Q. J. R. Meteorol. Soc. (2000), **126**, pp. 1143–1170

The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics

By F. RABIER, H. JÄRVINEN, E. KLINKER, J.-F. MAHFOUF* and A. SIMMONS
European Centre for Medium-Range Weather Forecasts, UK

(Received 17 November 1998; revised 25 May 1999)

SUMMARY

This paper presents results of a comparison between four-dimensional variational assimilation (4D-Var), using a 6-hour assimilation window and simplified physics during the minimization, and three-dimensional variational assimilation (3D-Var). Results have been obtained at 'operational' resolution T213L31/T63L31. (T defines the spectral triangular truncation and L the number of levels in the vertical, with the first parameters defining the resolution of the model trajectory, and the second the resolution of the inner-loop.) The sensitivity of the 4D-Var performance to different set-ups is investigated. In particular, the performance of 4D-Var in the Tropics revealed some sensitivity to the way the adiabatic nonlinear normal-mode initialization of the increments was performed. Going from four outer-loops to only one (as in 3D-Var), together with a change to the 1997 formulation of the background constraint and an initialization of only the small scales, helped to improve the 4D-Var performance. 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 1997 background formulation have been studied. The averaged scores show a small but consistent improvement in both hemispheres at all ranges. In the short range, each two- to three-week period has been found to be slightly positive throughout the troposphere. The better short-range performance of the 4D-Var system is also shown by the fits of the background fields to the data. More results are presented for the Atlantic Ocean area during FASTEX (the Fronts and Atlantic Storm-Track EXperiment), during which 4D-Var is found to perform better. In individual synoptic cases corresponding to interesting Intensive Observing Periods, 4D-Var has a clear advantage over 3D-Var during rapid cyclogeneses. The very short-range forecasts used as backgrounds are much closer to the data over the Atlantic for 4D-Var than for 3D-Var. The 4D-Var analyses also display more day-to-day variability. Some structure functions are illustrated in the 4D-Var case for a height observation inserted at the beginning, in the middle or at the end of the assimilation window. The dynamical processes seem to be relevant, even with a short 6-hour assimilation period, which explains the better overall performance of the 4D-Var system.

KEYWORDS: Adjoint models Analysis Variational assimilation

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 3D-Var analysis used operationally at ECMWF since January 1996 (Courtier *et al.* 1998; Rabier *et al.* 1998a; 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 (Lewis and Derber 1985; Le Dimet and Talagrand 1986; Courtier and Talagrand 1987; Talagrand and Courtier 1987), before being tested in the context of primitive-equation models (Thépaut and Courtier 1991; Rabier and Courtier 1992; Navon *et al.* 1992; Zupanski 1993). It is being developed 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 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.* (1998b). (The following convention is adopted: T106L31 is the resolution of the model trajectory and T63L31 is the resolution of the inner-loop.) In

* Corresponding author: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK. e-mail: mahfouf@ecmwf.int

[†] T106L31 is spectral triangular truncation 106 with 31 levels in the vertical.

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 and 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, with a view to operational implementation, before revisiting the 4D-Var with longer windows. Two additional two-week periods were then run, which showed a consistent improvement of 4D-Var over 3D-Var in the extratropics. Some problems emerged however in the tropical area.

This is the first of three papers describing the steps leading to the operational implementation of 4D-Var with a 6-hour window at ECMWF on 25th November 1997. As this is such an important achievement (a worldwide first, which involved many people and resources, and with much potential for better exploitation of existing and forthcoming sources of observations), it was decided to describe thoroughly the stages leading to the implementation. This paper (Part I) concentrates on the experimental results produced with 4D-Var using very simplified physics in the minimization (as in Buizza 1994); Part II describes a set of more elaborate linear physics, its validation and impact on 4D-Var performance; Part III deals more precisely with the diagnostics of the operational system chosen as the result of the main experimentation described in I and II.

After the work reported by Rabier *et al.* (1998b), the handling of observations in 4D-Var underwent some modifications. The observation screening and quality control which used to be done by the Optimum Interpolation are now performed within the variational assimilation (Järvinen and Undén 1997; Andersson 1996; Andersson and Järvinen 1999), and the observation operators are activated in their tangent-linear versions within the minimization (a 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. Further updates of 4D-Var include using a Lanczos algorithm for the evaluation of the analysis and background errors (Fisher and Courtier 1995; Derber and Bouttier 1999).

The validation of these changes was performed by re-running the January 1996 period previously run at T106L31/T63L31 resolution with the old set-up. As discussed in section 2, which describes the sensitivity to various 4D-Var set-ups, the improvement of 4D-Var over 3D-Var seen earlier is reproduced using the same amount of data in both schemes. The use of extra off-time observations is not found to be beneficial in the current set-up. Tropical wind scores are substantially poorer in 4D- than in 3D-Var in these new experiments, as they had been previously. Subsequent developments involving the reduction of the number of outer-loops from four to one, use of a new background term introduced in operational 3D-Var in 1997 (Derber and Bouttier 1999), and removal of initialization of large scales, resulted in better tropical behaviour as also explained in section 2. Section 3 presents the results of the baseline 4D-Var chosen on the basis of these experiments (one outer-loop, 1997 background term) over a total of twelve weeks, and discusses in detail results obtained over the Atlantic area during part of FASTEX (the Fronts and Atlantic Storm-Track EXperiment; Joly *et al.* 1997). Section 4

illustrates the structure functions used implicitly in 4D-Var. Discussion of results is given in section 5.

2. SENSITIVITY OF RESULTS TO DIFFERENT 4D-VAR SET-UPS

(*a*) *Validation of changes in resolution, observation processing and forecast-error specification*

One of the periods (16 to 29 January 1996) previously tested at resolution T106L31 was repeated at T213L31 on a new distributed memory computer, with some technical and scientific changes. As mentioned in the introduction, the observation screening and quality control are now performed within the variational assimilation; the evaluation of the analysis and background errors involves a Lanczos algorithm; and the observation operators are activated in their tangent-linear versions within the minimization. The version of the analysis used is with the 1996 formulation of the background term (Courtier *et al.* 1998), and with 4 outer-loops and 20 iterations within each minimization. The 4D-Var configuration uses the same amount of data as 3D-Var. Results are presented in Fig. 1: 3D-Var at low-resolution is shown as a solid line; 3D-Var at high-resolution as a dashed line; 4D-Var at low-resolution as a dotted line; and 4D-Var at high-resolution as a dash-dotted line. One can see the improvement of both 3D-Var and 4D-Var when going to the high-resolution version. The important point for our validation exercise is that the improvement 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 previous experiments, with scores markedly worse for 4D- than for 3D-Var. As an example, the averaged high-resolution tropical wind scores verified against their own analysis at 850 hPa are 4.1 m s^{-1} for 3D-Var and 4.8 m s^{-1} for 4D-Var at day 3.

(*b*) *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 individual frequently reporting stations can be used within one assimilation period. These extra observations are a resource that has not been fully utilized by 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, 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, preference is given to observations that are close to the middle of the time slot. The IFS (Integrated Forecasting System) is, however, designed so that either an hourly or a six-hourly observation screening can be performed for 4D-Var. In both cases, the background is compared with the observations in 1-hour time intervals (at the observation time), but the decision to keep off-time data is different. For instance, if at a SYNOP location, surface pressure is reported twice, at the main analysis time T and also 3 hours later, the hourly screening will keep both records while the 6-hourly screening will only keep the one at time T. The effect of this choice

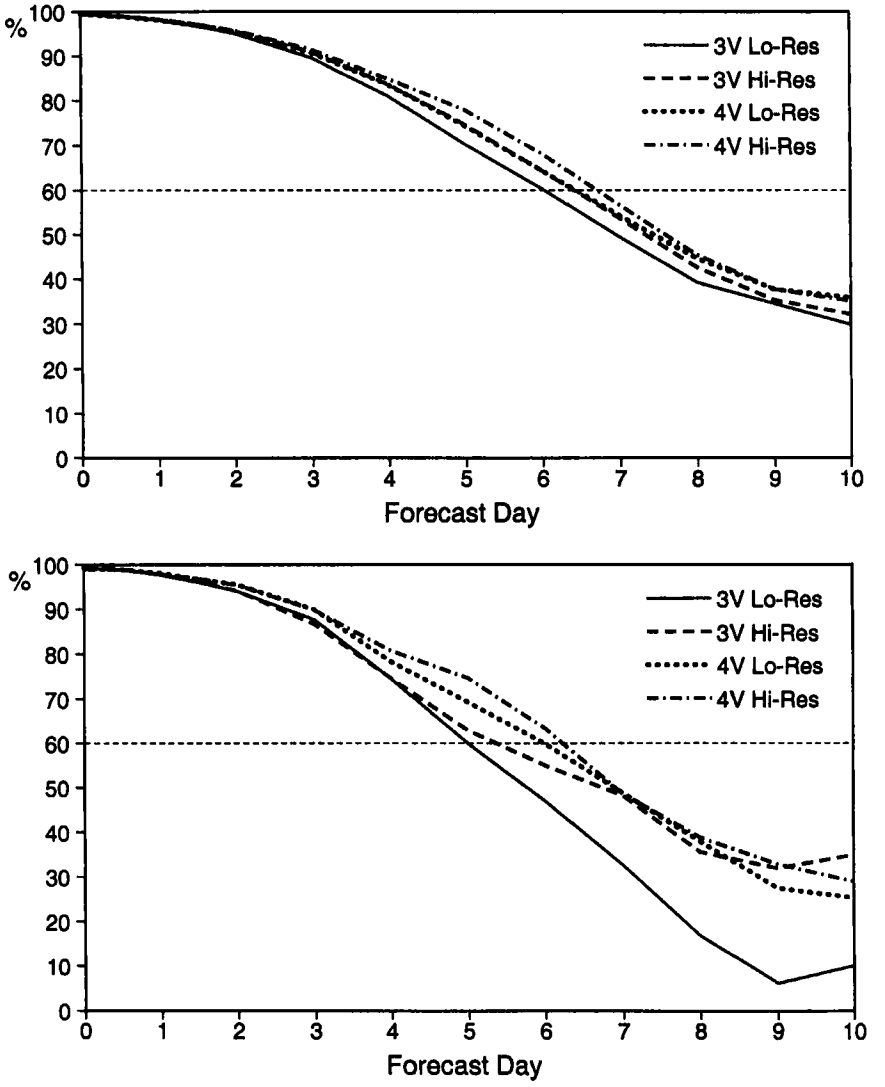

Figure 1. Anomaly correlation scores for 3D-Var (3V) and 4D-Var (4V) at T106L31 (Lo-Res) and at T213L31 (Hi-Res), using the same amount of observations, for the northern hemisphere (top panel) and North Pacific (bottom panel) for two weeks in January 1996. See text for discussion.

on the number of observations used in the assimilation is largest for observation types reported most frequently, like SYNOPs and DRIBUs. An illustrative example is given in Fig. 2 which displays the numbers of available observations in different time slots and the numbers 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 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 Fig. 3, both 4D-Var systems produce equally good forecast scores in the northern hemisphere. In the medium-range the scores are 6 to 12 hours

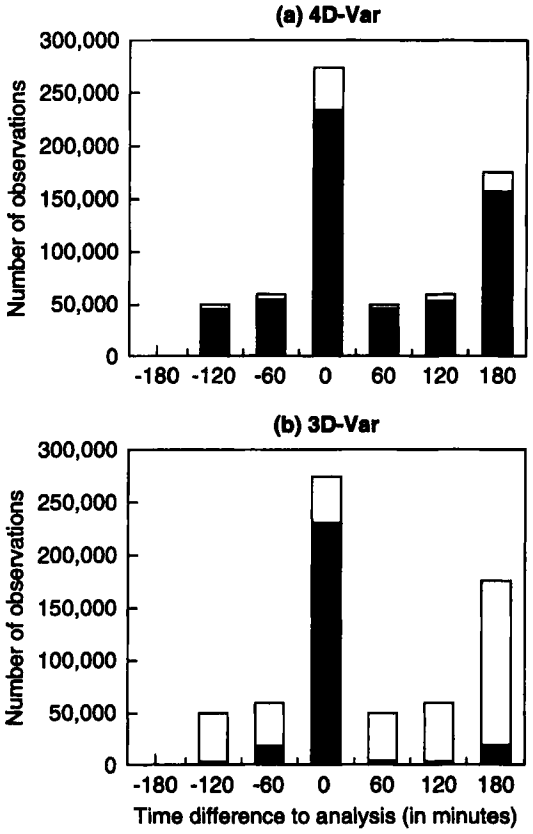

Figure 2. Time distribution of SYNOP surface pressure observations during a 2-week period. The column height gives the total number of observations available, while the shaded part displays those actually used by the analysis in (a) 4D-Var, and (b) 3D-Var.

ahead of the 3D-Var scores both at 1000 hPa and 500 hPa (not shown). The southern hemisphere scores are, however, better for the 4D-Var system without the off-time observations (dotted line). There the inclusion of the off-time observations decreased 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 compared with 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 observations, particularly in the southern hemisphere, we currently have no mechanism to prevent these unrealistic increments from appearing and developing into forecast errors. In the northern hemisphere, by 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

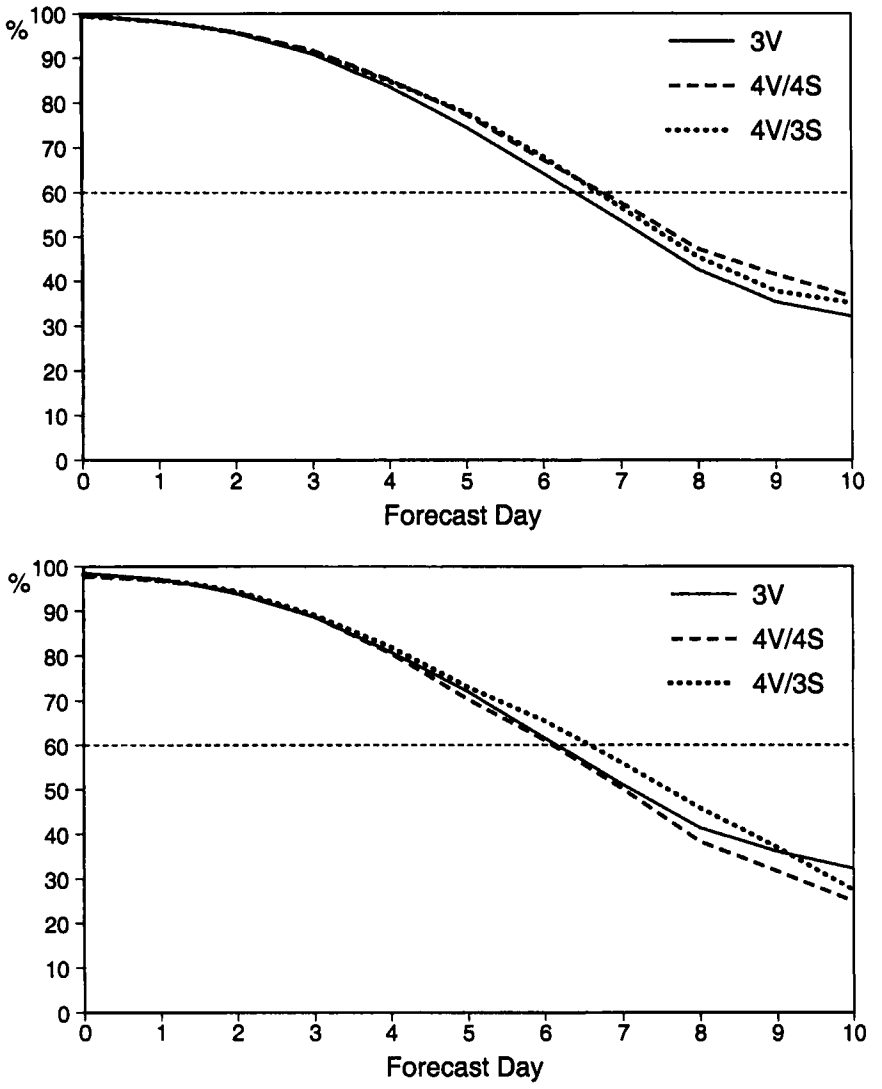

Figure 3. The anomaly correlation in the northern hemisphere (top) and the southern hemisphere (bottom) at 1000 hPa for two weeks in January 1996. 3V stands for 3D-Var; 4V/3S for 4D-Var with 3D-screening; and 4V/4S for 4D-Var with 4D-screening.

modelled. Work is in progress on this subject and recent results will be discussed in the summary.

(*c*) *Sensitivity of the tropical performance to the 4D-Var set-up*

In order to better understand the behaviour of the 4D-Var system, some analysis experiments were performed in which only a few simulated isolated observations were inserted. In particular a geopotential datum at 850 hPa was simulated for a given date (5 December 1996 at 0000 UTC) at 20°S, 80°W, with a height departure from the background of 10 m. The resulting mass and wind increments at 850 hPa are illustrated for both 4D-Var (Fig. 4(a)) and 3D-Var (Fig. 4(c)) with the use of the 1997 background constraint (Derber and Bouttier 1999). The comparison of 4D-Var and

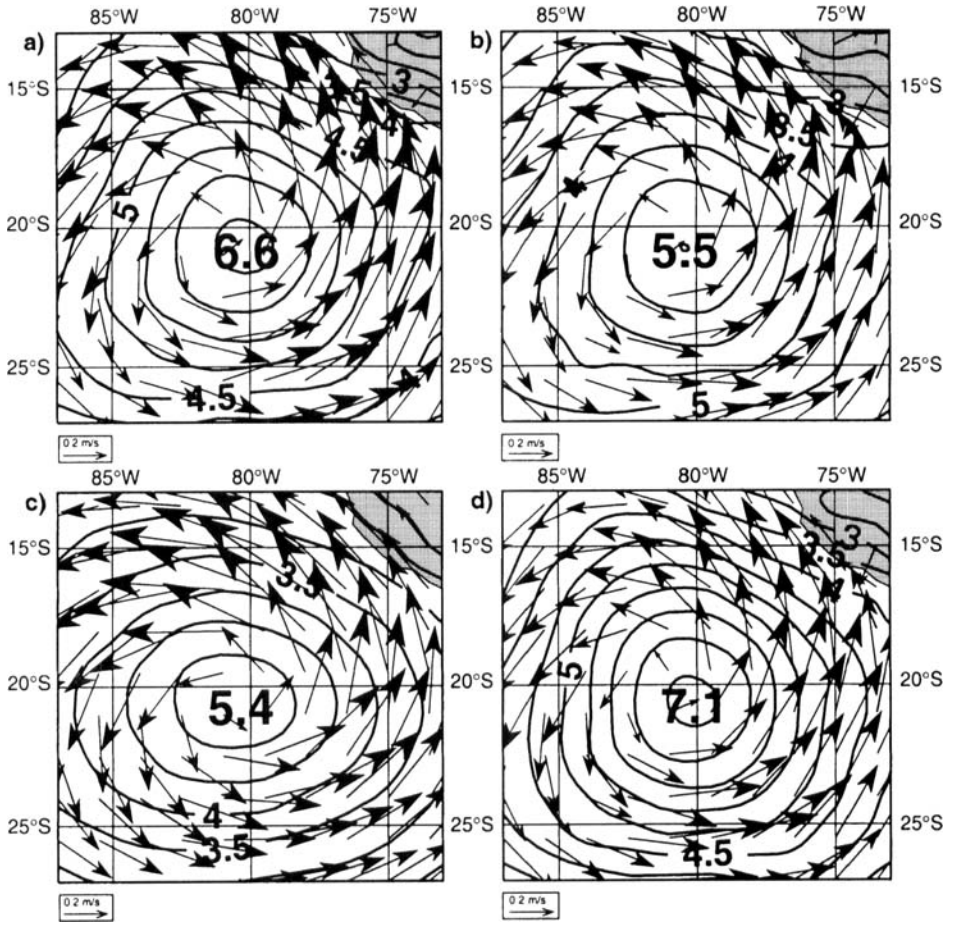

Figure 4. Mass and wind increments from an isolated mass observation at 850 hPa. (a) 4D-Var with 4 outer-loops and initialization; (b) 4D-Var with 1 outer-loop and initialization; (c) 3D-Var with initialization; and (d) 4D-Var with 1 outer-loop and without initialization. Contour intervals are 0.5 geopotential metres.

3D-Var geopotential increments shows a larger amplitude in the 4D-Var case (6.6 m versus 5.4 m). The 4D-Var wind increments also appear to be larger. The interpretation of this result requires us to refer back to the way the incremental variational assimilation is implemented.

An objective function J is defined for each minimization problem as:

$$J(\delta\mathbf{x}_1^n) = \frac{1}{2}(\delta\mathbf{x}_1^n + \mathbf{x}_1^{n-1} - \mathbf{x}_1^b)^T\mathbf{B}^{-1}(\delta\mathbf{x}_1^n + \mathbf{x}_1^{n-1} - \mathbf{x}_1^b) + \frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i\delta\mathbf{x}_1^n(t_i) - \mathbf{d}_i^{n-1})^T\mathbf{R}_i^{-1}(\mathbf{H}_i\delta\mathbf{x}_1^n(t_i) - \mathbf{d}_i^{n-1}), \quad (1)$$

with subscript 1 indicating that fields are at low resolution, subscript i the time index, superscript n the minimization index. $\mathbf{x}_1^b$ is the background field truncated at low resolution and $\mathbf{x}_1^{n-1}$ the current estimate of the atmospheric flow (it is equal to the background for the first minimization). $\delta\mathbf{x}_1^n$ is the increment at low resolution at initial time, and $\delta\mathbf{x}_1^n(t_i)$ the increment evolved according to the tangent-linear model from

the initial time to time index i. **R** and **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 by integrating the model at high resolution from our current $n-1$ estimate. The way the increment $\delta \mathbf{x}_1^n$ is then added to the current estimate can be written

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

where NMI stands for adiabatic nonlinear normal-mode initialization. The original purpose of this use of 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 wave numbers 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 Fig. 4(a) it has thus 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 panels (b) and (c) in Fig. 4. The increment in Fig. 4(b) is now only 5.5 m (versus 6.6 m in Fig. 4(a)), and the wind increments are only marginally larger than in 3D-Var. The impact of this initialization of the increments can be seen by comparing panel (b) (4D-Var with one update and NMI) with panel (d) (4D-Var with one update and no NMI). Initialization reduces the amplitude of the geopotential increment and creates more rotational wind increments associated with mass observations. The actual increment at low resolution created by the first minimization (not shown) is quite large: 7.3 m in the case of the 4D-Var experiments and 7.6 m in the 3D-Var case. Not using NMI when going to high resolution alters this value slightly to 7.1 m in the 4D-Var case (the resulting increment is displayed in Fig. 4(d)). The impact of initialization on this low-resolution increment is dramatic. It reduces the value to 5.5 m in the 4D-Var case (Fig. 4(b)) and 5.4 m in the 3D-Var case (Fig. 4(c)). When several outer-loops are used in 4D-Var, the effect of imposing initialization several times, and minimizing several times, is that the 4D-Var algorithm minimizes as much in terms of mass as when no initialization is performed, and creates rotational winds associated with these mass increments. From an increment value of 5.5 m after the first outer-loop, the second minimization increases the increment to 6.7 m. Then, the second initialization reduces it to 6.5 m, before reaching a stable 6.6 m value after 4 outer-loops. The cost-function is better minimized by performing several outer-loops in that way, but the enforced balance constraint creates large 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 from 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 nonlinear normal-mode initialization is enforced quite strongly in the tropical area. To evaluate whether this is beneficial or detrimental in 4D-Var, it was decided to compare the behaviour of 4D-Var with 4 outer-loops and with one outer-loop over a two-week period (1 to 14 February 1997). The experiments used the 1997 background constraint. The tropical wind scores are shown

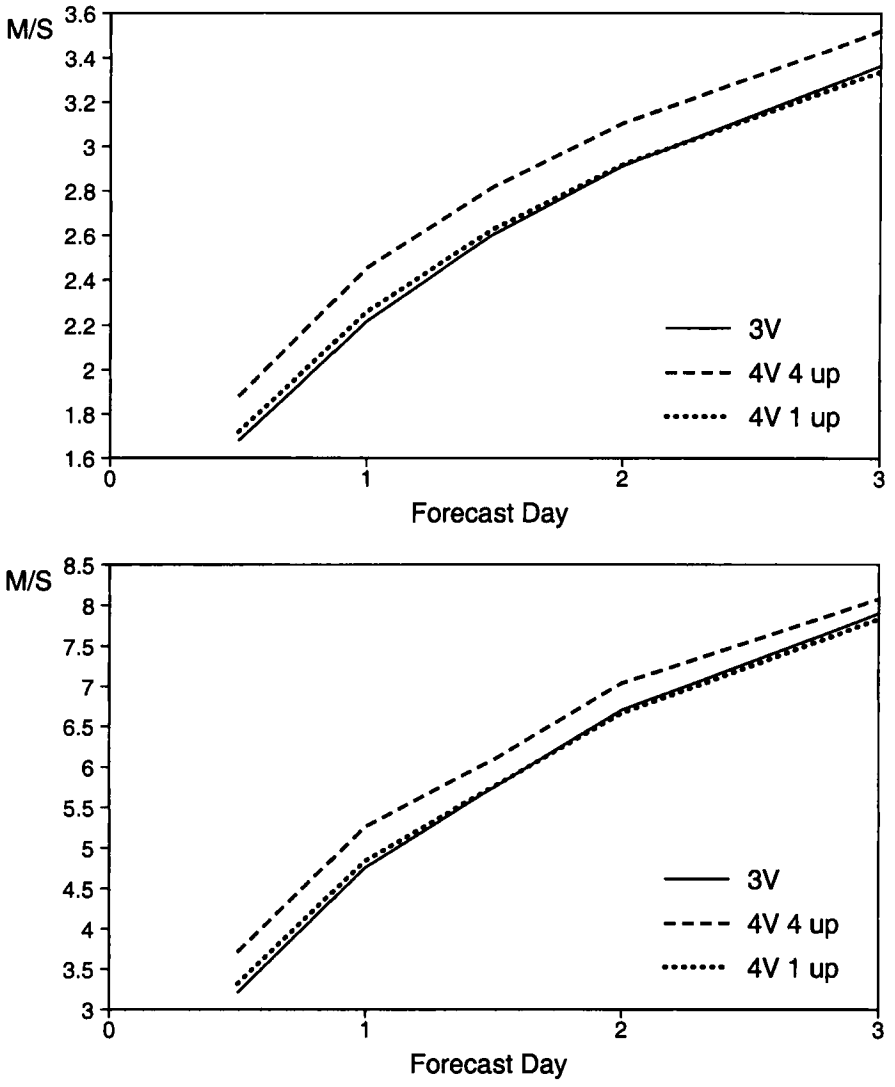

Figure 5. Tropical wind scores (root-mean-square errors in m s^{-1}) verified against own analysis at 850 (top) and 200 hPa (bottom). 3D-Var is shown as a solid line, 4D-Var with 4 outer-loops as a dashed line, and 4D-Var with one outer-loop as a dotted line.

in Fig. 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 competitive with the 3D-Var system in the Tropics. The hydrological budgets are presented for the tropical band (30°N to 30°S) for the three systems in Fig. 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 Part II that this can be remedied by using more physical processes in the 4D-Var minimization.

Besides the number of outer-loops performed in 4D-Var, two other factors were found to affect the tropical performance of both 3D- and 4D-Var. Firstly, the change to the 1997 background formulation dramatically improved the tropical scores (Derber

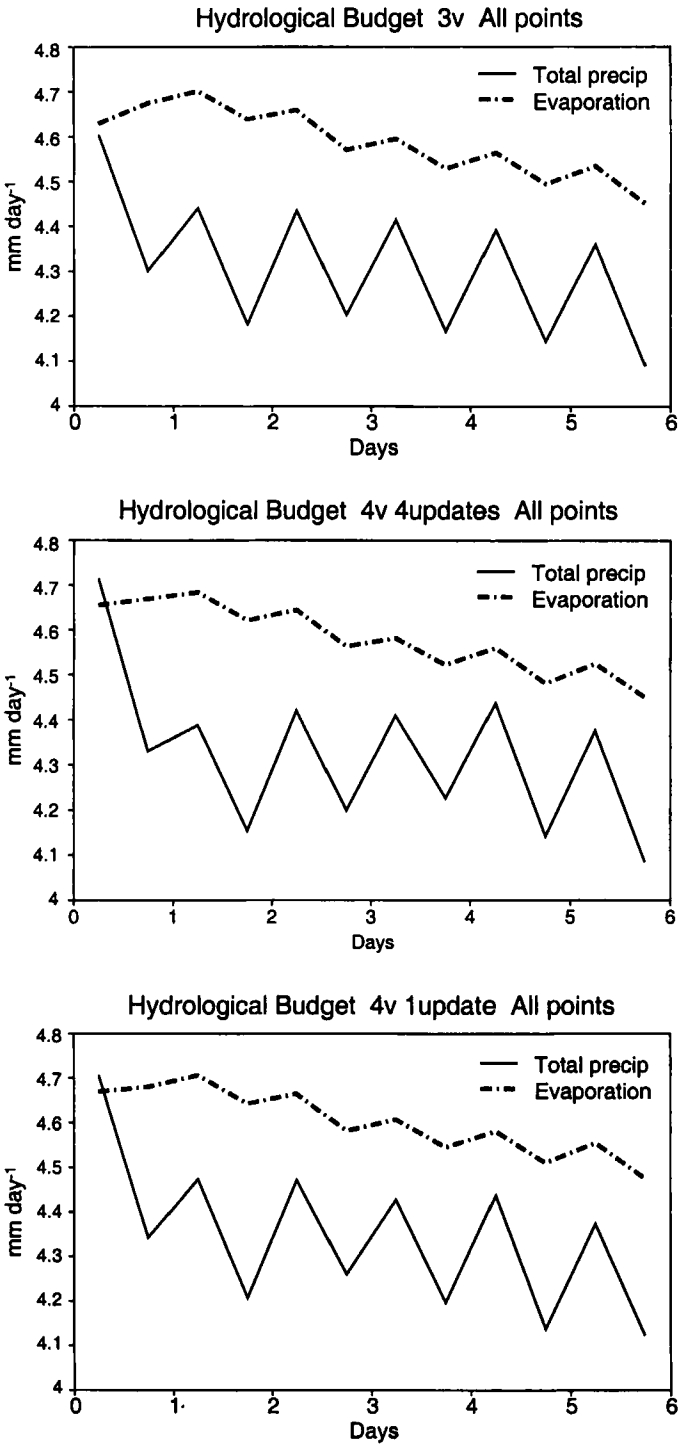

Figure 6. Hydrological budgets (mm day^{-1}) for the tropical band (30°N to 30°S) averaged over 14 days for 3D-Var (top panel), 4D-Var with 4 outer-loops (middle panel) and 4D-Var with one outer-loop (bottom panel). Total precipitation is shown as a solid line, and evaporation as a broken line.

and Bouttier 1999) in both 3D-Var and 4D-Var. The balance provided by the 1997 background constraint appeared to be sufficiently good to allow the initialization of total wave numbers below 20 to be abandoned (Simmons and Rabier 1997). We shall re-examine the impact of the number of outer-loops in 4D-Var later. For the time being, experimentation was resumed using 4D-Var with only one outer-loop. This baseline configuration has the advantage that it is relatively cheap, and ensures a reasonable behaviour in the Tropics.

3. BASELINE EXPERIMENTATION WITH 4D-VAR WITH ONE OUTER LOOP

(*a*) *Results over 12 weeks*

Several two- or three-week periods were run using 4D-Var with one outer-loop and with the 1997 formulation of the background term. The periods are: 24 August to 6 September 1995; 25 June to 15 July 1996; 15 to 28 January 1997; 1 to 21 February 1997; and 27 June to 10 July 1997.

In each experiment period, the 3D-Var control and the experimental 4D-Var system were based on the same model version (mainly IFS cycle 16r3), the same background-error term and the same data usage. Averaged scores verified against operations and averaged over the 12 weeks are presented in Fig. 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 5. In Europe, scores are positive in the medium range. The scatter in the medium range indicates that 4D-Var generally performs better in the cases of relatively bad forecast performance, while having at least as many very good forecasts as 3D-Var (not shown). In the northern hemisphere, for the anomaly correlation of geopotential height at 500 hPa at day 5, 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-Var. Similarly for the same score at day 7, 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. A selection of scores computed with respect to radiosonde observations in the northern hemisphere is presented in Fig. 8. One can see a better performance of the 4D-Var forecast at all ranges for parameters very relevant for the bench forecaster (for instance temperature and wind at low levels).

Since short-range forecast scores show less scatter than medium-range forecast scores, the short-range performance differences have a larger statistical significance than those for the medium range. For each individual 2- to 3-week period, 4D-Var was found to behave better in the short range (up to day 3) than 3D-Var. The difference in performance at day 1, averaged over the total of 12 weeks, is illustrated in Fig. 9. Cross-sections of differences between root-mean-square (r.m.s.) error at day 1 between 4D-Var and 3D-Var are presented for both hemispheres. They are computed for the 40° to 70° mid-latitude band. The values are predominantly negative throughout the atmospheric depth, which means that 4D-Var has a smaller error than 3D-Var almost everywhere. Note that the contour interval is doubled in the southern hemisphere cross-section where the impact of 4D-Var is larger. In the northern hemisphere, the largest improvement is over the oceans (160–240°E for the North Pacific and 280–340°E for the North Atlantic). The r.m.s. errors are also reduced over Europe (350–30°E). The picture is not so clear over Asia, with a large positive impact of 4D-Var in the mid-troposphere at around 100°E, but negative impact at 50°E and 150°E.

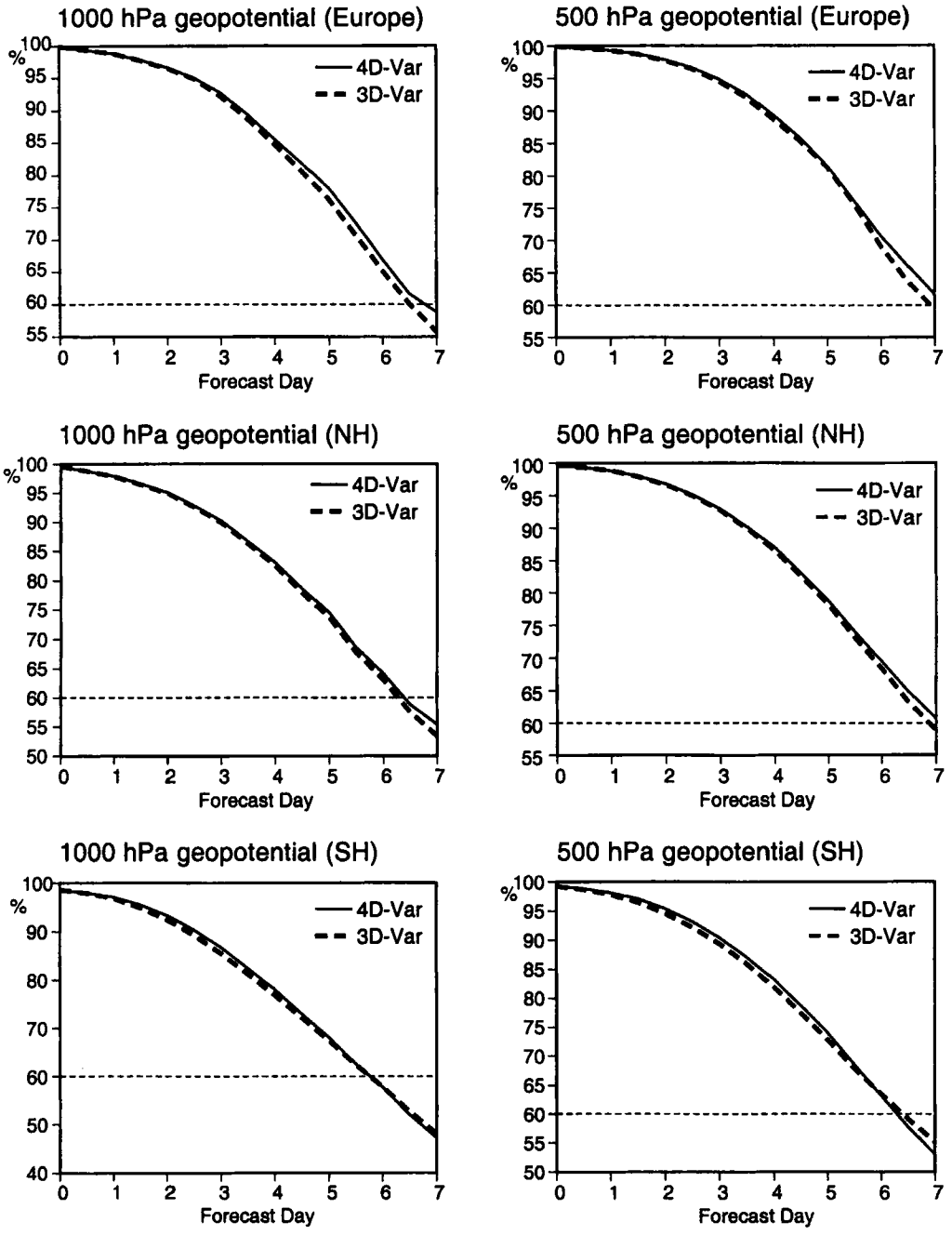

Figure 7. Anomaly correlation scores for forecasts from 4D-Var (solid line) and from 3D-Var (dashed line), averaged over 12 weeks. Scores are shown for geopotential height at 1000 hPa (left) and 500 hPa (right), for: Europe (top); the northern hemisphere (centre); and the southern hemisphere (bottom).

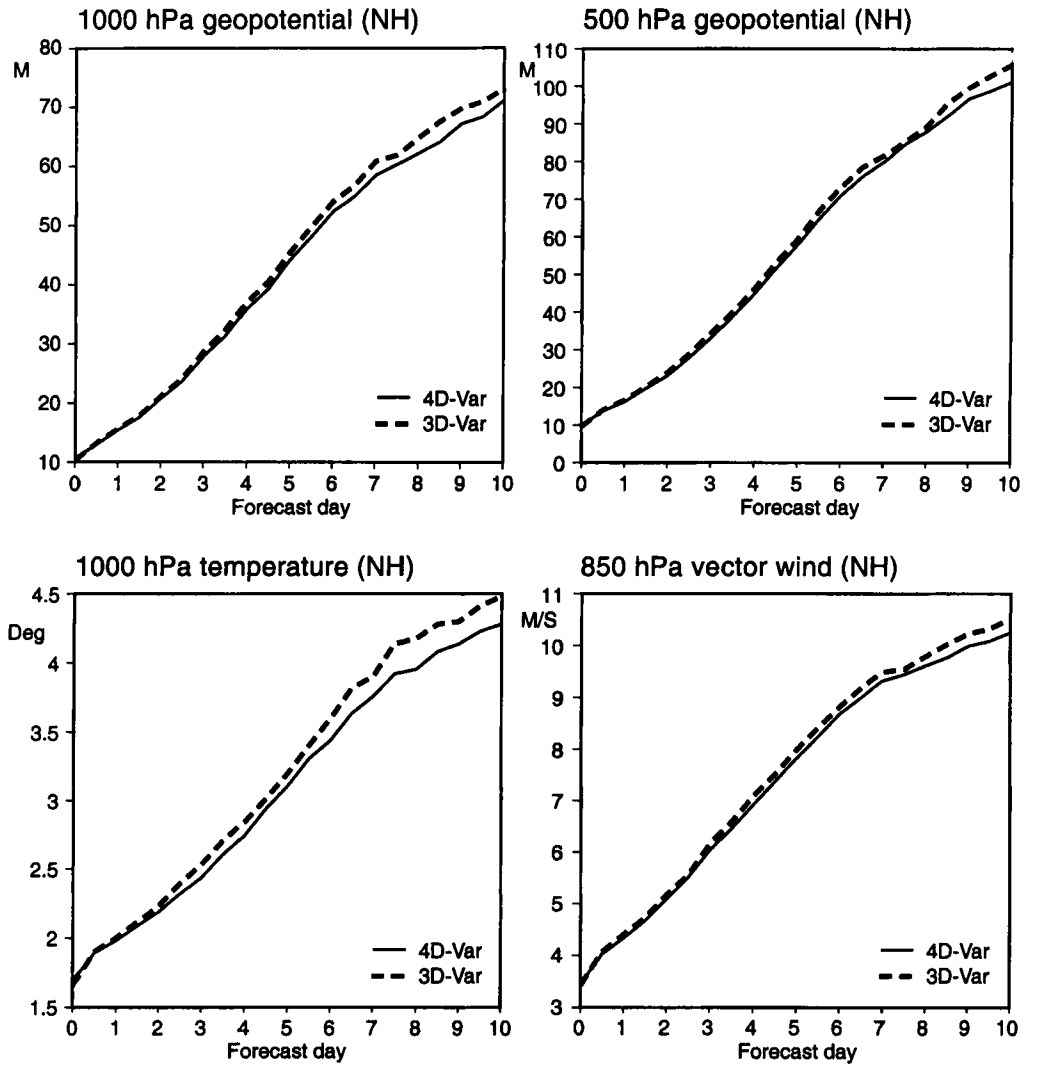

Figure 8. Root-mean-square differences between forecasts from 4D-Var and observations (solid lines) and forecasts from 3D-Var and observations (dashed lines), averaged over 12 weeks in the northern hemisphere for: geopotential height at 1000 hPa (m, top left); and 500 hPa (m, top right); temperature at 1000 hPa (degC, bottom left); vector wind at 850 hPa (m s^{-1}, bottom right).

(*b*) *Results over the February 1997 period*

The very short-range forecasts can be investigated by computing the r.m.s. fit of the background field to the observations for both systems. Figure 10 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 judging the quality of the assimilation system, as the short-range forecast errors and the observation errors are generally uncorrelated. For these (solid lines) 4D-Var clearly outperforms 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, a Fisher's test at 90% confidence level, was also carried

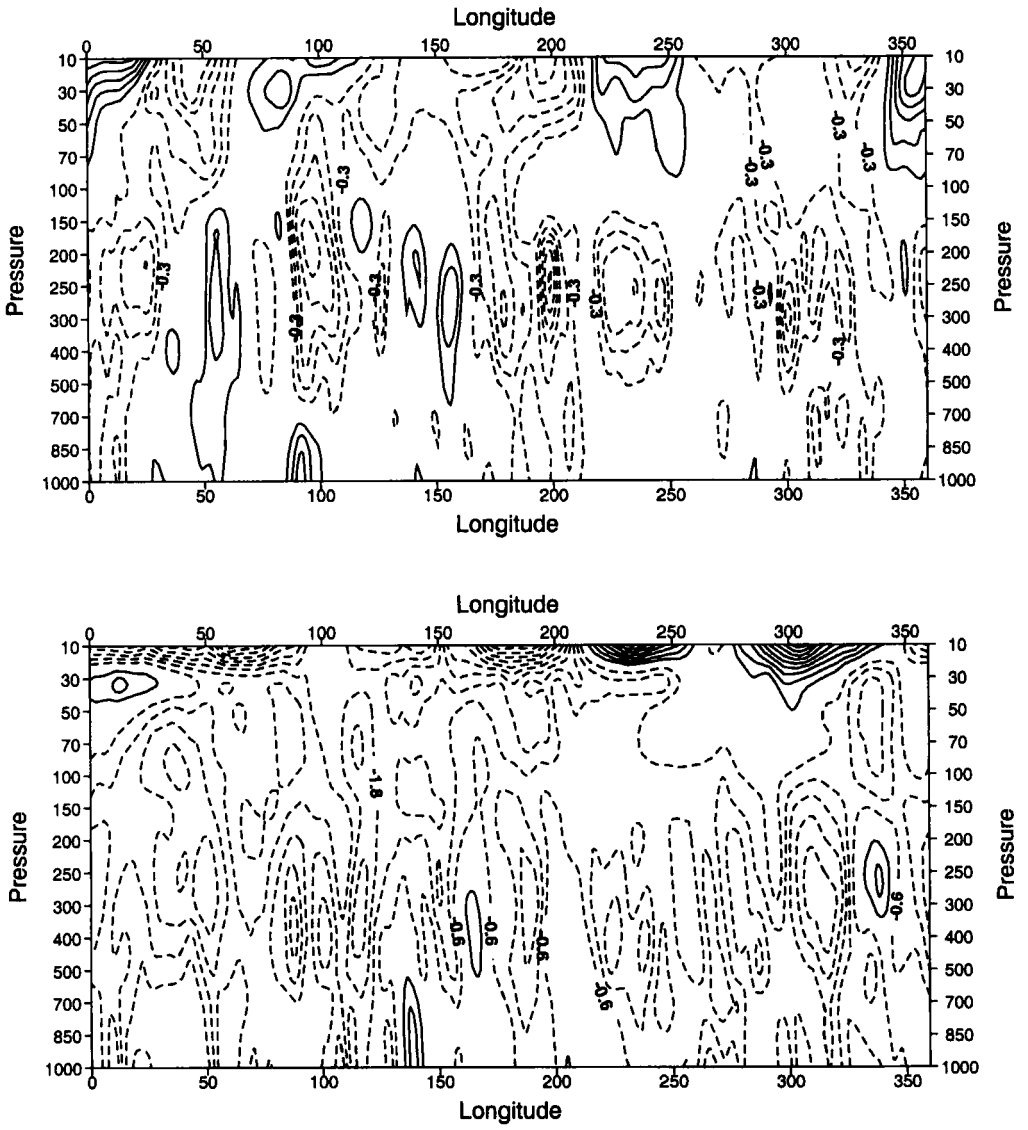

Figure 9. Cross-sections of the difference between the root-mean-square (r.m.s.) error of forecasts from 4D-Var and from 3D-Var at day 1, averaged over 12 weeks and over the 40° to 70° latitude band. Contour interval is 0.3 m for the top panel representing the northern hemisphere and 0.6 m for the bottom panel representing the southern hemisphere. Negative contours (dashed) mean that 4D-Var has a smaller r.m.s. error.

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, are 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. 4D-Var analysis increments are generally smaller than those of 3D-Var, which is consistent with an improved short-range forecast.

This period of February 1997 is particularly interesting because of FASTEX (Joly *et al.* 1997) which took place in the Atlantic Ocean storm-track. In the remainder of

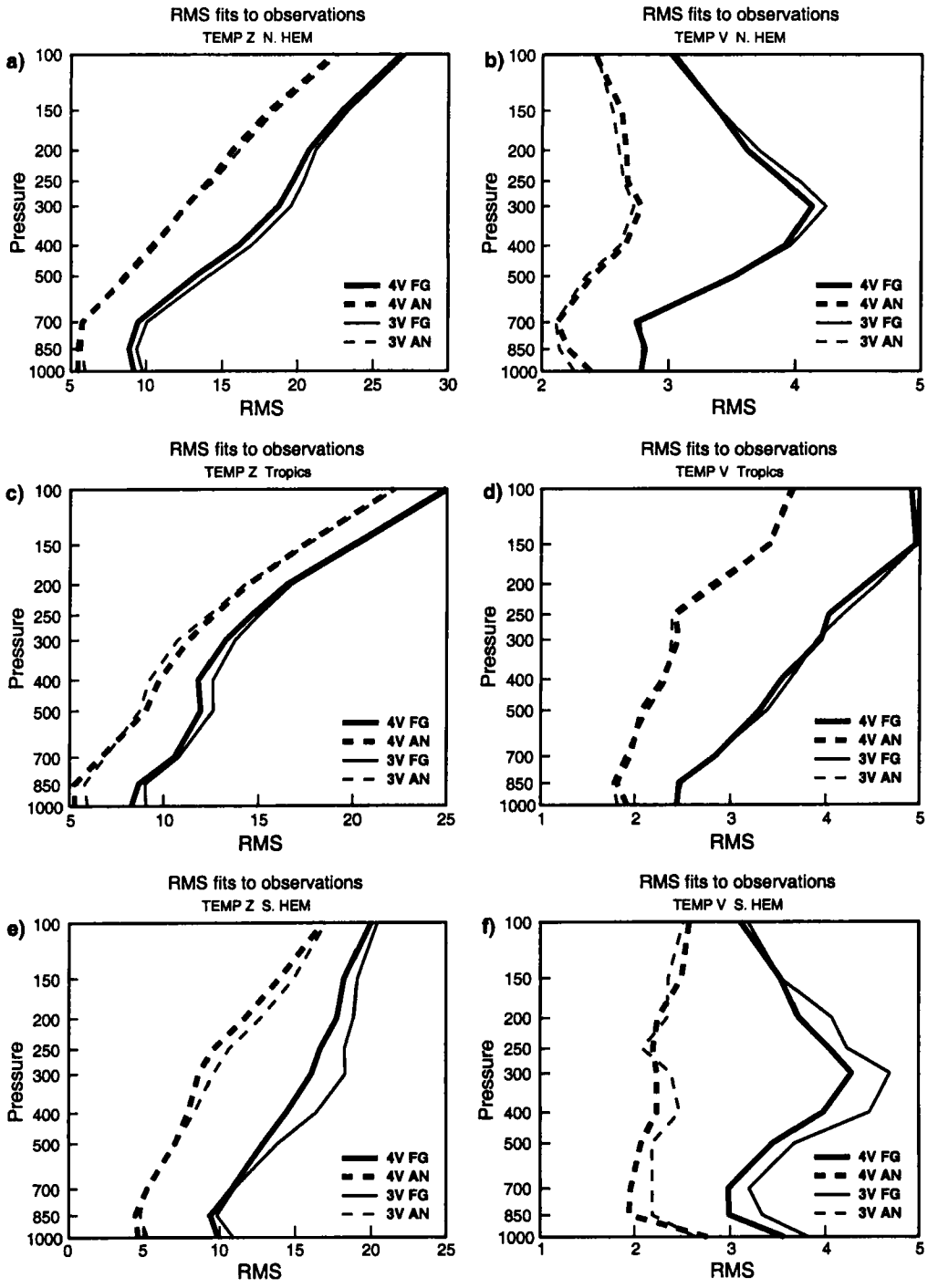

Figure 10. Root-mean-square (RMS) fits to the radiosonde height (TEMP Z) and meridional component of the wind data (TEMP V) produced over the northern hemisphere (top), the Tropics (centre) and the southern hemisphere (bottom), averaged over 1 to 14 February 1997. The solid lines represent the RMS fits of the backgrounds to the observations, and the dashed lines the RMS fits of the analyses to the observations. 4D-Var is shown as thick and 3D-Var as thin lines. The abscissa is the RMS in geopotential units and m s^{-1}; the ordinate is the pressure in hPa.

this section, we will concentrate on the discussion of the performance of 4D-Var in the Atlantic area over this period. Both 3D-Var and 4D-Var used the same observations: on top of the standard set of observations, the extra FASTEX radiosonde measurements have been taken into account. As far as synoptic cases are concerned, two examples are shown in Figs. 11 and 12. They correspond to interesting dates during FASTEX (Joly *et al.* 1997). The first synoptic case is the Intensive Observing Period (IOP) 12 during which a cyclone evolved in the vicinity of Iceland, with a deepening of 19 hPa in 24 hours seen in the Met. Office manual analysis of surface pressure: the surface pressure fell from 966 hPa at 1200 UTC 9 February to 947 hPa at 1200 UTC 10 February 1997. The deepening in the 24-hour forecast from the 4D-Var analysis is very similar to that analysed, the pressure dropping from 970 hPa to 951 hPa in the same period. 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 Fig. 11, together with the verifying analyses. The second synoptic case (IOP17) is a depression, with a manually analysed central pressure of 937 hPa on 20 February 1997 at 0000 UTC. In the medium-range, the forecast from 1200 UTC 12 February 1997 is significantly better using 4D-Var, as shown in Fig. 12 by 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 quality of the 4D-Var forecasts for this particular event, we computed r.m.s. errors of the intensity of the low, averaged over all forecasts from 11th to the 19th verifying at 0000 UTC 20 February 1997. These values are 15 hPa for 4D-Var, versus 19 hPa for 3D-Var; this corresponds to a 20% improvement. One can also notice a better analysed surface pressure minimum with 4D-Var (940 hPa versus 943 hPa).

Over the whole period for which both 3D-Var and 4D-Var were run (1 to 21 February), 4D-Var performed better on average in the Atlantic area. Figure 13 represents the cross-sections of differences in r.m.s. error against its own analysis at day 1 between 4D-Var and 3D-Var for the 40° to 70° mid-latitude band over the Atlantic area (310–350°E). The values are predominantly negative throughout the atmospheric depth for both height (top panel) and zonal component of the wind (bottom panel), which means that 4D-Var has a smaller error than 3D-Var. The largest differences can be seen in the higher troposphere and lower stratosphere, with a maximum around 300 hPa (this is also where the largest improvement was found in terms of temperature and meridional component of the wind, not shown). In terms of values of 500 hPa geopotential scores over the Atlantic area, at the 12-hour range the r.m.s. errors are 10.8 m and 8.8 m for 3D-Var and 4D-Var, respectively, while at the 24-hour range they are 13.2 m and 11.9 m. The same scores computed over the European area for slightly longer ranges are, consequently, improved by 4D-Var: at the 24-hour range, the r.m.s. errors are 14.3 m and 13.5 m for 3D-Var and 4D-Var, respectively, while at the 48-hour range they are 27.3 m and 25.6 m. The better results for 4D-Var are also confirmed when comparing forecasts with available observations (not shown).

(*c*) *Analysis diagnostics over the February period*

The Atlantic ocean is a particularly 'sensitive area' over this period as indicated by the operational 'key analysis errors'. As explained in Klinker *et al.* (1998), these are obtained by finding some increments to the analysis which would significantly reduce the two-day error of the ensuing forecast. They usually highlight the areas which are both dynamically unstable and not so well-resolved by observations. In this case, the key

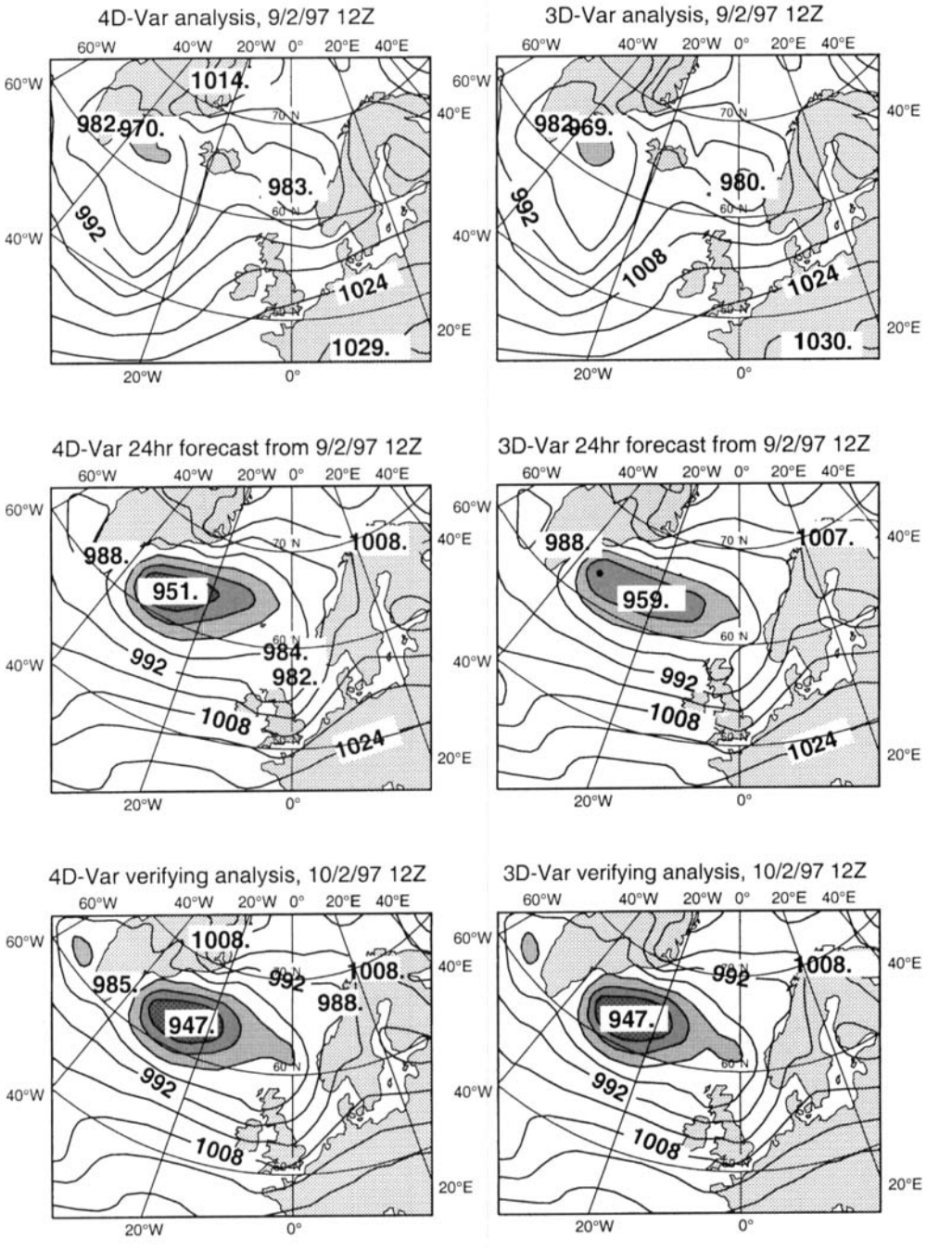

Figure 11. Maps of mean-sea-level pressure (hPa) for the synoptic case corresponding to IOP12 of FASTEX (see text). The top panels show the analyses for 1200 UTC 9 February 1997; the middle panels the 24-hour forecasts from 1200 UTC 9 February 1997; and the bottom panels the verifying analyses. 4D-Var charts are on the left, 3D-Var on the right.

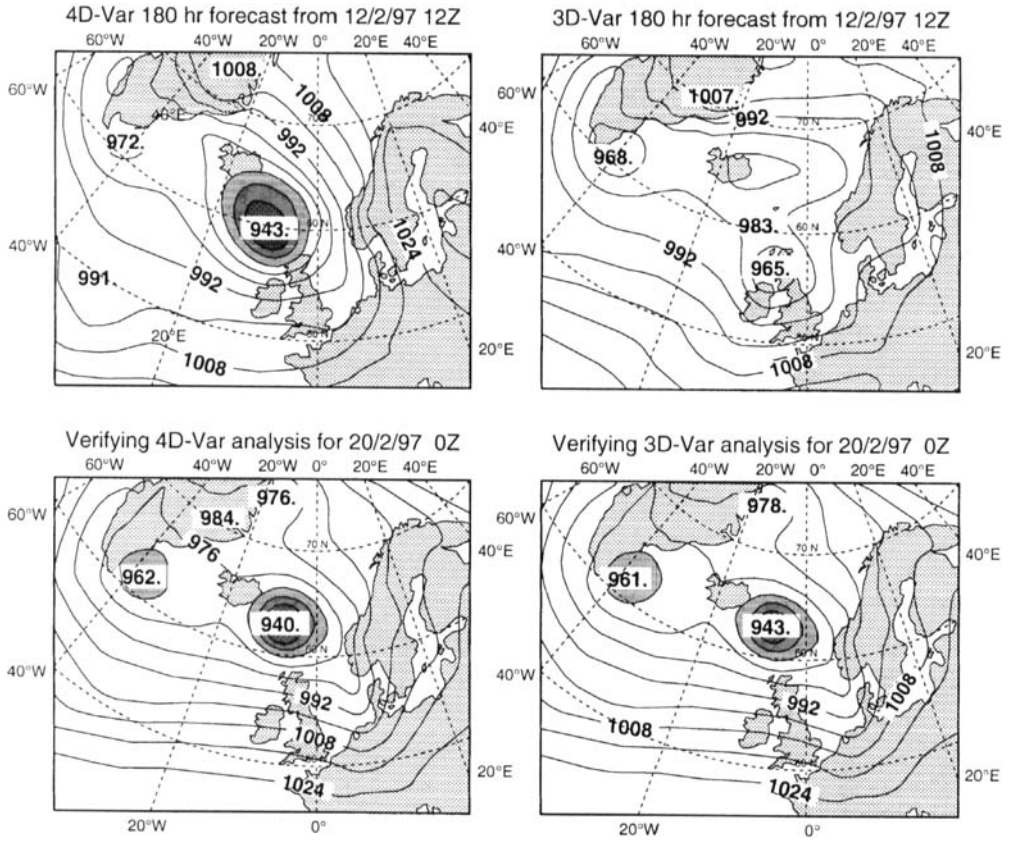

Figure 12. Maps of mean-sea-level pressure (hPa) for the synoptic case corresponding to IOP17 of FASTEX (see text). The top panels show the forecasts from 1200 UTC 12 February 1997 verifying at 0000 UTC 20 February 1997; the bottom panels give the verifying analyses. 4D-Var charts are on the left, 3D-Var on the right.

analysis errors (not shown) have a relatively large amplitude of more than 2 m south-west of Iceland (around 20–40°W, and 50–70°N). One can then compute the difference between the 3D-Var and 4D-Var analyses in this area: the r.m.s. of the differences between the analyses is largest at around 20–30°W and 50–60°N (not shown). The analyses are mostly different at the same longitudes as the ones highlighted by the key analysis errors (20–40°W), although the maximum is shifted towards lower latitudes. The maximum differences between 3D-Var and 4D-Var analyses are around 15 m in terms of r.m.s. error, which is significantly larger than the corresponding 2 to 3 m values for the key analysis errors. This is perfectly understandable, as the difference between the two assimilation systems gives an estimate of the total uncertainty within the analysis, whereas the key analysis errors only describe the fast-growing part of this uncertainty. In any case, the two sets of analyses appear to be significantly different over an area known to be sensitive as regards the quality of the resulting forecasts.

For a given meteorological situation (9 February 1997 already studied above), 3D-Var and 4D-Var are compared to a 'sensitivity analysis'. This sensitivity analysis is obtained by subtracting the key analysis errors from the 4D-Var analysis, and is the best estimate of the truth at our disposal (it was shown in Klinker *et al.* (1998) that the sensitivity analysis can provide both a better fit to the data and a better ensuing forecast). The difference between 3D-Var and the sensitivity analysis on the

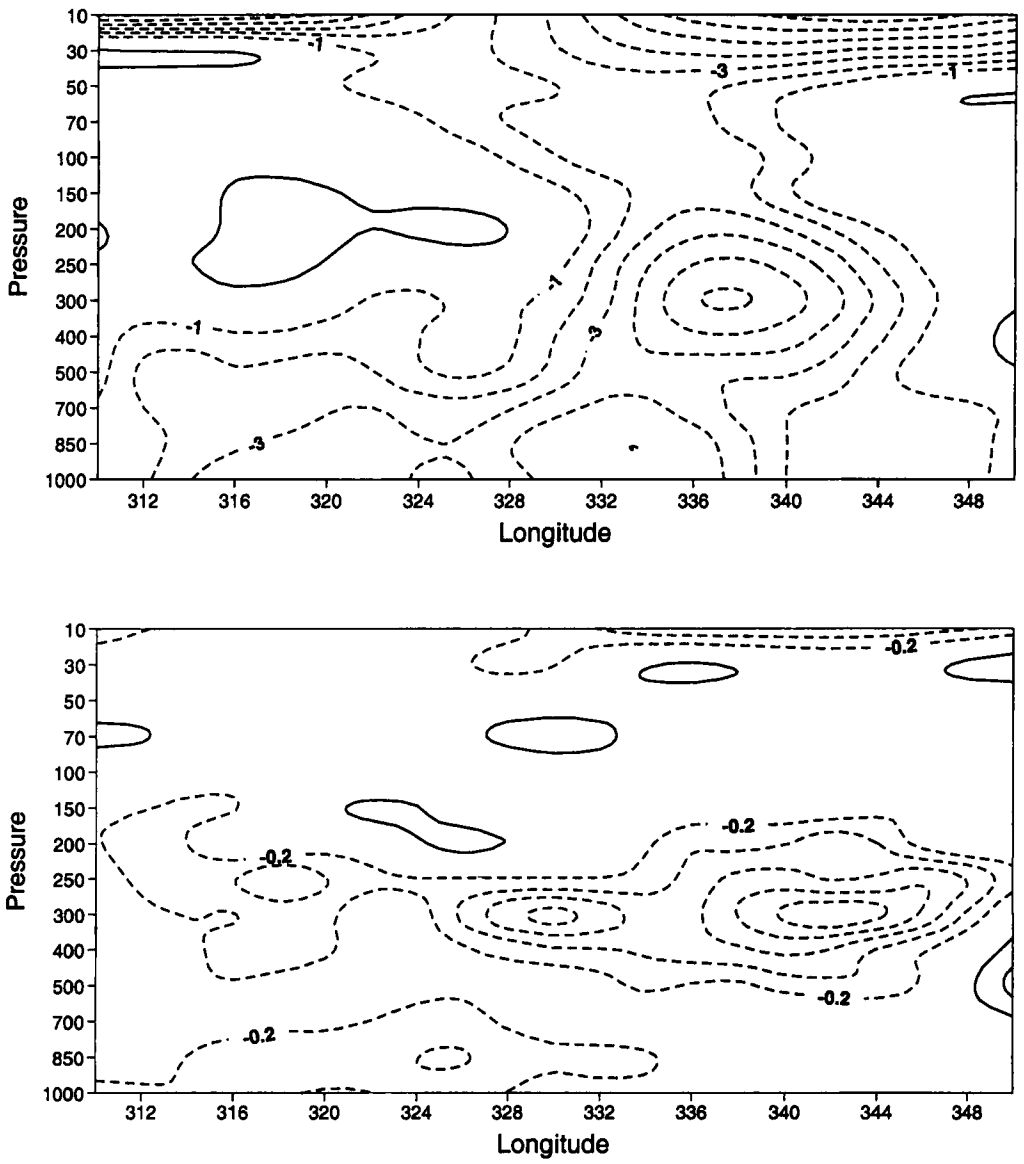

Figure 13. Cross-sections of the difference between the root-mean-square (r.m.s.) error of forecasts from 4D-Var and from 3D-Var at day 1, averaged over 3 weeks and over the 40° to 70° latitude band over the North Atlantic. Contour interval is 3 m for the top panel representing the geopotential height, and 0.2 m s^{-1} for the bottom panel representing the zonal component of the wind. Negative contours mean that 4D-Var has a smaller r.m.s. error.

one hand, and between 4D-Var and the sensitivity analysis on the other hand, are then considered as 3D-Var and 4D-Var errors. These errors are projected on the first singular vectors describing the unstable manifold (Buizza 1994). The first three singular vectors contributing to the Atlantic area (80–20°W; 30–90°N) are chosen. The energy norm of the 4D-Var and 3D-Var errors, projected onto this unstable manifold are 2.43 kg s^{-2} and 3.22 kg s^{-2}, respectively. The 4D-Var error is thus 25% smaller than the 3D-Var error on the unstable manifold defined by the first three singular vectors over the Atlantic area for this particular case. Of course, this computation is performed using only one

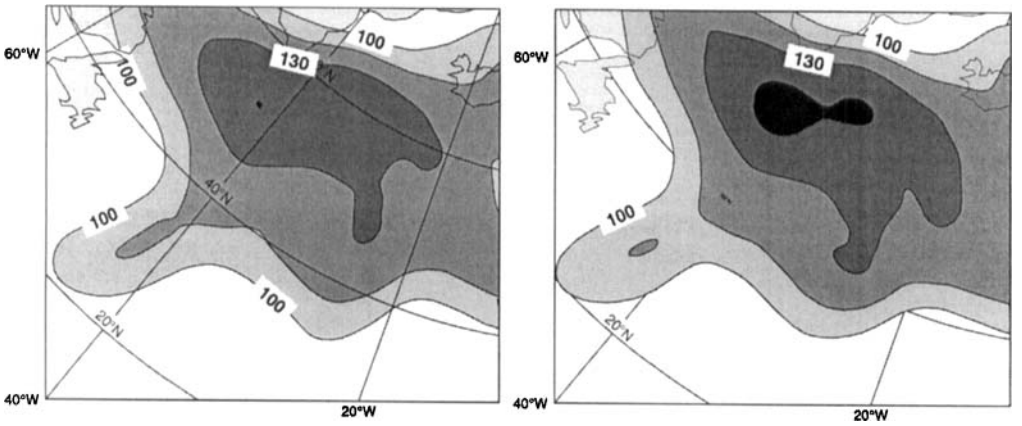

Figure 14. Maps of quantities related to the geopotential height at 500 hPa averaged over the period 1 to 21 February 1997 over the Atlantic Ocean. The left panel shows the standard-deviation of the 3D-Var analysis computed over all the 1200 UTC analyses of the period, and the right panel shows the standard-deviation of the 4D-Var analysis computed over all 1200 UTC analyses over the period. The units are geopotential metres. Contour intervals are 15 m starting at 100 m.

meteorological situation, and the choice of the 'truth' is not unique. However, the results indicate that 4D-Var is behaving better than 3D-Var on the unstable manifold. Another indication will be given in Part III (Klinker *et al.* 2000) in which key analysis errors are shown to be smaller than for 3D-Var over a 40-day parallel experimentation.

When decomposed into mean and standard-deviation components, the large values of the r.m.s. of the difference between 3D-Var and 4D-Var analyses are seen to come mainly from the standard-deviation part (not shown). In the mean difference, 4D-Var analyses exhibit slightly higher geopotential values, and a slight reduction of the Eady index estimating the instability of the flow (not shown). These differences seem not to be very well structured or significant. More importantly, one can compare the standard-deviations of analysis for both systems. These are shown in Fig. 14. The standard-deviation of the 4D-Var analysis (right panel) is quite noticeably larger than that for 3D-Var (left panel). Differences between the two can reach up to 6 m around 50°N, 30°W, where a typical value of standard-deviation of the analysis is 115 m. It thus corresponds to a 5% increase in the variability of the analysis.

The larger variability does not appear to be caused by larger increments at each analysis cycle. Similar to what was shown in a previous paper by Rabier *et al.* (1998b), in the Atlantic area 4D-Var actually produces smaller increments than 3D-Var (not shown). To clarify the use of observations in the area, the fit to the data was computed explicitly for the area 40–70°N and 50–10°W. The number of radiosondes in the area is slightly higher than normal for the period, due to the extra observations taken during FASTEX. In total, there are around 300 data at each pressure level entering the r.m.s. calculations illustrated in Fig. 15. Even though this is a limited sample, it is obvious from Fig. 15 that the background fields used in the 4D-Var assimilation fit the data much more closely than those used in the 3D-Var assimilation. In particular, at the 300 hPa level the r.m.s. fit to the height data is 17.6 m for 4D-Var versus 21.6 for 3D-Var, and the equivalent values for the meridional component of the wind are 4.3 m s^{-1} versus 5.4 m s^{-1}, respectively. The very short-range forecasts from 4D-Var agree much better with the synoptic observations than do the equivalent forecasts from 3D-Var. Even though the 4D-Var analysis produces smaller increments than 3D-Var, it can produce an analysis

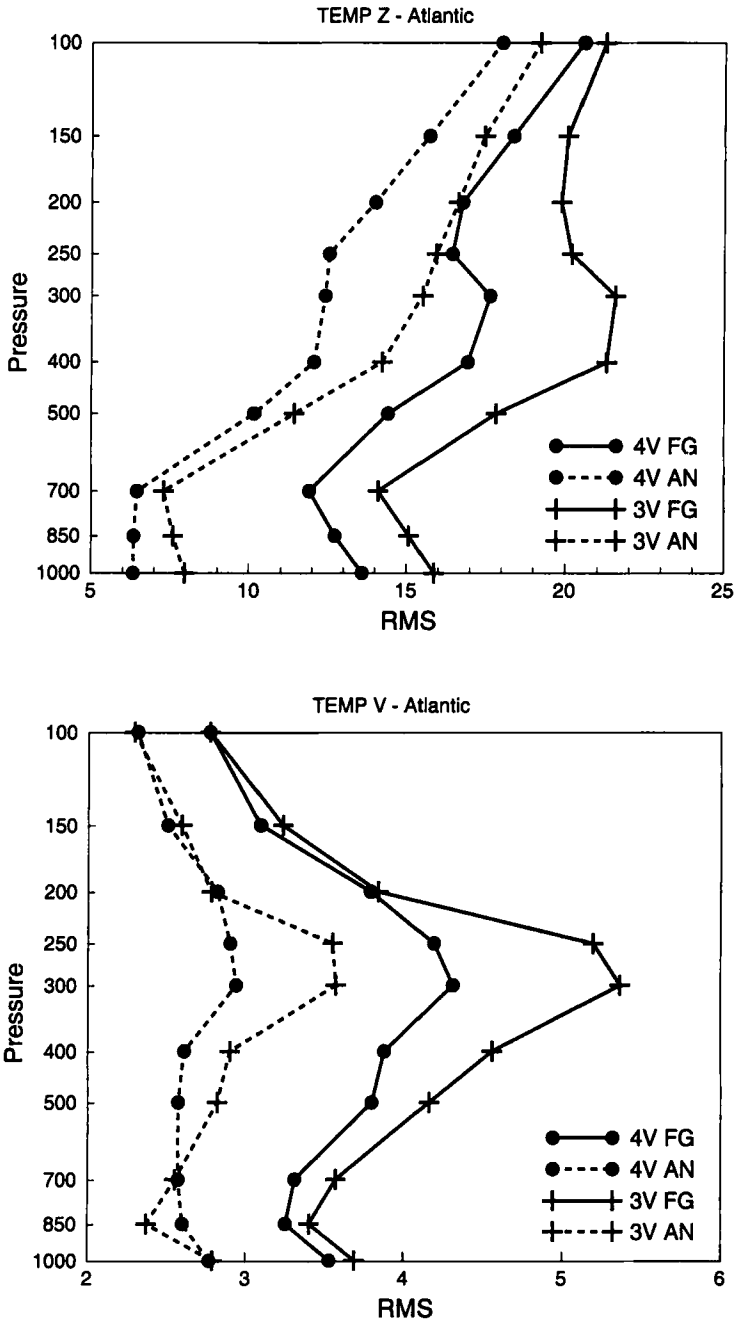

Figure 15. Root-mean-square (RMS) fits to the radiosonde height (TEMP Z) and meridional component of the wind data (TEMP V) produced over the Atlantic area, averaged over 1 to 21 February 1997. The solid lines represent the RMS fits of the backgrounds to the observations, the dashed lines the RMS fits of the analyses to the observations. 4D-Var is shown as circles, 3D-Var as plus signs. The abscissa is the RMS in geopotential units and $m\ s^{-1}$. The ordinate is the pressure in hPa.

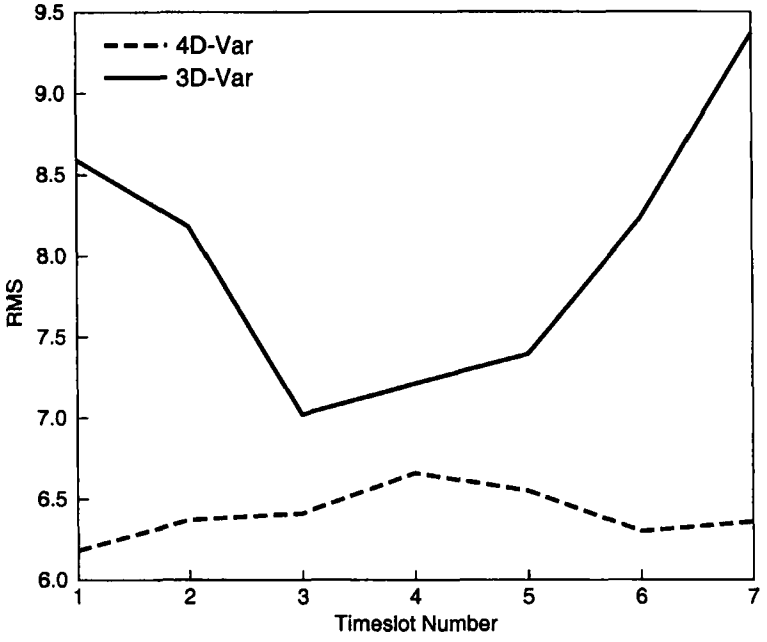

Figure 16. Root-mean-square (RMS) of the innovation vector wind for aircraft data produced over the Atlantic area, averaged over 1 to 21 February 1997, every 12 hours. 4D-Var is shown as a dashed line, 3D-Var as a solid line. The abscissa is the time slot number. The ordinate is the RMS in m s^{-1}.

of more variance, and presumably better quality. Comparing the background fields of both systems with asynoptic observations is more problematic, as 4D-Var backgrounds are compared with the observations at their valid time within the assimilation window, whereas 3D-Var backgrounds are compared with the observations gathered at the central analysis time only. This is illustrated for aircraft observations over the Atlantic area in Fig. 16. This figure represents the fit of the backgrounds to the data, in r.m.s. terms, averaged over all levels between 150 hPa and 400 hPa, by time slots. There are 7 time slots: the first and seventh are half-hour slots; while the remaining 5 are one-hour slots. Time slot number 4 is then centred around the main synoptic time. Data are accumulated every 12 hours, over the entire 21-day assimilation period. The average number of observations is 1000 per time slot. The r.m.s. fit to the data is shown as a solid line for 3D-Var and as a dashed line for 4D-Var. For 4D-Var, the fit is rather constant with respect to time slot number, around 6 to 6.5 m s^{-1} in vector wind. For 3D-Var, there is a convex curve, showing a better fit around the middle of the assimilation window (around 7 to 7.5 m s^{-1}) and a much degraded fit towards the sides of the assimilation window, reaching values around 9 m s^{-1}. At the central time, when the comparison between 3D-Var and 4D-Var is fair, one can note that the 4D-Var background is closer than the 3D-Var background to the observations, which is consistent with what has been observed with the radiosonde data. Obviously, the sample is not large enough to obtain smooth curves, but Fig. 16 shows quite clearly how the innovation vector can be degraded if the observations are not used at their appropriate time. Although this difference in the innovation vectors is quite spectacular, it is presented mainly to illustrate the importance of dynamical processes in a 6-hour window. It is not believed that the main improvement brought about is only due to a better computation of the innovation vector, but primarily to using the tangent-linear dynamics in evolving the increments. However, this is indeed

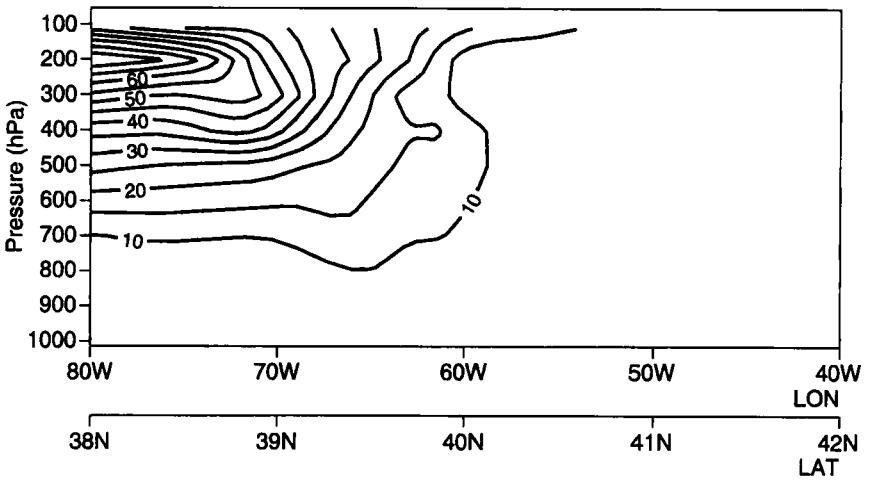

Figure 17. Cross-section of the zonal wind component (m s^{-1}) of the background for 0000 UTC 5 December 1996. See text for details.

a particular area where we expect an improvement from 4D-Var: the proper use of asynoptic data.

4. STRUCTURE FUNCTIONS IN 4D-VAR

The results presented above point to a better short-range forecast from 4D-Var than from 3D-Var in the Atlantic area during February 1997, probably implying a better analysis. In order to understand the better behaviour of 4D-Var in cyclogenetic situations, it is useful to illustrate the impact of the dynamics on the increments caused by one observation in this sort of meteorological situation. Structure functions can easily be illustrated by single-observation experiments, as in Thépaut *et al.* (1996). For a particular date (0000 UTC 5 December 1996), a baroclinic area was chosen in the West Atlantic. 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 Fig. 17 in a cross-section of the zonal wind component. One observation of geopotential height was inserted at location 60°W, 40°N, at 850 hPa in 4D-Var, either at 2100 UTC (3 hours before the main synoptic time 0000 UTC), or at 0000 UTC, or at 0300 UTC (3 hours after the main synoptic time).

Each time, the initial departure from the background is equal to 10 m. The structure functions can, therefore, be illustrated for three different scenarios: an observation at the beginning of the assimilation window (2100 UTC), in the middle (0000 UTC) and at the end (0300 UTC). These are shown at the time of the observation in Fig. 18 (which shows cross-sections along the same line as Fig. 17). This particular meteorological situation is not rapidly developing, which allows comparison of the increments even if their validity time can be up to 6 hours different. The top panel in Fig. 18 corresponds to the increments at 2100 UTC created from an observation at 2100 UTC. 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. 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 Fig. 18 are tilted in the vertical, in a way consistent

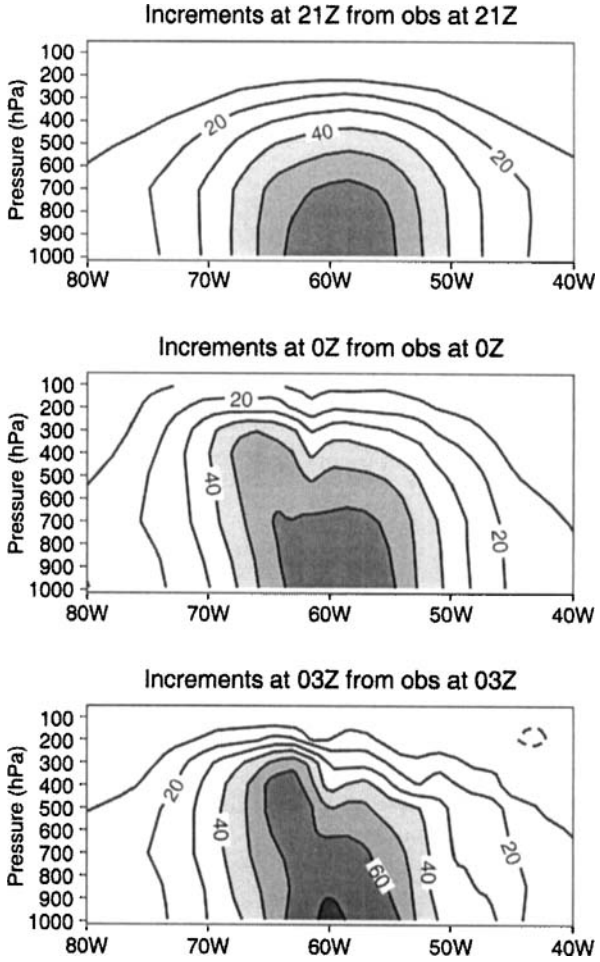

Figure 18. Structure functions for a height observation at 850 hPa, 40°N, 60°W. Isolines show the resulting increment, in geopotential units. The top panel corresponds to an observation at 2100 UTC, the middle panel to an observation at 0000 UTC, and the bottom panel to an observation at 0300 UTC.

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, 0000 UTC, from these three configurations. These are shown in Fig. 19. Although Fig. 18 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 alike. 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. However, the structure functions illustrated in Fig. 18 still appear to be constrained to a large extent by the initial static covariance matrix illustrated in the top panel. This indicates that 4D-Var will still benefit from a better specification of the initial covariance matrix, by a Kalman filter for example.

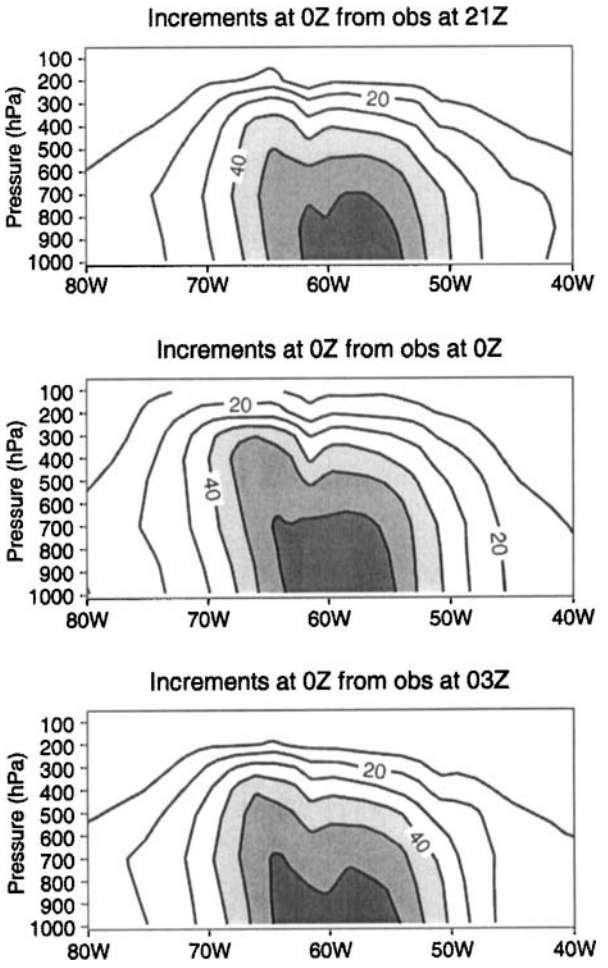

Figure 19. Increments using 4D-Var at 0000 UTC for a height observation at 850 hPa, 40°N, 60°W. Isolines show the resulting increment, in geopotential units. The top panel corresponds to an observation at 2100 UTC, the middle panel to an observation at 0000 UTC, and the bottom panel to an observation at 0300 UTC.

5. SUMMARY AND DISCUSSION

This article is the first of three describing the steps leading to the operational implementation of incremental 4D-Var at ECMWF on 25 November 1997. This first paper concentrates on results produced with 4D-Var on a 6-hour window at 'operational' resolution T213L31/T63L31, using very simplified physics during the minimization. Firstly, the sensitivity to different 4D-Var set-ups is investigated. It is found that using additional asynoptic data is not useful at the present stage. Later results (Järvinen *et al.* 1999) showed that the assimilation of observations from frequently reporting surface stations, with serial observation error correlation and joint quality control, resulted in a small but systematic increase in the short-range forecast accuracy in 4D-Var. These developments were not available for operational implementation but were part of later improvements to the system. Investigation of the poor performance of 4D-Var in the Tropics revealed some sensitivity to the way the adiabatic nonlinear 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 1997 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 one outer-loop in 4D-Var and with the 1997 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, computed either with respect to a verifying analysis or with respect to observations. In the short range, each two- to three-week period has been found to be slightly positive throughout the troposphere. The largest improvement is seen to come from the southern hemisphere as a whole and from the northern hemispheric mid-latitude oceanic areas. The better short-range performance of the 4D-Var system is also shown by the fits of the background fields to the data. Detailed results were presented over the Atlantic Ocean area during the FASTEX period, when 4D-Var is found to perform better. In individual synoptic cases corresponding to interesting IOPs during FASTEX, 4D-Var has a clear advantage over 3D-Var during rapid cyclogeneses. This advantage does not appear to stem from differences in mean analysis fields, but rather from relatively large day-to-day differences between the 4D-Var and 3D-Var analyses in a sensitive area. In one IOP, 4D-Var behaves better on the unstable manifold than 3D-Var. 4D-Var analyses display more day-to-day variability than their 3D-Var counterparts. This is believed to be an improvement as, in general, the 4D-Var system does not produce larger analysis increments than the 3D-Var system. The very short-range forecasts used as backgrounds are much closer to the data over the Atlantic area for 4D-Var than for 3D-Var, which explains why smaller increments can lead to more variable analyses. The small adjustments made by 4D-Var at each analysis cycle are sufficient to depict accurately the rapid evolution of the flow in this storm-track area.

Structure functions were illustrated in the 4D-Var case for a height observation inserted either at the beginning of the assimilation window, or in the middle or at the end. As anticipated from previous results, the dynamical processes seem to be relevant, even in a short 6-hour assimilation period. More influence of the dynamics could be taken into account by a proper cycling of 4D-Var using a simplified Kalman filter which is currently being developed (Fisher 1998). To be cycled in a cost-effective way, the simplified Kalman filter would need longer 4D-Var windows, which is an option to be studied in the near future.

As far as cost issues are concerned, the main analysis computing task using 4D-Var with only one outer-loop and no physics (our baseline 4D-Var) costs about twice as much as that for 3D-Var. The additional costs of 4D-Var are mainly due to the extra model integrations which run efficiently. 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, in line with expectations.

Although this paper has emphasized the role of the dynamics in 4D-Var results, there is scope for improvement in both the extratropics and the Tropics by including more elaborate physical processes within the 4D-Var minimization. Part II (Mahfouf and Rabier 2000) describes a set of physical processes developed, tested, and included in the 4D-Var experimentation. The resulting 4D-Var system was then tested over 80 days of experimentation with positive results. An additional 6-week parallel suite prior to implementation clearly showed the benefit of 4D-Var, with significant improvements with respect to 3D-Var. They are described extensively in Part III (Klinker *et al.* 2000).

ACKNOWLEDGEMENTS

It is a pleasure to acknowledge all the colleagues at ECMWF who have contributed to this 4D-Var experimentation. This includes all the staff of the data assimilation section (P. Undén, Erik Andersson, F. Bouttier, M. Fisher, J. Haseler and D. Vasiljevic), of the migration project (M. Hamrud, L. Isaksen and S. Saarinen) as well as of the satellite section (R. Saunders, G. Kelly, B. Harris, D. Lemeur, T. McNally). The success of the 4D-Var project relies heavily on its history and in particular on the contributions by P. Courtier and J.-N. Thépaut. A. Hollingsworth has been extremely supportive, providing us with all the resources and incentive to achieve this work. C. Temperton is acknowledged for his co-ordination of the project.

The minimization package used is provided by the Institut National de Recherche en Informatique et Automatique (INRIA, France). It is described in Gilbert and Lemaréchal (1989).

REFERENCES

Andersson, E. 1996 'Implementation of variational quality control'. In Proceedings of the ECMWF workshop on nonlinear aspects of data assimilation, 9–11 September 1996. Available from ECMWF, Shinfield Park, Reading, UK

Andersson, E. and Järvinen, H. 1999 Variational quality control. *Q. J. R. Meteorol. Soc.*, **125,** 697–722

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). III: Experimental results. *Q. J. R. Meteorol. Soc.*, **124,** 1831–1860

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

Courtier, P. and Talagrand, O. 1987 Variational assimilation of meteorological observations with the adjoint vorticity equation. 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). I: Formulation. *Q. J. R. Meteorol. Soc.*, **124,** 1783–1807

Derber, J. and Bouttier, F. 1999 A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51A,** 195–221

Fisher, M. 1998 'Development of a simplified Kalman filter'. ECMWF Research Department Technical Memorandum No. 260. Available from ECMWF, Shinfield Park, Reading, UK

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. Available from ECMWF, Shinfield Park, Reading, UK

Gilbert, J.-C. and Lemaréchal, C. 1989 Some numerical experiments with variable-storage quasi-Newton algorithms. *Math. Programming*, **B25,** 407–435

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. Available from ECMWF, Shinfield Park, Reading, UK

Järvinen, H., Andersson, E. and Bouttier, F.	1999	Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus*, **51A,** 469–488
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. and Wakimoto, R. M.	1997	Definition of the Fronts and Atlantic Storm-Track Experiment (FASTEX). *Bull. Am. Meteorol. Soc.*, **78,** 1917–1940
Klinker, E., Rabier, F. and Gelaro, R.	1998	Estimation of key analysis errors using the adjoint technique. *Q. J. R. Meteorol. Soc.*, **24,** 1909–1934
Klinker, E., Rabier, F., Kelly, G. and Mahfouf, J.-F.	2000	The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration. *Q. J. R. Meteorol. Soc.*, **126,** 1191–1215
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
Mahfouf, J.-F. and Rabier, F.	2000	The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics. *Q. J. R. Meteorol. Soc.*, **126,** 1171–1190
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. Weather 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., McNally, A., Andersson, E., Courtier, P., Undén, P., Eyre, J., Hollingsworth, A. and Bouttier, F.	1998a	The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Q. J. R. Meteorol. Soc.*, **124,** 1809–1829
Rabier, F., Thépaut, J.-N. and Courtier, P.	1998b	Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Q. J. R. Meteorol. Soc.*, **124,** 1861–1887
Simmons, A. and Rabier, F.	1997	'Removal of the incremental initialization of medium and large scales in the ECMWF analysis'. ECMWF Research Department Memorandum, May 1997. Available from ECMWF, Shinfield Park, Reading, UK
Talagrand, O. and Courtier, P.	1987	Variational assimilation of meteorological observations with the adjoint vorticity equation. 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 assimilation: a case-study. *Q. J. R. Meteorol. Soc.*, **122,** 535–561
Zupanski, M.	1993	Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Weather Rev.*, **121,** 2396–2408