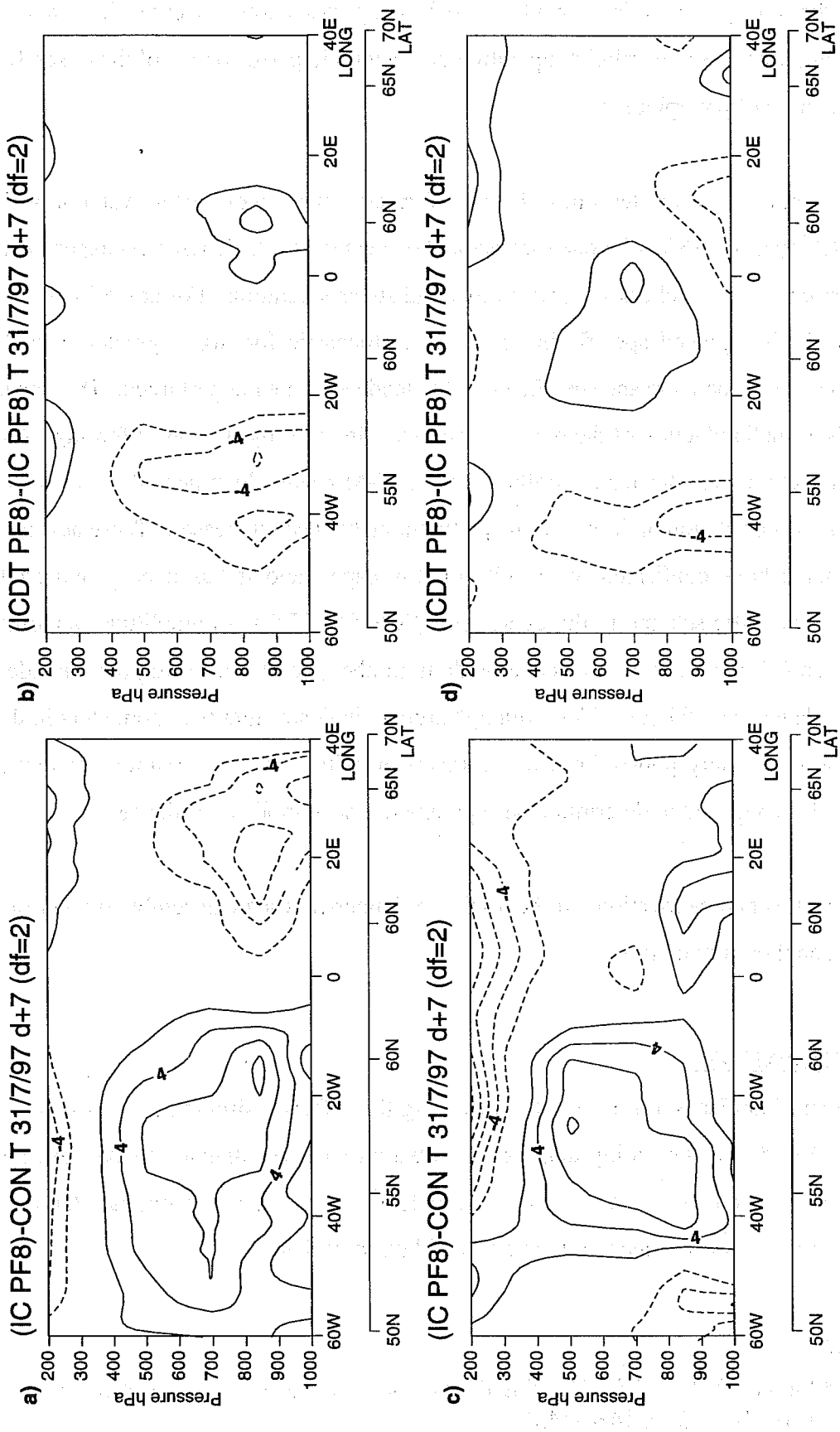

Fig. 13 Temperature vertical cross section from the Labrador Peninsula to Sardinia at forecast day of (a) difference between the 8th member of configuration IC and the control, and (b) difference at forecast day 7 between the 8th member of configuration ICDT M-M-M and configuration IC. (c-d): as (a-b) but for a vertical cross section from Newfoundland to the North Cape. Contour interval 2 degrees.

application of the scheme is for the European Centre for Medium-Range Weather Forecasts Ensemble Prediction System. It is shown how stochastic physics both increases the spread of the ensemble and improves the relative operating characteristic performance of the ensemble, particularly in terms of precipitation.

Following these results, further tests have been performed using a more recent version of the ECMWF model (cycle CY18R6) for one summer and one winter week. These new experiments have been performed with a check introduced to avoid super-saturation. For any grid point at any time step, if the updated specific humidity after stochastic forcing is greater than the saturation value, the temperature and specific humidity tendencies are not perturbed. The impact of this check is a small reduction of the overall divergence induced by stochastic forcing. As a consequence, to achieve a performance similar to the M-M-M results documented in this paper, the spatio-temporal correlation of the stochastic perturbation has been increased. These new two-week experiments have confirmed the results of this paper, and it has been proposed to implement the stochastic scheme in the operational ECMWF EPS with amplitude randomly selected between 0.5 and 1.5, with the same random number used for all grid points inside a 10°x10° box updated every 6 hours. This "tuning" process indicates that the parameters in this stochastic scheme are only poorly known. Comparison with either observations or cloud-resolving models is required to determine these parameters with mode confidence.

It is possible that this parametrization will be useful for longer-timescale ensemble predictions, e.g. on the seasonal or climate timescale.

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 diabatic tendencies, and David Gregory for producing Fig. 4. We acknowledge the advise of Christian Jakob to implement a super-saturation check on the perturbed humidity field.

REFERENCES

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 Hollingsworth, A, 1998. Probability precipitation prediction using the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, submitted.

Buizza, R, and Palmer, T N, 1998. Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, in press.

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.

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.

Harrison, M S J, Palmer, T N, Richardson, and Buizza, R, 1998. Analysis and model dependencies in medium-range ensembles: two transplant case studies. *Q. J. R. Meteorol. Soc.*, submitted.

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, 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.

Richardson, D, 1998. The relative effect of model and analysis differences on ECMWF and UKMO operational forecasts. In this volume, 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.

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

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

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