TECHNICAL MEMORANDUM

357

On the use of a predictor-corrector scheme to couple the dynamics with the physical parametrizations in the ECMWF model

M. J. P. Cullen and D. J. Salmond

Research Department

Submitted for publication in Q. J. Roy. Met. Soc.

January 2002

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

For additional copies please contact

The Library
ECMWF
Shinfield Park
Reading
RG2 9AX
library@ecmwf.int

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/pressroom/publications/

Abstract

Methods of coupling the time integration of the resolved dynamics with the parametrized processes in atmospheric models are an active development area. Many centres have demonstrated strong sensitivity to the methods of coupling in their models. ECMWF has recently introduced a revised method of coupling which gives significant forecast benefits. Theoretically optimal methods are difficult to establish because of the mixture of time-scales represented within the parametrizations. Implicit methods are not practical because of the nonlinear switching present in most parametrization schemes. In this paper we show that a predictor-corrector scheme can give some of the advantages of a fully implicit scheme. We show that the use of more than one physics evaluation per time-step significantly improves the accuracy in a model problem. We also demonstrate the effect of further iterations, which in principle would converge towards a fully implicit scheme. A second iteration has only a small effect on overall performance, but gives a large reduction in the amount of convection. this indicates that the current formulation of convection is not compatible with this type of integration scheme.

1 Introduction

The efficient integration of the dynamical equations in atmospheric models is well-explored. Increasing attention is now being paid to the accurate and efficient integration of the augmented equations, including the additional terms representing parametrized effects. A number of methods are described and reviewed by Beljaars (1991), McDonald and Haugen (1992) and Moorthi et al. (1995). All these are essentially first order accurate. Ideally, one would like to include the parametrized terms in a fully implicit formulation, McDonald (1998). However, this is not practicable due to the nonlinear nature of the parametrized equations, and in particular due to the logical switches often employed in the formulations. Wedi (1999) developed an explicit second-order scheme for including the parametrized terms. This reduced the short time-scale noise in the ECMWF model and improved the overall forecast performance. Further experience has been obtained as a result of including a linearised version of the parametrizations in variational data assimilation, Mahfouf and Rabier (2000). It is found that a linearisation can achieve much of the potential benefit of incorporating the full nonlinear parametrizations in the variational procedure, suggesting that a partially implicit treatment of the physics may be feasible.

Cullen (2001) shows that use of a predictor-corrector scheme could give benefits to the accuracy of the ECMWF model by allowing an approximation to an implicit treatment of the semi-Lagrangian trajectory. He also showed that the predictor-corrector method was more accurate on a simple analytic problem than an explicit second-order method. It is then natural to explore the use of a predictor-corrector method for the parametrized processes as well. In principle, this would give a practical and not-too-expensive approximation to a fully implicit scheme. It also allows full time-level values to be input to the highly nonlinear calculations of transfer coefficients, while still allowing coupling between the different processes.

There are a number of analyses of the multiple time-scale problem in the literature. Browning and Kreiss (1994) show that the accuracy with which the dynamical equations are solved is determined by the accuracy with which a 'reduced system' from which the fast time-scales have been removed is solved. The potential vorticity is a generic example of a slow variable in this context. Sportisse (2000) uses a similar technique to analyse reaction-diffusion equations, which can be parabolic in character and therefore more closely related to the problem of fast time-scales in the physics. The techniques described in both papers rely on splitting fast and slow variables. We carry out a version of this analysis to show how the predictor-corrector scheme can be applied in the presence of physics with multiple time-scales.

It is much harder to work out how to interpret physical parametrizations in terms of fast and slow time-scales than for the dynamical equations. Much of the physics will have a strong diurnal time-scale. Within this, there

are much faster time-scales, such as those involved in the generation of precipitation. Some of the physical processes will be strongly coupled to dynamical time-scales; for instance latent heat release will be closely coupled to vertical motion. In the dynamical equations, the natural split is between the potential vorticity controlled dynamics as 'slow' and the inertio-gravity waves as 'fast'. This can only be illustrative, as on some scales the inertio-gravity wave frequency is less then the diurnal frequency which is a main time-scale for the physics. However, by making such a scale separation, we can clarify the options for time integration of the physics.

We apply the predictor-corrector integration of the physics to the current version of the ECMWF model. This model is part of the IFS (Integrated Forecasting System) developed jointly with Meteo-France. The current implementation at ECMWF uses the two-time-level semi-Lagrangian integration scheme, Temperton et al. (2001). Recent developments to the parametrizations are described by Gregory et al. (2000). We show how each of the current parametrizations can be interfaced with the predictor-corrector scheme. However, we also show that there could be advantages if there was more consistency between the formulations of the different schemes.

2 Time Integration schemes for dynamics and parametrizations

2.1 Analysis of schemes with multiple time-scales

We first analyse a generic system containing multiple time-scales. Options for applying this framework to real physical parametrizations are discussed in a later subsection. We use a simplification of the models discussed by Sportisse (2000). Consider the model problem

$$\begin{aligned} \frac{dx}{dt} &= Ax + By \\ \varepsilon\frac{dy}{dt} &= Cx - Dy + \varepsilon Ey \end{aligned} \tag{1}$$

Here x, y are respectively slow and fast variables, where the ratio of slow to fast time-scales is given by $\varepsilon \ll 1$. A, B, C, D and E are time-dependent coefficients which may be functions of x.(1) is to be solved with initial data $x = x_0, y = y_0$. Efficient numerical integration requires choosing a time step such that $\varepsilon^{-1}\delta t \gg 1$. Both Browning and Kreiss (1994) and Sportisse (2000) analyse numerical approximations to (1) by first deriving a reduced system, whose solution is an accurate approximation to (1) for small ε. They then extract a numerical approximation to the reduced system from the algorithm proposed for the full system, and analyse its error in the normal way. An appropriate reduced system for (1) is

$$\begin{aligned} \frac{dx}{dt} &= Ax + By \\ 0 &= Cx - Dy \end{aligned} \tag{2}$$

Consider the following numerical approximation to (1):

$$x^* - x^t = \delta t(A^t x^t + B^t y^t)$$

$$
\begin{aligned}
y^* - y^t &= \varepsilon^{-1}\delta t\left(C^* x^* - D^* y^* + \varepsilon E^t y^t\right) \\
x^{t+\delta t} - x^t &= \frac{1}{2}\delta t(A^t x^t + A^* x^* + B^t y^t + B^* y^*) \\
y^{t+\delta t} - y^t &= \varepsilon^{-1}\delta t\left(C^{t+\delta t} x^{t+\delta t} - D^{t+\delta t} y^{t+\delta t} + \varepsilon E^* y^*\right)
\end{aligned}
\tag{3}
$$

This implies the following numerical approximation to the reduced system, eq. (2):

$$
\begin{aligned}
x^* - x^t &= \delta t(A^t x^t + B^t y^t) \\
0 &= C^* x^* - D^* y^* \\
x^{t+\delta t} - x^t &= \frac{1}{2}\delta t(A^t x^t + A^* x^* + B^t y^t + B^* y^*) \\
0 &= C^{t+\delta t} x^{t+\delta t} - D^{t+\delta t} y^{t+\delta t}
\end{aligned}
\tag{4}
$$

Substituting for y in terms of x, the final update of x becomes

$$
x^{t+\delta t} - x^t = \frac{1}{2}\delta t(A^t x^t + A^* x^* + (BD^{-1}C)^t x^t + (BD^{-1}C)^* x^*) \tag{5}
$$

We see that this is simply a predictor-corrector approximation to the evolution equation for x derived from (2), and is thus second order accurate. According to Sportisse (2000), we can therefore expect it to be second order accurate in practice for approximations to (1) with $\delta t \gg \varepsilon$.

2.2 Comparison of integration schemes for a problem with multiple time-scales

We illustrate with a simple example. The problem to be solved is

$$
\begin{aligned}
\frac{dx}{dt} &= \cos(t + .4y) \\
\frac{dy}{dt} &= C(x - y(1 + .5\sin(3x)))
\end{aligned}
\tag{6}
$$

Here, C is a constant chosen much greater than 1. The coefficients are chosen to introduce nonlinearity, but the coefficients in the equation for the fast variable only depend on the slow variable. The reduced system for (6) replaces the second equation by the diagnostic relation

$$
0 = x - y(1 + .5\sin(3x)) \tag{7}
$$

We compare the following schemes:

i) *Simple split*

$$
\begin{aligned}
x^{t+\delta t} - x^t &= \delta t\cos(t + .4y^t) \\
y^{t+\delta t} - y^t &= C\delta t(x^{t+\delta t} - y^{t+\delta t}(1 + .5\sin(3x^{t+\delta t}))
\end{aligned}
\tag{8}
$$

This scheme is only first order accurate in x.

(ii) *Predictor-corrector with implicit solution for y*

$$\begin{aligned} x^* - x^t &= \delta t \cos(t + .4y^t) \\ y^* - y^t &= C\delta t (x^* - y^*(1 + .5\sin(3x^*)) \\ x^{t+\delta t} - x^t &= \frac{1}{2}\delta t (\cos(t + \delta t + .4y^*) + \cos(t + .4y^t)) \\ y^{t+\delta t} - y^t &= C\delta t (x^{t+\delta t} - y^{t+\delta t}(1 + .5\sin(3x^{t+\delta t})) \end{aligned} \tag{9}$$

This scheme follows the form of (4) and is thus second-order accurate in the x equation.

(iii) *Predictor-corrector with analytic solution for y*

This scheme is illustrated because it is the natural method of integrating the ECMWF prognostic cloud scheme. The equation for y is solved analytically, assuming that x is constant during the time-step.

$$\begin{aligned} x^* - x^t &= \delta t \cos(t + .4y^t) \\ y^* = y^t \exp(-C\delta t(1 + .5\sin(3x^*))) &+ \frac{x^*(1 - \exp(-C\delta t(1 + .5\sin(3x^*)))}{(1 + .5\sin(3x^*)} \\ x^{t+\delta t} - x^t &= \frac{1}{2}\delta t (\cos(t + \delta t + .4y^*) + \cos(t + .4y^t)) \\ y^{t+\delta t} = y^t \exp(-C\delta t(1 + .5\sin(3x^{t+\delta t}))) &+ \frac{x^{t+\delta t}\left(1 - \exp(-C\delta t(1 + .5\sin(3x^{t+\delta t}))\right)}{(1 + .5\sin(3x^{t+\delta t})} \end{aligned} \tag{10}$$

(iv) *Extrapolated second-order scheme*

We calculate a provisional tendency using (8) at each step:.

$$\begin{aligned} x^o - x^{t-\delta t} &= \delta t \cos(t - \delta t + .4y^{t-\delta t}) \\ y^o - y^{t-\delta t} &= C\delta t \, (x^o - y^o(1 + .5\sin(3x^o)) \\ x^* - x^t &= \delta t \cos(t + .4y^t) \\ y^* - y^t &= C\delta t \, (x^* - y^*(1 + .5\sin(3x^*)) \end{aligned} \tag{11}$$

and then achieve a second-order accurate tendency by the standard extrapolation:

$$\begin{aligned} x^{t+\delta t} - x^t &= \frac{1}{2}\left(3(x^* - x^t) - (x^o - x^{t-\delta t})\right) \\ y^{t+\delta t} - y^t &= \frac{1}{2}\left(3(x^* - x^t) - (y^o - y^{t-\delta t})\right) \end{aligned} \tag{12}$$

This scheme only requires one evaluation of the nonlinear functions per time-step.

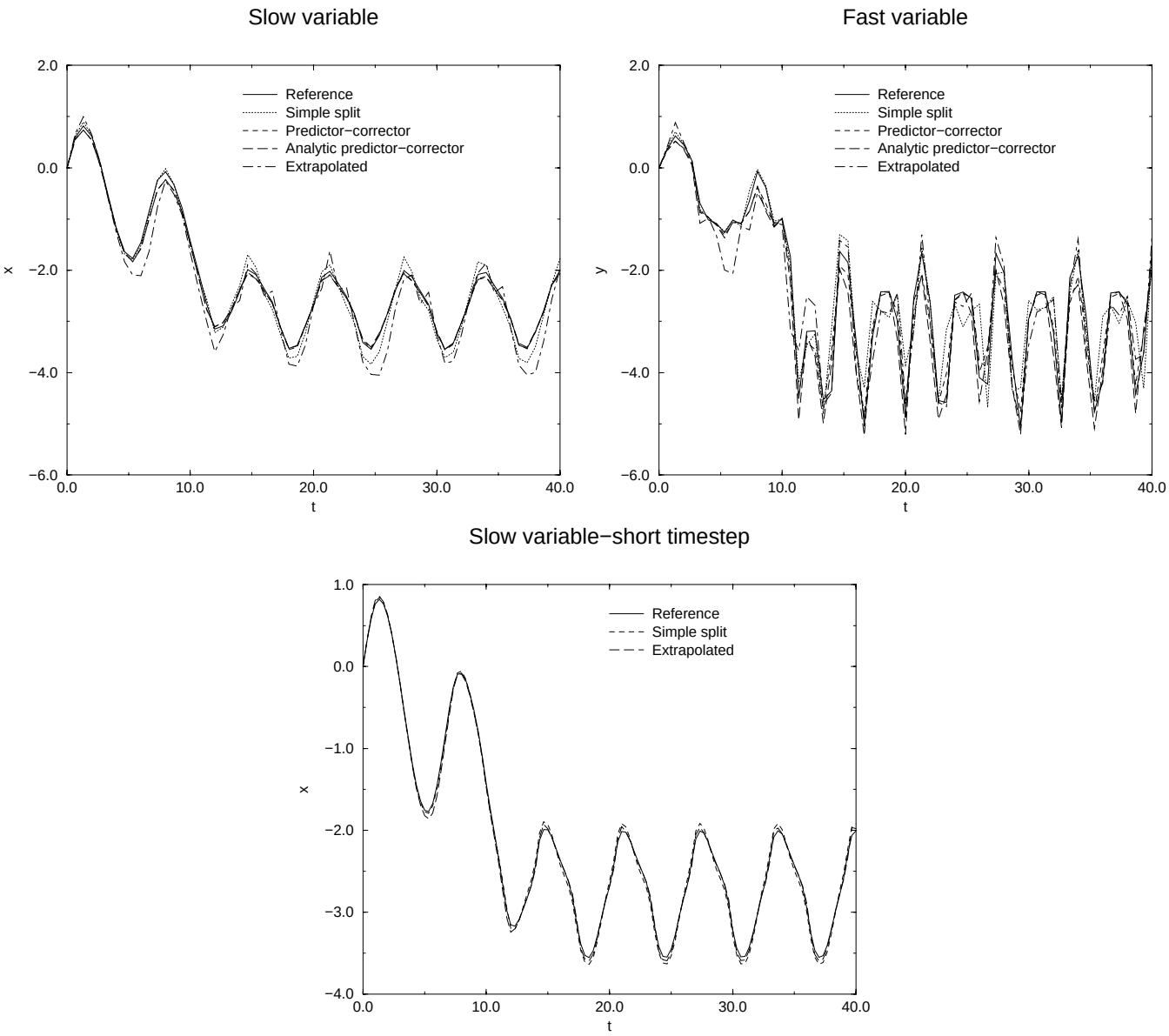

Figure 1: Graphs of x (top) and y (middle) against time for schemes (8) to (13) using time-step 0.66. Bottom graph is of x against time for schemes (8),(12) only with time-step 0.33. For details see text.

We illustrate results for this problem with initial data $x = y = 0$ and $C = 50$. The reference solution is provided using scheme (9) with a time-step of 0.066. Results for schemes (8) to (12) with a time-step of 0.66 are shown in Fig. 1. Results for schemes (8) and (12) with a time-step of 0.33 are also shown.

The results show that, early in the integration, the extrapolated scheme (12) is much the furthest from the reference, and the other schemes are all quite close. This is somewhat surprising, because scheme (8) should only be first order accurate while the extrapolated scheme is second-order accurate. However, beyond time 15, the errors in scheme (8) become as large as scheme (12). There is little to choose between schemes (9) and (10) in predicting the slow variable. However, scheme (10) is less good at predicting the fast variable. Experiments with the time-step reduced to 0.33 show that the errors in schemes (8) and (12) become comparable to those of scheme (9) and (10) with the longer time-step. Table 1 shows that reducing the time-step by a factor of 2 reduces the errors by more than the factor of 4 expected for first order accuracy and, in the case of schemes (9) and (12), by more than the factor of 16 expected for second-order accuracy. In terms of overall efficiency, scheme (12) is slightly ahead of the two predictor-corrector schemes since the gain in accuracy from reducing the time-step is more than that given by the more accurate schemes.

Table 1: Mean square differences from the reference integration of the slow variable averaged over time 0 to 40.

Time-step	Scheme (i)	Scheme (ii)	Scheme (iii)	Scheme (iv)
0.66	1.392	0.245	0.270	6.57
0.33	0.194	0.011	0.027	0.107

2.3 Interaction of parametrized forcing with the large-scale flow

In the previous subsection we illustrated the performance of different time integration schemes using a simple analytic problem. In this subsection, we use another simple model to illustrate how 'real' dynamics and parametrizations can be written in terms of multiple time-scales. The standard approach for the resolved dynamics is to treat inertio-gravity waves as 'fast' and potential vorticity controlled 'balanced' dynamics as 'slow'. We therefore use a simple balanced model with physics to illustrate which aspects of the parametrizations couple to the slow time-scale and which form part of the 'fast' dynamics.

In reality the situation is much more complex. The distinction between 'fast' and 'slow' dynamics becomes blurred on small vertical scales, where the inertio-gravity wave speed is no more than a typical advection velocity. Not all 'slow' parametrized effects couple to the balanced flow. For instance, the large-scale diurnal forcing couples most strongly to tidal motions. Similarly, some physics is much faster than the 'fast' dynamics; for instance, the physical processes within clouds excited by an internal gravity wave train act much faster than the time-scale of the waves themselves.

We use a semi-geostrophic model as a simple illustrative model. This is chosen as an example because a number of published solutions with physics are available, and the formulation of the model in primitive variables is well-suited to the inclusion of parametrized effects. Examples are given by Cullen et al. (1987) and Shutts et al. (1988). Similar conclusions would probably result from using any other model of purely balanced flow.

Following Hoskins (1975), the semi-geostrophic equations can be written in Cartesian coordinates (x,y,z), where z is a function of pressure, as

$$\begin{aligned}
\frac{D}{Dt}(u_g, v_g) + \left(\frac{\partial p}{\partial x}, \frac{\partial p}{\partial y}\right) + (-fv, fu) &= 0 \\
\frac{D\theta}{Dt} &= 0 \\
\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{u}.\nabla & \\
\nabla.\mathbf{u} = \quad 0 & \\
(fv_g, -fu_g, g\theta/\theta_0) = \quad \nabla p &
\end{aligned} \tag{13}$$

Here, $\mathbf{u} = (u, v, w)$ is the velocity, the suffix g denotes the geostrophic value, θ is the potential temperature with reference value θ_0, p the geopotential, f the Coriolis parameter, assumed constant, and g the acceleration due to gravity. As discussed by Cullen et al. (1987), one of the features of this model is that the solution is required to be statically, symmetrically and inertially stable at all times, in addition to the geostrophic and hydrostatic balance requirement. If forcing is applied to a stable state, geostrophic balance is maintained by the generation of an ageostrophic circulation. However, if forcing is applied to a marginally stable state, the response will take the form of dry mixing. This view can be formalised by writing the semi-geostrophic equations in the form of Schubert (1985). Write $\mathbf{U}_{ag}$ for the column vector with components $(u - u_g, v - v_g, w)$. Then

$$\begin{aligned}
\mathbf{Q}\mathbf{U}_{ag} + \frac{\partial}{\partial t}\nabla p = \quad & \mathbf{H} \\
\nabla.\mathbf{u} = \quad & 0 \\
(fv_g, -fu_g, g\theta/\theta_0) = \quad & \nabla p
\end{aligned} \tag{14}$$

where

$$\mathbf{Q} = \begin{pmatrix} fv_{gx} + f^2 & fv_{gy} & fv_{gz} \\ -fu_{gx} & f^2 - fu_{gy} & -fu_{gz} \\ g\theta_x/\theta_0 & g\theta_y/\theta_0 & g\theta_z/\theta_0 \end{pmatrix} \tag{15}$$

and

$$\mathbf{H} = \begin{pmatrix} -f\mathbf{u}_g \cdot \nabla v_g \\ f\mathbf{u}_g \cdot \nabla u_g \\ -g\mathbf{u}_g \cdot \nabla\theta/\theta_0 \end{pmatrix}. \tag{16}$$

We can eliminate the geostrophic pressure tendency from (14) to obtain an equation for $\mathbf{U}_{ag}$:

$$\nabla \times (\mathbf{Q}\mathbf{U}_{ag}) = \nabla \times \mathbf{H} \tag{17}$$

This is a generalised omega equation, and can be considered as a reduced equation for the full primitive equations in the sense of (2) and Browning and Kreiss (1994). It shows that the atmosphere responds to the forcing $\mathbf{H}$ by an ageostrophic circulation $\mathbf{U}_{ag}$. The amplitude of the response is determined by the 'potential vorticity' matrix $\mathbf{Q}$, being largest in the directions where $\mathbf{Q}$ has the smallest eigenvalues. If $\mathbf{Q}$ has a zero eigenvalue, the ageostrophic transport is achieved by mixing over the extent of the region where $\mathbf{Q}$ has a zero eigenvalue.

Now consider how some of the main parametrized terms can be expressed in this framework. We include a simple vertical diffusion scheme to represent boundary layer friction. The effect of latent heating is included by using equivalent potential temperature θ_E in saturated regions. The total water content r is changed by a source/sink term R. We include a radiative forcing S, some of which may take the form of a surface flux. We do not include a thermodynamic mixing term, because if static instability is generated by S, the necessary vertical mixing will be generated automatically by solving the equations. The (small) vertical mixing of thermodynamic quantities in the statically stable case is not included. Equations (13) then become

$$\begin{aligned}
\frac{D}{Dt}(u_b, v_b) + f(v_b - v, u - u_b) &= \frac{\partial}{\partial z} K_b \frac{\partial}{\partial z}(u_b - u, v_b - v) \\
\frac{D\theta_E}{Dt} + Gw &= S; r \geq r_{SAT} \\
\frac{D\theta}{Dt} + Gw &= S; r < r_{SAT} \\
\frac{Dr}{Dt} &= R \\
\nabla.\mathbf{u} &= 0 \\
(f v_b + \frac{\partial}{\partial z} K_b \frac{\partial u_b}{\partial z}, -f u_b + \frac{\partial}{\partial z} K_b \frac{\partial v_b}{\partial z}, g\theta/\theta_0) &= \nabla p.
\end{aligned} \tag{18}$$

The suffix b denotes balanced velocity components, as defined by the last equation of (18) and r_{SAT} is the saturation value of specific humidity. The boundary layer friction coefficient K_b depends only on the balanced variables. G is a non-local coefficient expressing the fact that precipitation processes generated by vertical motion at one level create thermodynamic changes at other levels. If the last equation of (18) is written in the form

$$\nabla p = \mathbf{F}(v_b, -u_b, \theta) \tag{19}$$

Then the remaining equations can be written in a form analogous to (14) as

$$\begin{aligned}
\mathbf{Q}\mathbf{U}_{ab} + \frac{\partial}{\partial t}\mathbf{F}^{-1}\nabla p &= \mathbf{H} \\
\nabla.\mathbf{u} &= 0
\end{aligned} \tag{20}$$

where $\mathbf{U}_{ab}$ is the column vector with components $(u_{ab}, v_{ab}, w) = (u - u_b, v - v_b, w)$. We interpret $\mathbf{u}$ as a total transport velocity, including both parametrized and large-scale transport. This in the spirit of Gent and McWilliams (1996), though with a different method of determining the total transport. Here

$$\mathbf{Q} = \begin{pmatrix} v_{bx} + f & v_{by} + \frac{\partial}{\partial z} K_b \frac{\partial}{\partial z} & v_{bz} \\ -u_{bx} - \frac{\partial}{\partial z} K_b \frac{\partial}{\partial z} & f - f u_{by} & -u_{bz} \\ \Theta_x & \Theta_y & \Theta_z + G \end{pmatrix} \tag{21}$$

where Θ denotes θ in unsaturated regions and θ_E in saturated regions, and

$$\mathbf{H} = \begin{pmatrix} -\mathbf{u}_b \cdot \nabla v_b \\ \mathbf{u}_b \cdot \nabla u_b \\ -\mathbf{u}_b \cdot \nabla \Theta + \frac{\partial}{\partial z} K_b \frac{\partial \Theta}{\partial z} S \end{pmatrix}. \tag{22}$$

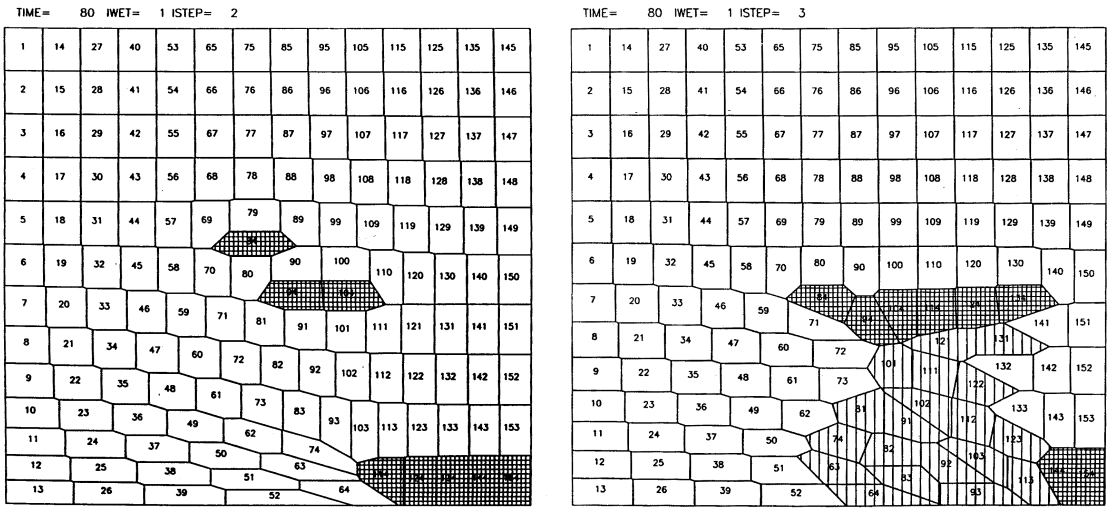

Figure 2: Fluid element pictures showing a vertical cross-section of a frontal zone with moisture. Left: the hatched elements are moist. Right: the striped elements have been cooled by precipitation falling from the convecting hatched elements.

We can then derive a generalised omega equation

$$\nabla \times (\mathbf{FQU}_{ab}) = \nabla \times (\mathbf{FH}) \tag{23}$$

Equation (23) shows that the unbalanced wind is a response to the forcing **H** which includes advection by the balanced wind and the thermal forcing. The magnitude of the response again depends on the potential vorticity matrix **Q**. We can see from (21) that the effect of friction is to increase the positive definiteness of **Q**, and thus reduce the amplitude of the response to forcing. The effect of latent heating is to increase the response. If **Q** has a region with zero eigenvalues, the response to forcing will take the form of dry mixing, moist mixing (shallow convection), or immediate penetrative convection according to the nature of the current state. The overall ageostrophic circulation includes the net transport due to the mixing or convection. An example of the penetrative convection case is shown in Fig. 2, generated by the code used in Shutts et al.(1988). The data represent a low-level frontal discontinuity. The evolution is driven by a large-scale deformation field which compresses the vertical cross-section illustrated. The hatched elements in the boundary layer are moist, and potentially unstable. In the left-hand picture, the deformation drives ascent at the front, which results in several elements convecting to a new equilibrium level. The right-hand picture shows the effect of including the cooling effect of the resulting precipitation, an example of the term G in (18). The descent of elements under the convecting elements causes enhanced low-level convergence which increases the amount of convection. However, the total mass exchange is still determined by having to maintain geostrophic balance against the effect of the large-scale convergence. This total transport is the unbalanced circulation $\mathbf{U}_{ab}$ that appears in (20).

Following the methods of the previous subsection, we now suppose that (20) acts as a reduced system for the full equations including physics. A key property of this equation is that the total unbalanced circulation is uniquely determined by the large-scale forcing. In a primitive equation model, the vertical motion, boundary layer mixing coefficients, and convective mass fluxes are all separate variables. There is thus a possibility that the same overall mass transport can be achieved by many different combinations of resolved and parametrized transports. This can be avoided if the parametrized transports are functions of 'slow' variables only, so that the total transport needed to maintain large-scale balance is then made up by resolved vertical motion. However,

if the parametrized fluxes depend on the large-scale vertical motion, or the horizontal divergence, there is a danger of multiple solutions and unstable feedback loops. Some other, more detailed requirements are listed below:

i) The boundary layer vertical diffusion of momentum must be implemented in a way that maintains Ekman balance (last equation of (18)) as a limit solution. The current ECMWF implementation, Beljaars (1991) includes this.

ii) The effect of large-scale latent heating should be incorporated in a way that is consistent with the effect of vertical motion on the dry air. This is achieved at ECMWF by averaging the increments from the cloud scheme, which are largely driven by the local vertical motion, along the trajectory in the same way as the explicit dynamics, Wedi (1999).

iii) The parametrized transports should be determined implicitly. This can be partly achieved in the predictor-corrector scheme, which is a first iteration towards a fully implicit scheme.

2.4 Implementation of a predictor-corrector algorithm using the current ECMWF parametrizations

We now seek to apply the principles of the previous two subsections to the current ECMWF parametrizations. Details of these parametrizations are set out in Gregory (2000) and references therein, together with ECMWF internal documentation available on request.

All the parametrizations contain fast processes, but the final increment obtained from each parametrization scheme should only vary on a slow time-scale, which is well-resolved by the model time-step. We use schemes (9) or (10) to integrate fast and slow processes, and use the structure of (20) to guide the separation of time-scales. The resulting scheme has to be more complex than (9) or (10) because of the presence of fast time-scales in the resolved dynamics. Essentially the procedure has to be carried through twice, though only two evaluations of each function are needed in each predictor-corrector time-step. The general procedure is:

i) The first and third equations of (9) or (10) are used to estimate increments from the dynamics and radiation.

ii) The second and fourth equations of (9) or (10) are then used to integrate the vertical diffusion, gravity wave drag, cloud and (ideally) convection. However, an implicit form of the ECMWF convection scheme is not currently available.

iii) These increments are then combined with the resolved dynamics, using the first and third equations.

iv) The fast dynamics is accounted for by using the second and fourth equations to solve for the vertical motion.

Care is required in applying this procedure with semi-Lagrangian advection. Consider the simple equation

$$\frac{Du}{Dt} = U + V \tag{24}$$

where U and V are large, so that u is a fast variable and the reduced equation is $U+V=0$. Then the natural implicit scheme is

$$u_a^{t+\delta t} - u_d^t = \frac{1}{2}\delta t \left((U+V)_d^t + (U+V)_a^{t+\delta t} \right) \tag{25}$$

The form of this equation suggests that we treat $\mu \equiv U+V$ as a slow variable, which we have to estimate

separately at time t and time $t+\delta t$. The overall scheme will then be second-order accurate for the reduced system, as required. If we assume that U is given, but that V depends on u, these estimates can be made by solving

$$\begin{aligned}(u^o-u^t)_a &= \delta t(U+V^o)_a \\ (u^O-u^{t+\delta t})_a &= u_a^O-u_d^t=\delta t(U+V^O)_a\end{aligned} \tag{26}$$

for u_a^o, u_a^O respectively, and setting $\mu_a^t=(\delta t)^{-1}(u^o-u^t)_a, \mu_a^{t+\delta t}=(\delta t)^{-1}(u^O-u^{t+\delta t})_a$. The replacement of $u_a^{t+\delta t}$ by u_d^t follows because all the contributions to Du/Dt are accounted for in $U+V$. The implicit scheme (25) is then approximated by the predictor-corrector scheme as above. Equation (26) respects the formulation of the parametrizations, represented by V, as calculations at single columns of grid-points.

We illustrate this first with the scheme used for the vertical diffusion of the u component of the wind:

$$\begin{aligned}(u^o-u^t)_a &= \delta t\left(U^t+\frac{\partial}{\partial z}K_b^t\frac{\partial u^o}{\partial z}\right)_a \\ u_a^*-u_d^t &= \frac{1}{2}\delta t\left((u^o-u^t)_d+(u^o-u^t)_a+\text{other physics}+Pu_a^*\right) \\ u_a^O-u_d^t &= \delta t\left(U^*+\frac{\partial}{\partial z}K_b^*\frac{\partial u^O}{\partial z}\right)_a \\ u_a^{t+\delta t}-u_d^t &= \frac{1}{2}\delta t\left((u^o-u^t)_d+(u_a^O-u_d^t)+\text{other physics}+Pu_a^{t+\delta t}\right)\end{aligned} \tag{27}$$

Here, U denotes increments from the resolved dynamics and/or other parametrized processes (calculated in step (i)), and P represents the semi-implicit part of the model (calculated in step (iv)). This scheme approximates the reduced equation

$$U+\frac{\partial}{\partial z}K\frac{\partial u}{\partial z}=0 \tag{28}$$

In this paper, U only includes increments from the resolved dynamics so that, in the formulation illustrated in (18), we would have $U=-\frac{\partial p}{\partial x}+fv$. We implement the gravity wave drag scheme in the same way, since part of that scheme represents low level drag due to unresolved hills. The equivalent term to U now includes the vertical diffusion increments, to prevent possible double-counting.

Similar equations to (27) represent the thermodynamic equation in the boundary layer. The first equation takes the form for the dry case

$$\theta^o-\theta^t=W^t+\frac{\partial}{\partial z}K_b^t\frac{\partial\theta^o}{\partial z} \tag{29}$$

In (20), we expect that any instability generated by the source term S will have to be compensated by vertical mixing, so that the term W needs to include radiative and surface flux terms. Since the resolved dynamics in the ECMWF model expresses conservation of θ along trajectories, it does not contribute to W.

The prognostic cloud scheme takes the form of evolution equations of the form $\frac{Dl}{Dt}=C-Dl$ for cloud water, l, and cloud amount. Here, C, D are respectively the rate of generation of cloud water and the rate of destruction

by precipitation. C is dominated by a term proportional to the vertical motion and is associated with latent heat release. We can write this as a source term Lw in the thermodynamic equation, with L being the local latent heat (water or ice). The cloud sink term Dl generates a (non-local) source term V in the thermodynamic equation associated with evaporation of precipitation. The thermodynamic equation can thus be written as $\frac{D\theta}{Dt} = L(C)w + V(Dl)$. The cloud liquid water and thermodynamic equations are integrated as follows:

$$\begin{aligned}
z_a - z_d &= \frac{1}{2}\delta t(w_d^t + w_a^t) \\
l_d^* &= l_d^t \exp(-D_d^t \delta t) + \frac{C_d^t}{D_d^t}(1 - \exp(-D_d^t \delta t)) \\
L^t &= L(C^t), V^t = V(D^t l^t) \\
\theta_a^* - \theta_d^t &= \frac{1}{2}\delta t \left(L_d^t w_d^t + L_a^t w_a^t + V_d^t + V_a^t + P\theta_a^*\right) \\
z_a^{t+\delta t} - z_d^t &= \frac{1}{2}\delta t(w_d^t + w_a^*) \\
l_a^O &= l_d^t \exp(-D_a^* \delta t) + \frac{C_a^*}{D_a^*}(1 - \exp(-D_a^* \delta t)) \\
\theta_a^{t+\delta t} - \theta_d^t &= \frac{1}{2}\delta t \left(L_d^t w_d^t + L_a^* w_a^* + V_d^t + V_a^* + P\theta_a^{t+\delta t}\right)
\end{aligned} \tag{30}$$

where the equations for calculating the vertical coordinate z_d of the departure point from the vertical velocity are included to demonstrate the overall structure, and P represents the semi-implicit correction. The reduced equation approximated by this scheme is

$$\begin{aligned}
C &= Dl \\
\frac{D\theta}{Dt} &= Lw + V(Dl)
\end{aligned} \tag{31}$$

Using the first of these equations and the fact that C is largely a function of w, the second equation can be written as $\frac{D\theta_E}{Dt} = V(w)$, where V is a non-local function of w. This is consistent with (21). The scheme (30) thus implies a second-order accurate approximation to (31) as desired.

The equations solved in the convection scheme take the form

$$u^c - u^t = \frac{1}{\rho}\delta t \frac{\partial}{\partial z}\left(M_{up}(u_{up} - \bar{u}) + M_{down}(u_{down} - \bar{u})\right) \tag{32}$$

There are similar equations for the other model variables. The mass-fluxes M_{up}, M_{down} are complex nonlinear functions of the atmospheric state. $\bar{u}$ represents an environmental value of u, while u_{up} and u_{down} are updraught and downdraught values. This scheme is formulated explicitly, so has to be integrated differently from the other parametrizations. The natural method consistent with the predictor-corrector scheme for the resolved dynamics is

$$u^* - u^t = \frac{1}{2}\delta t \left[U^t + \frac{1}{\rho^t}\frac{\partial}{\partial z}\left(M_{up}^t(u_{up}^t - \bar{u}^t) + M_{down}^t(u_{down}^t - \bar{u}^t)\right)\right]_d +$$

$$\frac{1}{2}\delta t\left[U^t+\frac{1}{\rho^t}\frac{\partial}{\partial z}\left(M^t_{up}(u^t_{up}-\overline{u}^t)+M^t_{down}(u^t_{down}-\overline{u}^t)\right)\right]_a+\delta t P u^* \tag{33}$$

$$u^{t+\delta t}-u^t=\frac{1}{2}\delta t\left[U^t+\frac{1}{\rho^t}\frac{\partial}{\partial z}\left(M^t_{up}(u^t_{up}-\overline{u}^t)+M^t_{down}(u^t_{down}-\overline{u}^t)\right)\right]_d+$$
$$\frac{1}{2}\delta t\left[U^*+\frac{1}{\rho^*}\frac{\partial}{\partial z}\left(M^*_{up}(u^*_{up}-\overline{u}^*)+M^*_{down}(u^*_{down}-\overline{u}^*)\right)\right]_a+\delta t P u^{t+\delta t}$$

where U and P play the same role as in (27). This allows the explicit vertical motion to combine with the convective environmental mass flux in achieving a total vertical transport $(w+M_{up}+M_{down})$ of each variable, though there is still an inconsistency between the semi-Lagrangian treatment of the resolved dynamics and the flux form treatment of the convective transports. The use of a predictor-corrector scheme gives the opportunity for an implicit adjustment of the convective mass flux.

We note that this formulation assumes that the mass fluxes M are slow variables, with the same time scale as the explicit vertical motion. However, M is actually chosen to allow instability in the model profile to be removed rapidly, and would thus be more naturally formulated as a fast process like the vertical diffusion. This would require a change in the formulation of the convection scheme to allow the fast changes due to convection to balance the other processes which create instability, giving an equation of the form (24). Such a scheme would give behaviour closer to the balanced solution illustrated in Fig. 2 and be more consistent with the other parametrizations.

3 Examples of performance

The analysis of the previous section shows that a more accurate time integration of the combined dynamics and parametrizations can be obtained using the predictor-corrector approach. However, the overall performance of the combined system depends also on the formulation of the parametrization schemes, and so the cost-effectiveness of the predictor-corrector scheme cannot be judged until the parametrizations have been optimised to work within this context. The results shown in this paper are direct comparisons between the operational extrapolated scheme of Wedi (1999) and the predictor-corrector scheme. They use a version of the model with T_L511 horizontal resolution and 60 levels (Cycle 23R4) tested on a set of 14 cases distributed between August 1998 and December 1999. The analyses were reruns of the operational analyses using a T_L511 model. These cases formed the second set used by Cullen (2001). The forecasts were run with both the operational 15-minute time-step and a 20-minute time-step, to see if the improved accuracy of the predictor-corrector scheme reduces the sensitivity to the time-step and, in particular, would allow a longer time-step. The cases were also run using a second iteration of the predictor-corrector scheme. If the predictor-corrector scheme is iterated to convergence, the limit scheme would be a centred fully-implicit scheme. By comparing the results from one and two iterations, we can see if convergence is occurring, and, if so, estimate the benefit that could be obtained from a fully implicit scheme. Non-convergence of the iteration would suggest that there are undesirable features of the model formulation.

Tables 2-4 shows the averaged statistics obtained over the 14 cases. The diagnostics chosen are those that are expected to be the most sensitive to model changes. We show the effect of increasing the time-step from 15 to 20 minutes in both the operational scheme and the predictor-corrector scheme; also the effect of the second iteration of the predictor-corrector scheme with a 20 minute time-step. The 'imbalance' statistic measures the ratio of the r.m.s. second time difference $D^{t+\delta t}-2D^t+D^{t-\delta t}$ to the r.m.s. divergence itself. It is rescaled by a factor $9/16$ for the 20 minute time-step so that the two time-steps can be compared as second time derivatives.

Tables 2 and 4 show that the short time variability of the divergence is reduced by using the predictor-corrector

Table 2: Northern hemisphere (30°N-90 °N) diagnostics for 10 day forecasts averaged over 14 cases. Balance and vertical motion are averaged over two tropospheric levels. Ctl denotes the control, PC the predictor-corrector integrations.

	Ctl (15m)	Ctl (20m)	PC (15m)	PC (20m)	PC (2 its)
Imbalance (%)	18	16	15	12	14
r.m.s. vertical motion (pa s^{-1})	0.14	0.14	0.15	0.15	0.15
s.d. of PMSL (hpa)	8.9	9.0	9.2	9.0	8.9
Large-scale precip (12hr accum)	0.83	0.83	0.88	0.88	0.84
Convective precip (12hr accum)	0.30	0.30	0.29	0.27	0.24

Table 3: Tropical (30°N-30 °S) diagnostics . Details as Table 2.

	Ctl (15m)	Ctl (20m)	PC (15m)	PC (20m)	PC (2 its)
Imbalance (%)	18	14	17	12	14
r.m.s. vertical motion (pa s^{-1})	0.12	0.11	0.14	0.14	0.14
Large-scale precip (12hr accum)	0.82	0.84	0.99	1.07	0.98
Convective precip (12hr accum)	1.40	1.34	1.26	1.20	1.09

scheme. This is a good feature, because if the model is properly formulated it should not have too much variability on the time-scale close to the time-step. The difference is larger with a 20-minute than a 15-minute time-step. This is also a good result. The variability is increased when a second iteration is used, though it is still less than the control in the extra-tropics. This shows that a single iteration does not fully solve for the shorter time-scales.

The amount of vertical motion is not changed in the extra-tropics. In the tropics, Table 3 shows that the convective precipitation is decreased by about 10% while the large-scale precipitation and vertical motion are increased by a similar amount to compensate. This is to be expected, since in the operational scheme the convection and cloud schemes are called sequentially, so that the convection has an advantage over the large-scale precipitation. Both the increased time-step and the second iteration decrease the large-scale as well as the convective precipitation. With the convection scheme as currently formulated, it is likely that most of the instability in the profiles will be removed by the first iteration, thus leading to an underestimate of convection at the second iteration. This illustrates the point made at the end of section 2 that the iteration is designed for a 'slow' formulation of the convective mass transport and is not well-suited to a scheme that behaves like a convective adjustment scheme. The level of synoptic activity, as measured by the standard deviation of PMSL, is increased by about 5% in the Southern hemisphere, but is hardly changed in the Northern.

The second illustration uses seasonal climate integrations using a T_L95L60 version of the model. These integrations are routinely used to test the effect of parametrization changes, as in Gregory et al. (2000). Each of the seasons is simulated using an ensemble of 3 forecasts from starting dates a day apart. In general, the total

Table 4: Southern hemisphere (30°S-90°S) diagnostics. Details as Table 2.

	Ctl (15m)	Ctl (20m)	PC (15m)	PC (20m)	PC (2 its)
Imbalance (%)	21	19	17	14	15
r.m.s. vertical motion (pa s^{-1})	0.16	0.15	0.16	0.16	0.16
s.d. of PMSL (hpa)	16.5	16.7	17.1	17.5	17.2
Large-scale precip (12hr accum)	1.08	1.08	1.18	1.15	1.12
Convective precip (12hr accum)	0.23	0.23	0.21	0.20	0.19

precipitation is increased by about 1%, but the convective precipitation is reduced by almost 10%, consistent with the high resolution results shown above. The results for the Northern hemisphere winter are shown in Fig.3. The ITCZ is shifted southwards over the East Pacific and the total rainfall is increased over the Atlantic and near the dateline. Southern Africa and an area to the east of Australia are drier. The western parts of the sub-tropical oceans in the southern hemisphere are wetter. The Northern hemisphere storm-tracks are displaced southwards. In the Northern hemisphere summer (not shown), the ITCZ is again more intense near the dateline. It is displaced southwards over the Atlantic abut northwards north of Australia. Western Europe and central and eastern North America are wetter. The southern hemisphere storm-track is displaced northward in the Australian sector. Verification against GPCP data (not shown) indicates that the differences form a significant fraction of the difference between model and climatology in a number of places, showing that the effects are significant. However, the level of agreement between model and climatology is not very different between the two sets of integrations.

The third illustration shows the convective rainfall prediction for an area in the tropics. We compare the operational scheme with the predictor-corrector scheme and the predictor-corrector scheme with a second iteration. All integrations use a 20 minute time-step. The results shown in Fig. 4 are a 24 hour accumulation between T+12 and T+36 of a forecast starting from 10 August 1999. The overall patterns are similar, since sufficient time has not elapsed to allow the large-scale patterns to diverge. The predictor-corrector scheme gives smoother fields in regions where the total convective amounts are small. This is consistent with the exclusion of the resolved dynamics from the term W in equation (29). The second iteration reduces the amount of convection substantially, especially over the ocean in the eastern half of the area shown. The reason is again the way the convection is formulated, as discussed at the end of section 2.

We finally illustrate the effect of the scheme on Northern hemisphere performance using scatter plots of the anomaly correlation of 7-day forecasts of the 500hpa geopotential. The choice of day 7 is to allow the changes to the model to become clear, as short-range forecast errors are usually dominated by the analysis. However, there is still substantial deterministic skill at this time range. The results are shown in Fig. 5. The change from the operational scheme to the predictor-corrector scheme produces on average a positive impact. The spread is quite large, so that the difference is only statistically significant at the 5% level. Results from other time-ranges (not shown) show that the predictor-corrector scheme is worse at day 3, but that statistically significant improvements are also obtained at 1000hpa at day 5. A proper assessment of the effect on 3-day forecasts cannot be made without performing a data assimilation experiment.

The effect of changing the time-step is much smaller than the change of scheme if measured in terms of spread of impacts. This is true whether the operational or the predictor-corrector scheme is used. The operational scheme is slightly more sensitive to the change in time-step, as indicated by the significance test. Results from other time ranges show that the increased time-step gives significantly worse results at day 3 in both the control and predictor-corrector experiments. In the control experiment, there are significant improvements as well as deteriorations at later times. In the predictor-corrector experiment there are no further significant impacts till day 9.

The second iteration has a larger effect than changing the time-step, though not as large as the replacement of the operational scheme by the predictor-corrector scheme. The effect is consistently good at day 3 (not shown), removing the deterioration of the predictor-corrector against the control at this time mentioned above. At day 7, the effect is not consistently good or bad and so does not suggest that a fully converged scheme would perform better than a single iteration.

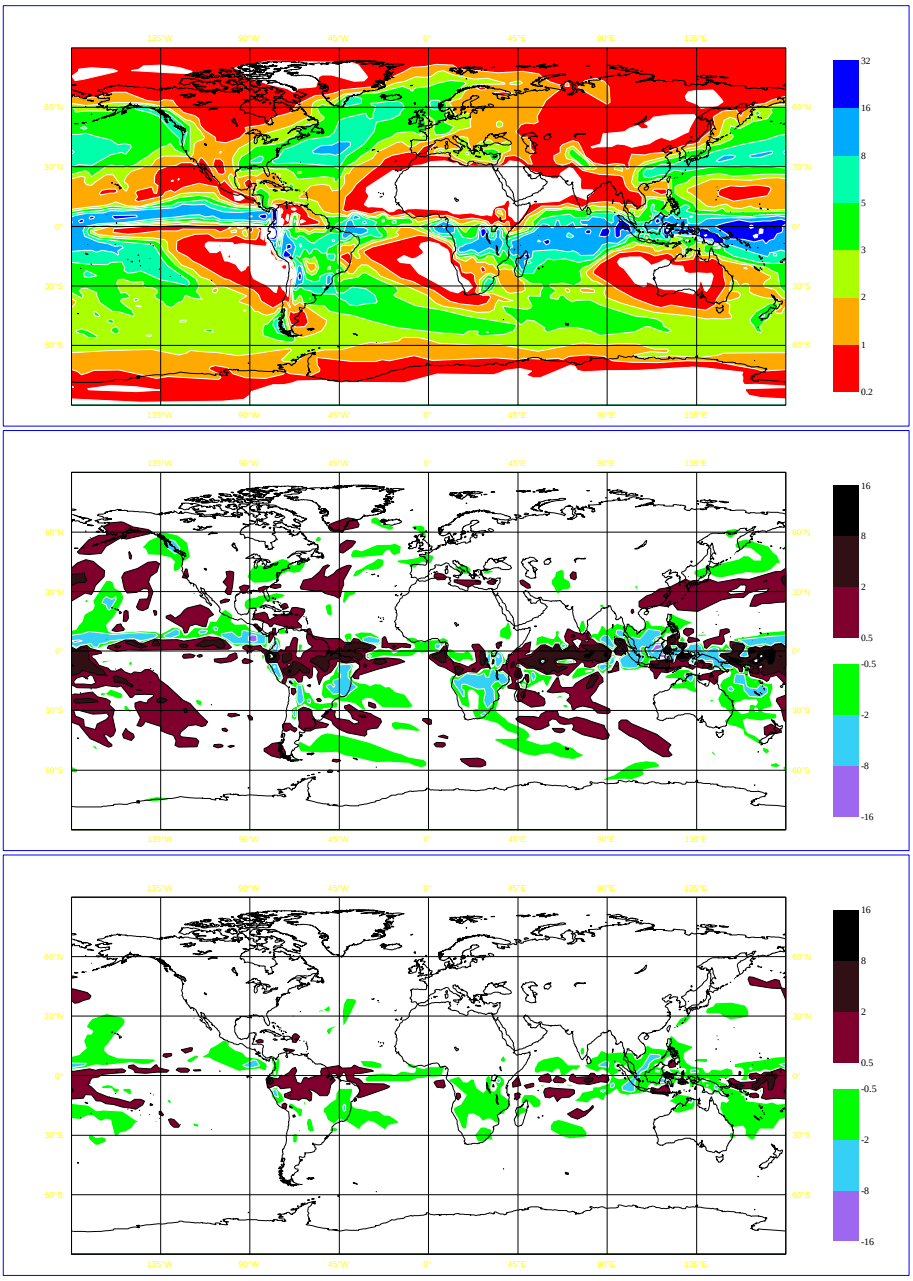

Figure 3: 3-month mean precipitation for Northern hemisphere winter. Forecasts used mean of 3 simulations for each season. Top: predictor-corrector, middle: difference predictor-corrector minus control, bottom: difference in convective precipitation predictor-corrector minus control. Units are mm/day. Contours as shown on each panel.

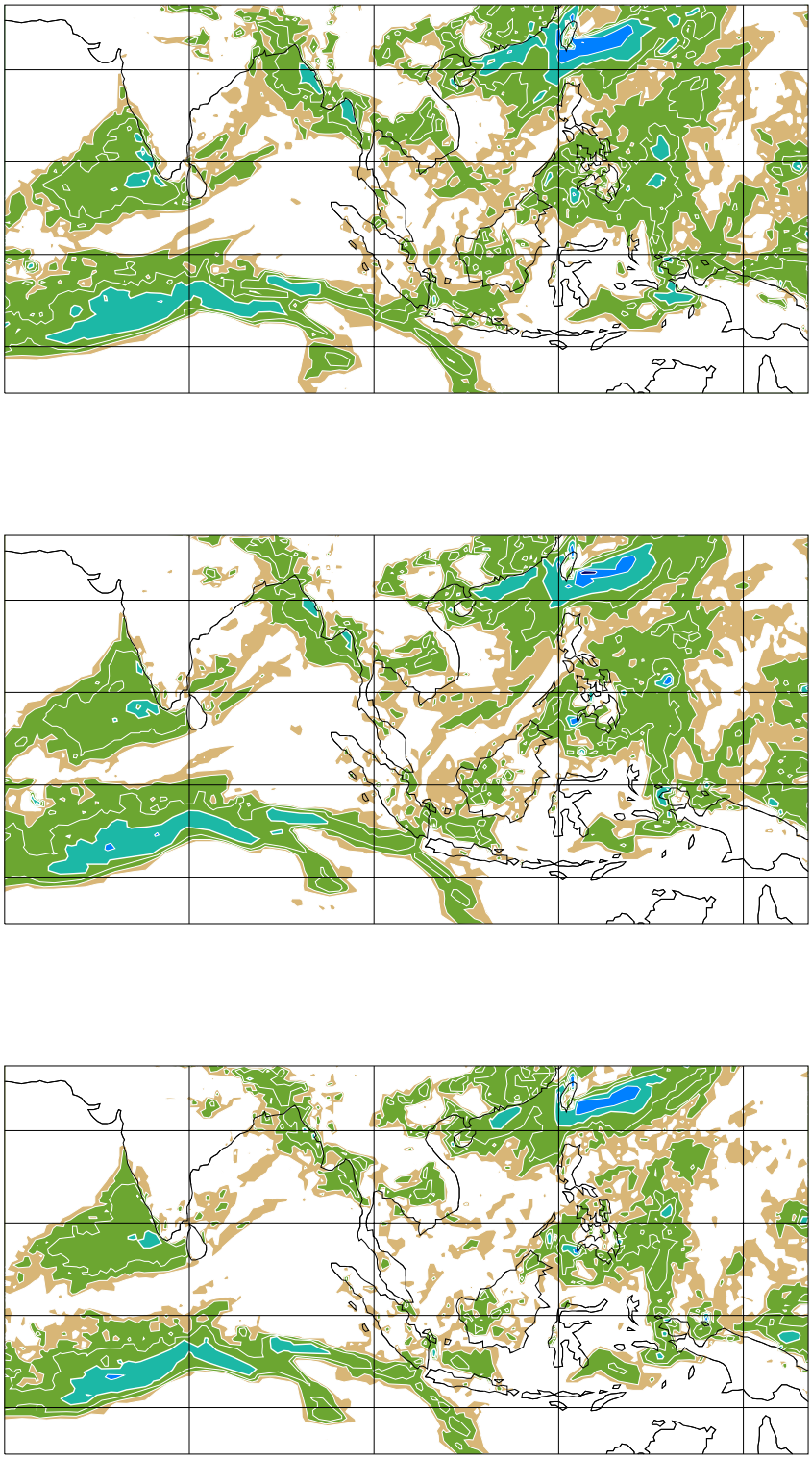

Figure 4: 24 hour accumulation between T+12 and T+36 of convective precipitation for forecasts starting from 10 August 1999. Contours at 1,2,4,8,16,32,64mm. Top: control, middle: predictor-corrector, bottom: predictor-corrector with 2 iterations.

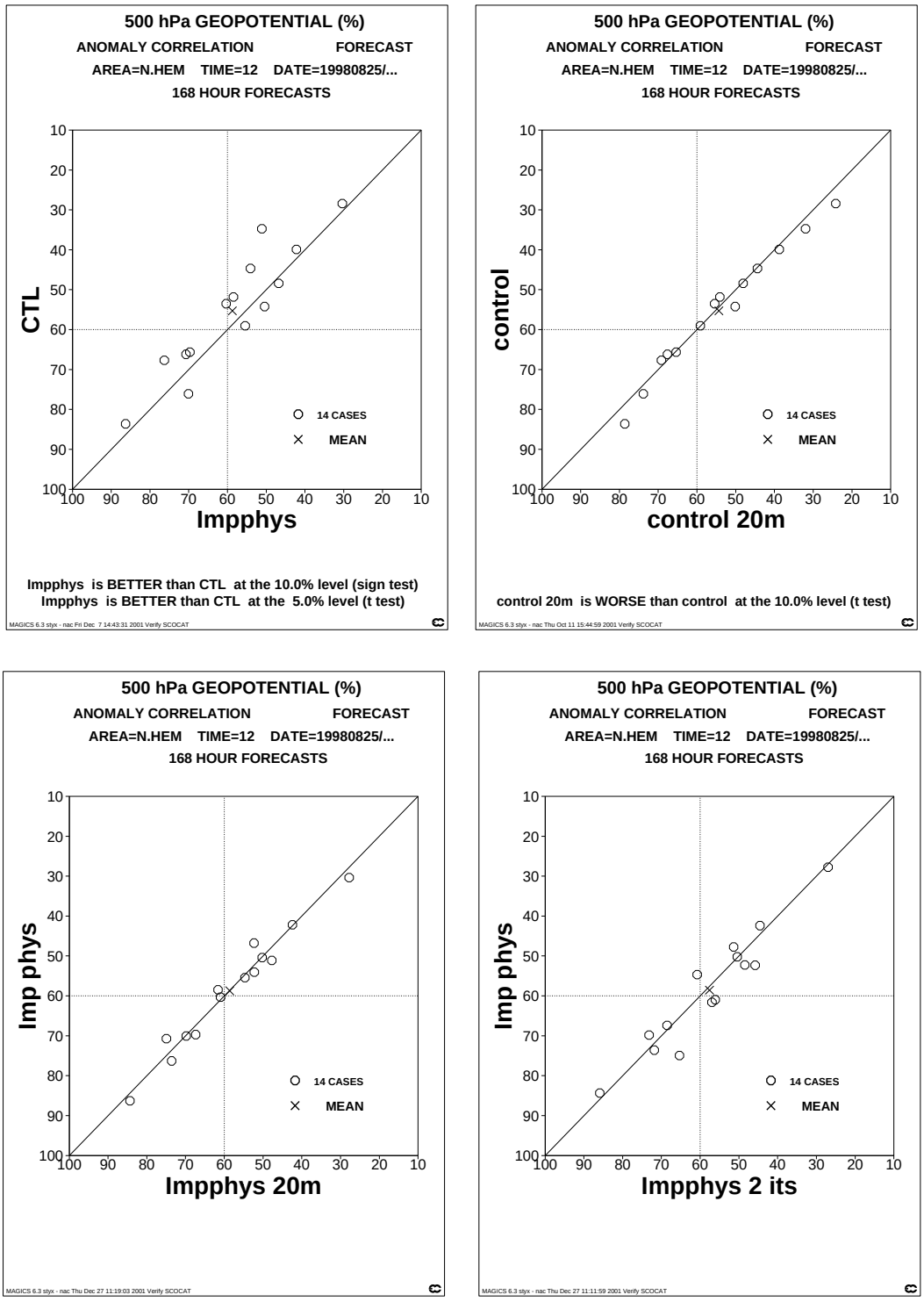

Figure 5: Scatter plots of anomaly correlations (percent) of 500hpa height over the Northern hemisphere for 14 test cases. Top left: Predictor-corrector scheme plotted against the control (15 minute time-step).Top right: 20 minute time-step control plotted against 15 minute time-step control. Bottom left: 20 minute predictor-corrector plotted against 15 minute predictor-corrector. Bottom right: Predictor-corrector with 2 iterations plotted against single iteration (15 minute time-step).

4 Discussion

We have shown that the predictor-corrector scheme is competitive in accuracy for given cost for a simple problem with multiple time-scales. We have also shown that, because parametrized processes form an integral part of the equations describing large-scale balance, it is desirable to determine the parametrized transports implicitly, along with the resolved transports. The predictor-corrector scheme provides a framework for doing this which may give it an advantage over single-step schemes with a shorter time-step.

We illustrate an implementation using current ECMWF parametrization schemes. It is clear that the replacement of the operational scheme by the predictor-corrector scheme has a significant impact on large-scale performance, which demonstrates that the integration of resolved and parametrized inputs to a model is an important area to get right. In general, large-scale scores in the Northern hemisphere are improved and others are neutral. The short-time variability is reduced. The results with a second iteration show that the results from a single iteration are still some way from those that would be obtained by a fully implicit scheme. A major factor in this is that the convection scheme is not well-suited to this framework. A different formulation of convection is required if this sort of integration scheme is to perform at its best. More generally, better performance of an iterative scheme needs parametrizations which vary smoothly with input data, as already found desirable for data assimilation applications.

Acknowledgements

The authors wish to thank many colleagues at ECMWF, particularly Christian Jakob, with help and explanations of the parametrization code and supply of diagnostic programmes. The integrations shown in Fig.2 were performed by M.W.Holt.

References

Beljaars,A.C.M. (1991): Numerical schemes for parametrizations; *Proc. ECMWF seminar on 'Numerical methods in atmospheric models*, 308-334.

Browning,G. and Kreiss,H-O. (1994): Splitting methods for problems with different time-scales; *Mon. Weather Rev.*, **122**, 2614-2622.

Cullen,M.J.P. (2001): Alternative implementations of the semi-Lagrangian semi-implicit scheme in the ECMWF model; *Quart. J. Roy. Meteorol. Soc.*, **127**, 2787-2802.

Cullen,M.J.P., Norbury,J., Purser,R.J. and Shutts,G.J. (1987): Modelling the quasi-equilibrium dynamics of the atmosphere; *Quart. J. Roy. Meteorol. Soc.*, **113**, 735-758.

Gent,P.R. and McWilliams,J.C. (1996): Eliassen-Palm fluxes and the momentum equation in non-eddy-resolving ocean circulation models; *J.Phys. Oceanog.*, **26**, 2539-2546.

Gregory,D., Morcrette,J.-J., Jakob,C., Beljaars,A.,C.,M. and Stockdale,T. (2000): Revision of convection, radiation and cloud schemes in the ECMWF Integrated Forecasting System; *Quart. J. Roy. Meteorol. Soc.*, **126**, 1685-1710.

Hoskins,B.J. (1975): The geostrophic momentum approximation and the semi-geostrophic equations; *J. Atmos. Sci.*, **32**, 233-242.

MacDonald,A. and Haugen,J. (1992): A two-time-level, three dimensional, semi-Lagrangian, semi-implicit, limited area, grid-point model of the primitive equations; *Mon. Weather Rev.*, **120**, 2603-2621.

MacDonald,A. (1998): The origin of noise in semi-Lagrangian integrations; *Proc. ECMWF seminar on 'Recent developments in numerical methods for atmospheric modelling*, 308-334.

Mahfouf,J.-F. and Rabier,F. (2000): The ECMWF operational implementation of four-dimensional variational data assimilation. II: Experimental results with improved physics; *Quart. J. Roy. Meteorol. Soc.*, **126**, 1171-1190.

Moorthi,S., Higgins,R.W. and Bates,J.R. (1995): A global multilevel atmospheric model using a vector semi-Lagrangian finite-difference scheme. Part II: version with physics; *Mon. Weather Rev.*, **123**, 1523-1541.

Schubert,W.H. (1985): Semi-geostrophic theory; *J. Atmos. Sci.*, **42**, 1770-1772.

Shutts,G.J., Cullen,M.J.P. and Chynoweth,S. (1988): Geometric models of balanced semi-geostrophic flow; *Ann. Geophysicae*, **6(5),** 493-500.

Sportisse, B. (2000): An analysis of operator splitting techniques in the stiff case; *J. Comput. Phys.*, **161**, 140-168.

Temperton,C., Hortal,M. and Simmons,A.J. (2001): A two-time-level semi-Lagrangian global spectral model; *Quart. J. Roy. Meteorol. Soc.*, **127**, 111-128.

Wedi,N.P. (1999): The numerical coupling of the physical parametrizations to the "dynamical" equations in a forecast model; *ECMWF Tech. Memo.*, no. 274..