TECHNICAL MEMORANDUM

240

Recent experimentation on 4D Var and first results from a Simplified Kalman Filter

F. Rabier, J F. Mahfouf, M. Fisher,
H. Järvinen, A. Simmons, E. Andersson,
F. Bouttier, P. Courtier, M. Hamrud,
J. Haseler, A. Hollingsworth, L. Isaksen,
E. Klinker, S. Saarinen, C. Temperton,
J-N. Thépaut, P. Undén and D.Vasiljević

Research Department

October 1997

This paper has not been published and should be regarded as an Internal Report from ECMWF.
Permission to quote from it should be obtained from the ECMWF.

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

Recent experimentation on 4D-Var and first results from a simplified Kalman filter

F. Rabier, J-F. Mahfouf, M. Fisher, H. Järvinen, A. Simmons, E. Andersson, F. Bouttier, P. Courtier[1], M. Hamrud, J. Haseler, A. Hollingsworth, L. Isaksen, E. Klinker, S. Saarinen, C. Temperton, J N. Thépaut[2], P. Undén and D. Vasiljevic

[1] Current affiliation: CNES, Paris, France

[2] Current affiliation: Météo-France, Toulouse, France

Abstract

Four-dimensional variational assimilation (4D-Var) has been extensively tested at high resolution on the FUJITSU, and compared with three-dimensional variational assimilation (3D-Var). To validate the migration a 2-week period studied on the CRAY was repeated at higher resolution on the FUJITSU, with a similar improvement brought by 4D-Var. Then, a more thorough investigation of the poor performance of 4D-Var in the Tropics revealed some problems in the way the adiabatic non-linear normal mode initialization of the increments was performed. Going from four outer loops to only one (as in 3D-Var) helped to reduce the problem, together with a change to the new background formulation and an initialization of only the small scales. Tropical scores then became only marginally worse for 4D-Var than for 3D-Var.

Twelve weeks of experimentation with the one outer-loop 4D-Var and the new background formulation have been studied. In the medium range, the improvement is more pronounced in the Southern Hemisphere. In the short-range, each two to three-week period has been found to be positive. The better short-range performance of the 4D-Var system was also shown by the fits of the background fields to the data. The comparison of forecasts to data in the Northern Hemisphere up to day 10 confirms the better scores for 4D-Var. In individual synoptic cases corresponding to interesting IOPs during the FASTEX period, 4D-Var is seen to perform better than 3D-Var during rapid cyclogenesis.

A first comprehensive set of linear physics has been developed for 4D-Var applications. First, it has been evaluated by comparing the evolution of analysis increments with respect to non-linear integrations including the full physics. A better agreement of the evolved increments is found when the physics is included. The inclusion of this package in a 4D-Var "2-update" configuration has a positive impact on the performance of the analysis in the tropics, with a reduction of the spin-down of precipitation in the subsequent forecasts, and improved wind scores. The averaged extratropical scores averaged over 8 weeks show a slight improvement brought by the physics.

Some structure functions were illustrated in the 4D-Var case for a height observation inserted at the beginning of the assimilation window, in the middle or at the end. The dynamical processes seem to be relevant, even on a short 6-hour assimilation period. More influence of the dynamics could be taken into account by properly cycling 4D-Var using a simplified Kalman filter (SKF), which is currently being developed, and whose feasibility has been demonstrated.

ECMWF LIBRARY
17 NOV 1997
No. C

1. Introduction

Four-dimensional variational assimilation (4D-Var) minimizes a cost-function measuring the distance between a model trajectory and the available information (background, observations) over an assimilation interval or window. The 4D-Var system is the temporal extension of the three-dimensional variational analysis operational since January 1996 (Courtier et al., 1998, Rabier et al., 1998, Andersson et al., 1998). The 4D-Var algorithm uses the adjoint equations for the computation of the gradient of the cost-function. 4D-Var was first applied to simple models (Le Dimet and Talagrand, 1986, Lewis and Derber, 1985, Courtier and Talagrand, 1987, Talagrand and Courtier, 1987), before being tested in the context of primitive equations models (Thépaut and Courtier, 1991, Rabier and Courtier, 1992, Navon et al, 1992, Zupanski, 1993). It is being tested at ECMWF in its incremental formulation (Courtier et al, 1994) which comprises running a high-resolution model with the full physical parametrization package to compare the atmospheric states with the observations as part of the evaluation of the cost-function and a low resolution model with simplified physics to minimize the cost-function. Results obtained on the CRAY at resolution T106L31/T63L31 with 3 to 4 outer-loops and 15 to 25 inner-loops, and very simplified physics (horizontal and vertical diffusion and a surface drag) are described in Rabier et al, 1997. In summary, it was found that 4D-Var using a 6 or 12-hour window performed better than 3D-Var over a 2-week assimilation period, whereas 4D-Var using a 24-hour window did not. The poorer performance of 4D-Var with a relatively long assimilation window could be partly explained by the fact that, in these experiments, the tangent-linear and adjoint models used in the minimization were only approximations of the assimilating model (lower resolution, crude physics). The error these approximations introduced in the time evolution of a perturbation affected the convergence of the incremental 4D-Var , with larger discontinuities in the values of the cost-function when going from low to high resolution for longer assimilation windows. While the tangent-linear and adjoint of a more accurate physical parametrization package was being developed, the strategy was to concentrate on the 4D-Var with a 6-hour window, the goal being its operational implementation, before revisiting the 4D-Var with longer windows. Two additional two-week periods were then run on the CRAY, which showed a consistent improvement of 4D-Var over 3D-Var in the extratropics. Some problems existed however in the Tropical area.

With the change to a distributed memory FUJITSU computer, some additional work was needed to migrate 4D-Var following the migration of 3D-Var in September 1996. Experimentation with 4D-Var resumed in early 1997, with some new features. The main changes are that the observation screening and quality control are now performed within the variational assimilation (Järvinen and Undén, 1997, Andersson, 1996) (they used to be performed by the Optimum Interpolation), the evaluation of the analysis and background errors involves a Lanczos algorithm (Fisher and Courtier, 1995, Bouttier et al, 1997), and the observation operators are activated in their tangent-linear versions within the minimization (the finite-difference approximation was previously used). Furthermore, there is now the possibility of using more observations from frequently reporting stations in the 4D-Var scheme, and of performing the experiments at higher resolution T213L31/T63L31.

The validation of the migration was performed by re-running the January 1996 period previously run on the CRAY C90. As discussed in section 2, the improvement of 4D-Var over 3D-Var seen earlier is reproduced using the same amount of data in both schemes. However, the use of extra off-time observations is not found to be beneficial. Tropical wind scores are substantially poorer in 4D- than in 3D-Var in these new experiments, as they had been previously. Subsequent tests involving reduction of the number of outer loops from four to one, use of a new background term, and removal of initialization of large scales, resulting in a better tropical behaviour, are explained in section 3. Section 4 presents the results of the baseline 4D-Var (one outer loop, new background term) on a total of twelve 2-week periods. A comprehensive linearized physics and its adjoint has been implemented in 4D-Var in an efficient configuration. Details and results are given in section 5. Section 6 illustrates the structure functions used implicitly in 4D-Var and explains how these can be further modified by using a simplified Kalman filter. Discussion of results and of cost issues is given in section 7.

2. Validation of migration to FUJITSU

2.1 Repeat of the January 1996 period

One of the periods tested at resolution T106L31 on the CRAY was repeated at T213L31 on the FUJITSU (16 to 29 January 1996). The version of the analysis which is used is with the "old" formulation of the background term, as operational in early 1997 at ECMWF. 4D-Var is implemented with 4 outer loops and 20 iterations within each minimization. 4D-Var's configuration uses the same amount of data (no extra off-time data) as 3D-Var. 4D-Var tested on this period on the FUJITSU does not include any variational quality control, because this had not been tested yet in the 4D-Var context. However, it is not believed to change significantly the scores. Results are presented in Figure 1. 3D-Var on the CRAY is shown as a solid line, 3D-Var on the FUJITSU as a dashed line, 4D-Var on the CRAY as a dotted line and on the FUJITSU as a dash-dotted line. One can notice the improvement of both 3D-Var and 4D-Var when going to the high-resolution FUJITSU version. The important point for our validation exercise is that the improvement when going from 3D- to 4D-Var is retained in the later version. This is true for the Northern Hemisphere as a whole (top panel), and also for the dramatic improvement in the North Pacific area at least up to day six (bottom panel). The tropical scores are also consistent with the CRAY experiments, with scores markedly worse for 4D- than for 3D-Var. As an example, the averaged FUJITSU tropical wind scores verified against their own analysis at 850 hPa are 4.1 m/s for 3D-Var and 4.8m/s for 4D-Var at day 3.

2.2 Influence of extra off-time data

Unlike a static assimilation scheme, 4D-Var assimilates the observations along the trajectory that extends over the assimilation window. There are two related benefits. First, the observations are used at the appropriate time. Second, many observations from frequently reporting stations can be used within one assimilation period. These extra observations are a resource that has not been fully utilized with earlier static assimilation schemes. The observations are selected for the assimilation during the initial high resolution trajectory integration. At this stage, all the necessary information is available for the quality assessment of the observations. From the subset of good quality observations, all redundant information is rejected so that a unique set of observations is left for the assimilation. In 3D-Var the preference is given to the observations that are close to the middle of the assimilation period. In 4D-Var, the observations are organized into one hour time slots, and the comparison of the trajectory with observations is accordingly done once per hour. Within a time slot, the preference is given to observations that are close to the middle of the time slot. The IFS is however designed so that either an hourly or a six-hourly observation screening can be performed for 4D-Var. The effect of this choice on the number of observations used in the assimilation is largest for observation types that report frequently, like SYNOPs and DRIBUs. An illustrative example is given in Fig. 2 which displays the number of available observations in different time slots and the number actually used. The number of SYNOP surface pressure observations used in a two-week period in the 4D-Var assimilation is roughly twice the number in a corresponding 3D-Var assimilation for the same period. The difference arises from the observations made at times other than the main observing times.

The impact of these off-time observations was tested for the January 1996 period in addition to the previous experiments: 3D-Var and 4D-Var without the off-time observations. As shown in Figure 3, in the Northern Hemisphere both 4D-Var systems produce equally good forecast scores. In the medium-range the scores are 6 to 12 hours ahead of the 3D-Var scores both at 1000hPa and 500hPa (not shown). The Southern Hemisphere scores are however the best for the 4D-Var system without the off-time observations (dotted line). There the inclusion of the off-time observations deteriorated the 4D-Var scores (dashed line) to the level of, or even below, the 3D-Var scores (solid line).

Investigation of the observation statistics revealed certain stations with significant biases against the background for all time slots. Some of these stations got an increased weight in the analysis as there was an observation contributing in each time slot and therefore large analysis increments were produced in the vicinity of those stations. For isolated

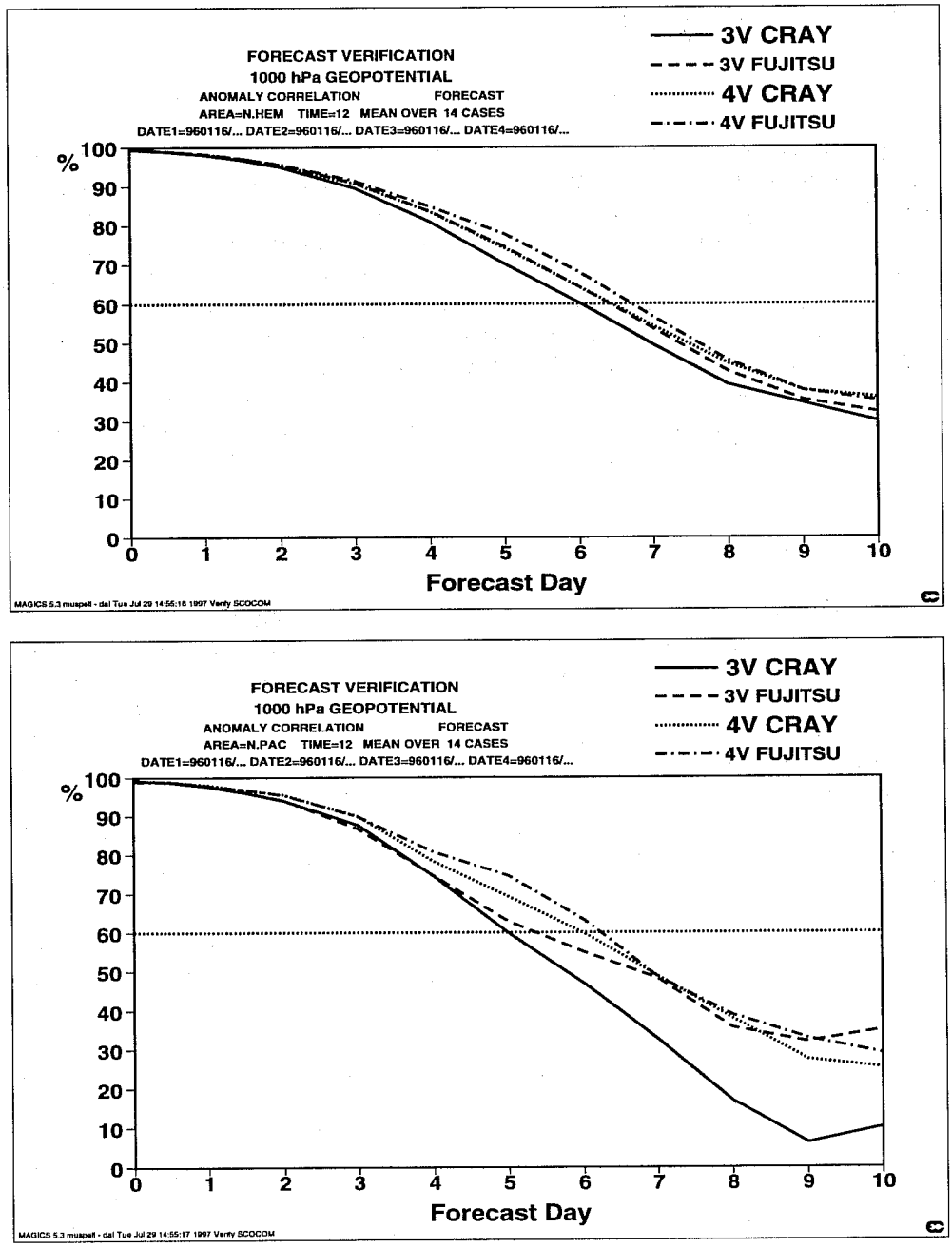

Figure 1 : Anomaly correlation scores for 3D-Var and 4D-Var on the CRAY at T106L31 and on the FUJITSU at T213L31, using the same amount of observations, for the Northern Hemisphere (top panel) and North Pacific (bottom panel) for two weeks in January 1996.

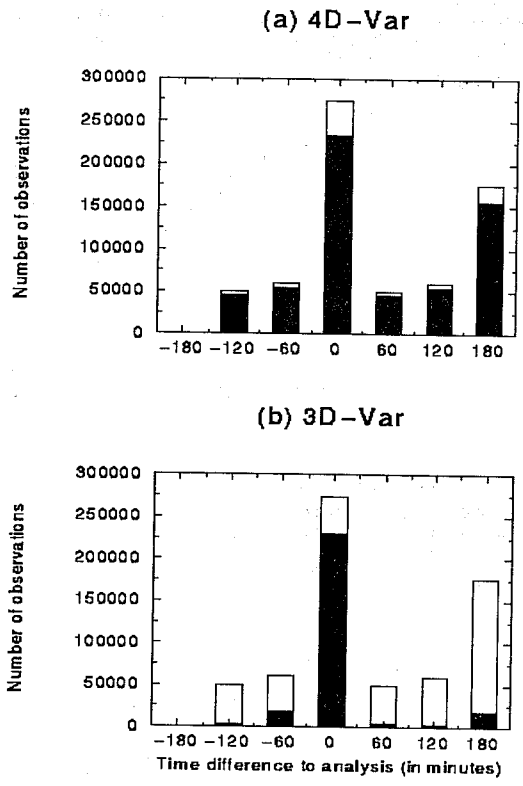

Figure 2 : Time distribution of SYNOP surface pressure observations during a 2-week period. The column height gives the number of all the available observations while the shaded part displays those actually used by the analysis. a) 4D-Var with extra off-time data and b) 3D-Var.

observations, particularly in the Southern Hemisphere, there is currently no mechanism to prevent these unrealistic increments from appearing and developing into forecast errors. In the Northern Hemisphere, in contrast, there are many more observations to constrain the analysis and therefore the forecast scores there are equally good for the two 4D-Var systems. It is clear now that time-correlated observation errors have to be taken into account in 4D-Var in order to make proper use of the off-time observations. This involves changes to the way the observation term of the cost function is calculated for these observations. It is also necessary to perform the variational quality control simultaneously for all the observations from the same station. The actual form of the time-correlation of the observation errors has to be studied and modelled. Work is in progress on this subject.

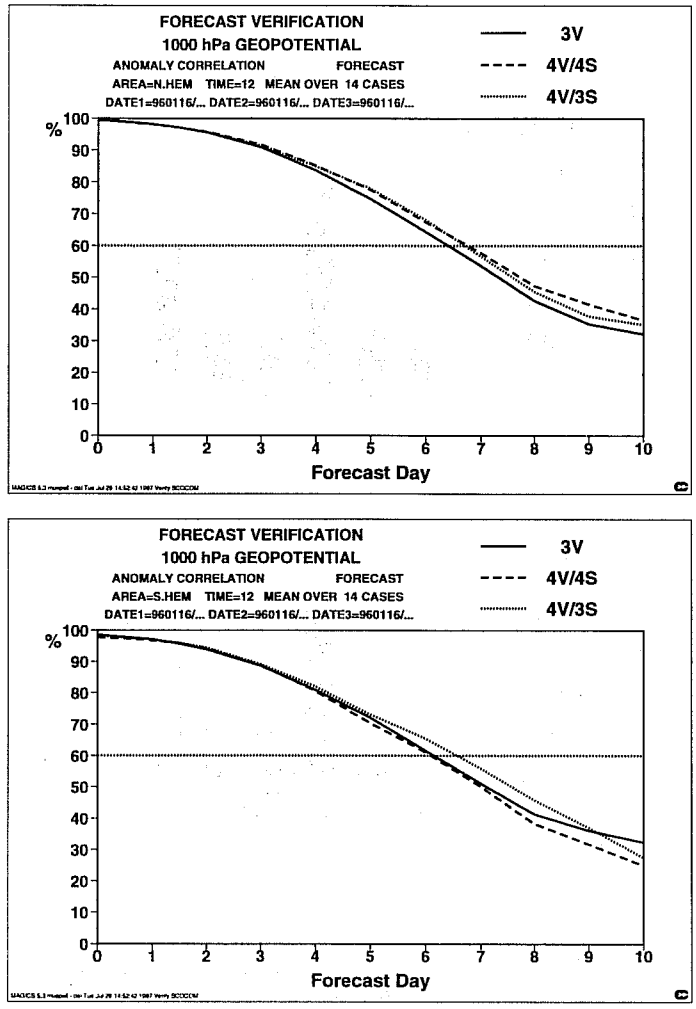

Figure 3 : The anomaly correlation in the Northern Hemisphere (top) and the Southern Hemisphere (bottom) at 1000hPa for two weeks in January 1996. 3V stands for 3D-Var, 4V/3S for 4D-Var with 3D-screening and 4V/4S for 4D-Var with 4D-screening

3. Tropical problem

3.1 Incremental formulation and initialization

In order to understand better the behaviour of the 4D-Var system, some analysis experiments were performed in which only some simulated isolated observations were inserted. In particular a few geopotential data at 850 hPa were simulated at 0Z for a particular date (5/12/96), with a departure from the background of 10m. The resulting mass and wind increments at 850hPa are illustrated for both 4D-Var and 3D-Var in Figure 4 (top left panel and bottom left panel respectively) with the use of the new background constraint (Bouttier et al, 1997). One can clearly notice the locations of the observations on the figure: two are located in the extratropical Northern Hemisphere, and two are located in the Southern Hemisphere, in the tropical band. The comparison of 4D- and 3D-Var increments shows a larger amplitude in the 4D-Var case, especially in the Tropics. The interpretation of this result requires us to go back to the way the incremental variational assimilation is implemented.

Each minimization problem can be written

$$J(\delta\mathbf{x}_l^n) = \frac{1}{2}(\delta\mathbf{x}_l^n + \mathbf{x}_l^{n-1} - \mathbf{x}_l^b)^T \mathbf{B}^{-1}(\delta\mathbf{x}_l^n + \mathbf{x}_l^{n-1} - \mathbf{x}_l^b) \tag{1}$$

$$+\frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i\delta\mathbf{x}_l^n(t_i) - \mathbf{d}_i^{n-1})^T \mathbf{R}^{-1}(\mathbf{H}_i\delta\mathbf{x}_l^n(t_i) - \mathbf{d}_i^{n-1})$$

with subscript l indicating that fields are at low resolution, subscript i the time index, superscript n the minimization index. $\mathbf{x}_l^b$ is the background field truncated at low resolution and $\mathbf{x}_l^{n-1}$ the current estimate of the atmospheric flow (it is equal to the background for the first minimization). $\delta\mathbf{x}_l^n$ is the increment at low resolution at initial time, and $\delta\mathbf{x}_l^n(t_i)$ the increment evolved according to the tangent linear model from the initial time to time index i. $\mathbf{R}_i$ and $\mathbf{B}$ are the covariance matrices of observation and background errors respectively. $\mathbf{H}_i$ is a suitable linear approximation at time index i of the observation operator H_i. The innovation vector is given at each time step by $\mathbf{d}_i^{n-1} = \mathbf{y}_i^o - H_i\mathbf{x}^{n-1}(t_i)$, where $\mathbf{y}_i^o$ is the observation vector at time index i. This innovation vector is computed integrating the model at high resolution from our current *n-1* estimate. The way the increment is then added to the current estimate can be written

$$\mathbf{x}^n = \mathbf{x}^{n-1} + NMI(\mathbf{x}^{n-1} + \delta\mathbf{x}^n) - NMI(\mathbf{x}^{n-1}) \tag{2}$$

where NMI stands for adiabatic non-linear normal mode initialization. The original purpose of this initialization was to ensure that the analysis was adjusted to the high-resolution orography of the forecast model. As adjustment was likely to be needed predominantly on smaller scales, the initialization was restricted to total wavenumbers 20 and above. During the pre-operational development of 3D-Var it became desirable to initialize all scales of motion in the final incremental initialization step, and this form of initialization was used for the first operational version of 3D-Var. It was also used for each outer loop (or, equivalently, each update of the trajectory) in 4D-Var. In the top left panel in Figure 4, it has then been used 4 times in 4D-Var. If one uses for 4D-Var a set-up similar to 3D-Var, i.e. one outer loop and only one initialization of the increments, results come much closer, as can be seen by comparing the top right and bottom left panels in Figure 4. In particular, for the observation located at 10 S, there is now only one height isoline in both cases (there were two previously for 4D-Var), and only slightly larger wind increments in 4D-Var. The impact of this initialization of the increments can be seen by comparing the top right panel (4D-Var with one update and NMI) with the bottom right panel (4D-Var with one update and no NMI). Initialization reduces the amplitude of the large-scale component of the increment. It also creates more rotational wind associated with mass observations in the Tropics. When several outer loops are performed, the effect of imposing initialization several times and minimizing several times is that

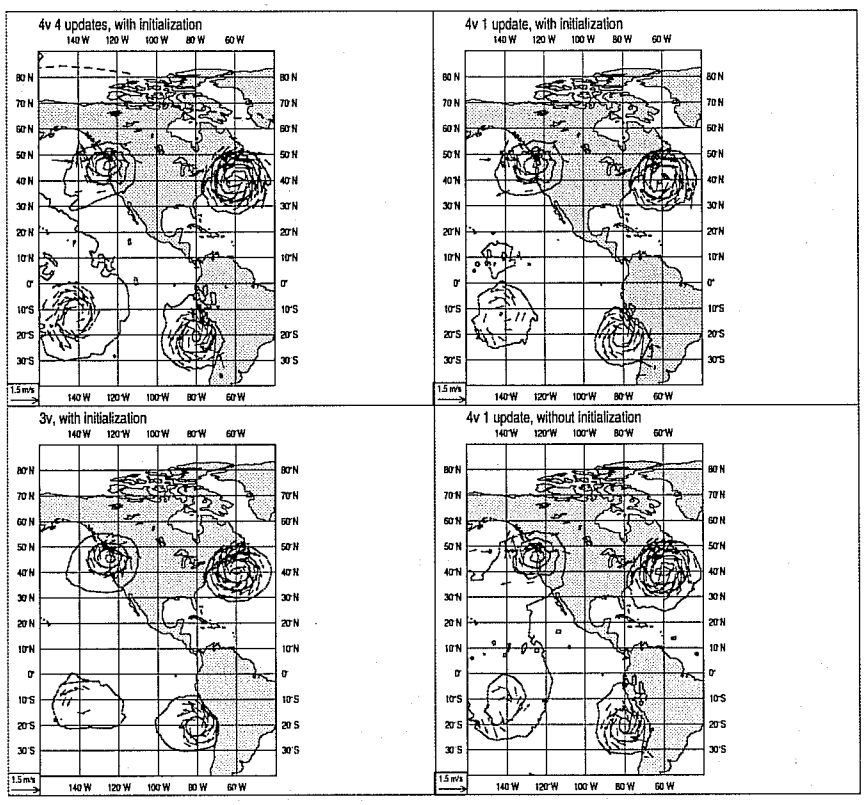

Figure 4 : Mass and wind increments from isolated mass observations at 850 hPa. The top left panel is 4D-Var with 4 updates and initialization, the bottom left panel is 3D-Var with initialization, the top right panel is 4D-Var with 1 update and initialization, and the bottom right panel is 4D-Var with 1 update and without initialization. Contour intervals are every 20 geopotential units.

the 4D-Var algorithm manages to minimize as much in terms of mass as when no initialization is performed, and to create rotational winds associated with these mass increments. The cost-function is better minimized by performing several outer loops in that way, but the drawback is that the enforced balance constraint creates unrealistic wind increments from mass observations. A test was run performing several outer loops without applying any initialization. The results were then very similar to when only one outer loop was performed without initialization. The main impact was thus confirmed to come from the NMI. The conclusion of these simulated observation experiments is that, with the initialization of the increments performed at each outer loop of the 4D-Var algorithm, the dynamical balance implied by adiabatic non-linear normal-mode initialization is enforced too strongly in the tropical area. To confirm these findings in a more realistic assimilation setting, it was decided to compare the behaviour of 4D-Var with 4 outer-loops and with one outer-loop on a two-week period (1 to 14 February 1997). The experiments are using the new background constraint. The tropical wind scores are shown in Figure 5. These scores clearly show that performing 4 outer loops is detrimental for the

4D-Var tropical wind scores. 4D-Var with only one outer loop is now competitive with the 3D-Var system in the Tropics. The hydrological budgets are presented for the tropical band (30N to 30S) for the three systems in Figure 6. 4D-Var has a smaller evaporation spin-up than 3D-Var, but a larger precipitation spin-down. Going to one outer loop slightly reduces the spin-down of precipitation, but it is still larger than that for 3D-Var. We will see in a later section that this can be remedied by using more physical processes in the 4D-Var minimization.

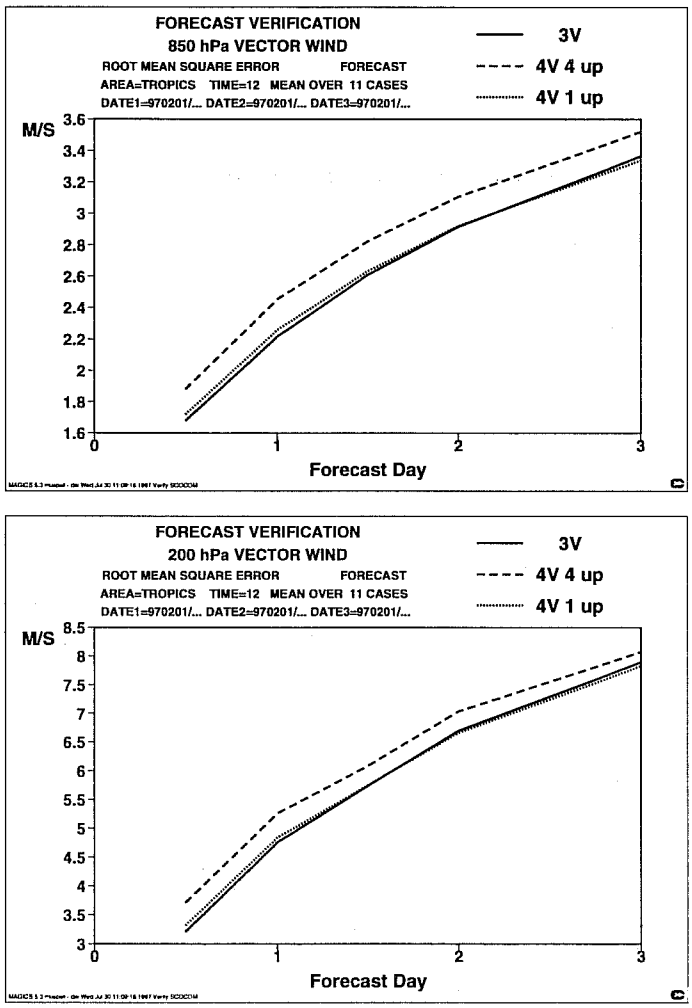

Figure 5 : Tropical wind scores verified against own analysis at 850 and 200 hPa. 3D-Var is shown as a solid line, 4D-Var with 4 outer loops as a dashed line and 4D-Var with one outer loop as a dotted line

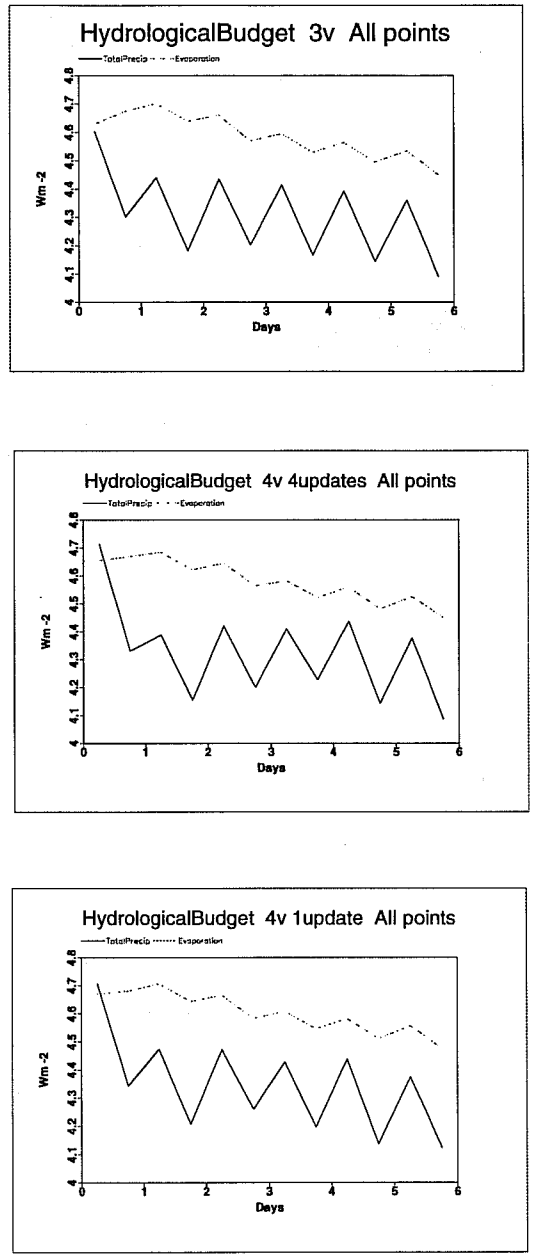

Figure 6 : Hydrological budgets for the tropical band (30N to 30S) averaged over 14 days for 3D-Var (top panel), 4D-Var with 4 outer loops (middle panel) and 4D-Var with one outer loop (bottom panel).

3.2 Other factors affecting the tropical scores

Apart from the number of outer loops performed in 4D-Var, there are two factors which were found to affect the tropical performance of both 3D- and 4D-Var. Firstly, the change to the new background formulation dramatically improved the tropical scores (Bouttier et al, 1997). Secondly, the initialization of medium and large scales was removed (Simmons and Rabier, 1997). The balance provided by the new background constraint appeared to be sufficiently good to allow the initialization of total wavenumbers below 20 to be abandoned. This will allow us to re-examine the impact of the number of outer loops in 4D-Var. But, for the time being, experimentation was resumed with 4D-Var with only one outer loop. This baseline configuration has the advantage that it is relatively cheap, and ensures a reasonable tropical behaviour.

4. Baseline experimentation with 4D-Var with one outer loop

Several two to three-week periods were run with 4D-Var with one outer loop and the new background term formulation. The list of periods and set-ups is given below.

25 June 1996 to 15 July 1996: experiments were performed using IFS cycle 16r1, i.e. including the new background formulation but still using the initialization of large and medium scales. Furthermore, the inflation of the errors specified for observations made far from the main synoptic times was included in this first set of experiments as well as for 3D-Var, but abandoned later on for the 4D-Var system as inappropriate for the 4D-Var problem.

1 to 21 February 1997: cycle 16r2, i.e. including both the new background term and the restriction of initialization to small scales only.

24 August to 6 September 1995: cycle 16r2 with additional changes to the processing of scatterometer data, as introduced in cycle 16r3.

15 to 28 January 1997: cycle 16r3, which includes changes in TOVS the bias-correction and use in the stratosphere of radiance data, and changes in quality control, particularly of AIREP data.

27 June 1997 to 10 July 1997: cycle 16r3.

In each experiment period, the 3D-Var control and the experimental 4D-Var system used the same model version, the same background error term and the same TOVS bias correction. Averaged scores verified against operations are presented in Figure 7. Impact in the medium range varies from neutral to significantly positive from one period to another, producing a slightly better overall performance from 4D-Var. In the Southern Hemisphere, scores are clearly positive up to day 6. In Europe, scores are positive in the medium-range. The extratropical scores are also better from 4D-Var than those from 3D-Var in the early part of the forecast range. The impact in the short-range is better seen when forecasts from a particular assimilation are compared with analyses from the same assimilation. Figures 8 to 10 present these scores up to day 3 for respectively geopotential, temperature and wind. These are generally better, except for tropical wind and stratospheric temperature scores, which are slightly worse in the first day or so. The main improvements can be seen in the Southern Hemisphere and at 1000 hPa in the Northern Hemisphere. The scatter for the day 3 geopotential scores is given in Figure 11. Scores are mostly in the upper left-hand side, which is positive for 4D-Var. Scores are systematically better in the Southern Hemisphere. The scatter in the medium-range confirms the impression that 4D-Var is generally performing better in the cases of relatively bad forecast performance, while having at least as much very good forecasts as 3D-Var (not shown). In the Northern Hemisphere, for the anomaly correlation of geopotential height at 500 hPa at day5, the number of very good forecasts scoring better than 85% is 17 for 3D-Var versus 20 for 4D-Var, while the number of very bad forecasts scoring worse than 65% is 6 for 3D-var and only 2 for 4D-

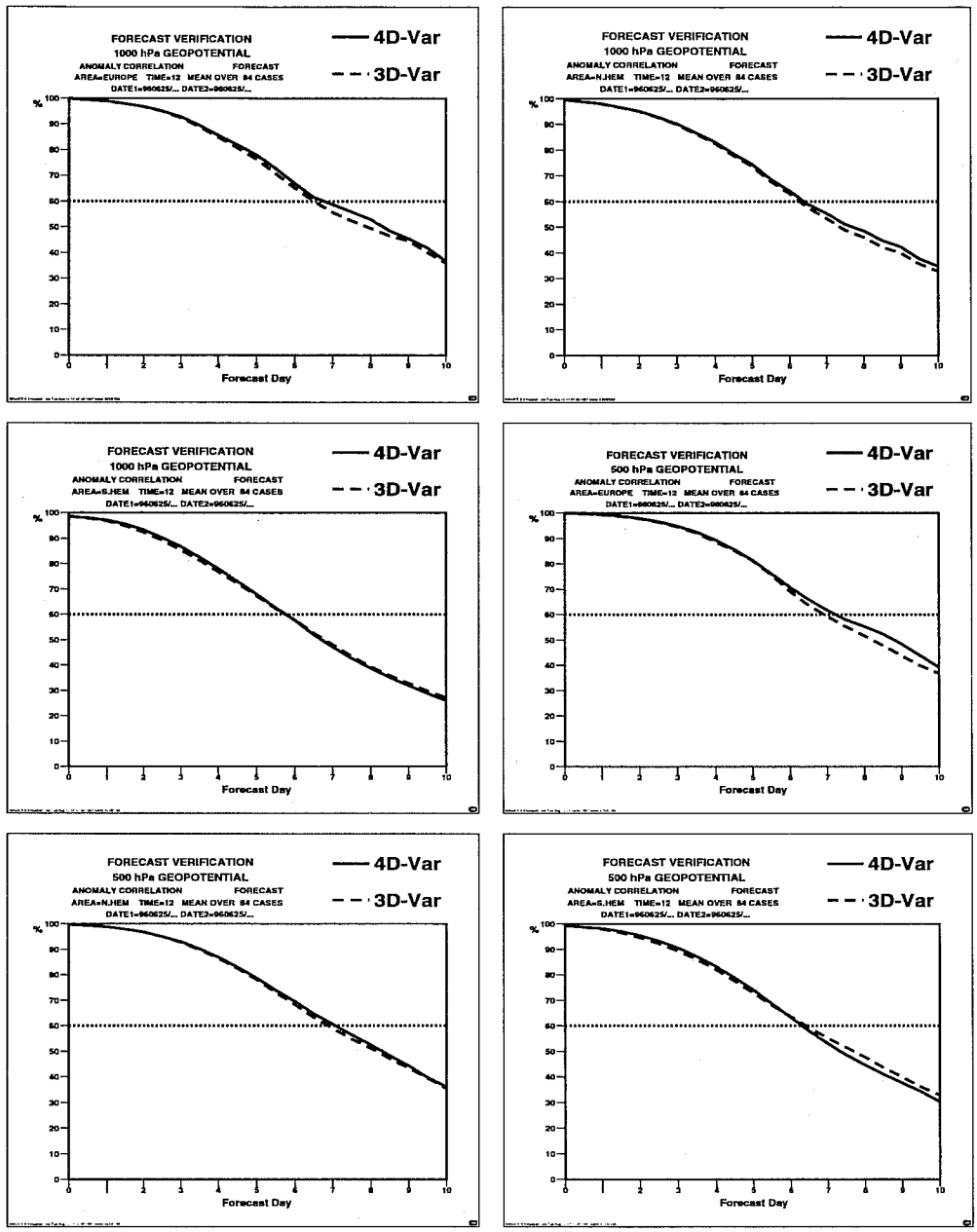

Figure 7 : Anomaly correlation scores for forecasts from 4D-Var (solid line) and from 3D-Var (dashed line), averaged over 12 weeks. Scores are shown for geopotential height at 1000 hPa and 500 hPa , for Europe, the Northern Hemisphere and the Southern Hemisphere.

Var. Similarly for the same score at day7, the number of very good forecasts scoring better than 70% is 9 for 3D-Var versus 12 for 4D-Var, while the number of bad forecasts scoring worse than 40% is 15 for 3D-var versus 12 for 4D-Var.

The very short range can further be investigated by computing the root-mean-square fit of the background field to the observations for both systems. Figure 12 shows the averaged fit to the radiosonde data over the first two weeks of February 1997. The fit of background to observations is very relevant to judge the quality of the assimilation system as the short-range forecast errors and the observation errors are uncorrelated. For these (solid lines) 4D-Var is clearly outperforming 3D-Var for both mass and wind (for wind mainly at the jet level in the Northern hemisphere) in all areas. A statistical test found a significantly better fit of the 4D-Var background fields to the observations for the tropospheric height in all areas, and for the wind at upper levels in the Northern hemisphere (the statistical test was also carried out for other periods, including summer periods, and the significance of a better fit for geopotential height data was confirmed in all areas.) Differences between 3D-Var and 4D-Var in the fit of the analyses to the data used in the analyses is not of themselves a performance criterion. The relevant conclusion which can be made from the analysis fits (dashed lines) is that both systems fit the data reasonably well.

As far as synoptic cases are concerned, two examples are shown in Figures 13 and 14. They correspond to interesting dates during the FASTEX experiment (Joly et al, 1997). The first synoptic case is IOP 12, with a deepening of 19 hPa in 24 hours from the manual UKMO analysis of surface pressure: the surface pressure fell from 966 hPa the 9/2/97 at 12Z to 947 hPa the 10/2/97 at 12Z. The deepening in the 24-hour forecast from the 4D-Var analysis is identical to the analysed one, pressure dropping from 970 hPa to 951 hPa in the same period of time. In contrast, cyclogenesis in the corresponding forecast from 3D-Var is not intense enough, with a deepening of only 10 hPa, from 969 hPa to 959 hPa. The analyses and 24-hour forecasts for both 3D-Var and 4D-Var are shown in Figure 13, together with the verifying analyses. The second synoptic case is IOP 17, with a manually analysed depression of 937 hPa on 20/2/97 at 0Z. In the medium-range, the forecast from 12/2/97 at 12Z is significantly better for 4D-Var, as shown in Figure 14 comparing the forecasts from 3D-Var and 4D-Var with their verifying analyses. Of course, this only corresponds to one forecast, and this dramatic improvement was not found systematically for all forecasts verifying on this date. To judge the overall better quality of the 4D-Var forecasts for this particular event, one can compute some average error over all the forecast ranges. The averaged rms error of the intensity of the low, averaged over all the forecasts from the 11th to the 19th verifying on 20/2/97 at 0Z is 15 hPa for 4D-Var versus 19 hPa for 3D-Var, which corresponds to a 20% improvement. One can also notice a better analysed surface pressure low with 4D-Var (940 hPa versus 943 hPa).

5. Influence of physical processes

5.1 Introduction

The tangent-linear and adjoint versions of the ECMWF model used currently to produce perturbations for the Ensemble Prediction System and used for the baseline 4D-Var experimentation reported above are almost adiabatic. Buizza (1994) introduced a simplified vertical diffusion scheme in order to avoid the growth of unrealistic singular vectors near the surface. The inclusion of more physical processes in linearised versions of the model is a required step to get a more realistic time evolution of analysis increments and a better estimation of the gradient of the cost-function. This should provide more consistency between inner and outer loops of the incremental approach. The impact of the physics should be more important in the tropics where the general circulation is strongly driven by the distribution of diabatic heating. A comprehensive set of linearised physics will also be necessary when satellite observations related to the hydrological cycle (cloudiness, precipitation, liquid water contents) are assimilated in the variational context. However, development of tangent linear and adjoint physics is non trivial because physical processes are strongly non-linear and generally include a lot of conditionals.

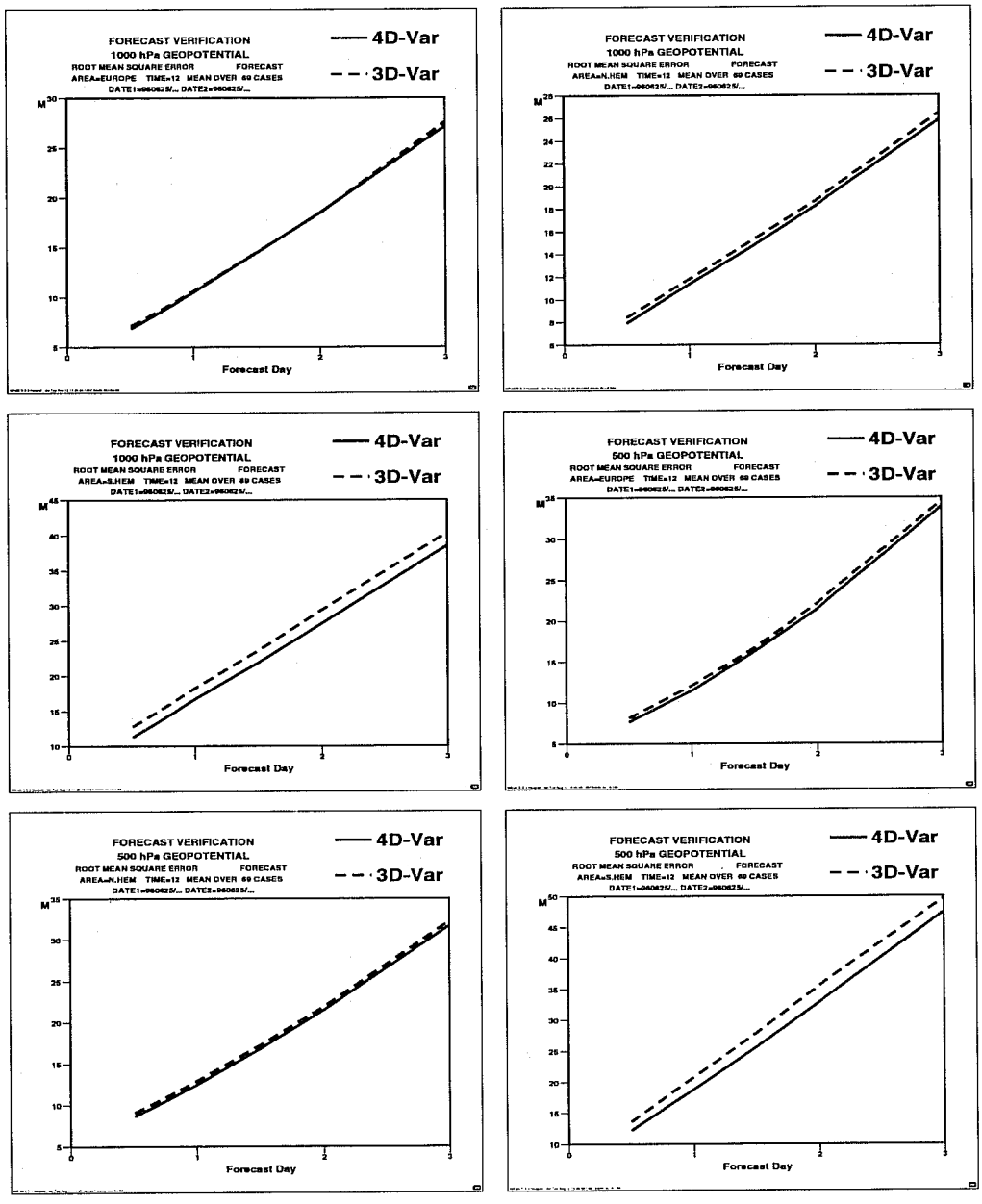

Figure 8 : Root-mean-square error scores for forecasts from 4D-Var (solid line) and from 3D-Var (dashed line), verified against own analysis, averaged over 12 weeks. Scores are shown for geopotential height at 1000 hPa and 500 hPa, for Europe, the Northern Hemisphere and the Southern Hemisphere.

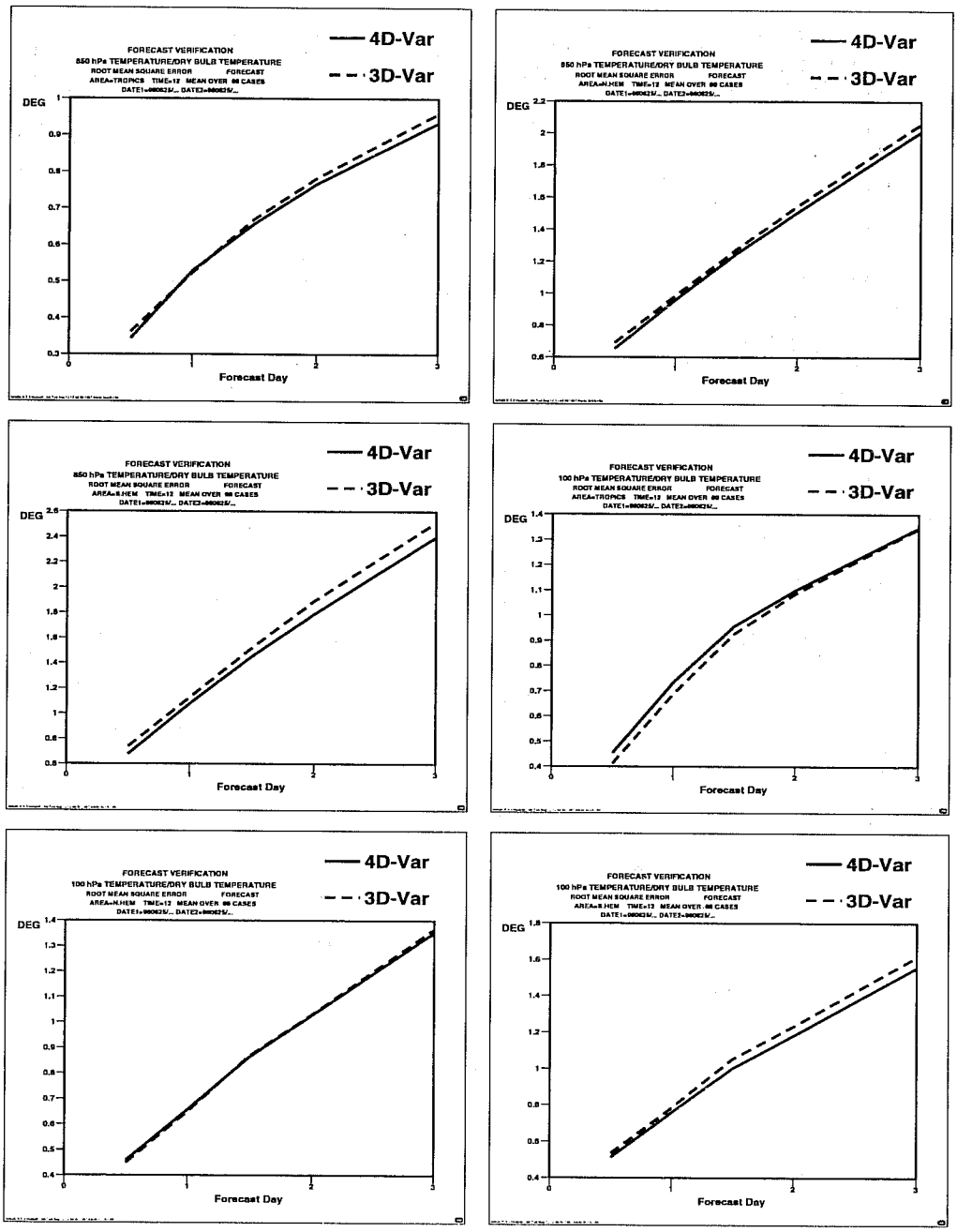

Figure 9 : Root-mean-square error scores for forecasts from 4D-Var (solid line) and from 3D-Var (dashed line), averaged over 12 weeks, verified against own analysis. Scores are shown for temperature at 850 hPa and 100 hPa, for the Tropics, the Northern Hemisphere and the Southern Hemisphere.

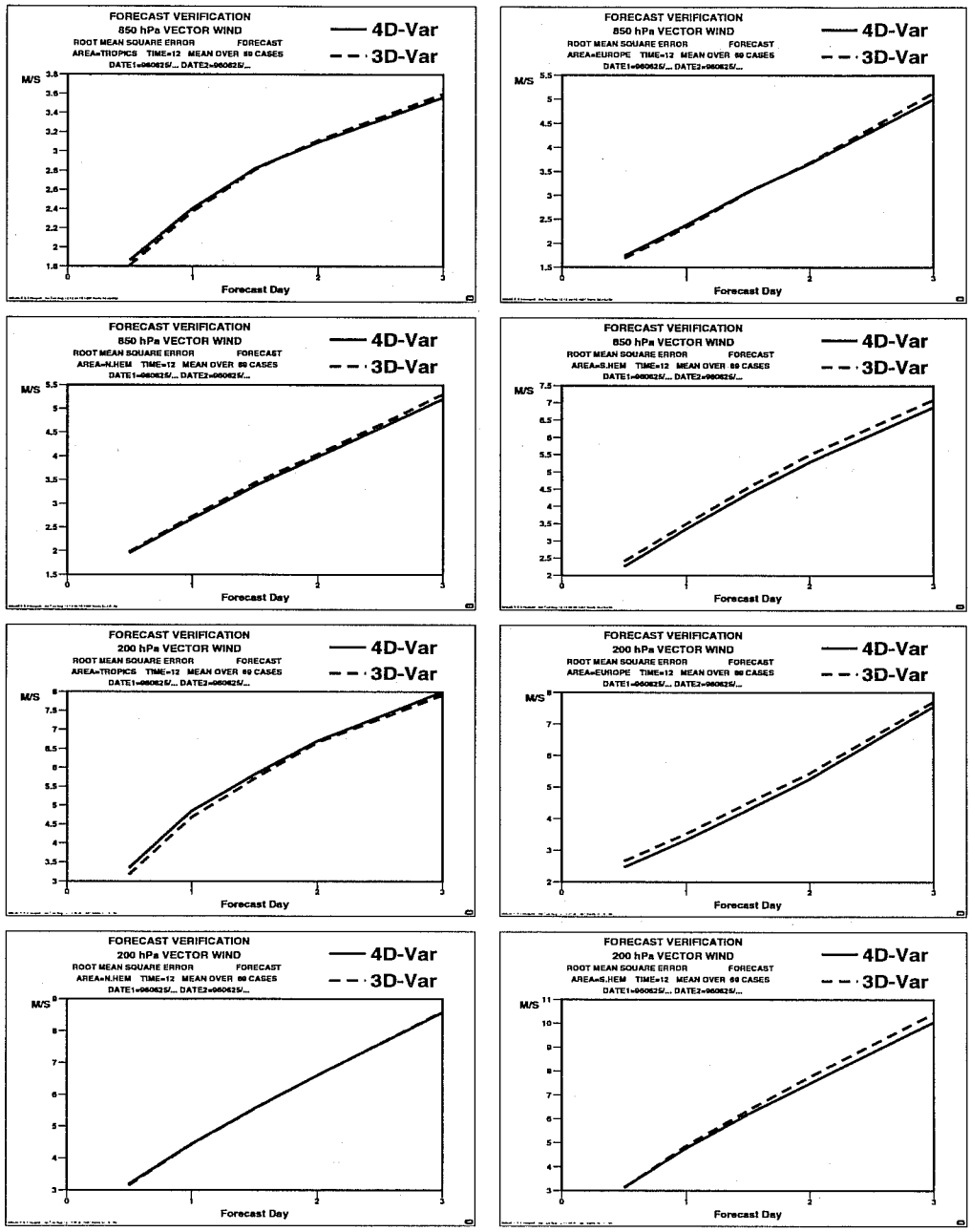

Figure 10 : Root-mean-square error scores for forecasts from 4D-Var (solid line) and from 3D-Var (dashed line), averaged over 12 weeks, verified against own analysis. Scores are shown for vector wind at 850 and 200hPa, for the Tropics, Europe, the Northern Hemisphere and the Southern Hemisphere.

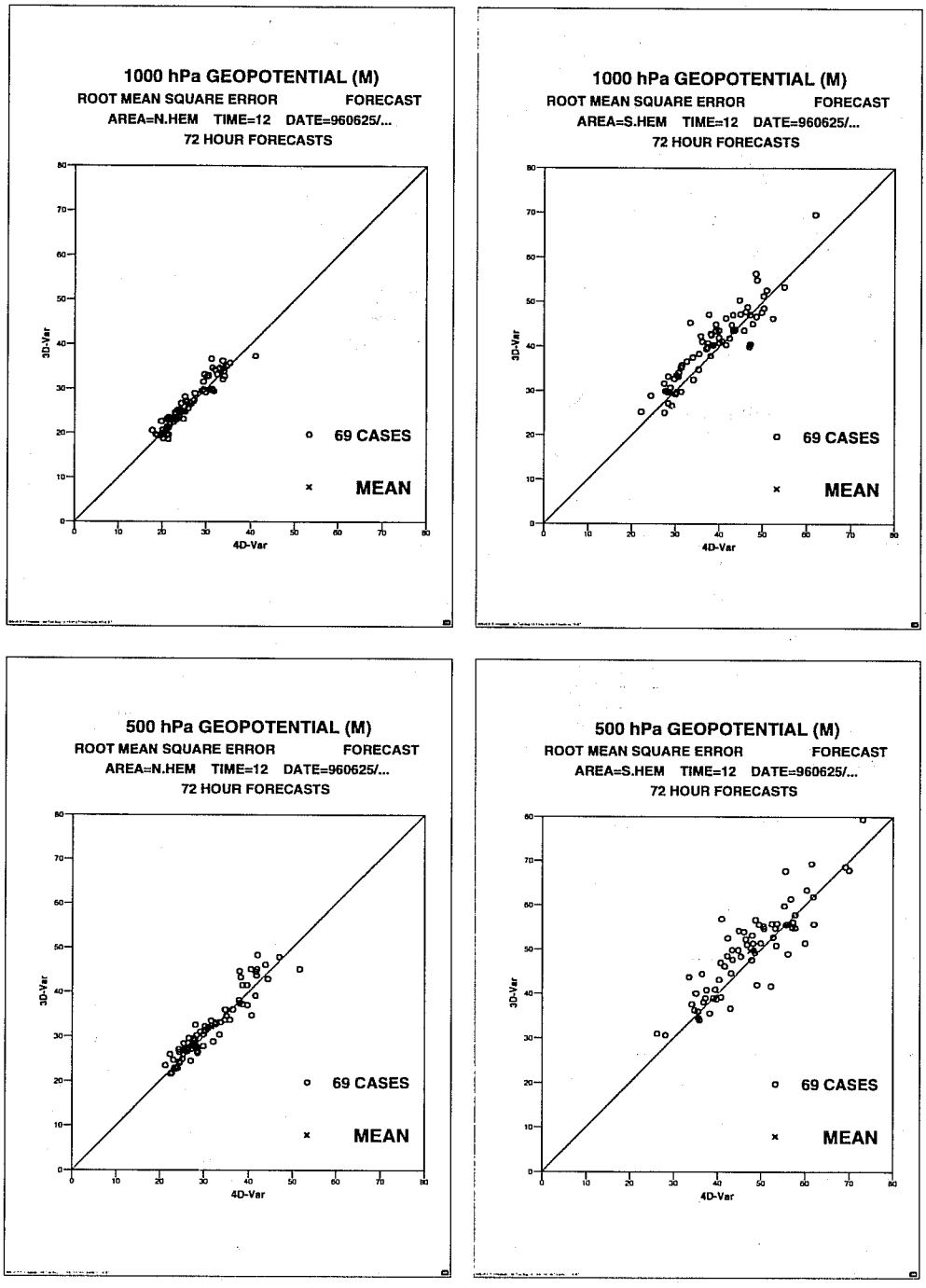

Figure 11 : Scatter diagram of root-mean-square error scores for forecasts from 4D-Var (solid line) and from 3D-Var (dashed line), averaged over 12 weeks, verified against own analysis. Scores are shown for geopotential at 1000 and 500hPa, for the Northern Hemisphere and the Southern Hemisphere .

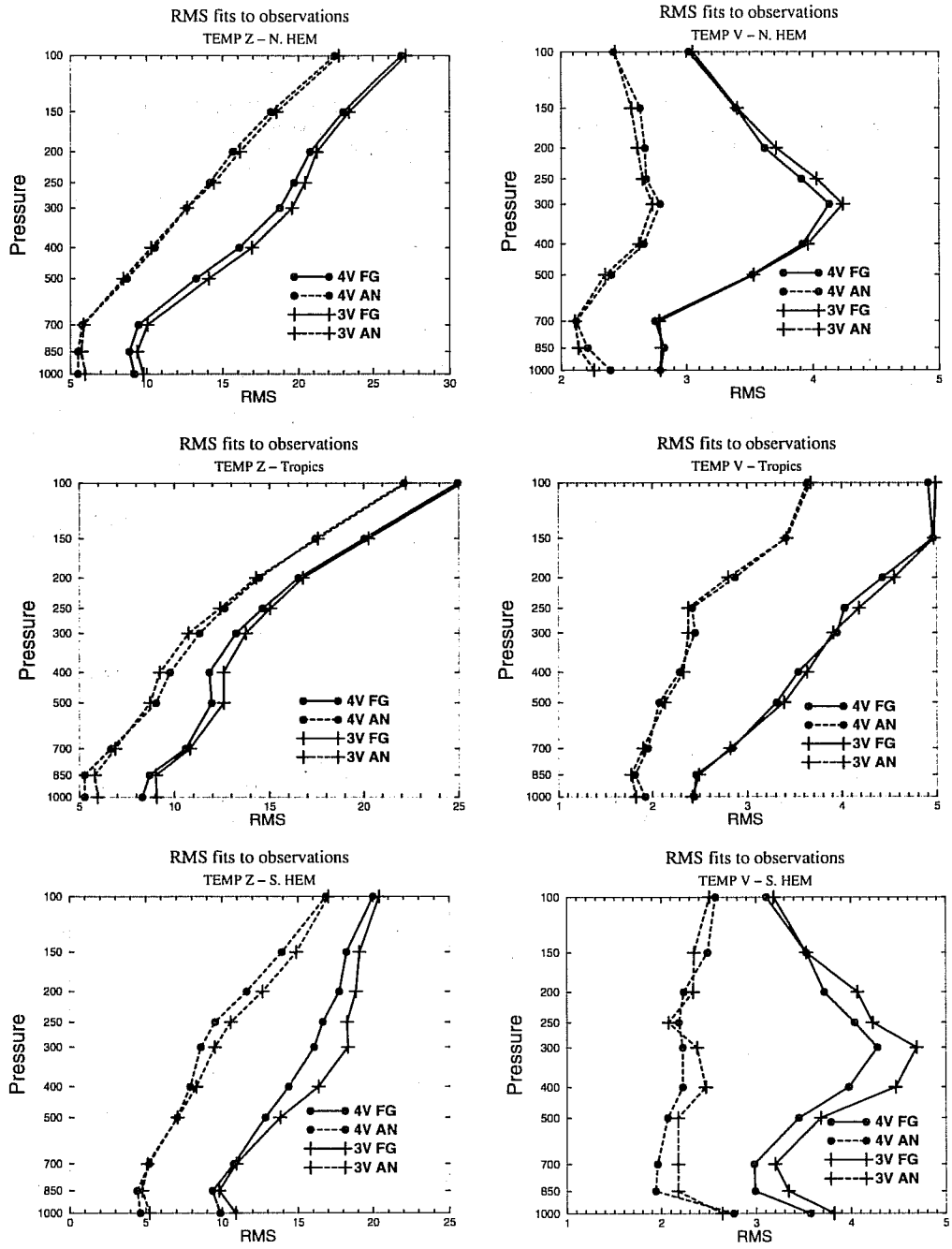

Figure 12 : Root-mean-square (RMS) fits to the radiosonde height and meridional component of the wind data produced over the Northern Hemisphere, the Tropics and the Southern Hemisphere, averaged over 1 to 14 February 1997. The solid lines represent the RMS fits of the backgrounds to the observations, the dashed ones the RMS fits of the analyses to the observations. 4D-Var is shown as circles, 3D-Var as plusses. The abscissa is the RMS in geopotential units and m/s. The ordinate is the pressure in hPa.

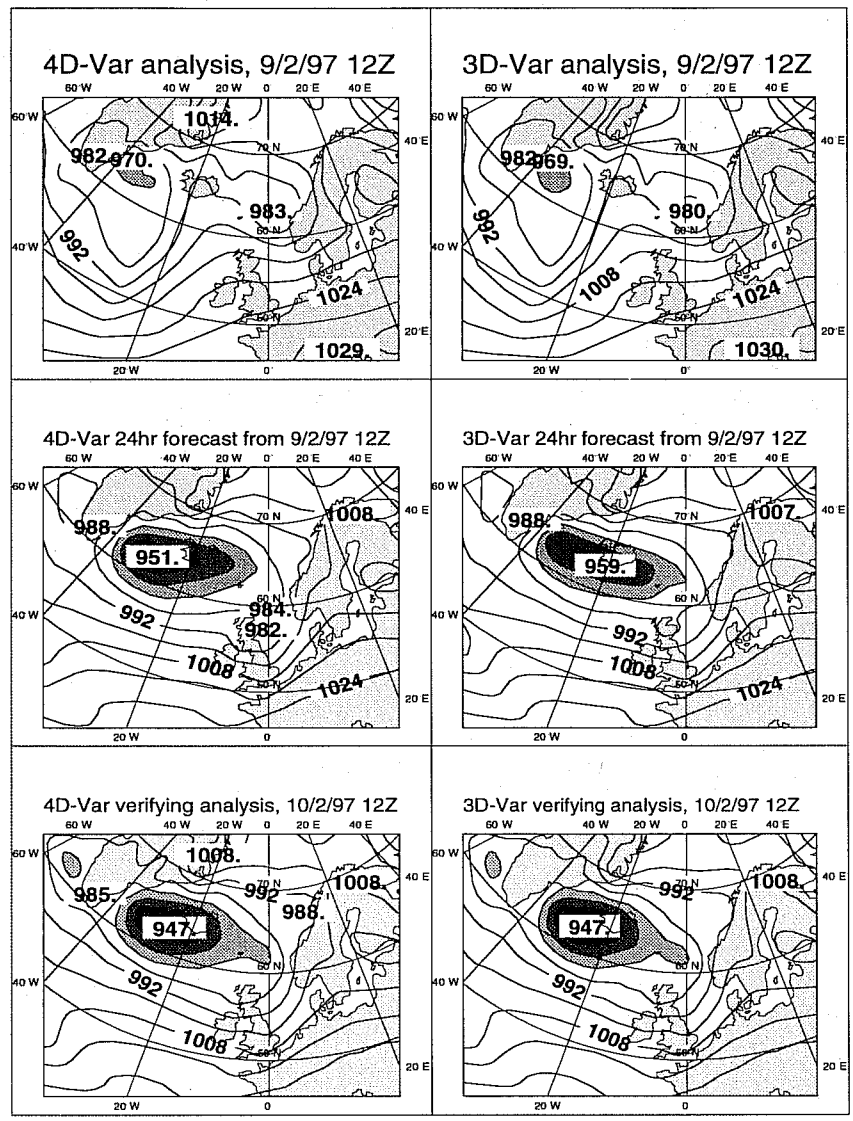

Figure 13 : Maps of surface pressure for the synoptic case corresponding to IOP 12 of FASTEX. The top panels show the analyses for the 9/2/97 at 12Z, the middle panels the 24-hour forecasts from 9/2/97 at 12Z, and the bottom panels the verifying analyses. 4D-Var charts are on the left, 3D-Var on the right.

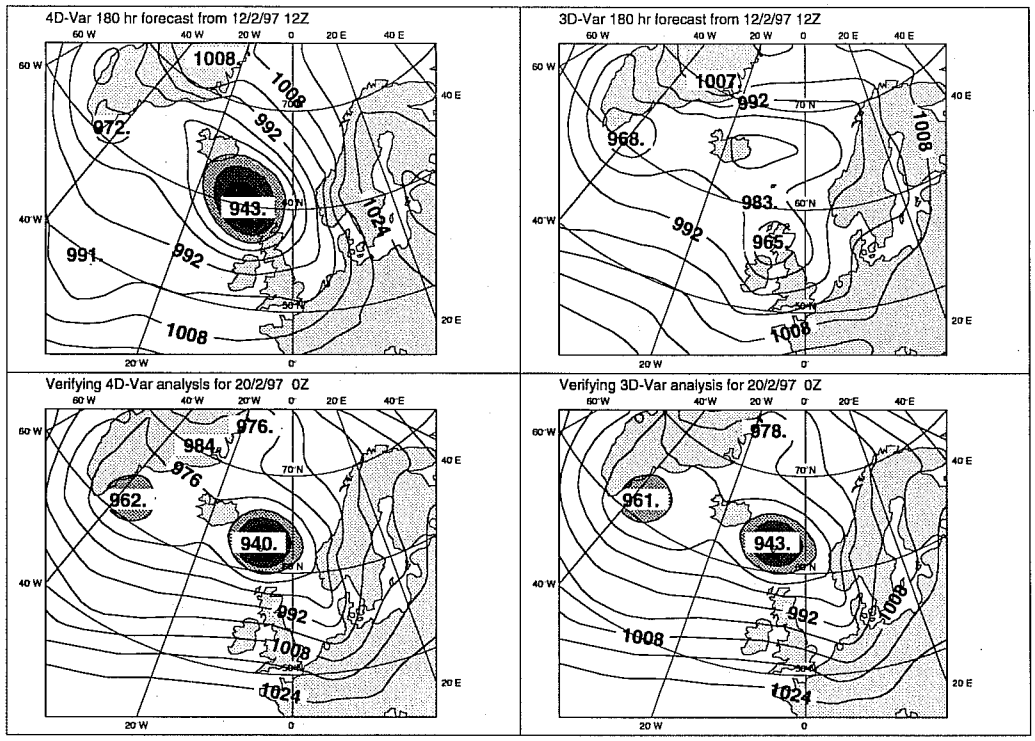

Figure 14 : Maps of surface pressure for the synoptic case corresponding to IOP 17 of FASTEX. The top panels show the forecasts from 12/2/97 at 12Z verifying 20/2/97 at 0Z, and the bottom panels the verifying analyses. 4D-Var charts are on the left, 3D-Var on the right.

5.2 Description of the linear physics

A first package of linear physics has been developed to be used in 4D-Var assimilation and singular vector computations (Mahfouf et al, 1996). For both scientific and technical reasons, simplifications have been devised with respect to the operational non-linear physics. The incremental approach allows a progressive improvement of physical processes in the inner loop of the minimisation.

A set of five physical processes has been linearised:

a) Vertical diffusion : The previous operational ECMWF scheme where eddy diffusivity coefficients are expressed as analytical functions of the local Richardson number, according to Louis et al. (1981), has been chosen. The tendency of a given perturbed conservative variable χ is written:

$$\frac{\partial \chi}{\partial t} = -g\frac{\partial}{\partial p}\left(\overline{K}\cdot\frac{\partial \chi}{\partial z}\right) \tag{3}$$

where $\overline{K}$ is the eddy diffusivity from the trajectory.

b) Sub-grid scale orographic effects : The low level blocking part of the operational ECMWF scheme has been linearised (Lott and Miller, 1997).

c) Large scale precipitation : A simple moist adjustment scheme removes supersaturation. Evaporation of precipitation in sub-saturated layers is not included.

d) Longwave radiation : A constant emissivity approach has been linearised. Given the emissivity $\bar{\varepsilon}$ from the trajectory computed from the full non-linear radiation scheme, the evolution of the perturbed temperature T is given by:

$$\frac{\partial T}{\partial t} = -\frac{g}{C_p}\frac{\partial}{\partial p}(4\sigma\bar{\varepsilon}\bar{T}^3 T) \tag{4}$$

e) Deep moist convection : A simplified linearisation of the ECMWF convection scheme has been performed, which accounts for the vertical transport of the perturbed variables by the mass-flux from the trajectory. The tendency of of a given perturbed variable χ is written:

$$\frac{\partial \chi}{\partial t} = \frac{1}{\rho}\left[(\bar{M}_u + \bar{M}_d)\frac{\partial \chi}{\partial z}\right] \tag{5}$$

where $\bar{M}_u$ and $\bar{M}_d$ are the mass-fluxes for the updrafts and the downdrafts (from the trajectory) respectively.

Future developments will include the improvement of the representation of moist convection and of cloud-radiation interactions.

5.3 Evaluation of the linear physics

The linear physics has been evaluated by comparing the time evolution of initial perturbations produced by the tangent-linear model with the difference of two non-linear integrations. In order to quantify the usefulness of the linear physics, non-linear integrations were performed with the comprehensive set of physical parametrizations and where the corresponding tangent-linear integrations included either the simplified vertical diffusion used in the EPS (Buizza et al, 1997) or the improved set described above. We then calculated differences between the full non-linear integration and each of the two linearised integrations. The validity of the tangent-linear approximation is examined for finite size perturbations (analysis increments) which are the ones of practical interest. Results obtained using a T42L31 version of the IFS integrated during 24 hours are summarised in Table 1 in terms of mean absolute error. The inclusion of improved linear physics improves the behaviour of the tangent-linear model with respect to the full non-linear model. The improvement reaches 20 % for the specific humidity field, and is about 10 % for the other variables.

TL version	U (m/s)	V (m/s)	q (g/kg)	T (K)
adiabatic	0.54	0.52	0.15	0.28
linear physics	0.49	0.47	0.12	0.26

Table 1: Mean absolute difference (in global mean) of two versions of the tangent-linear model for wind, specific humidity and temperature for a 24-hour forecast at T42L31 (07/02/97). "adiabatic" refers to the current simplified vertical diffusion scheme and "linear physics" to the improved package described above.

Singular vectors have been computed with the improved linear physics. Latent heat release is the dominant mechanism by which singular vectors are modified with respect to the adiabatic computations. Larger growth rates have been found

together with a displacement of the maximum of energy towards smaller scales. Singular vectors have also been useful to identify and remove spurious unstable modes present in preliminary versions of the linearised physics.

5.4 Impact on 4D-Var experiments

A first one-week 4D-Var assimilation was undertaken with the linear physics for the "1-update" configuration (1 to 7 February 97). Results were neutral in terms on quality of the analyses and of the forecasts but the cost of 4D-Var was increased by a factor of 2.5. However, we have learned from these preliminary experiments that, thanks to the incremental approach, the convergence of the minimisation is not slowed by the introduction of the physics, and that the extra-memory storage for the trajectory at t-dt is not an issue in the 4D-Var context. Another strategy was then proposed, in order to reduce the extra-cost of the physics and to better account for the non-linearities induced by the physics. A "2-update" configuration was developed, where the first minimisation includes 50 iterations performed without physics, and the second minimisation performed after updating the trajectory at high resolution comprises 20 iterations performed with physics. With such configuration, the increase in CPU time introduced by the linear physics is only 25 % (to which another 25% is added due to the introduction of a second update).

5.5 Results from a 14-day winter assimilation

Three configurations of 4D-Var were tested during a 2-week winter period (1 to 14 February 97) using the operational 3D-Var as a control:

- 4D-Var one-update adiabatic
- 4D-Var two-updates adiabatic
- 4D-Var two-updates linear physics

We compare the performance of three versions of 4D-Var in order to evaluate the impact of both the "2-update" configuration and of the physics. The impact of the physics on extra-tropical scores is neutral (Figure 15). Results are presented at 1000 hPa where the influence of an improved boundary layer should be the largest. Going from one to two updates is slightly detrimental for the Southern Hemisphere around day 3, but 4D-Var assimilation during other periods show that this feature is not systematic. The inclusion of the linear physics has the largest impact on the analysis of humidity. The zonal mean increments of specific humidity averaged over the two-week period show a significant reduction of the increments in the tropical belt, both in the boundary layer and above (Figure 16) when comparing 4D-Var with physics and standard 4D-Var. Mean increments are systematically larger in the standard 4D-Var system than in 3D-Var. The direct consequence of a drier initial atmospheric state when introducing the physics is a significant reduction of the spin-down for precipitation during the first day of the forecasts as shown on Figure 17. The short range tropical wind scores are improved with the inclusion of the physics, the improvement being larger at 850 hPa than at 200 hPa (Figure 18)

5.6 Results from 8 weeks of data assimilation

The comparison between the baseline 4D-Var and the 4D-Var with two outer-loops (with physics in the second outer-loop) has been performed on a total of 8 weeks. These include the periods 1 to 14 February 1997 , 24 August to 6 September 1995, 15 to 28 January 1997 and 27 June to 10 July 1997. The combined extratropical scores are shown in Figure 19. The scores over the Northern and Southern Hemispheres are marginally positive, with a larger positive signal over Europe. The scatter between 4D-Var and 4D-Var with physics is very small in the short-range and the fit of the background fields to the data very similar (not shown). The overall clearest improvement is seen in the Tropics for the low-level winds, similar to the one in Figure 18 for one two-week period (not shown).

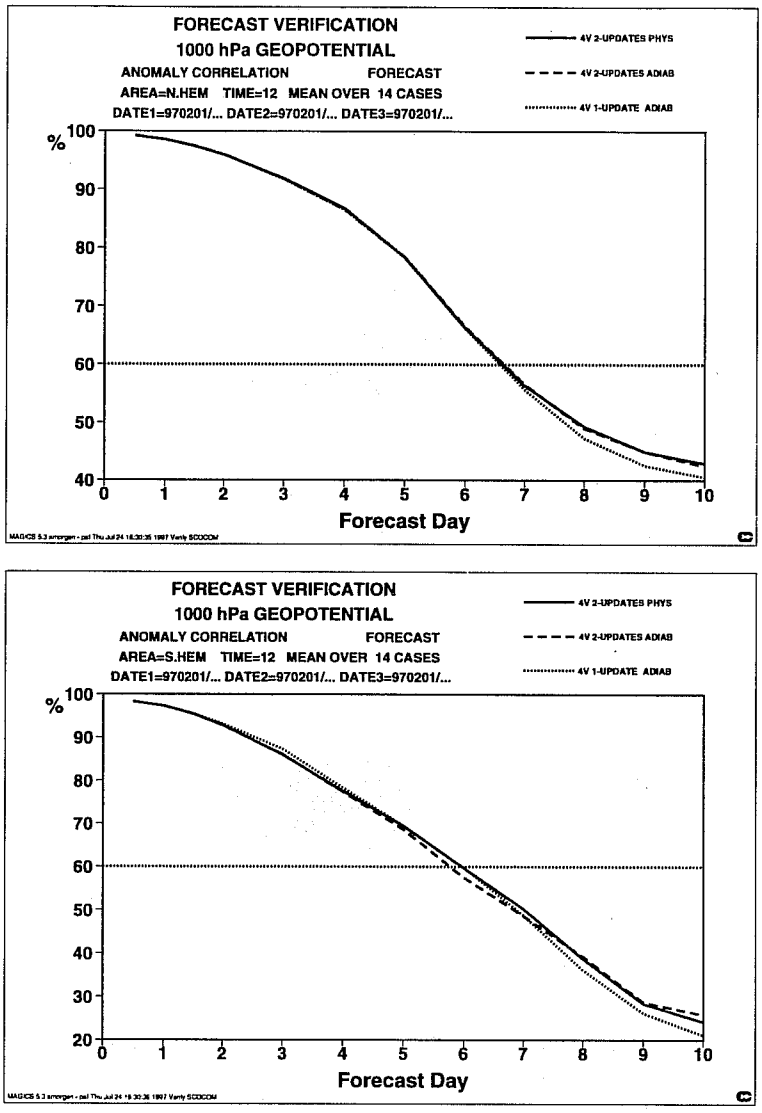

Figure 15 : Anomaly correlation for the geopotential at 1000 hPa averaged over 14-forecasts issued from 4D-Var assimaltion with physics (solid), 2-updates adiabatic (dashed) and 1-update adiabatic (dotted). Top : Northern Hemisphere, bottom : Southern Hemisphere.

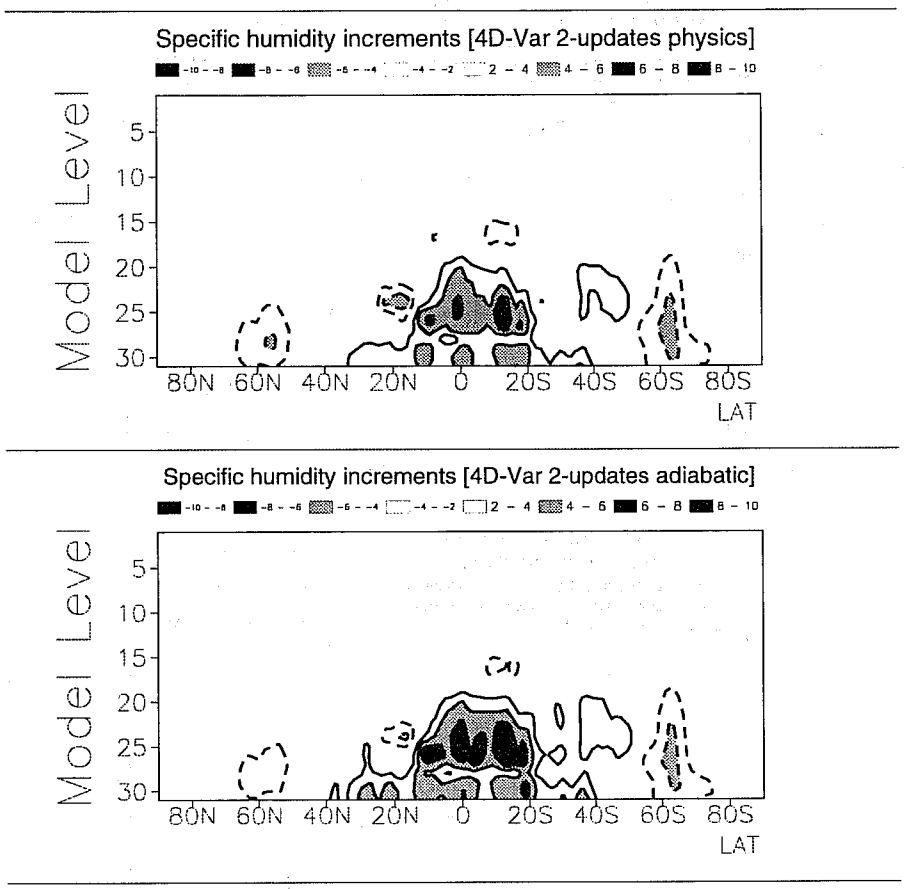

Figure 16 : Mean zonal increments for specific humidity (unit is 0.01 g/kg) averaged over a two-week period (01/02/97-14/02/97). Top : 4D-Var 2updates with physics, Bottom : 4D-Var 2updates adiabatic

6. Structure functions and simplified Kalman filter

6.1 Structure functions in 4D-Var

Structure functions can easily be illustrated by using some single observation experiments, as in Thépaut et al., 1996. For a particular date (5/12/96, 0Z), a baroclinic area was chosen in the West Pacific. The background for this date, which corresponds to a 6 hour forecast from the last operational analysis exhibits fields which are tilted in the vertical. This is illustrated in Figure 20 in a cross-section of the zonal wind component. One observation of geopotential height was inserted at location 60W, 40N, at 850 hPa in 4D-Var, either at time 21Z (3 hours before the main synoptic time 0Z), or at 0Z, or at 03Z (3 hours after the main synoptic time 0Z). Each time, the initial departure from the background is equal to 10m. The structure functions can then be illustrated for three different scenarios: an observation at the beginning of the

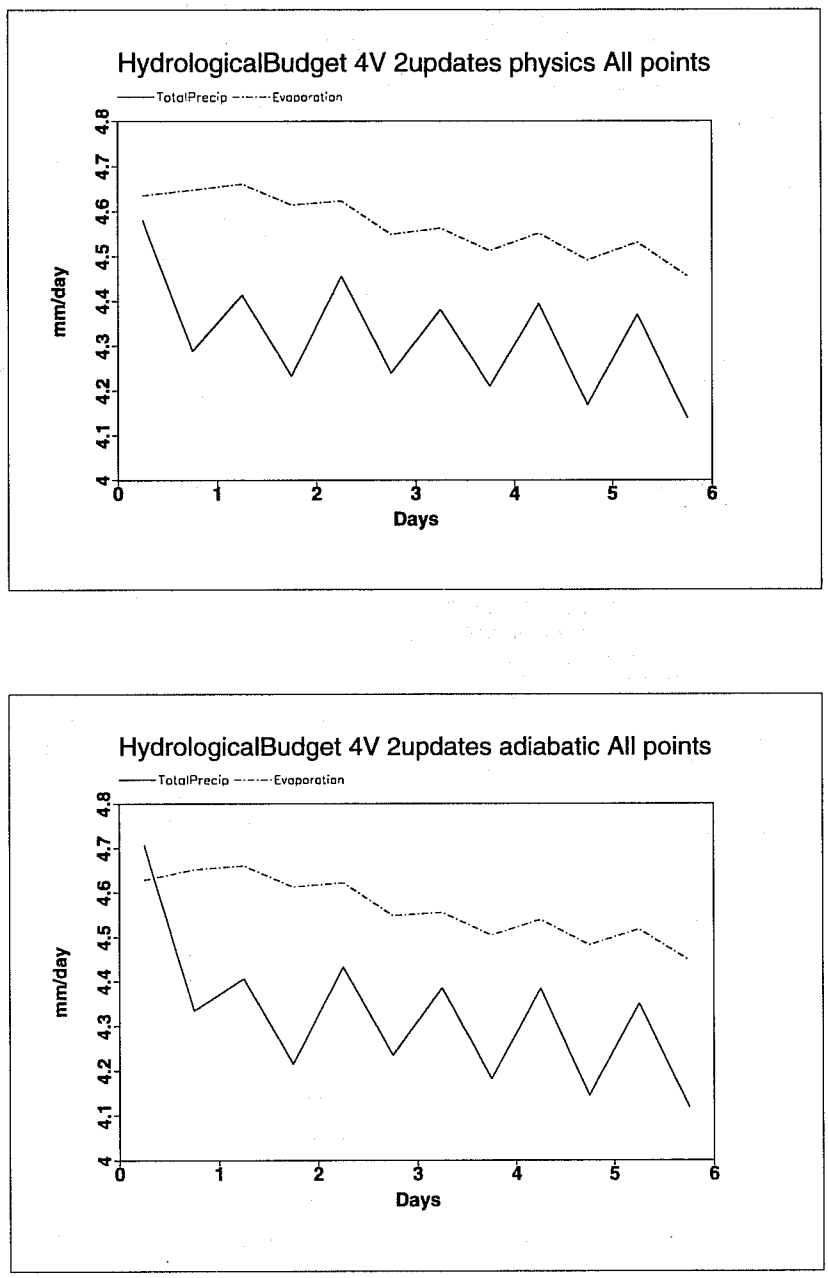

Figure 17 : Time evolution of global precipitation and evaporation in the tropical belt [-30,30] averaged over 14 forecasts (T213L31) issued from 4D-Var assimilation 2-updates with (top) and without (bottom) linear physics.

assimilation window (21Z), in the middle (0Z) and at the end (03Z). These are shown, at the time of the observation, in Figure 21.This particular meteorological situation is not rapidly developing which allows comparison of the increments

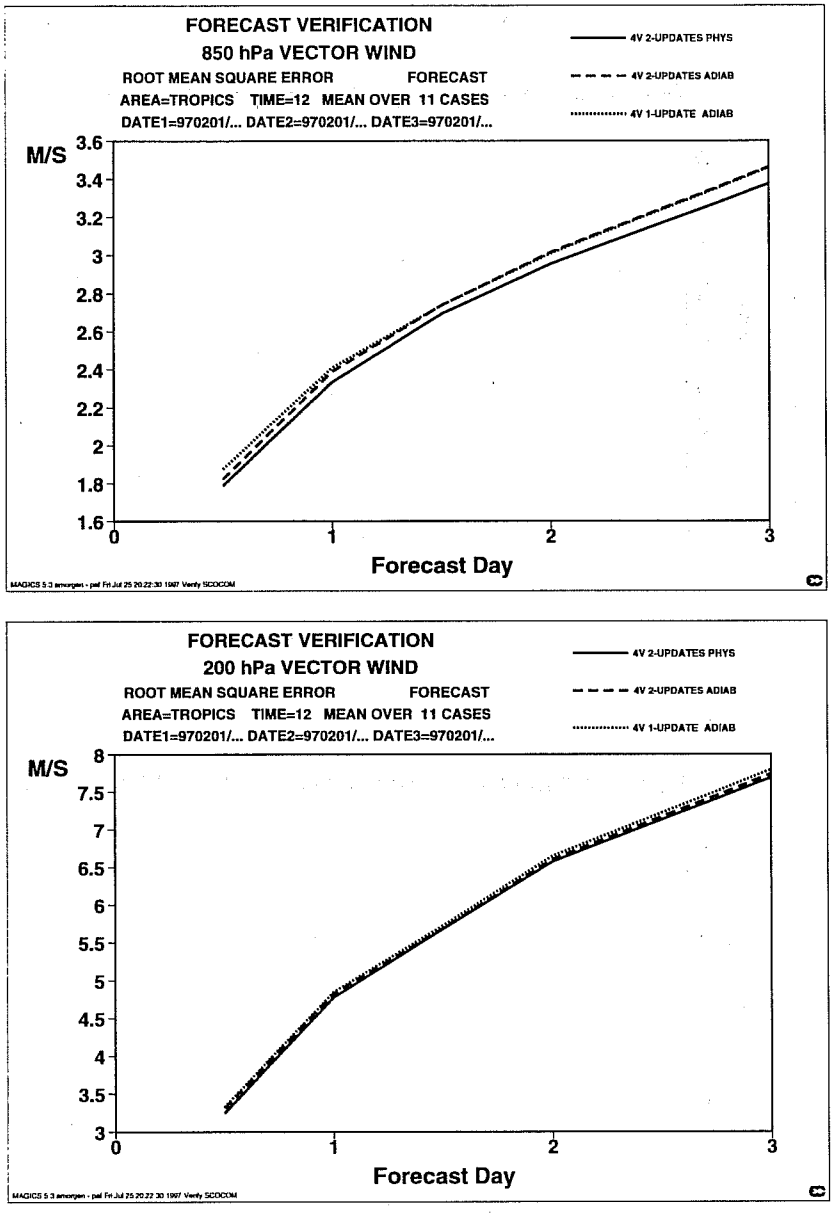

Figure 18 : Root mean square error of the short range forecast of the tropical wind verified against its own analysis from 4D-Var assimilation 1-update adiabatic (dotted), 2-updates adiabatic (dashed) and 2-updates with physics (solid).

even if their validity time can be up to 6 hours different. The top panel in Figure 21 corresponds to the increments at 21Z created from an observation at 21Z. These are similar to the increments which would have been created by the 3D-Var system. They are barotropic, the value decreasing with height and horizontal distance from the observation location.

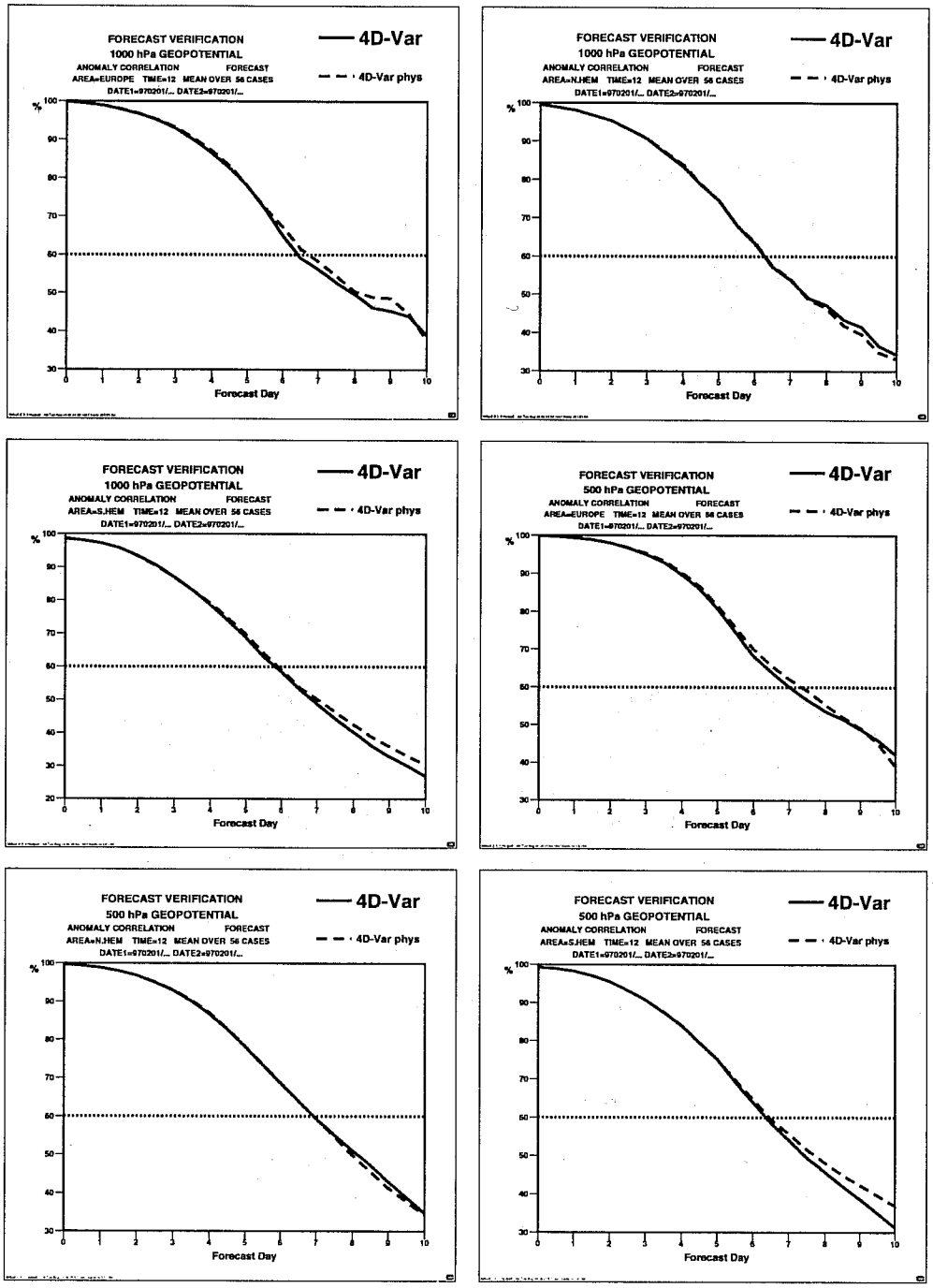

Figure 19 : Anomaly correlation scores for forecasts from baseline 4D-Var (solid line) and from 4D-Var with physics (dashed lines). Scores are shown for geopotential height at 1000 hPa and 500 hPa, for Europe, the Northern Hemisphere and the Southern Hemisphere.

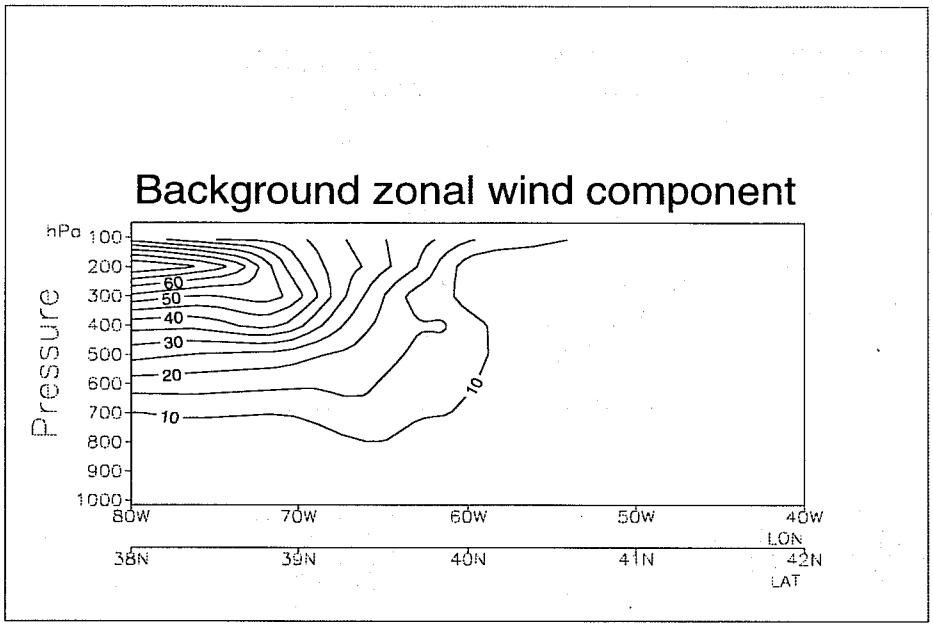

Figure 20 : Cross-section for the zonal wind component of the background for 5/12/96, 0Z. Unit is m/s.

When the observation is located 3 or 6 hours after the initial time of the assimilation window, some influence of the dynamics is noticeable. The increments in the middle and bottom panels of Figure 21 are tilted in the vertical, in a consistent way with the meteorological situation. The longer the time gap between the initial time and the observation time, the tighter the structure functions get at the surface and the more the impact of the observation spreads vertically (following the baroclinic tilt). One can also illustrate the increments produced by the 4D-Var system at the main synoptic time 0Z from these three configurations. These are shown in Figure 22. Although Figure 21 illustrates how different the structure functions can be, the actual increments at a given time (here, the main synoptic time, corresponding to the middle of the assimilation window) are much more similar. This can be regarded as reassuring, as it implies that there is no real conflict in the 4D-Var system between observations inserted at different times within the window.

The following section discusses how even more dynamics can enter the 4D-Var structure functions, through the use of a simplified Kalman filter.

6.2 Brief Description of the Simplified Kalman Filter

A simplified Kalman Filter (SKF) has been developed. The SKF accounts for the dynamical evolution of the forecast error covariance matrix for a low dimension subspace spanned by a set of singular vectors. By concentrating on the covariance evolution in this subspace, the SKF attempts to maximise the accuracy of the analysis for those components of analysis error which will grow rapidly during the early stages of the subsequent forecast.

The SKF has two distinct parts. The first part consists of a modification to the background cost function of 3D/4D-Var, which accurately represents the cost due to the component of the background error which lies in the subspace. The second part propagates the covariance matrix of forecast error using the tangent linear dynamics.

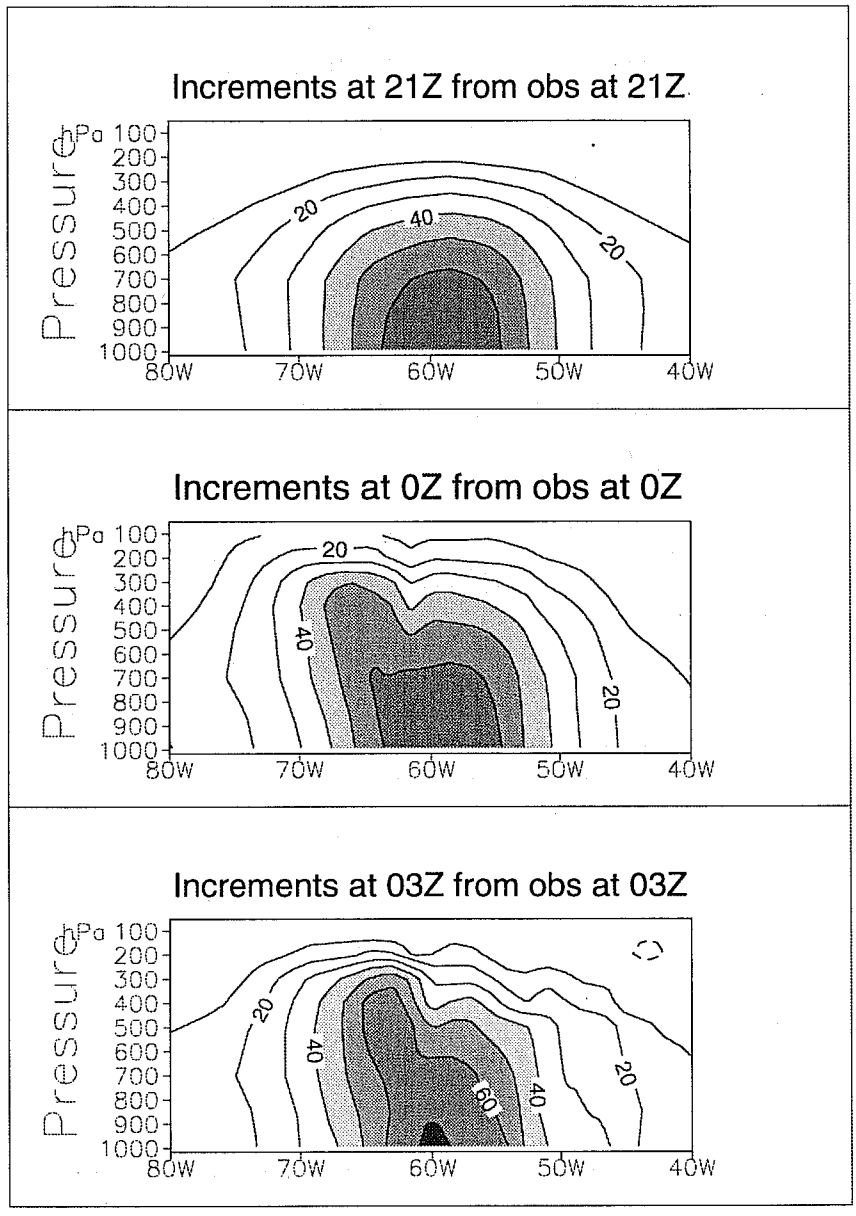

Figure 21 : Structure functions for a height observation at 850 hPa, 40N, 60W. Isolines show the resulting increment, in geopotential unit. The top panel corresponds to an observation at 21Z, the middle panel to an observation at 0Z, and the bottom panel to an obsveration at 03Z.

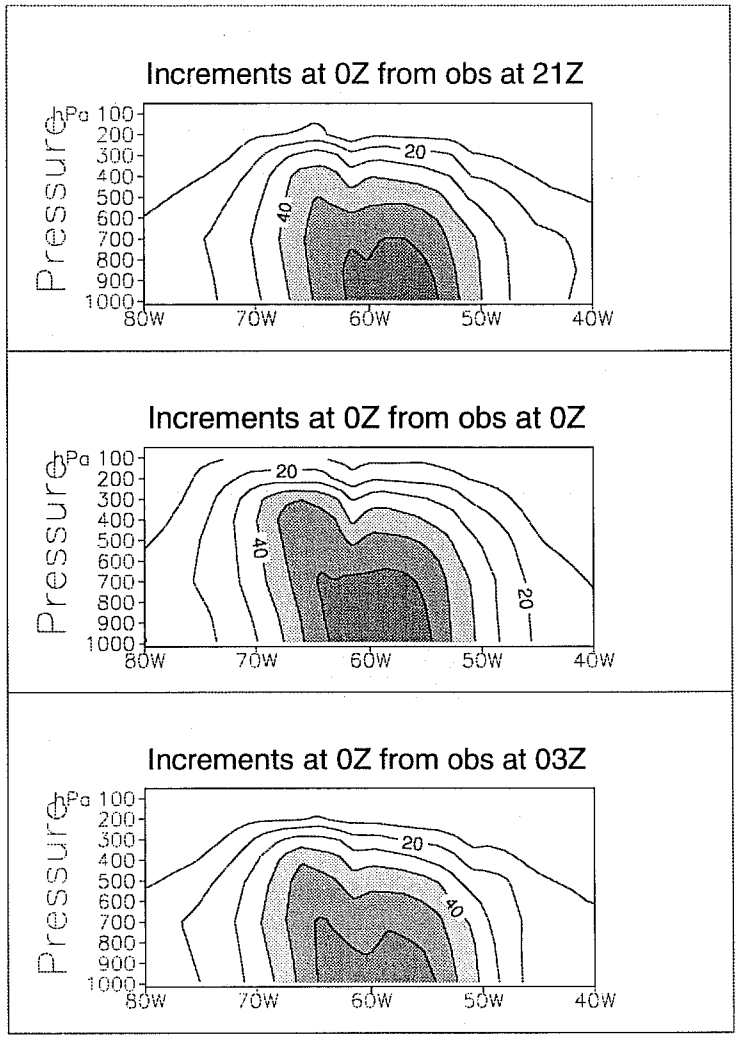

Figure 22 : Increments at 0Z for a height observation at 850 hPa, 40N, 60W. Isolines show the resulting increment, in geopotential unit. The top panel corresponds to an observation at 21Z, the middle panel to an observation at 0Z, and the bottom panel to an obsveration at 03Z.

The modified background cost function may be written as

$$J_b = \frac{1}{2}(\underline{x} - \underline{x}_b)L^T X B_\chi^{-1} X^T L(\underline{x} - \underline{x}_b) \qquad (6)$$

Here, L represents the usual change of variable of 3D/4D-Var. The background error covariance matrix of 3D/4D-Var is defined implicitly through the change of variable as $\bar{B}_x = L^{-1}L^{-T}$. (Subscripts x or χ will be used in the following to show whether a covariance matrix applies to model variables or to the control vector.) The matrix X is orthogonal and acts merely as a rotation of coordinates to place elements related to the subspace in the top left hand corner of the innermost matrix, B_χ^{-1}, which is the time-independent mean background error covariance matrix.

Note that $XX^T = I$, since X is orthogonal. Thus the background term used currently in 3D/4D-Var corresponds to replacing B_χ^{-1} by the identity matrix. The modified cost function of the simplified Kalman filter defines B_χ^{-1} to be a matrix of the form

$$B_\chi^{-1} = \begin{bmatrix} E & F^T \\ F & I \end{bmatrix}, \tag{7}$$

where E is a small square matrix and is the projection of the inverse of the covariance matrix of background error (for the control vector) onto the subspace. E determines the contribution to the background cost function from terms of the form

$$\frac{1}{2}(\delta \underline{x}_1)^T B_x^{-1} \delta \underline{x}_2 \tag{8}$$

where $\delta \underline{x}_1$ and $\delta \underline{x}_2$ both lie in the subspace. The matrix F ensures that the correct weight is given to contributions for which $\delta \underline{x}_1$ lies in the subspace and $\delta \underline{x}_2$ lies outside the subspace. The identity matrix in the lower right hand corner of B_χ^{-1} indicates that if both $\delta \underline{x}_1$ and $\delta \underline{x}_2$ lie outside the subspace, the associated background cost is the same as in 3D/4D-Var.

The modified background cost function described above requires two new inputs to the analysis system. First, the subspace must be defined. This is done by giving the analysis a set of vectors, $\{\underline{y}_i | i = 1 \ldots K\}$, which span the subspace. These vectors define the orthogonal matrix X. A second set of vectors, $\{\underline{z}_i | i = 1 \ldots K\}$ is also supplied to the analysis. Each vector $\underline{z}_i$ is assumed to be the result of multiplying the corresponding vector $\underline{y}_i$ by the inverse of the true covariance matrix of background error. These vectors provide sufficient information to diagnose the submatrices E and F of the true background error covariance matrix. Details of the construction of B_χ^{-1} and X from the vectors $\underline{y}_i$ and $\underline{z}_i$ are given in the appendix.

The true covariance matrix of background error is not known. However, a good approximation, which incorporates dynamical evolution, is given by

$$P_x^f = MP_x^a M^T \tag{9}$$

where M denotes the resolvent of the tangent linear dynamics, and P_x^a is an approximation to the covariance matrix of analysis error. It is shown in the appendix that if P_x^a is the inverse of the Hessian matrix of the analysis cost function, then pairs of vectors $\{(\underline{y}_i, \underline{z}_i) | i = 1 \ldots K\}$ satisfying $\underline{z}_i = (P_\chi^f)^{-1} \underline{y}_i$ may be generated at no additional cost from a singular vector calculation, provided that the inner product at initial time for the singular vector calculation is defined using the analysis Hessian. The ability to calculate this type of singular vector has already been developed as part of the research in predictability (Barkmeijer et al, 1997). Details of the calculation are given in the appendix.

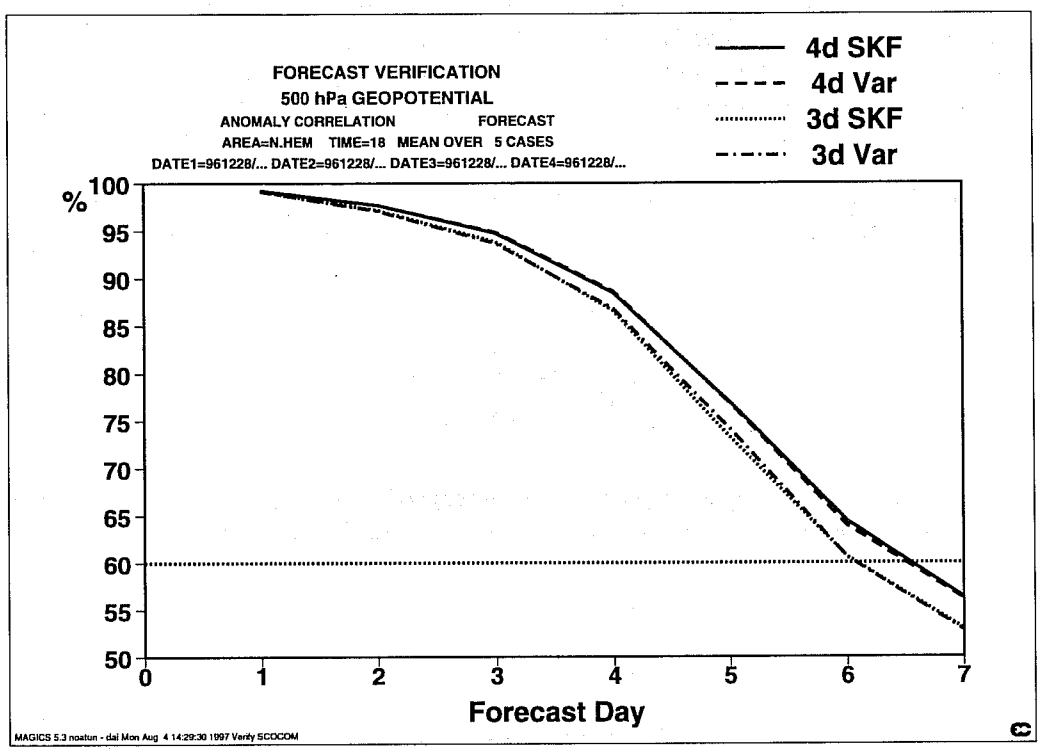

Figure 23 : Anomaly correlation scores for 500hPa geopotential for 7 pairs of 3D-Var analyses and 5 pairs of 4D-Var analyses.

6.3 Preliminary Experimentation

The simplified Kalman filter is at an early stage of development. At the time of writing, only two analyses have been performed using the scheme described above. More analyses have been performed with an earlier version of the simplified Kalman filter, which is algorithmically somewhat different to that described above, but is mathematically equivalent to setting the elements of the matrix F to zero. These analyses were run for various dates during winter 96/97. In each case, the analysis time was 18z. The singular vectors used to define the unstable subspaces were calculated at T42 resolution with an initial time of 12z (i.e. 6 hours before analysis time) and an optimization period of 48 hours. The inner product at final time was defined as energy integrated over the Northern Hemisphere.

The earlier version of the simplified Kalman filter had a neutral impact on forecast scores, as illustrated by Figure 23. There are several possible reasons for this. The matrix F, which was set to zero for these analyses, is likely to be important; the choice of optimization time and inner product for the singular vectors may be suboptimal; and the resolution of the singular vectors may be too low to capture the covariances associated with rapidly growing small-scale structures. Furthermore, these results were obtained on separate analyses, without any cycling from the SKF from one to the other.

A case study from 18Z on 28th December 1996 will now be discussed. The three panels on the left side of Figure 24 show the difference in the evolution of the 500hPa geopotential height field over the first two days of two forecasts whose initial data was provided in one case by a standard 4D-Var analysis and in the other case by the simplified Kalman filter.

The panels on the right side of Figure 24 show the difference between the forecast run from the standard 4D-Var analysis and the verifying ECMWF analysis. (Note that the contour interval varies from panel to panel, and is indicated by CI in the text above each plot. The units are m^2s^{-2}.) The analyses differ by only a few metres. However the main differences occur in a region of rapid cyclogenesis. After two days, the differences have grown rapidly. The forecast from the

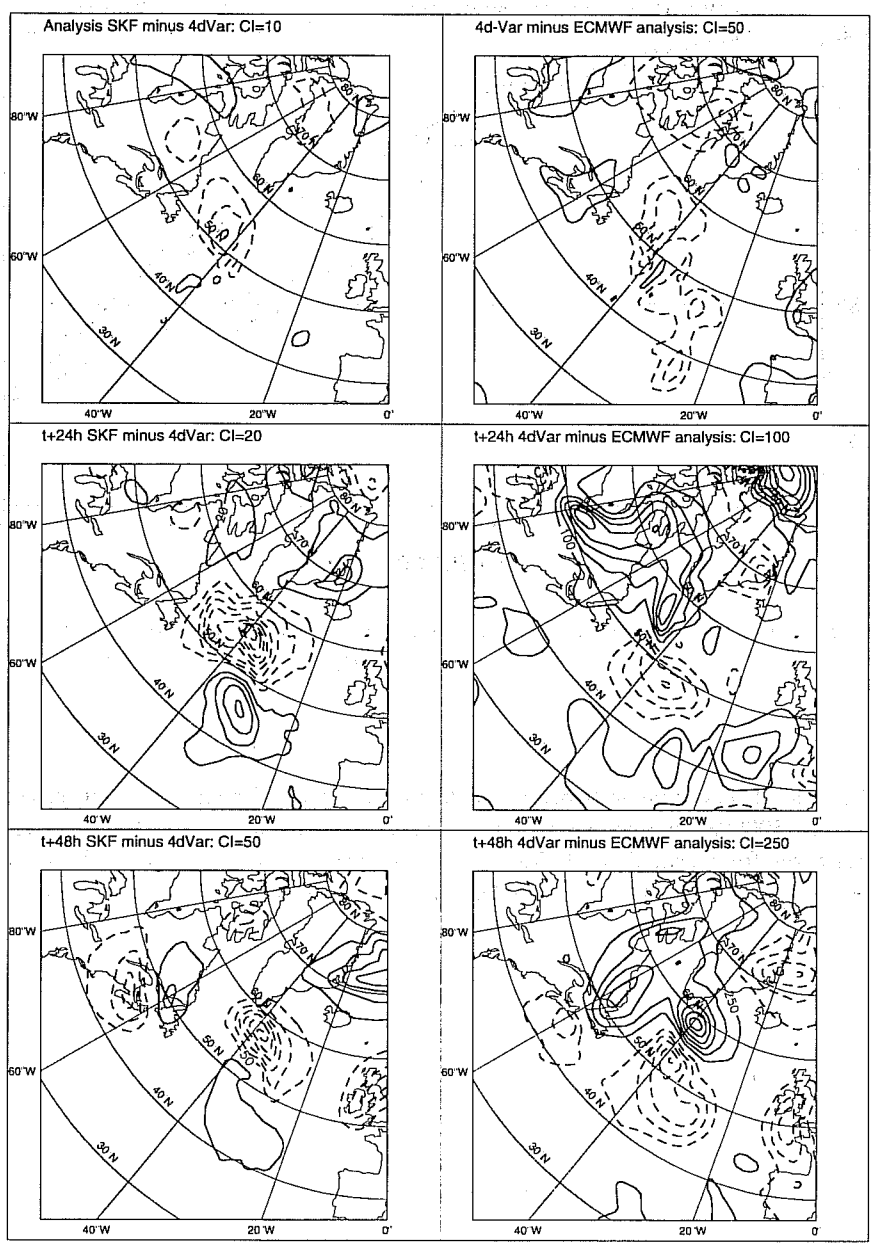

Figure 24 : Analysis and forecast differences: SKF minus 4D-Var (panels on the left); 4D-Var minus verifying analysis (panels on the right).

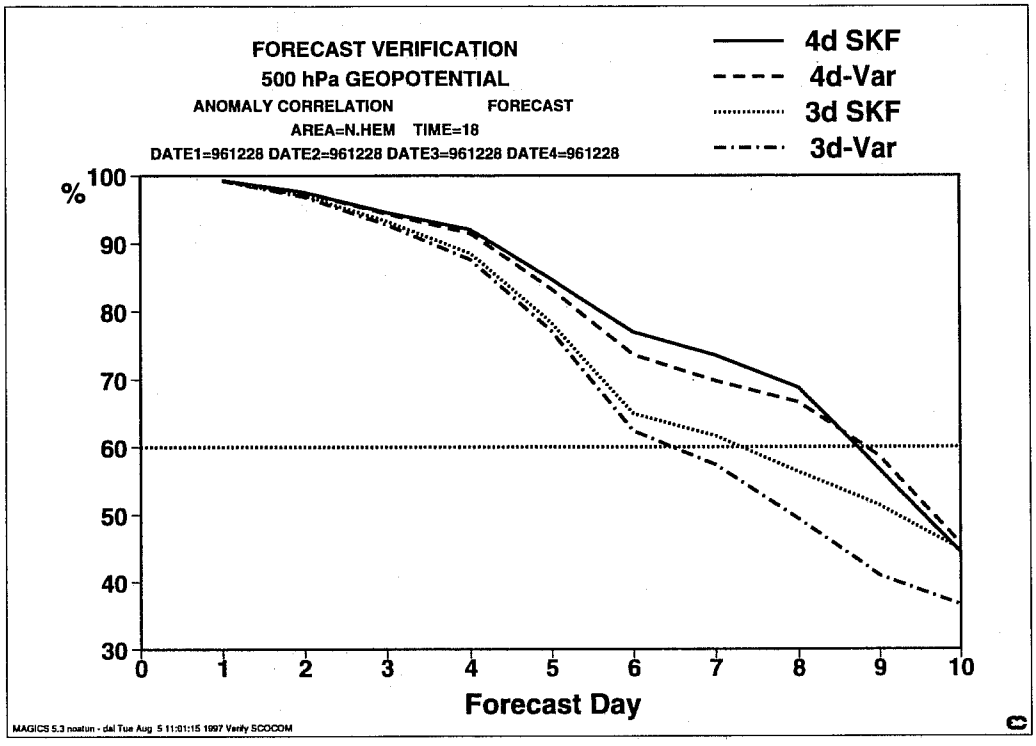

Figure 25 : Anomaly correlation scores for 500hPa geopotential for four forecasts from 18z on 28th December 1996.

simplified Kalman filter analysis is significantly better than that from the 4D-Var analysis in this case, as demonstrated by Figure 25, which shows the 500hPa geopotential anomaly correlation scores for the two forecasts. Also shown in Figure 25 are scores for forecasts run from a standard 3D-Var analysis and from a 3D-Var simplified Kalman filter analysis. It is clear that in this case, use of the flow-dependent background error covariance matrix has significantly improved the forecasts in both 3D-Var and 4D-Var.

7. Summary and discussion

Four-dimensional variational assimilation (4D-Var) has been extensively tested at high resolution on the FUJITSU, and compared with three-dimensional variational assimilation (3D-Var). Since previous work on the CRAY, a number of changes have been made to the system. Some technical and scientific changes involve the way the quality control is performed, the evaluation of the analysis and background errors computation and the way the observation operators are implemented in the minimization. To validate all these changes a 2-week period studied on the CRAY was repeated at higher resolution on the FUJITSU, with a similar improvement brought by 4D-Var. Then, a more thorough investigation of the poor performance of 4D-Var in the Tropics revealed some problems in the way the adiabatic non-linear normal mode initialization of the increments was performed. Going from four outer loops to only one (as in 3D-Var) helped to reduce the problem, together with a change to the new background formulation and an initialization of only the small scales. Tropical scores then became only marginally worse for 4D-Var than for 3D-Var.

Twelve weeks of experimentation with the one outer-loop 4D-Var and the new background formulation have been studied. These include 7 weeks of summer and 5 weeks of winter. In the medium range, each two to three-week period has been found to be either neutral or positive, resulting in slightly positive averaged scores. The improvement is more pronounced in the Southern Hemisphere. In the short-range, each two to three-week period has been found to be positive. The better short-range performance of the 4D-Var system was also shown by the fits of the background fields to the data. The comparison of forecasts to data in the Northern Hemisphere up to day 10 confirms the better scores for 4D-Var (see Figure 26). In individual synoptic cases corresponding to interesting IOPs during the FASTEX period, 4D-Var is seen to perform better than 3D-Var during rapid cyclogeneses.

A first comprehensive set of linear physics has been developed for 4D-Var applications. First, it has been evaluated by comparing the evolution of analysis increments with respect to non-linear integrations including the full physics. A better agreement of the evolved increments is found when the physics is included. The improvement is largest in the boundary layer and for specific humidity. The inclusion of this package in a 4D-Var "2-update" configuration has a positive impact on the performance of the analysis in the tropics, with a reduction of the spin-down of precipitation in the subsequent forecasts, and improved wind scores. The extratropical scores averaged over 8 weeks show a slight improvement brought by the physics. The positive impact is particularly large over Europe. In view of these results, it is planned to implement 4D-Var with physics and two outer loops in operations in the near future.

Some structure functions were illustrated in the 4D-Var case for a height observation inserted at the beginning of the assimilation window, in the middle or at the end. The dynamical processes seem to be relevant, even on a short 6-hour assimilation period. More influence of the dynamics could be taken into account by a proper cycling a 4D-Var using a simplified Kalman filter (SKF), which is currently being developed, and whose feasibility has been demonstrated.

To be cycled in a cost-effective way, the SKF would need longer 4D-Var windows, which is an option to be studied in the near-future. Future work will also concentrate on further developing the physics and studying and improving the convergence of the incremental approach.

As far as cost issues are concerned, the main analysis job of 4D-Var with only one outer loop and no physics (our baseline 4D-Var) costs about twice the main analysis job of 3D-Var. The extra-cost of the configuration "2-update with physics" is 50 % more than the "1-update adiabatic", which amounts to three times the cost of a 3D-Var main analysis job. The additional costs of 4D-Var are mainly due to the extra model integrations which run efficiently on the FUJITSU. Comparing the relative costs of a whole day of assimilation plus one 10-day forecast, the baseline 4D-Var costs 40% more than 3D-Var, and the 4D-Var with physics 65% more than 3D-Var. Viewed in the context of the overall cost of the operational suite, including the Ensemble Prediction System for instance, this increased cost is an acceptable small fraction, in line with expectations.

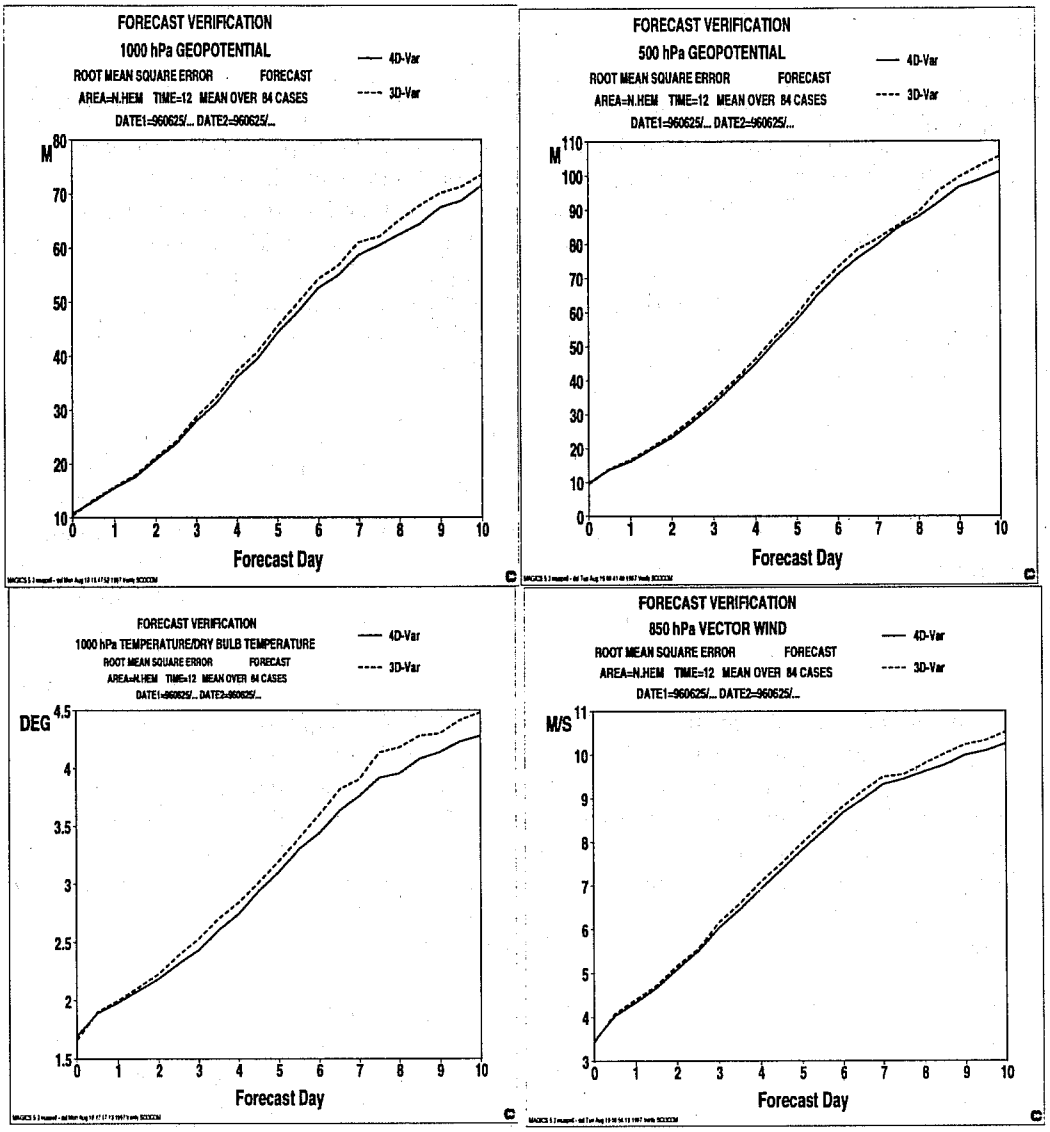

Figure 26 : Root mean square of the difference between forecasts from 4D-Var and observations (solid lines) and forecasts from 3D-Var and observations (dashed lines), averaged over 12 weeks in the Northern Hemisphere, for various parameters (geopotential height at 1000 hPa and 500 hPa, temperature at 1000 hPa, vector wind at 850 hPa).

Appendix

A1.Implementation of the Simplified Kalman Filter in 3D- and 4D-Var.

Suppose we have a set of vectors, $\{\underline{y}_i | i = 1...K\}$, which span a subspace of interest. Let X be an orthogonal matrix which rotates coordinates so that, for each $i = 1...K$, all but the first K elements of $X^T \underline{y}_i$ are zero. X is too large to be represented explicitly, but may be constructed as a sequence of Householder transformations which reduce to upper triangular form the matrix whose columns are the vectors $\underline{y}_i$. This allows operators which multiply arbitrary vectors by X or its transpose to be coded with a storage requirement of $N \times K$ where N is the dimension of the control vector.

Let us, without loss of generality, write the inverse of the true covariance matrix of background error for the control vector as

$$(P_\chi^f)^{-1} = X \begin{bmatrix} E & F^T \\ F & G \end{bmatrix} X^T \tag{A.1}$$

where E is a $K \times K$ matrix.

Since, by construction, X^T zeroes all but the first K elements of $\underline{y}_i$, we have

$$(P_\chi^f)^{-1} \underline{y}_i = X \begin{bmatrix} E \\ F \end{bmatrix} \underline{u}_i \tag{A.2}$$

where $\underline{u}_i$ is the vector comprising the first K elements of $X^T \underline{y}_i$.

Now suppose that we have a set of vectors $\underline{z}_i$, which satisfy $\underline{z}_i = (P_\chi^f)^{-1} \underline{y}_i$. Substituting this into equation 2 and multiplying to the left by X^T, gives

$$X^T \underline{z}_i = \begin{bmatrix} E \\ F \end{bmatrix} \underline{u}_i \tag{A.3}$$

Let U and Z be the matrices whose columns are, respectively, the vectors $\underline{u}_i$ and $X^T \underline{z}_i$, then we have the following system of linear equations:

$$U^T \begin{bmatrix} E & F^T \end{bmatrix} = Z^T \tag{A.4}$$

The matrix U is small and upper triangular, so this equation may be easily solved to determine the elements of the matrices E and F.

The vectors $\underline{y}_i$ contain no information about the matrix G. The inverse of the approximate background error covariance matrix used in the analysis is therefore constructed by replacing the matrix G by the identity matrix, as is done in the current formulation of 3D- and 4D-Var. Replacing G by the identity matrix does not guarantee that the inverse of the approximate background error covariance matrix is positive definite. However, this is easily tested, since $det(B_\chi^{-1}) = det(E - F^T F)$. If necessary, the cross-covariances represented by the matrix F may be reduced to restore positive definiteness.

The simplified Kalman filter is implemented as a modification to the change of variable. This requires that B_χ^{-1} is expressed in the form

$$B_\chi^{-1} = \hat{L}^T \hat{L} \tag{A.5}$$

A convenient decomposition is to define $\hat{L}$ as

$$\hat{L} = \begin{bmatrix} U & 0 \\ F & I \end{bmatrix} \tag{A.6}$$

where U is the upper triangular Cholesky square root of $E - F^T F$.

The modified change of variable is therefore

$$\hat{\chi} = \begin{bmatrix} U & 0 \\ F & I \end{bmatrix} X^T L(\underline{x} - \underline{x}_b) \tag{A.7}$$

where L denotes the static change of variable used currently in 3D- and 4D-Var.

A2.Definition of the subspace for the Simplified Kalman Filter.

The modified change of variable described above requires pairs of vectors $\{(\underline{y}_i,\underline{z}_i)|i = 1...K\}$ which satisfy $\underline{z}_i = (P_\chi^f)^{-1}\underline{y}_i$. The vectors $\underline{y}_i$ define a subspace for which we wish to accurately calculate the background cost. An attractive choice is to define $\underline{y}_i$ to be a set of singular vectors with optimization time a day or two after the analysis time. That is, we attempt to analyse as accurately as possible in directions in which analysis errors will amplify rapidly.

Neglecting model error, the true covariance matrix of background error is related to the covariance matrix of analysis error for the previous analysis cycle, P^a, through

$$P^f = M_{01}P^aM_{01}^T \tag{A.8}$$

where M_{01} is the resolvent of the tangent linear model integration over the period of the background forecast. In general, inversion of P^f to calculate $\underline{z}_i$ given a vector $\underline{y}_i$ is computationally expensive. The inverse of the tangent linear and adjoint dynamics are not available as operators and the inversion must be carried out iteratively. (The inverse of the analysis error covariance matrix, on the other hand, is available since it is equal to the Hessian matrix of the analysis cost function.)

Consider the singular vectors which maximize the ratio

$$\lambda = \frac{\langle M_{02}\underline{s}_0, M_{02}\underline{s}_0\rangle_2}{\langle \underline{s}_0, \underline{s}_0\rangle_0} \tag{A.9}$$

where subscripts 0 and 2 denote the initial and final time of the singular vector calculation, and where M_{02} is the resolvent of the tangent linear dynamics.

Suppose we choose the inner product at initial time to be defined by the covariance matrix of analysis error as $\langle \underline{x},\underline{y}\rangle_0 \equiv \underline{x}^T(P^a)^{-1}\underline{y}$. Suppose also that the inner product at final time may be written as $\langle \underline{x},\underline{y}\rangle_2 \equiv \underline{x}^TV^TV\underline{y}$. Then the singular vectors satisfy the following generalized eigenvector equation

$$M_{02}^TV^TVM_{02}\underline{s}_0 = \lambda(P^a)^{-1}\underline{s}_0 \tag{A.10}$$

The algorithmic developments required to solve this equation have been made as part of the research of the predictability section.

Now define M_{01} to be the resolvent of the tangent linear dynamics for an integration from the initial time of the singular vector calculation to some intermediate time between initial and final time. Similarly, define M_{12} to be the resolvent of the tangent linear dynamics for an integration from the intermediate time to final time, so that $M_{02} = M_{12}M_{01}$.

Consider the partially evolved singular vector $\underline{y} = M_{01}\underline{s}_0$. Assuming invertibility of the dynamics, equation 10 may be written as

$$M_{12}^TV^TVM_{02}\underline{s}_0 = \lambda(M_{01}P^aM_{01}^T)^{-1}\underline{y} \tag{A.11}$$

Now, $M_{02}\underline{s}_0$ is the evolved singular vector at final time, $\underline{s}_2$, so we have

$$M_{12}^T V^T V \underline{s}_2 = \lambda (P^f)^{-1} \underline{y} \tag{A.12}$$

We may therefore define $\underline{z} = \frac{1}{\lambda} M_{12}^T V^T V \underline{s}_2$ to get a pair of vectors which satisfy $\underline{z}_i = (P_\chi^f)^{-1} \underline{y}_i$, as required to define the modified background error covariance matrix described in the previous section. Pairs of vectors may be generated at no additional cost from the singular vector calculation, since both are already produced during the course of the computation.

References

Andersson, E., 1996: Implementation of variational quality control. In Proceedings of the ECMWF workshop on non-linear aspects of data assimilation, Shinfield Park, Reading, RG2 9AX, 9-11 September 1996.

Andersson, E., Haseler, J., Undén, P., Courtier, P., Kelly, G., Vasiljevic, D., Brankovic, C., Cardinali, C., Gaffard, C., Hollingsworth, A., Jakob, C., Janssen, P., Klinker, E., Lanzinger, A., Miller, M., Rabier, F., Simmons, A., Strauss, B., Thépaut, J.–N. and Viterbo, P., 1998: The ECMWF implementation of three dimensional variational assimilation (3D–Var). Part III: Experimental results. To appear in *Q. J. R. Meteorol. Soc.*

Barkmeijer, J., Van Gijzen, M., and Bouttier, F., 1997: Singular vectors and the analysis error covariance metric. Submitted to Q. J. R. Meteorol. Soc.

Bouttier, F., Derber, J., and Fisher, M., 1997: The 1997 revision of the J_b term in 3D/4D-Var. ECMWF Research Department Technical Memorandum No. 238.

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

Buizza, R., Petroliagis, T., Palmer, T., N., Barkmeijer, J., Hamrud, M., Hollingsworth, A., Simmons, A., and Wedi, N., 1997: Impact of model resolution and ensemble size on the performance of an ensemble prediction system. Submitted to Q. J. R. Meteorol. Soc.

Courtier P. and Talagrand, O., 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. Part II: Numerical results. Q. J. R. Meteorol. Soc., 113, 1329-1347.

Courtier, P. , Thépaut, J-N., and Hollingsworth, A., 1994: A strategy for operational implementation of 4D-Var, using an incremental approach.Q. J. R. Meteorol. Soc., 120, 1367-1388.

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F. and Fisher, M., 1998: The ECMWF implementation of three dimensional variational assimilation (3D–Var). Part I: Formulation. To appear in *Q. J. R. Meteorol. Soc.*

Fisher, M., and Courtier, P., 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. ECMWF Research Department Technical Memorandum No. 220.

Järvinen, H, and Undén, P., 1997: Observation screening and first-guess quality control in the ECMWF 3D-Var data assimilation system. ECMWF Research Department Technical Memorandum No. 236.

Joly, A., Jorgensen, D., Shapiro. M., A., Thorpe, A., Bessemoulin, P., Browning, K., A., Cammas, J-P., Chalon, J-P., Clough, S.,A., Emanuel, K.,A., Eymard, L., Gall, R., Hildebrand, P., H., Langland, R.,H., Lemaître, Y., Lynch, P., Moore, J., A., Persson, P. , O., G., Snyder, C., Wakimoto, R., M., 1997: Definition of the Fronts and Atlantic Storm-Track Experiment (FASTEX). Submitted to BAMS.

Le Dimet, F-X., and Talagrand, O., 1986:Variational algorithms for analysis and assimilation of meteorological observations. Tellus, 38A, 97-110.

Lewis, J., and Derber, J., 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. Tellus, 37, 309-327.

Lott, F., and Miller, M., J., 1997: A new subgrid-scale orographic drag parametrization: its formulation and testing. Q. J. R. Meteorol. Soc., 123, 101-127.

Louis, J-F., Tiedke, M., and Geleyn, J-F., 1981: A short history of the operational PBL parametrization at ECMWF. In Proceedings of the ECMWF workshop on boundary layer parametrization, Shinfield Park, Reading, RG2 9AX, November 1981.

Mahfouf, J-F., Buizza, R., and Errico, R., 1996: Strategy for including physical processes in the ECMWF variational data assimilation system. In Proceedings of the ECMWF workshop on non-linear aspects of data assimilation, Shinfield Park, Reading, RG2 9AX, 9-11 September 1996.

Navon, I-M., Zou, X., Derber, J. C., and Sela, J., 1992: Variational data assimilation with an adiabatic version of the NMC spectral model. Mon. Wea. Rev., 120, 1433-1446.

Rabier, F. and Courtier, P., 1992: Four-dimensional assimilation in the presence of baroclinic instability. Q. J. R. Meteorol. Soc., 118, 649-672.

Rabier, F., Thépaut, J-N., and Courtier, P. , 1997: Extended assimilation and forecast experiments with a Four-Dimensional variational assimilation system. Submitted to Q. J. R. Meteorol. Soc.

Rabier, F., McNally, A., Andersson, E., Courtier, P., Undén, P., Eyre, J., Hollingsworth, A. and Bouttier, F., 1998: The ECMWF implementation of three dimensional variational assimilation (3D–Var). Part II: Structure funtions. To appear in *Q. J. R. Meteorol. Soc.*

Simmons, A., and Rabier, F., 1997: Removal of the incremental initialization of medium and large scales in the ECMWF analysis. ECMWF memorandum Research Department, May 1997,

Talagrand and Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I: Theory. Q. J. R. Meteorol. Soc., 113, 1321-1328.

Thépaut, J-N., and Courtier, P., 1991:Four-dimensional variational assimilation using the adjoint of a multilevel primitive-equation model. Q. J. R. Meteorol. Soc., 117, 1225-1254.

Thépaut, J-N., Courtier, P., Belaud, G., and Lemaître, G., 1996: Dynamical structure functions in a four-dimensional variational assimilatio: a case study. Q. J. R. Meteorol. Soc., 122, 535-561.

Zupanski, M., 1993: Regional four-dimensional variatio nal data assimilation in a quasi-operational forecasting environment. Mon. Wea. Rev., 121, 2396-2408.