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

By J.-F. MAHFOUF* and F. RABIERT

European Centre for Medium-Range Weather Forecasts, UK

(Received 17 November 1998; revised 1 July 1999)

SUMMARY

A comprehensive set of physical parametrizations has been linearized for use in the European Centre for Medium-Range Weather Forecasts (ECMWF)'s incremental four-dimensional variational (4D-Var) system described in Part I. The following processes are represented: vertical diffusion, subgrid-scale orographic effects, large-scale precipitation, deep moist convection and long-wave radiation. The tangent-linear approximation is examined for finite-size perturbations. Significant improvements are illustrated for surface wind and specific humidity with respect to a simplified vertical diffusion scheme. Singular vectors computed over 6 hours (compatible with the 4D-Var assimilation window) have lower amplification rates when the improved physical package is included, due to a more realistic description of dissipative processes, even though latent-heat release contributes to amplify the potential energy of perturbations in rainy areas. A direct consequence is a larger value of the observation term of the cost-function at the end of the minimization process when improved physics is included in 4D-Var. However, the larger departure of the analysis state from observations in the lower-resolution inner-loop is in better agreement with the behaviour of the full nonlinear model at high resolution. More precisely, the improved physics produces smaller discontinuities in the value of the cost-function when going from low to high resolution. In order to reduce the computational cost of the linear physics, a new configuration of the incremental 4D-Var system using two outer-loops is defined. In a first outer-loop, a minimization is performed at low resolution with simplified physics (50 iterations), while in the second loop a second minimization is performed with improved physics (20 iterations) after an update of the model trajectory at high resolution. In this configuration the extra cost of the physics is only 25%, and results from a 2-week assimilation period show positive impacts in terms of quality of the forecasts in the Tropics (reduced spin-down of precipitation, lower root-mean-square errors in wind scores). This 4D-Var configuration with improved physics and two outer-loops was implemented operationally at ECMWF in November 1997.

KEYWORDS: Physical parametrizations Tangent-linear models Variational assimilation methods

1. INTRODUCTION

In the first of this series of three papers, preliminary results from the first pre-operational four-dimensional variational (4D-Var) assimilation at ECMWF (European Centre for Medium-Range Weather Forecasts) were described (Rabier et al. 2000). A first configuration, defined as the baseline 4D-Var system, with almost adiabatic tangent-linear and adjoint models was compared to the ECMWF operational 3D-Var assimilation system (Courtier et al. 1998; Rabier et al. 1998b; Andersson et al. 1998), and showed a better overall performance of the 4D-Var system. An incremental approach, as proposed by Courtier et al. (1994), had been chosen with one outer-loop and an 'operational' resolution T213L31/T63L31. (The following convention is adopted: T213L31 is the resolution of the model trajectory and T63L31 is the resolution of the inner-loop.) The assimilation window of the baseline 4D-Var system is 6 hours. The central part of the 4D-Var assimilation is the adjoint (transpose of the tangent-linear version) of a Numerical Weather Prediction (NWP) model. This Part focuses on improving the tangent-linear and adjoint versions of the ECMWF model through the inclusion of physical parametrization schemes.

Feasibility studies on 4D-Var data assimilation were first performed with highly simplified barotropic models (Courtier and Talagrand 1987, 1990) and gradually moved towards adiabatic primitive-equation models (Thépaut and Courtier 1991; Errico and

* Corresponding author: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK. e-mail: mahfouf@ecmwf.int
† T213L31 is spectral triangular truncation 213 with 31 levels in the vertical.
Vukičević 1992). The next improvement would be the inclusion of subgrid-scale physical processes. For example, Rabier et al. (1993) performed a 4D-Var assimilation with an almost adiabatic version of the ECMWF model for an explosive cyclogenesis situation (15 October 1987 storm). They showed that the lack of condensation processes was preventing 4D-Var from providing a realistic humidity analysis. Linearized physical processes are expected to be important near the surface through momentum friction, and in tropical regions where the large-scale circulation is mostly driven by diabatic heating. Introducing physical processes in linearized NWP models is not trivial since most of them are strongly nonlinear (NL) and include on/off switches.

The first results on the usefulness of linearized physical processes were presented during the 1st Workshop on adjoint applications in dynamic meteorology held in August 1992 in California (Zou et al. 1993; Zupanski 1993; Vukičević and Errico 1993; Bao and Warner 1993). In the context of incremental 4D-Var proposed by Courtier et al. (1994), the need for linearized physics was not a priority. By performing the minimization in terms of increments, simplified tangent-linear and adjoint versions of a NWP model can be used to provide an acceptable solution of the variational problem. Indeed, the short-range evolution of analysis increments in mid latitudes is well described by adiabatic low-resolution models.

The incremental approach presented in Part I uses a low spectral resolution model (T63) with a simple vertical diffusion scheme for the minimization, and a high spectral resolution model (T213) with a full physical package for the comparison with observations. The simplified vertical diffusion scheme was introduced by Buizza (1994) for the computation of singular vectors, in order to avoid non-meteorological structures growing near the surface. The need for physical processes in the incremental 4D-Var was implicitly demonstrated by some results of Rabier et al. (1998a) where different assimilation windows were compared. Results appeared to be better for 6-hour and 12-hour assimilation windows than for a 24-hour period. The validity of the tangent-linear approximation is certainly reduced with an adiabatic model, going from 6 hours to 24 hours. The implicit assumption of 4D-Var, that the NWP model is perfect, also explains such degradation. Rabier et al. (1998a) also pointed out problems in the Tropics and in mountainous areas with the 6-hour 4D-Var, due partly to a lack of physics in the linearized versions of the ECMWF model. Therefore, even over a 6-hour window assimilation, it is reasonable to expect some improvements of the incremental 4D-Var through a better description of physical processes. Moreover, these processes will gradually become more important when longer assimilation windows and higher spatial resolutions are considered. Besides, the future assimilation of new satellite products related to the hydrological cycle (cloudy radiiances, rainfall rates, liquid- and ice-water contents, etc.) in the 4D-Var framework, although theoretically possible, will require the development of new linearized observation operators provided by subgrid-scale physical parametrizations. For operational applications, a trade-off is necessary between the high computational cost of the linearized physics and the benefit it can provide in terms of quality of analyses and forecasts.

Section 2 briefly describes the methodology adopted for developing the first set of linearized physical parametrizations in the ECMWF model. Section 3 examines how the linear physics actually improves the tangent-linear approximation for propagating analysis increments over a 6-hour window. Singular vectors produced using an almost adiabatic model are compared to those produced with the linearized physics in section 4. Section 5 presents preliminary 4D-Var experiments over a few assimilation cycles, designed to define a first optimal configuration with linearized physics. Results from a 4D-Var incremental assimilation with two outer-loops, including physical processes in
the second minimization, are compared in section 6 to an almost adiabatic configuration of 4D-Var over a two-week period. The main results are summarized in section 7.

2. LINEARIZATION OF PHYSICAL PARAMETRIZATION SCHEMES

The incremental approach allows for a progressive inclusion of physical processes in 4D-Var, and for some simplifications with respect to the NL physical processes. Differences between the inner- and outer-loops of the baseline 4D-Var described in Rabier et al. (2000) concern both horizontal resolution and physical processes. In order to reduce such inconsistencies, several updates of the trajectory during the minimization can be performed, but it is also possible to improve the description of the physics in the inner-loop. The methodology defined to introduce linearized physical parametrization schemes in the ECMWF 4D-Var is explained below.

We developed a set of simplified linearized physical parametrizations that represent the most important diabatic sources and sinks taking place in the atmosphere, given the time-scale of data assimilation (one day). For both technical and scientific reasons explained in Mahfouf (1999), it was not possible to develop the complete linearized versions of the full NL physical package described in the ECMWF operational forecast model. Physical processes are characterized by strong nonlinearities, as transitions from unstable to stable regimes in the boundary layer, and on/off switches (e.g. condensation takes place when humidity is above saturation). This means that the range of validity of the tangent-linear approximation can be reduced by including physical parametrizations in linearized models. Therefore, the potential of linearized physical parametrizations has to be assessed first, at least for some processes, before performing a comprehensive linearization. To reduce the validation part of the simplified parametrizations within the NL model, previous operational schemes used in the ECMWF model have been chosen.

The package of linearized physical parametrizations developed for the ECMWF model has been described in detail previously by Mahfouf (1999). The following processes are represented: vertical diffusion by turbulence in the boundary layer; momentum dissipation by subgrid-scale orographic effects; large-scale precipitation; vertical transport by mass-flux cumulus convection and long-wave radiative cooling. In consequence, the main feedback loops between the processes are described, except for cloud-radiation interactions which will require important future developments. Some strong nonlinearities that could lead to noise problems, already mentioned by several authors (Errico and Raeder 1999; Janísková et al. 1999; Mahfouf 1999), have been removed through a partial linearization of the schemes (e.g. vertical diffusion, moist convection). Despite the fact that a number of thresholds have been removed for some schemes (e.g. evaporation of precipitation, wave breaking), important switches still remain for the onset of moist physical processes.

3. EVALUATION OF THE TANGENT-LINEAR APPROXIMATION

In this section analysis increments evolved using a tangent-linear version of the ECMWF global spectral model are compared to differences of NL integrations from the model including a full physical package. The goal of this study is to demonstrate that when linearized physical processes are included, the tangent-linear model is in better agreement with the NL model. Similar comparisons have already been made by Mahfouf (1999) with a T42L31 version of the ECMWF model for 24-hour simulations. We present here a study at a higher horizontal resolution (T63) and over a time window compatible with the framework of 4D-Var experimentation (i.e. 6 hours). It is important
to examine the generality of results from Mahfouf (1999), since the validity of the
tangent-linear approximation strongly depends on the horizontal scales and on the
time periods of interest. Such analysis will also help in understanding the influence
of physical processes in a 4D-Var system using a similar resolution and time window.

Figure 1(a) displays analysis increments for the zonal wind at the lowest model level
(\(\approx 30\) m above the surface) over the northern hemisphere. This level is chosen because
turbulent processes are dominant close to the surface. The increments are computed
from the operational 3D-Var system (0000 UTC 5 December 1996) and have values
around 1 m s\(^{-1}\). Larger values are seen over ocean areas where SHIP and BUOY surface
wind observations are assimilated. Maximum values can reach up to 5 m s\(^{-1}\). The
analysis increments evolved nonlinearly are strongly reduced after 6 hours (Fig. 1(b)).
The high surface drag over continents leads to non-negligible values of wind increments
mainly over the Atlantic and Pacific oceans. But, even over the oceans, the initial
perturbations are reduced by a factor of between two and three through momentum
friction.

Figure 2 shows the evolution of zonal wind increments obtained from two versions
of the tangent-linear model. The first version is almost adiabatic and referred to hereafter
as Simplified Physics (SP); it only represents surface drag and vertical diffusion assuming
a neutral boundary layer as described in Buizza (1994). This simplified physics
was used in the tangent-linear and adjoint models presented in Part I. The second ver-

cion includes the physical parametrization schemes described in Mahfouf (1999), and
is referred to hereafter as Improved Physics (IP). The new vertical diffusion scheme
accounts for stability dependencies of both the exchange coefficients, the surface drag
and the surface heat and moisture fluxes (Louis et al. 1981). Surface roughness lengths
have geographical variations depending upon vegetation cover and orography over con-
tinents, and upon surface stress over oceans (Charnock’s formula). Roughness lengths
for momentum have values of about 1 m over vegetated areas and greater than 10 m
over mountainous areas. Both linearizations are performed around a reference trajectory
obtained from an integration of the NL model with the full physical parametrization
package. Although the 6-hour evolution with SP has significantly reduced the size of the
initial increments (Fig. 2(a)), they appear to be overestimated with respect to the NL
reference (Fig. 1(b)). This is particularly true over continents where momentum dissipa-
tion is clearly too weak with SP. This can be understood by reference to the low value of
surface roughness length (5 cm) chosen in the simplified vertical diffusion scheme.
This leads to an underestimation of the surface stress of around 0.5 N m\(^{-2}\) over most of the
continents. Figure 2(b) shows that the evolution produced using IP is in much better
agreement with the NL evolution. Increments are very small over continents. They are
also reduced over the western Pacific (Kuroshio current) where strong unstable condi-
tions prevail. The blocking effect of the subgrid-scale orography scheme also acts to
reduce the low-level wind speed near high mountain ranges with IP. Although the moist
physics is activated in the experiment with IP, these parametrization schemes play a
negligible role in explaining the wind structure near the surface. A detailed examination
of the influence of individual physical processes is presented in Mahfouf (1999). Over
the Atlantic ocean only few significant differences are observed between SP and IP.
Since the most spectacular differences between SP and IP are found over land, where no
SYNOP wind observations are used in the ECMWF data assimilation system (mostly
due to representativeness problems), it is likely that the positive impact of the improved
physics will not be reflected in terms of quality of wind analyses.
Figure 1. Analysis increments of the zonal wind at model level 31 ($\approx 30$ metres) for 5 December 1996 at 0000 UTC (upper panel) and after a 6-hour evolution obtained from the difference between two nonlinear integrations with full physics (lower panel). The isolines are plotted every $0.5 \text{ m s}^{-1}$. 
Figure 2. Evolution of analysis increments for the zonal wind at model level 31 (≥30 metres) during 6 hours, produced with a tangent-linear model with simplified physics (upper panel), and with a tangent-linear model with improved physics (lower panel). The isolines are plotted every 0.5 m s$^{-1}$. 
In order to get a global picture of the improvement coming from the linearized physical package, the mean absolute error in specific humidity is shown in Fig. 3 as zonal means. This error is computed with respect to pairs of NL integrations. The global error with IP compared to the NL evolution is $0.055 \text{ g kg}^{-1}$, and is $0.065 \text{ g kg}^{-1}$ with SP. Largest errors are located in the lowest equatorial troposphere. The inclusion of IP in the tangent-linear model improves the evolution of specific-humidity increments in the boundary layer, since the simple vertical diffusion scheme of Buizza (1994) is not applied to this variable. Significant reduction of errors are also noticed at mid latitudes where the large-scale condensation scheme is active in baroclinic disturbances. The inclusion of the simplified linear convection scheme also improves the behaviour of the lower troposphere in the tropics. However, there is a slight degradation higher up where detrainment effects make less valid the assumption of the dominance of the mass-flux transport used to derive the simplified linearized convection scheme.

Conclusions obtained when comparing tangent-linear evolutions of analysis increments with NL evolutions over a 6-hour period, with a T63L31 version of the ECMWF
forecast model, are very similar to those presented by Mahfouf (1999) with a lower-resolution model and over a longer time integration. This gives confidence in the generality of his results, which are also consistent with those obtained by Janísková et al. (1999) in a similar framework. It appears that even for a 6-hour evolution of analysis increments, the tangent-linear approximation can be improved when diabatic processes are described. The improvement is noticeable near the surface where momentum dissipation is more active (through stability effects on turbulent transfers and low-level blocking by subgrid-scale orography). Moist processes (large-scale condensation and deep convection) also lead to significant improvements of evolved increments for specific humidity.

4. SINGULAR VECTORS WITH LINEARIZED PHYSICS

The singular vectors of a tangent-linear operator are the initial perturbations producing a maximum localized energy growth over an optimization time interval (Buizza 1994). They are computed for this study over a 6-hour window with a T63L31 version of the ECMWF model for one of the intensive observation periods (IOPs) of the field campaign of the Fronts and Atlantic Storm-Track EXperiment (FASTEX), IOP7 on 7 February 1997. Although the usual period over which singular vectors are computed is 48 hours, comparison over a 6-hour optimization period allows one to examine how optimal perturbations can grow over the assimilation window of the first 4D-Var configuration. It is also possible to examine if the linear physics produces either different amplification rates or different geographical locations with respect to adiabatic singular vectors. The spectrum of amplification factors for the first 16 singular vectors is presented in Fig. 4 both for SP and IP. They correspond to the ratio of total-energy norm at optimization time over total-energy norm at initial time, and have been computed over the northern hemisphere extratropics (north of 30°N). A moist-energy norm defined by Ehrendorfer et al. (1999) was used for the computation of growth rates in Mahfouf (1999), in order to identify spurious growth rates produced by the linearized large-scale condensation scheme. In the present study, a dry-energy norm similar to the one chosen by Buizza and Palmer (1995) allows us to isolate the effects of diabatic processes in the linearized models. Apart from the first singular vector, the amplification factors are lower with IP than with SP. This feature is consistent with the increased dissipation produced by both the stability-dependent vertical diffusion scheme and the subgrid-scale orographic scheme. The projection of the first 10 singular vectors obtained with IP on the first 10 singular vectors obtained with SP leads to a 73% similarity of the two sub-spaces, which means that the improved physics does not strongly modify the structure of the most unstable perturbations. The extratropical singular vectors using IP have a high degree of similarity with singular vectors obtained using SP. This result is not surprising since the growth rate of such unstable modes is governed by the baroclinic instability, which is essentially a dynamical process accurately described by adiabatic models linearized around a realistic mean state. It appears that the amplification rates are lower with IP due to more active dissipative processes, which are in better agreement with the full NL model, as previously shown. With improved physics, kinetic energy is reduced by turbulent processes, whereas potential energy is increased through latent-heat release. However, potential energy is only 5% of the total energy at optimization time. This can be seen more clearly when examining the first singular vector evolved after 6 hours and located in the middle of the Atlantic ocean. It is depicted for specific humidity near 700 hPa (model level 22) in Fig. 5. Specific-humidity perturbations are larger with IP and the patterns are significantly different. For example, the negative anomaly along the
50°N parallel with SP is almost non-existent with IP. Indeed, latent-heat release is a mechanism by which static stability is reduced in baroclinic systems leading to higher growth rates of unstable modes.

This result can be made more obvious when using a moist-energy norm allowing non-zero perturbations of specific humidity at initial time. Amplification rates are then systematically larger when moist processes are included (Mahfouf et al. 1996; Ehrendorfer et al. 1999). In terms of data assimilation, it is likely that the fit to data will be more difficult using IP than SP, since the sensitivity to initial conditions is reduced for dissipative systems. This could be different in precipitating areas where latent-heat release amplifies initial perturbations. However, current assimilation systems do not sample properly the vertical thermodynamical structure of rainy systems, since satellite radiances are only used in clear-sky situations. The only data assimilated in cloudy areas are SATOB winds from geostationary satellites, and surface winds from the scatterometer on board the ERS-2 satellite. However, they do not provide information on temperature and humidity profiles, which are the relevant parameters for latent-heat release in cloudy systems. Therefore it is likely that the differences noticed regarding the behaviour of singular vectors with IP will not affect significantly the 4D-Var analyses in the extratropics.

5. DESIGN OF A 4D-VAR CONFIGURATION WITH PHYSICS

A first experimental 4D-Var assimilation with linearized physical processes was undertaken with a one-outer-loop configuration. Three configurations were tested over a few assimilation cycles starting from 0000 UTC 17 February 1997. The first configuration corresponds to the baseline 4D-Var system described in Part I, which is almost adiabatic but still includes a very simple vertical diffusion scheme without stability effects (SP). A second configuration of 4D-Var includes the linear physics previously evaluated (IP). A third 4D-Var configuration corresponds to experiments run with adiabatic tangent-linear and adjoint models (referred to as AD). From these three experiments, the
progressive impact of including physical processes in the 4D-Var system is evaluated. The main lesson we learned from this exercise concerns the influence of physical processes on the convergence of the minimization. Although forecasts were run from these three analyses, they did not show significant differences in a case of rapid cyclogenesis over Western Europe (IOP17 of FASTEX).

Figure 6 presents the evolution of the total cost-function for the 6-hour cycle around 1200 UTC 17 February 1997 for the three 4D-Var configurations, together with the square of the gradient norm. The norm of the gradient is estimated in a space defined
Figure 6. Total cost-function (upper panel) and square of the gradient (lower panel) during the minimization process of an assimilation cycle (1200 UTC 17 February 1997) for three configurations of the 4D-Var system. '4V AD' is a 4D-Var one update adiabatic; '4V SP' is a 4D-Var one update with simplified physics; '4V IP' is a 4D-Var one update with improved physics. See text for details.

by the change of variable \( \chi = B^{-1/2}x \), where \( B \) is the covariance matrix of background errors (Derber and Bouttier 1999) and \( x \) the control vector.

The convergence process is very similar between the three experiments, a minimum being reached in about 50 iterations (upper panel). Around iteration 30, the variational quality control is activated, leading to the rejection of some data and therefore to a reduction of the cost-function. The square of the gradient norm (lower panel) decreases by four orders of magnitude, which is also an indication that the physics does not produce convergence problems. Earlier 4D-Var experiments with physics (Zou et al. 1993;
Zupanski 1993; Tsuyuki 1996; Zou 1996) were not performed with the incremental approach. In the incremental 4D-Var the full problem is linearized around the background state, and the actual cost-function to be minimized is quadratic in the control variables. Therefore, convergence problems identified by earlier authors cannot occur in the inner-loop of the incremental 4D-Var. However, as shown hereafter, inconsistencies can be identified between the low- and high-resolution problems. Figure 6 (upper panel) shows that the value of the cost-function reached at the minimum is the lowest with the AD configuration and the highest with the IP configuration. The fit to the observations is therefore easier with an adiabatic model, since only dynamical constraints are imposed on the time evolution of increments. For example, as demonstrated in the singular-vector study, the growth rate of unstable perturbations is higher when the model contains only simplified physics, which indicates that the improved linear physics is more dissipative. As a consequence, larger initial increments would be required with improved physics to fit observations in the middle of the assimilation window. In fact, the background term imposes a similar size of analysis increments leading to a larger departure from the observations when the improved physics is included in 4D-Var.

The various terms of the cost-function at the minimum are depicted in Fig. 7 for the low-resolution problem ($J_0(L)$). The observation term is also computed at high resolution with the T213L31 model including full physical parametrizations ($J_0(H)$). The background ($J_b$) and penalty ($J_p$) terms are very similar between the 4D-Var configurations but do not represent the major contribution to the total cost-function. The mismatch between the observation term at low- and high-resolution, measures the degree of inconsistency in the incremental approach between the problem to be solved and the linear problem actually solved. An interesting feature of the 4D-Var IP configuration is that the discontinuity arising in $J_0$ when going from the low- to the high-resolution is smaller than for the SP configuration or without physics. As a consequence, the background term is larger for the AD configuration. Therefore, the better convergence of the low-resolution adiabatic problem is produced by finding a low-resolution model trajectory which overfits some observations in contradiction to the behaviour of the high-resolution NL model. More specifically, the root-mean-square (r.m.s.) error with respect to TEMP data is largest for relative humidity in the 4D-Var IP configuration. However, the reduction of $J_0$ discontinuity indicates that the inclusion of physical processes in 4D-Var provides a better consistency between the inner- and outer-loops of the incremental approach.

The positive impact of the linear physics on the convergence of the incremental 4D-Var must be set against the fact that the minimization of the variational problem is 2.5 times more expensive than the adiabatic (or SP) solution. We therefore proposed an alternative strategy to reduce the cost of the physics and also to better take into account the nonlinearities it introduces. This is a 'two-update' configuration (or, equivalently, a two outer-loop configuration) where a first minimization is performed with SP, followed by a second minimization with IP after updating the trajectory at high-resolution. The underlying idea is similar to the multi-incremental approach proposed by Veersé and Thépaut (1998). We assume that the minimization first adjusts the larger scales of the atmospheric flow which are correctly described by a model without physics, and then the minimization adjusts the smaller scales which depend more on physical processes. With this configuration, the additional computing time introduced by the linear physics has been reduced to 50% in a '40/40' configuration (40 iterations with SP followed by 40 iterations with IP). For a few assimilation cycles, we have checked that the behaviour of the minimization when the linear physics is activated after a first update of the trajectory is very similar to a minimization where the physics is always activated.
Figure 7. Contributions to the total cost-function for three configurations of the 4D-Var system at 1200 UTC 17 February 1997. '4V AD' is a 4D-Var one update adiabatic; '4V SP' is a 4D-Var one update with simplified physics; '4V IP' is a 4D-Var one update with improved physics. $J_0(H)$ is the observation term computed at high resolution, $J_0(L)$ is the observation term computed at low resolution, $J_b$ is the background term and $J_c$ the penalty term. See text for further details.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$J_0(L)$</th>
<th>$J_0(H)$</th>
<th>$J_0(FG)$</th>
<th>CPU (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1UP SP</td>
<td>40256</td>
<td>45245</td>
<td>86441</td>
<td>21258</td>
</tr>
<tr>
<td>2UP IP/IP</td>
<td>40650</td>
<td>44321</td>
<td>85320</td>
<td>61620</td>
</tr>
<tr>
<td>2UP SP/IP</td>
<td>40539</td>
<td>44342</td>
<td>85336</td>
<td>37811</td>
</tr>
</tbody>
</table>

$J_0(H)$ is the observation term at high resolution; $J_0(L)$ is the observation term at low resolution and $J_0(FG)$ is the observation term at the beginning of the minimization (fit to the first-guess). '1UP SP' is a one-update SP configuration (70 iterations); '2UP IP/IP' is a two-update IP configuration with two sets of 40 iterations; '2UP SP/IP' is a two-update configuration with 40 SP iterations and 40 IP iterations with IP. CPU is central-processor utilized time.

as shown in Table 1. Since results are presented for the last 6-hour window from a set of three assimilation cycles the $J_0$ term for the first-guess is different in the various configurations tested. The two-updates configuration provides a lower value of the cost-function at the beginning of the minimization (the first-guess is thus closer to the observations), and also at the end of the descent process after the observation term is recomputed at high resolution. A "50/20" configuration (50 iterations with SP followed by 20 iterations with IP) was finally chosen in order to reach a better convergence during the first minimization. Since the quality control is activated after 30 iterations, it needs more than 10 iterations to stabilize the convergence with the reduced set of observations.
(Fig. 6 (upper panel)). The reduction of the number of iterations with IP to 20, leads to an extra-cost of only 25% imposed by the physics. The additional trajectory at high resolution also imposes a 25% increase with respect to the one-outer-loop configuration. The impact of the improved physics on this 4D-Var configuration is evaluated in the next section.

6. EXPERIMENTATION USING 4D-VAR WITH TWO OUTER-LOOPS

Three configurations of 4D-Var were tested during a two-week winter period (1 to 14 February 1997) using the operational 3D-Var as a control:

- 4D-Var one-update SP (70 iterations).
- 4D-Var two-updates SP (50 iterations with SP and 20 iterations with SP).
- 4D-Var two-updates IP (50 iterations with SP and 20 iterations with IP).

We compared the performance of these three versions of 4D-Var in order to evaluate the impact of the two-outer-loops configuration and of the physics. The extratropical forecast scores are presented in Fig. 8 in terms of anomaly correlation for the geopotential at 1000 hPa. The impact of including improved physical processes is rather neutral, whereas going from one to two outer-loops is slightly detrimental in the southern hemisphere around day 3. A more extended comparison of 4D-Var with one outer-loop and 4D-Var with two outer-loops over 56 days has demonstrated that this feature is not systematic. The neutral impact of the linearized physics over mid latitudes has been better understood by examining the r.m.s. errors of short-range forecast differences between the various systems. The differences between 3D-Var and 4D-Var shown in Mahfouf and Rabier (1999), for the temperature at 850 hPa, mostly occur over the Atlantic and Pacific oceans along the storm tracks. Although some differences are present with the inclusion of the physics, they are located over Siberia north of 60°N which is not a sensitive region for the growth of unstable perturbations. Therefore, these short-range forecast differences are not likely to influence the medium range. In terms of weather parameters, the agreement with observations from SYNOP stations is slightly better over Europe when the linear physics is included in 4D-Var. The fit to observed two-metre temperature, which is improved by 4D-Var SP with one outer-loop compared with 3D-Var, is further improved when adding IP in the second outer-loop (Mahfouf and Rabier 1999).

The inclusion of the linearized physics has most impact on the analysis of specific humidity. The zonal mean increments of specific humidity averaged over the two-week period show a global moistening of the lower tropical troposphere and a drying in mid latitudes with both configurations (Fig. 9). The vertical profile of the increments is imposed by the statistics of background error for specific humidity derived from the 'NMC method', described by Rabier et al. (1998b) and Derber and Bouttier (1999). The standard-deviations of forecast errors are a maximum around 850 hPa, where the increments are the largest. Around level 28, vertical correlations are very small leading to minimum values of increments. 4D-Var IP produces smaller increments in the tropical belt, both in the boundary layer and above, compared to the baseline 4D-Var. As shown before, the inclusion of physical processes weakens the amplification of initial perturbations. This effect is more dominant for specific humidity, since no diabatic processes are applied to this model variable when using SP (dry vertical diffusion). The fit to data is therefore more difficult when the improved physics is included in 4D-Var. It is important to emphasize that mean increments of specific humidity were systematically larger in the baseline 4D-Var system than in 3D-Var, explaining part of the worse performance of 4D-Var in the Tropics (Rabier et al. 1998a).
The direct consequence of a drier initial atmospheric state when introducing improved physics in 4D-Var is a significant reduction of the spin-down in precipitation during the first day of the forecasts; this is shown in Fig. 10 where the time evolution of total precipitation in the tropical belt (between 30°S and 30°N) averaged over 14 forecasts is presented. Since the tropical circulation is strongly driven by the distribution of diabatic heating, the analysis of humidity has a direct influence on the quality of wind forecasts. The short-range tropical wind scores are improved with the inclusion of physical processes, whereas the impact of the two-update SP configuration is almost neutral. The improvement is larger at 850 hPa than at 200 hPa (Fig. 11) in agreement with the stronger impact of the linearized physics on the time evolution of analysis.
increments of specific humidity in the tropical belt, described in section 3. Although the inclusion of physics is beneficial to the analysis of humidity, the fact that less weight is given to the observations in the assimilation indicates that further improvements are needed to make the system more optimal; for example, by improving the statistics of forecast errors, which are currently global and univariate, and tuning the background variances using statistics of first-guess departures.

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 period of eight weeks. These include the periods 1 to 14 February 1997, 26 August to 6 September 1995, 15 to 28 January 1997 and 27 June to 10 July 1997. The scores over the northern and southern hemispheres are marginally positive, with a larger positive signal over Europe. The scatter between the baseline 4D-Var and the 4D-Var with physics is very small in the short range, and the fit of the background fields to the data is very similar. Overall, the clearest improvement is seen in the Tropics for the 850 hPa wind field, similar to that in Fig. 11 for one two-week period. Part III describes a more extensive evaluation against the 3D-Var system (Klinker et al. 2000).
Figure 10. Time evolution of global precipitation in the tropical belt (30°S to 30°N) averaged over 14 forecasts based on 4D-Var assimilation two-updates, with simplified and improved physics. See text for details.

7. CONCLUSIONS

The baseline configuration of the incremental 4D-Var developed at ECMWF and described in Part I (Rabier et al. 2000) has been modified to include physical processes in the tangent-linear and adjoint versions of the model. The objective was to improve the time evolution of increments and the estimation of the gradient of the cost-function. A comprehensive package of linearized physics (Mahfouf 1999) has been evaluated, by comparing the time evolution of analysis increments produced by the tangent-linear model with differences between two NL integrations. The inclusion of the linearized physics significantly improves the behaviour of the tangent-linear model with respect to the NL model. Improvements of a T63L31 version of the ECWMF model over a 6-hour time window with respect to a simple vertical diffusion scheme were illustrated for both surface wind and specific humidity. The time evolution of analysis increments is closer to the NL behaviour when the tangent-linear model includes linearized physical processes. These results are consistent with those obtained by Mahfouf (1999) for a different experiment with a T42L31 version of the ECMWF model over 24 hours. Singular vectors have been computed with linear physics over 6 hours in the extratropics. Amplification rates are generally smaller with improved physics, except in areas where latent-heat release takes place. The inclusion of improved physics in 4D-Var has led to a better consistency between the minimization at high and low resolutions. The discontinuity of the cost-function imposed by the change of resolution and the change of physical parametrization is smaller with linearized physics.

The cost of the linear physics has been reduced by defining a new 4D-Var configuration with two outer-loops, where a first minimization with only simplified physics
Figure 11. Tropical wind scores at 850 (upper panel) and 200 hPa (lower panel) averaged over 11 forecasts, each verified against its own analysis, from 4D-Var assimilation one-update with simplified physics (dotted), two-updates with simplified physics (dashed) and two-updates with improved physics (solid). See text for details.

(50 iterations) is followed by a second minimization with improved physical processes activated (20 iterations). Indeed, the minimization first adjusts the large-scale structures of the atmospheric flow which are correctly described by an adiabatic model. Then, the second minimization modifies the smaller scales which depend more on physical processes. For such a configuration, the increase in computational cost is 50% with respect to the baseline configuration presented in Part I (half of it comes from the physics and the other half comes from the extra-trajectory computation). Medium-range forecasts are improved over Europe for weather parameters. In the Tropics, the r.m.s. error for wind is significantly reduced in the short range when data assimilation systems are compared to their own analysis. This improvement in the forecast winds is a direct consequence of the reduced spin-down in precipitation noticed in the revised 4D-Var with physics. This 4D-Var configuration, with two outer-loops including improved physics in the second
minimization, is extensively evaluated against the operational 3D-Var system in Part III (Klinker et al. 2000).

ACKNOWLEDGEMENTS

We would like to thank R. Buizza for his help with the singular vectors including linearized physics. R. M. Errico provided much advice and encouragement in the early stages of the developments of the linearized physics during his visit to ECMWF. The technical support from M. Hamrud and L. Isaksen was appreciated. A. Hollingsworth and A. Simmons were extremely supportive and enthusiastic during the pre-operational evaluation of the various 4D-Var configurations. The authors also acknowledge the anonymous reviewers of this manuscript for their useful comments.

REFERENCES


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


Louis, J.-F., Tiedtke, M. and Geleyn, J.-F.

Mahfouf, J.-F.

Mahfouf, J.-F. and Rabier, F.

Mahfouf, J.-F., Buizza, R. and Errico, R. M.

Rabier, F., Courtier, P., Pailleux, J., Talagrand, O., Thépaut, J.-N. and Vasiljević, D.

Rabier, F., Thépaut, J.-N. and Courtier, P.


Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A.

Thépaut, J.-N. and Courtier, P.

Tsuyuki, T.

Veersé, F. and Thépaut, J.-N.

Vukičević, T. and Errico, R. M.

Zou, X.

Zou, X., Navon, I. M. and Sela, J. G.

Zupanski, D.


1999 Influence of physical processes on the tangent-linear approximation. Tellus, 51A, 147–166

1999 'The ECMWF operational implementation of four dimensional variational assimilation. Part II: Experimental results with improved physics'. ECMWF Technical Memorandum No 272. Available from ECMWF, Shinfield Park, Reading, UK


1998a Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. Q. J. R. Meteorol. Soc., 124, 1861–1887


1993 Linearization and adjoint of parameterized moist diabatic processes. Tellus, 45A, 493–510


1993 Variational data assimilation with moist threshold processes using the NMC spectral model. Tellus, 45A, 370–387

1993 The effects of discontinuities in the Betts–Miller cumulus convection scheme on four-dimensional variational data assimilation. Tellus, 45A, 511–524