Interaction of parametrised processes with resolved dynamics

Mike Cullen, Deborah Salmond and Nils Wedi

European Centre for Medium-Range Weather Forecasts

1. Introduction

This paper discusses the interaction of the part of the ECMWF model which is 'parametrised' with the part that consists of explicitly resolved solutions of the equations of motion. Though the parametrisations are often collectively referred to as 'physics', they include a number of different elements, as listed below.

(i) The sub-grid model, which represents the effect of unresolved dynamics on resolved dynamics.

(ii) Additional source/sink terms in the equations for temperature and moisture, resulting from radiation, cloud and precipitation physics in particular.

(iii) Boundary flux terms, which include terms that couple different sub-models, such as atmosphere to soil or atmosphere to ocean.

Each of the main processes into which the parametrisations are divided contains elements of all the above. In this paper we first discuss the sub-grid modelling problem, in particular for model resolutions similar to the T511 spectral resolution of the ECMWF model. We then discuss the interaction of parametrised processes with large-scale dynamics by using 'balanced' models, which only describe large scales. This leads to a number of useful principles which can be applied in primitive equation models. We then illustrate how integration schemes can be designed which improve the coupling of different physical processes, and in particular allow the treatment of a mixture of time-scales. We demonstrate this in the operational model. Finally we show how the formulations of the sub-grid model and resolved dynamics can be combined to give a more numerically accurate formulation.

2. The sub-grid modelling problem

The ECMWF operational model has a resolution of about 40km in the horizontal, and a timestep of 15 minutes. Figure 1 shows a plot due to Smagorinsky (1974) which shows the horizontal space and time-scales of typical atmospheric motions. We can see that the operational resolution cuts across the internal gravity wave spectrum. It allows fronts and cyclones to be well-resolved, but cuts across the scales of organised convection systems. Sub-grid modelling can only be done accurately if there is a 'spectral gap' between resolved and unresolved motions, in which case it can then be expected that a statistical treatment of unresolved motions will be accurate.

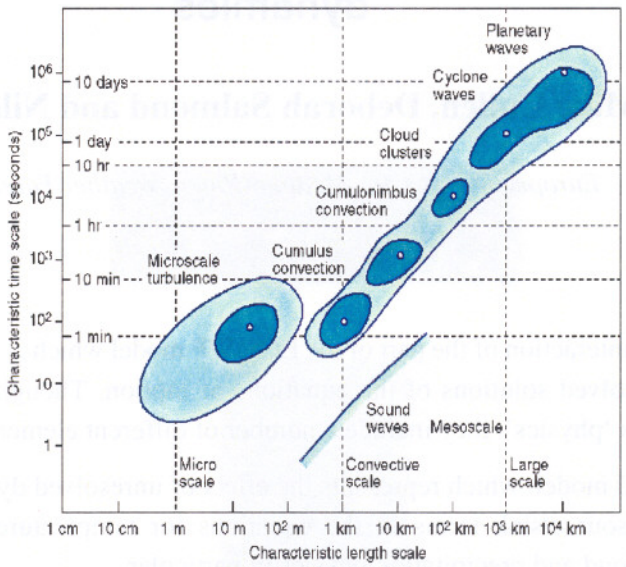

Figure 1 Characteristic scales of atmospheric motions.

However, Figure 2, from Gage and Nastrom (1985), shows that there is no spectral gap. This means that the sub-grid modelling can only be carried out imperfectly, and unsatisfactory results are inevitable for motions with scales close to the resolution of the model. Thus it is always desirable to try and increase the model resolution as much as available computer power will permit.

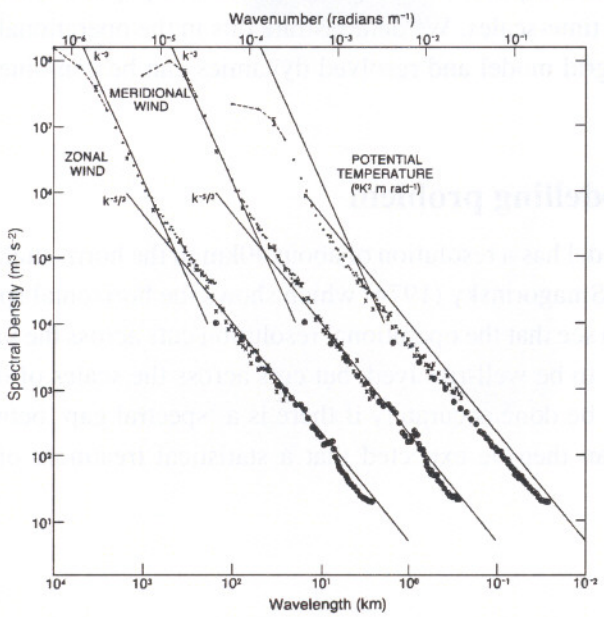

Figure 2 Graph of spectral density against wavenumber for various quantities derived from aircraft data (after Gage and Nastrom (1986).

3. Interaction of parametrised processes with large-scale dynamics

3.1 General principles

The basic principles of the interaction can be studied using the three-dimensional anelastic equations. These exclude sound waves, but describe all motions of interest to meteorologists. They can be written as

$$\frac{D\mathbf{u}}{Dt} + C_p\theta_0\nabla\Pi' + (-fv, fu, -g\theta'/\theta_0) = \mathbf{F} \tag{1}$$

$$\nabla\cdot(\rho_0\mathbf{u}) = 0$$

$$\frac{D\theta}{Dt} = \mathbf{H}$$

$$\Pi = \left(\frac{p}{p_{ref}}\right)^{\frac{R}{C_p}}$$

The flow variables $\mathbf{u}, \theta, p, \Pi$ are assumed to represent space-time averages of the velocity, potential temperature, pressure and Exner pressure. p_{ref} is a constant reference pressure. ρ_0, θ_0, Π_0 represent the density, potential temperature and Exner pressure of a horizontally uniform reference state in hydrostatic balance and ρ', θ', Π' deviations from this state. $\mathbf{F}$ and $\mathbf{H}$ represent the sub-grid model. These equations can be used to derive a second-order gravity wave equation. If we take f as constant, this is:

$$\frac{\partial^2}{\partial t^2}\left(\frac{\partial}{\partial z}\left(\frac{1}{\rho_0\theta_0}\frac{\partial}{\partial z}(\rho_0 w)\right)+\frac{1}{\theta_0}\nabla_h^2 w\right)+f^2\frac{\partial}{\partial z}\left(\frac{1}{\rho_0\theta_0}\frac{\partial}{\partial z}(\rho_0 w)\right)+\nabla_h^2\frac{gw}{\theta_0^2}\frac{\partial\theta_0}{\partial z} \tag{2}$$
$$= \mathbf{N}-\frac{\partial^2}{\partial t\partial z}\frac{\nabla_h\cdot\mathbf{F}}{\theta_0}+\frac{\partial}{\partial t}\nabla_h^2\mathbf{H}$$

$\mathbf{N}$ represents a function of the nonlinear advection terms. We can write equation (2) symbolically in the form

$$\frac{\partial^2}{\partial t^2}\mathbf{L}w + \mathbf{M}w = \mathbf{N}+\mathbf{P} \tag{3}$$

$\mathbf{L}$ and $\mathbf{M}$ are linear operators. Equation (3) is the equation for forced inertio-gravity waves. The wave frequency is determined by the eigenvalues of $\mathbf{L}$ and $\mathbf{M}$. The eigenvalues of $\mathbf{L}$ are of order $(H^{-2}+L^{-2})/\theta_0$, where H and L are vertical and horizontal scales. The eigenvalues of $\mathbf{M}$ are of order $(f^2H^{-2}+N^2L^{-2})/\theta_0$, where N^2 is the Brunt-Vaisala frequency $\frac{g}{\theta_0}\frac{\partial\theta_0}{\partial z}$. The inertio-gravity wave frequency is thus of order

$$\sqrt{\frac{(f^2H^{-2}+N^2L^{-2})}{(H^{-2}+L^{-2})}} \tag{4}$$

Equation (3) shows what the nature of the response to a physical forcing $\mathbf{P}$ will be. If the forcing frequency is comparable with the eigenvalues of $\mathbf{L}^{-1}\mathbf{M}$, then the response will be inertio-gravity waves. If it is much lower, then the response will be a 'balanced' vertical motion satisfying the equation

$$\mathbf{M}w = \mathbf{N} + \mathbf{P} \tag{5}$$

The dependence of the inertio-gravity wave frequency on horizontal and vertical length scale can be deduced from (4) and is illustrated in Figure 3 If the Brunt-Vaisala frequency is greater than the inertial frequency f, the normal case, the inertio-gravity wave frequency is very sensitive to the aspect ratio H/L if $H/L < 1$. Almost all resolved motions satisfy $H/L < 1$ in the operational ECMWF model. Also, if the Brunt-Vaisala frequency is greater than the inertial frequency f then (4) shows that the inertio-gravity wave frequency is also greater than f, approaching f for large horizontal and small vertical scales. Examples of the former are tidal motions and of the latter are quasi-inertia waves excited by the diurnal variation of the boundary layer top. In the tropics, where f is small, almost any forcing on large horizontal or small vertical scales will excite inertio-gravity waves. Only forcing acting directly on the time-mean state will elicit a balanced response. In a numerical model, it is important that the scale on which parametrised forcing is applied to the model is realistic, or the wrong type of response will result. It is easy to exaggerate the high frequency component if parametrisations 'switch' modes in a discontinuous way between time steps. Similarly, small horizontal scales may be exaggerated because parametrised increments are traditionally calculated in vertical columns, with an independent calculation for each column.

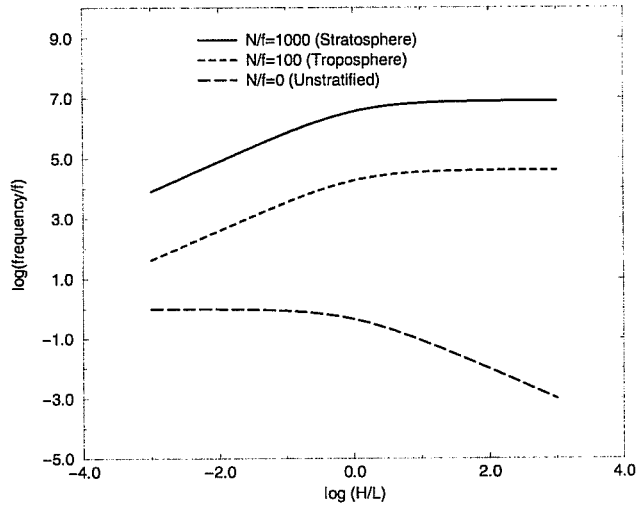

Figure 3 Logarithm of the ratio of the inertio-gravity wave frequency to the inertial frequency plotted against the logarithm of aspect ratio.

3.2 Interaction with large-scale flow

The interaction of parametrised forcing with the balanced part of the flow is difficult to get right because it relies on explicit geostrophic adjustment of increments added separately to temperature and wind fields. It is therefore useful to study the interaction directly using a balanced model. This gives valuable information about the desired solution, and also suggests ways of making this interaction more effective in operational models. We study this with a semi-geostrophic approximation to the anelastic equations (1):

$$\frac{D\mathbf{u}_g}{Dt} + C_p\theta_0\nabla_h\Pi' + (-fv, fu) = \mathbf{F}_h \tag{6}$$

$$\nabla \cdot (\rho_0\mathbf{u}) = 0$$

$$\frac{D\theta}{Dt} = \mathbf{H}$$

$$(fv_g, -fu_g, -g\theta'/\theta_0) = C_p\theta_0\nabla\Pi'$$

The final equation of this set defines the geostrophic wind and imposes hydrostatic balance. The total (including ageostrophic) wind is used in the D/Dt operator. These equations can always be solved, and require the physical forcing to interact only with the geostrophic part of the flow. As shown by Shutts and Cullen (1987), the solutions are always inertially and statically stable. This is needed for them to be a reasonable approximation to large-scale solutions of the full equations (1). An unstable state would rapidly self-destruct on small space scales under the full equations, so cannot represent a large-scale solution.

We illustrate this point in Figure 4, which is a very simple idealised model of a sea-breeze developing across a straight coastline. We aim to show that a solution of (5) describes the large-scale part of the solution of (1). The pictures show a (x, z) cross-section across the coast. The elements represent parcels of air. Those adjacent to the land surface on the left are heated with a constant sensible heat flux. There is no heat flux through the sea surface. The developing thermal contrast sets up a pressure gradient across the coast because of the hydrostatic relation. This has to be balanced by a geostrophic wind parallel to the coast. Since there is no pressure gradient in the y direction, the y component of the momentum equation in (6) becomes

$$\frac{Dv_g}{Dt} + fu = F_y \tag{7}$$

This shows that the geostrophic wind parallel to the coast has to be set up by movement of air parcels across the coastline. Figure 4shows a state generated this way, neglecting the effect of friction parallel to the coast. As the amount of heating builds up, the wind parallel to the coast increases and the depth of the circulation increases. This represents the balanced response to the thermal contrast as shown in the middle panels in Figure 4, which show the solution of (6) including friction. The actual gravity current dynamics by which the sea-breeze propagates cannot be seen by a balanced model. The solution of (essentially) (1) with the same friction is shown in the lower panels of Figure 4. We can now also see the gravity current, and inertio-gravity waves propagating upwards. These solutions are discussed in more detail in Cullen (1989).

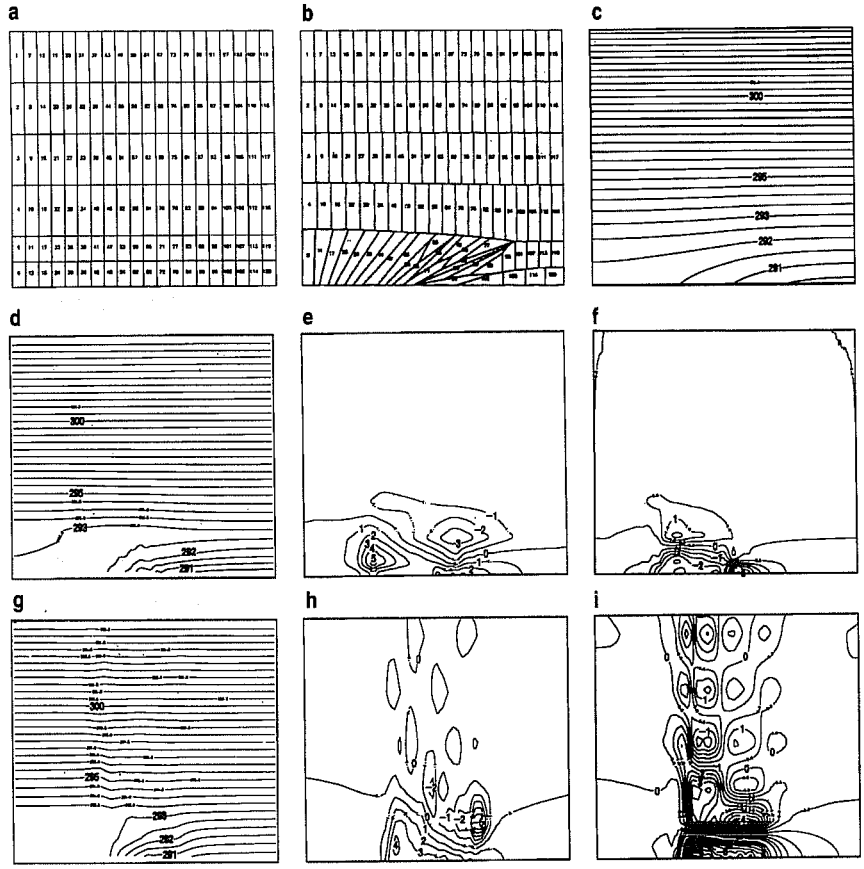

Figure 4 Solutions of idealised sea-breeze problem. (a) to (c): solution of (6) without friction in the y direction. (a) Initial data shown as fluid elements, (b) element positions after 6 hours, (c) potential temperature contours after 6 hours. Panels (d) to (f): solutions of (6) using finite-difference method after 6 hours; (d) potential temperature, (e) wind parallel to coast, (f) wind across coast. Panels (g) to (i): as (d) to (f) using hydrostatic version of equations (1).

We now study the balanced response to physical forcing in more detail by considering some key terms in $\mathbf{F}$ and $\mathbf{H}$. We assume a simple vertical diffusion scheme with mixing coefficient K, a mass-flux convection scheme with convective mass transport M_c, a diagnostic cloud and precipitation scheme where latent heat release in saturated air results from the change of saturation vapour pressure with pressure, and a simple radiative source term R. Write E for the entrainment/detrainment from the convective plumes. The potential temperature equation can then be expanded as

$$\frac{\partial \theta}{\partial t} + u\frac{\partial \theta}{\partial x} + v\frac{\partial \theta}{\partial y} + \left(w + M_c - \frac{\partial K}{\partial z}\right)\frac{\partial \theta}{\partial z} + L\frac{dq_{SAT}}{dp}\frac{\partial p}{\partial z}\left(w + M_c - \frac{\partial K}{\partial z}\right) - K\frac{\partial^2 \theta}{\partial z^2} + E = R \qquad (8)$$

Equations (6) can then be rewritten following Schubert (1985) as

$$\mathbf{Q}\begin{bmatrix} u \\ v \\ w + M_c - \dfrac{\partial K}{\partial z} \end{bmatrix} + C_p \theta_0 \frac{\partial}{\partial t} \nabla \Pi' = \mathbf{S} \tag{9}$$

where

$$\mathbf{Q} = \begin{bmatrix} f^2 + f v_{gx} & f v_{gy} + f \dfrac{\partial}{\partial z}\left(K \dfrac{\partial}{\partial z}\right) & f v_{gz} \\ (-f u_{gx}) - f \dfrac{\partial}{\partial z}\left(K \dfrac{\partial}{\partial z}\right) & f^2 - f u_{gy} & -f u_{gz} \\ \theta^{-2}\theta_{Ex} & \theta^{-2}\theta_{Ey} & \theta^{-2}\theta_{Ez} \end{bmatrix} + \mathbf{Q}_E \tag{10}$$

and

$$\mathbf{S} = \begin{bmatrix} f^2 u_g + 2fK \dfrac{\partial^2 v_g}{\partial z^2} \\ f^2 v_g - 2fK \dfrac{\partial^2 v_g}{\partial z^2} \\ R - \theta^{-2} K \dfrac{\partial^2 \theta}{\partial z^2} \end{bmatrix} \tag{11}$$

where θ_E is the moist equivalent potential temperature and $\mathbf{Q}_E$ represents a non-local matrix describing the convective plumes.

Equation (9) has solutions which can be expressed as a time-dependent geostrophic pressure Π'. The potential vorticity (PV) matrices $\mathbf{Q}$ and $\mathbf{Q}_E$ are calculated from it, together with the moisture distribution and boundary layer mixing coefficients. The matrix $\mathbf{Q}_E$ defines the convective 'geometry'. The combined matrix $\mathbf{Q}$ and $\mathbf{Q}_E$ always has non-negative eigenvalues, Shutts and Cullen (1987). This reflects the fact that the solutions of (6) must always be stable to inertial and convective instability. A 'correct' convective parametrisation for use in a balanced model must ensure this property.

Equation (9) shows that the rate of change of the geostrophic flow is driven by the forcing term $\mathbf{S}$ which includes both dynamical and physical forcing. The response to the forcing, expressed as the total wind (u, v, w), is determined by the PV matrix $\mathbf{Q}$. $\mathbf{S}$ contains geostrophic advection, radiation, which have slow time-scales, and a slowly varying part of the vertical diffusion. If $\mathbf{S}$ contains fast time scales (compared to f^{-1}), this decomposition is inappropriate, as discussed in 3.1.

We illustrate how this works in the case of convection by an example drawn from Shutts et al. (1988). The illustration, Figure 5, shows a cross-section of a frontal zone. Large-scale convergence intensifies the front, leading to local ascent. This results in saturation of the boundary layer air parcels ahead of the front. The data is chosen to be conditionally convectively unstable, so that once the air saturates it 'jumps' to a new equilibrium

position in mid-levels. The right-hand panel shows the additional effect if precipitation is taken into account. Precipitation cools the air through which it falls. The air parcels below the cloud then sink to a new equilibrium position, increasing the convergence at the front and triggering more convection. Thus in the right-hand panel, twice as much air has convected as in the left-hand panel.

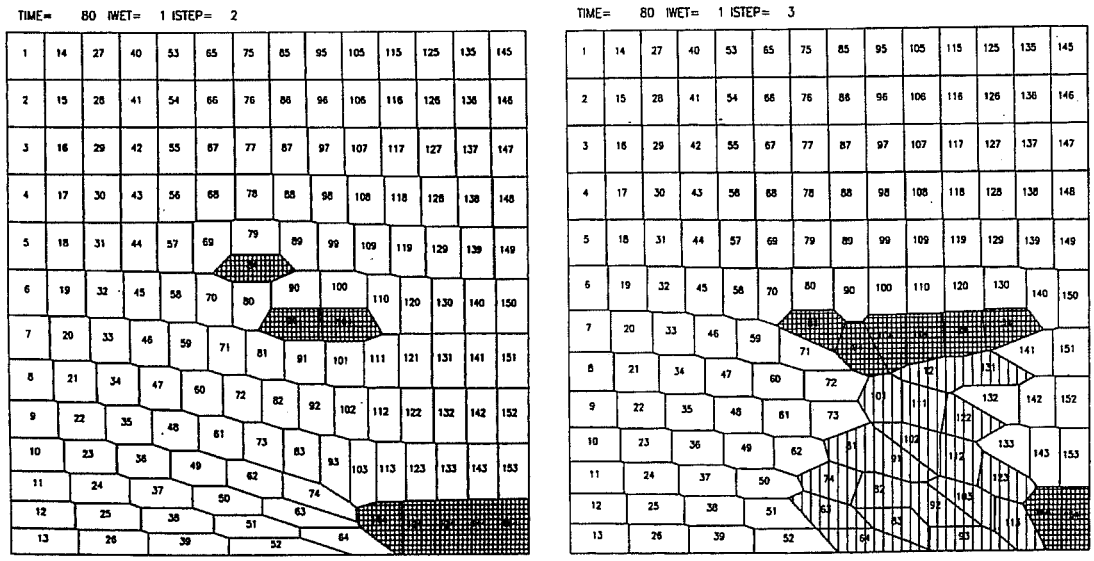

Figure 5 Fluid element pictures showing balanced cross-section of a frontal zone with moisture. Left: the hatched elements are moist. Right: the striped elements have been cooled by precipitation falling from the convecting hatched elements.

In this description, we can see that the convective mass transport is essentially determined by the large-scale forcing. Thus the convective closure should take account of the forcing terms on the right hand side of (9) as well as the convective geometry which is determined by the temperature, moisture and wind profile and expressed in the matrix $\mathbf{Q}_E$. This is not the same as the closure used in Kuo-type schemes where the closure depends on the resolved divergence. In fact, the idea is to replace, rather than supplement, the large-scale transport associated with the vertical motion w. In section 6.we will show how this can be achieved in an operational model.

Other lessons which can be learned from (9) are the need to treat terms which play a similar role in the same way numerically. In equation (8), the main effect of moisture in non-convecting cases is to reduce the static stability from $\partial\theta/\partial z$ to

$$\frac{\partial\theta}{\partial z}+L\frac{dq_{SAT}\partial p}{dp}\frac{}{\partial z}. \tag{12}$$

In a semi-Lagrangian model, the static stability is averaged along the trajectory in the dynamics, and effectively becomes $\theta_d-\theta_a$, so that the second term in (12) should be represented as $L(q_{SATd}-q_{SATa})$. The vertical motion w is replaced by the total vertical transport

$$\left(w + M_c - \frac{\partial K}{\partial z}\right) \tag{13}$$

In the dynamics, w is averaged along the trajectory, and treated implicitly in time as far as possible. Thus the convective mass transport and boundary layer mixing coefficients should also be averaged in time along the trajectory and treated implicitly.

The final example in this section shows that equation (9) can be used to predict large scale quasi-steady circulations driven by physical forcing. Figure 6is from an idealised monsoon simulation of Mawson and Cullen (1992). The circulation is forced by an upper-level heat source over Tibet. The response is constrained by the east African mountain barrier. This leads to a strong cross-equatorial jet which forms part of the circulation maintaining geostrophic balance in response to the heat source. The solution of equation (9) is illustrated, together with a simulation by a full climate model. The latter also has a mountain wave response, which is excluded by the balance approximation, but the large-scale circulation is the same.

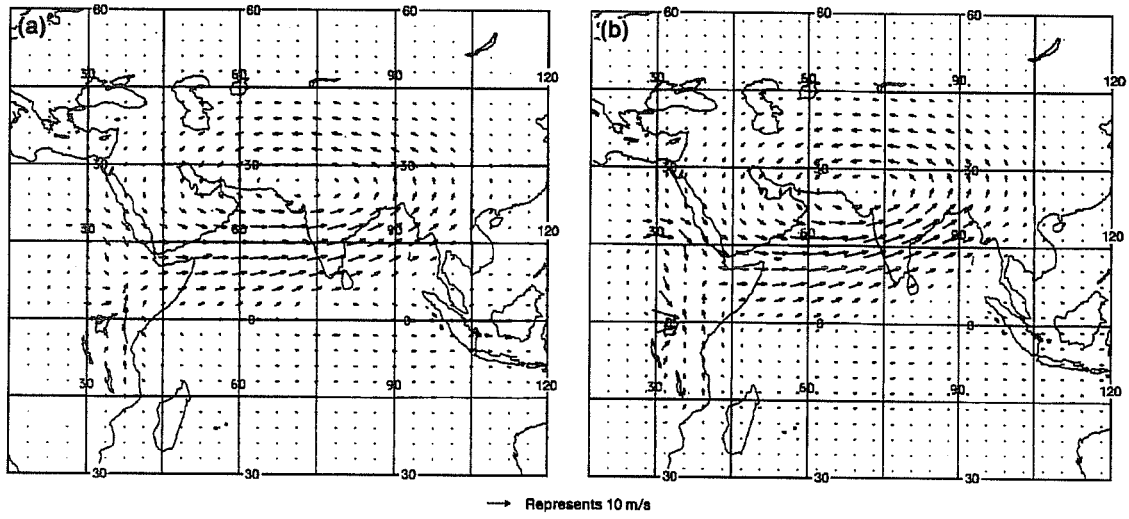

Figure 6 Surface winds after 5 days of forecast driven by idealised heat source. Left: semi-geostrophic model with physics. Right: full climate model.

4. Coupling of the physics in an operational model

We now illustrate how the coupling of the existing operational ECMWF parametrisations and the explicit dynamics already uses the ideas developed in section 3. Two key principles suggested there are:

(i) All terms in the vertical transport (13) should be calculated implicitly, using consistent space and time-averaging.

(ii) All terms in the potential vorticity matrix (10), which includes the effective static stability (12), should be calculated consistently using real atmospheric states, not partly updated ones.

The following are examples of use in the current operational ECMWF parametrisations:

(i) The vertical diffusion and gravity wave drag schemes are implicit. They take as input an estimate of the actual atmospheric profile, and also the rates of change of winds, temperatures and moisture. The output is an updated rate of change, thus ensuring consistency between the boundary layer 'transport' term $\partial K/\partial z$ and the explicit vertical motion.

(ii) The cloud and precipitation scheme also takes in an estimate of the actual atmospheric profile, together with the cloud source terms. The most important of these is the large-scale vertical motion. The temperature increments from the cloud scheme include the large term $wL\frac{dq_{SAT}}{dp}\frac{\partial p}{\partial z}$. The increments are averaged in space and time along the trajectory as described in Wedi (1999), which is consistent with the way the resolved term $w\frac{\partial \theta}{\partial z}$ is calculated.

(iii) The increments from the convection scheme are also averaged in space and time along the trajectory. Thus there is consistency in the treatment of w and M_c in (13). However, there is an inconsistency in the spatial differencing used. The resolved term is treated by semi-Lagrangian methods and the convective term by Eulerian flux differencing.

Creating a fully implicit version of the model would be a major undertaking. We therefore demonstrate the impact obtainable by using a predictor-corrector version of the model, which can be viewed as a first iteration of a fully implicit scheme. This method is described by Cullen (2000). For a simple equation

$$\frac{\partial u}{\partial t} = F \tag{14}$$

the operational integration scheme can be represented as

$$u^{t+\delta t} = u^t + \delta t\left(\frac{3}{2}F^t - \frac{1}{2}F^{t-\delta t}\right) \tag{15}$$

The predictor-corrector scheme can be represented as

$$\begin{aligned} u^* &= u^t + \delta t F^t \\ u^{t+\delta t} &= u^t + \frac{1}{2}\delta t(F^t + F^*) \end{aligned} \tag{16}$$

This scheme assumes that the time-scale of F is comparable to or greater than the timestep. However, the physics includes some very fast processes which are calculated implicitly within the current algorithms. We illustrate how to deal with this by considering the generic equation for a single physical process:

$$\frac{\partial s}{\partial t} + D + F + S = 0 \tag{17}$$

Here s is a prognostic variable, D represents all other processes, F represents the fast part of a process and S the slow part. A time-discretisation of this consistent with (16) is:

$$
\begin{aligned}
& s^{+} - s^{t} + D^{t} + F^{+} + S^{t} = 0 \\
& s^{*} - s^{t} + \frac{1}{2}(D_d^t + D_a^t + F_d^+ + F_a^+ + S_d^t + S_a^t) = 0 \\
& s^{*+} - s^{t} + D^{*} + F^{*+} + S^{*} = 0 \\
& s^{t+\delta t} - s^{t} + \frac{1}{2}(D_d^t + D_a^* + F_d^+ + F_a^{*+} + S_d^t + S_a^*) = 0
\end{aligned}
\tag{18}
$$

The first and third equations of (18) are solved at individual grid points, using implicit time integration. The latest available estimate of the slow increments D and S is used. Increments F calculated this way then become 'slow', because the fast time-scales have been accounted for. The result is then averaged in a second-order way in space and time according to the last equation in (18), which has the same form as the second equation of (16). Note that the first and third equations are like conventional local grid-point calculations of physical effects. All the internal consistency built into modern parametrisation systems is thus preserved. Only the final totals are averaged in a way that is consistent with the dynamics, thus obtaining overall second-order accuracy.

We now demonstrate the effect of this using a simple analytic example. Consider an equation for a single variable s as follows:

$$
\begin{aligned}
& \frac{\partial s}{\partial t} + A + Bs = 0 \\
& A = \cos(t + 4s) \\
& B = 10 + 5\cos(3s) \\
& s(0) = 0.1
\end{aligned}
\tag{19}
$$

B is chosen so that it always implies a fast time scale compared with A. We illustrate integrating (19) with the following schemes:

(i) Centred implicit

$$
s^{t+\delta t} = s^{t} - \delta t(A^{t} + B^{t} s^{t+\delta t}) \tag{20}
$$

(ii) Predictor-corrector

$$
\begin{aligned}
& s^{*} = s^{t} - \delta t(A^{t} + B^{t} s^{*}) \\
& s^{+} = s^{t} - \delta t(A^{*} + B^{*} s^{+}) \\
& s^{t+\delta t} = \frac{1}{2}(s^{*} + s^{+})
\end{aligned}
\tag{21}
$$

(iii) Analytic predictor-corrector

$$s^* = s^t \exp(-B^t \delta t) - \frac{A^t}{B^t}(1-\exp(-B^t \delta t)) \tag{22}$$
$$s^+ = s^t \exp(-B^* \delta t) - \frac{A^*}{B^*}(1-\exp(-B^* \delta t))$$
$$s^{t+\delta t} = \frac{1}{2}(s^* + s^+)$$

(iv) Extrapolated

$$s^- = s^{t-\delta t} - \delta t(A^{t-\delta t} + B^{t-\delta t} s^-) \tag{23}$$
$$s^* = s^t - \delta t(A^t + B^t s^*)$$
$$s^{t+\delta t} = s^t + \frac{3}{2}(s^* - s^t) - \frac{1}{2}(s^- - s^{t-\delta t})$$

Scheme (ii) is essentially that used in the vertical diffusion scheme and scheme (iii) is that used in the cloud scheme. Figure 7shows the results with a timestep of 1.0. The reference solution uses scheme (i) with a timestep of 0.1.

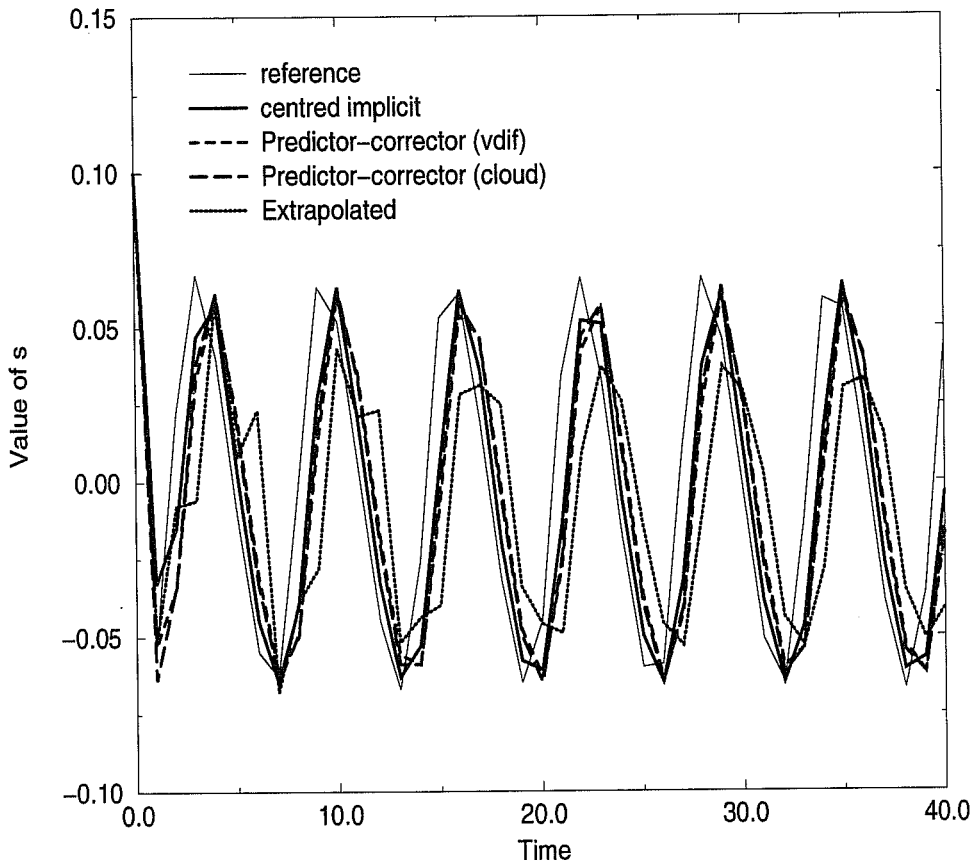

Figure 7 Solutions of (19) with different algorithms.

These results show that there is little to choose between schemes (i), (ii) and (iii). The error is dominated by the choice of timestep. Scheme (i) is difficult to use in an operational model because it leads to coupling between all variables at all levels and requires all processes to be computed simultaneously. However, the inability to use new time-level values of the nonlinear coefficients A and B in this scheme does not appear to introduce much error. Scheme (iv) achieves second-order accuracy in time without iteration by extrapolating the tendencies of s. However, it is clearly less accurate.

Some implementations of parametrisations work on the basis that the increments are not closely tied to the starting values and can be interpolated or extrapolated between time steps. We show that this is not the case for equation (19) by solving with scheme (iii) with the last two equations of (22) replaced by

$$s^{+} = s^{*}\exp(-B^{*}\delta t) - \frac{A^{*}}{B^{*}}(1-\exp(-B^{*}\delta t)) \tag{24}$$

$$s^{t+\delta t} = s^{t} + \frac{1}{2}(s^{*} - s^{t} + s^{+} - s^{*})$$

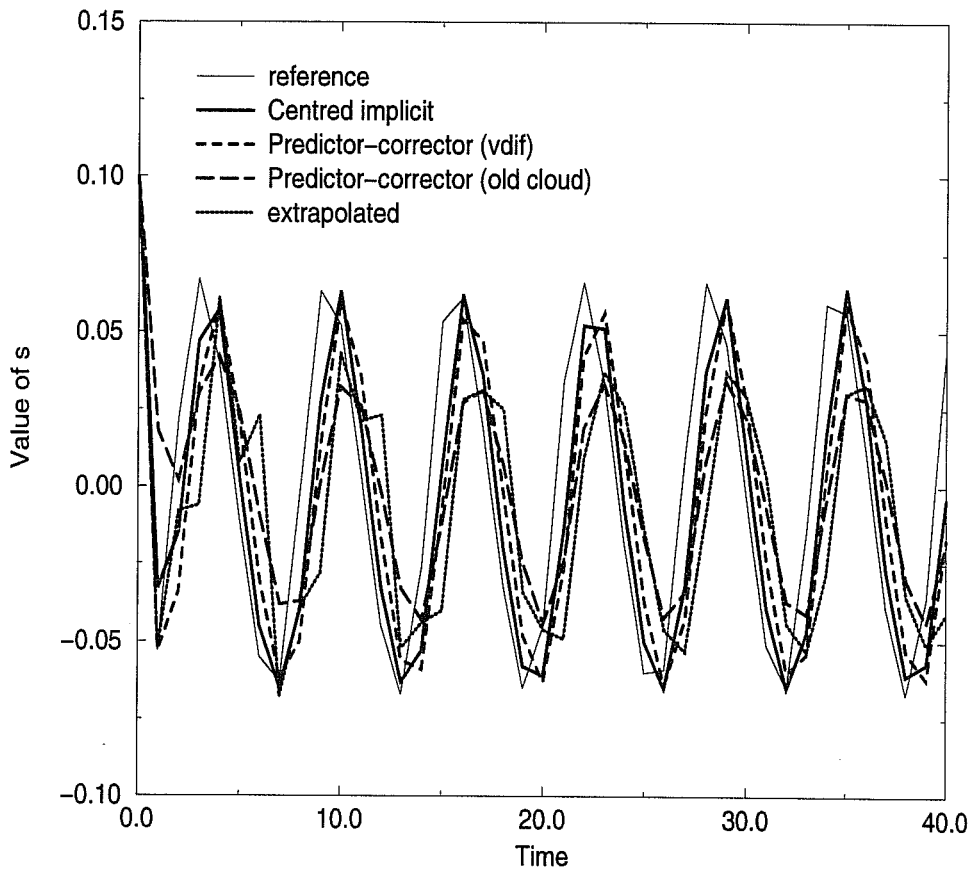

Figure 8 As Figure 7 with equation (24) replacing equation (22).

The result is illustrated in Figure 8, in which scheme (iii) is modified while the remainder of the schemes are as in Figure 7 The results are clearly inferior to those when (22) is used, showing that, in this case, increments

cannot be detached from their starting values. This may also explain the inferior performance of scheme (iv), and is because equation (19) is diffusive with a fast time scale. Scheme (24) is the same as (22) for an equation that can be solved explicitly rather than implicitly.

We now illustrate how to implement (18) within the context of the ECMWF physics, taking account of the results shown in Figure 7 and Figure 8. Only a few of the main equations can be illustrated here. A typical equation to be solved in the vertical diffusion scheme is

$$u^{t+\delta t}-u^t-\delta t\left(U+\frac{\partial}{\partial z}K(z)\frac{\partial u^{t+\delta t}}{\partial z}\right)=0 \tag{25}$$

This equation has to be solved implicitly because K is large relative to the vertical grid length. The vertical diffusion is called after the radiation but before the rest of the physics. This is to ensure that the correct inputs and outputs are available. Scheme (ii) can then be implemented as follows: in the predictor step set

$$u^+-u^t-\delta t\left(U^t+\frac{\partial}{\partial z}K^t(z)\frac{\partial u^+}{\partial z}\right)=0 \tag{26}$$

$$V^+=u^+-u^t-\delta tU^t$$

U^t should be the tendency from all other processes calculated at time t evaluated at the grid point. Since this are not all available when the scheme is called, only the pressure gradient, Coriolis term, and radiation increment are used. These are much the largest contributions. V^+ is the vertical diffusion increment. In the corrector step set

$$u^{*+}-u^t-\delta t\left(U^*+\frac{\partial}{\partial z}K^*(z)\frac{\partial u^{*+}}{\partial z}\right)=0 \tag{27}$$

$$V^{*+}=u^{*+}-u^t-\delta tU^*$$

$$u^{t+\delta t}-u^t-\frac{1}{2}\delta t(W_d^t+W_a^*+V_d^++V_a^{*+})=0$$

where U^* is the sum of the pressure gradient, Coriolis term, and radiation increment calculated at the grid point at time $*$, V^{*+} is the vertical diffusion increment estimated at time $t+\delta t$, and W represents increments from all processes other than vertical diffusion.

The implementation of the gravity wave drag scheme is similar. The first equations of (26) and (27) are replaced by

$$\begin{aligned}u^+-u^t-\delta t(U^t+A^t|u|^tu^+)&=0\\ u^{*+}-u^t-\delta t(U^*+A^*|u|^*u^{*+})&=0\end{aligned} \tag{28}$$

and the remaining equations are similar. A is the drag coefficient.

A typical equation from the convection scheme can be written

$$\frac{\partial u}{\partial t}=\frac{1}{\rho}\frac{\partial}{\partial z}[M_{up}u_{up}+M_{down}u_{down}-(M_{up}+M_{down})\bar{u}] \tag{29}$$

M_{up}, M_{down} represent updraught and down draught mass fluxes. u_{up}, u_{down} are updraught and down draught horizontal velocities, and $\bar{u}$ is the environmental horizontal velocity. If M_{up} and M_{down} are regarded as velocities, the tendencies in (29) can be evaluated explicitly. Since they are advective, rather than diffusive, they can be safely averaged in time as in equation (24), experiments show that this gives the best results. An equivalent term from the explicit vertical momentum equation in flux form would be $\frac{\partial}{\partial z}(wu)$. This is averaged in space and time along the trajectory using the semi-Lagrangian dynamics. A consistent method for (29) would therefore be to define convection increments

$$C^{+} = \frac{1}{\rho^{t}}\frac{\partial}{\partial z}[M^{t}{}_{up}u^{t}{}_{up} + M^{t}{}_{down}u^{t}{}_{down} - (M^{t}{}_{up} + M^{t}{}_{down})\overline{u^{t}}] \tag{30}$$

$$C^{*+} = \frac{1}{\rho^{t}}\frac{\partial}{\partial z}[M^{*}{}_{up}u^{*}{}_{up} + M^{*}{}_{down}u^{*}{}_{down} - (M^{*}{}_{up} + M^{*}{}_{down})\overline{u^{*}}]$$

These are then applied averaged in space and time as in (27). The convective calculation determines M_{up}, M_{down} from the CAPE in the profiles at time t and time * respectively. In order to maintain a steady state, the amount of CAPE present in the profiles must be related to the amount generated in a timestep. Thus if the scheme is designed to remove all the CAPE in a time interval T, and an amount C is generated during a timestep, the profiles must contain an amount of CAPE equal to $CT/\delta t$. There is a danger that this could allow explicit convection to take over. A remedy is to allow the closure to use also the amount of CAPE generated during the timestep. In the current operational scheme this is included, but at the expense of using partly updated profiles to calculate the cloud geometry.

A typical equation in the cloud scheme is the evolution equation for cloud liquid water

$$\frac{\partial l}{\partial t} = C - Dl \tag{31}$$

where C, D are respectively the rate of generation of cloud water and the rate of destruction by precipitation. The time scale associated with D is very short and the equations are therefore integrated analytically in time as illustrated in equation (22), giving

$$l^{*} = l^{t}\exp(-D^{t}\delta t) + \frac{C^{t}}{D^{t}}(1 - \exp(-D^{t}\delta t)) \tag{32}$$

$$l^{+} = l^{t}\exp(-D^{*}\delta t) + \frac{C^{*}}{D^{*}}(1 - \exp(-D^{*}\delta t))$$

The increments $l^{*} - l^{t}, l^{+} - l^{t}$ are then averaged in space and time as for the convection increments. The source term C is dominated by the term $wL\frac{dq_{SAT}}{dp}\frac{\partial p}{\partial z}$ which combines with $w\frac{\partial \theta}{\partial z}$ to represent the reduced static stability in the presence of cloud. The space-time averaging of (32) is consistent with the space time averaging of $w\frac{\partial \theta}{\partial z}$ in the dynamics.

We illustrate the performance of this type of interfacing by showing the effect on various diagnostics from the model. 14 forecasts at T511L60 resolution were run from T511 analyses for various dates between August 1998 and December 1999. The diagnostics shown in Table 5 . were computed from day 10 forecasts and averaged over all the cases. The 'balance' diagnostic is the ratio expressed as a percentage of

$D^{t+\delta t} - 2D^t + D^{t-\delta t}$, where D is the horizontal divergence, to D^t itself. The control and predictor-corrector schemes were tested with 15 and 20 minute time-steps in order to see if the time-step sensitivity is reduced by the predictor-corrector scheme. The balance diagnostic is rescaled by the square of the time step to make the 15 and 20 minute results comparable.

TABLE 5 . DIAGNOSTICS FROM 14 CONTROL AND PREDICTOR-CORRECTOR 10-DAY FORECASTS

	Control 15m time step	Control 20m time step	Predictor-corrector 15m time step	Predictor-corrector 20m time step	2nd corrector iteration 15m time step
		NORTHERN	HEMISPHERE		
Balance (%)	18	16	16	13	18
r.m.s. vertical motion (pa s^{-1})	.14	.14	.15	.15	.15
SD of PMSL (hpa)	8.9	9.0	9.1	9.2	9.0
Large-scale precip (12hr accum)	.83	.83	.86	.88	.83
Convective precip (12 hr accum)	.30	.30	.29	.28	.25
		TROPICS			
Balance (%)	18	14	18	12	22
r.m.s. vertical motion (pa s^{-1})	.12	.11	.13	.12	.14
Large-scale precip	.82	.84	.95	.99	.91
Convective precip	1.40	1.34	1.24	1.20	1.15
		SOUTHERN	HEMISPHERE		
Balance (%)	21	19	18	15	19
r.m.s. vertical motion (pa s^{-1})	.16	.15	.16	.16	.16
SD of PMSL (hpa)	16.5	16.7	17.2	17.6	17.3
Large-scale precip	1.08	1.08	1.08	1.10	1.10
Convective precip	.23	.23	.22	.21	.20

Table 5 . shows that the predictor-corrector scheme does reduce the small time-scale variability, as measured by the balance diagnostic. Balance is also improved with a longer time-step as not as many small time-scales are resolved. The change in the results between 15 and 20 minute time-steps is small for both schemes.

Considering all the diagnostics together, neither scheme is more sensitive. There is a transfer of about 10% of the tropical convective precipitation to large-scale precipitation when the predictor-corrector scheme is used. This is probably because of the difference in the convective closure discussed above. This change is within acceptable bounds. There is greater synoptic activity in the extra-tropics, as measured by the standard deviation of PMSL. This is particularly true in the southern hemisphere winter, and suggests that some damping present in the operational scheme has been removed. The effect of a second corrector iteration is mostly seen in the precipitation and tropical dynamics. This is most likely to be caused by switching in the precipitation parametrisations, and suggests that a smoother formulation is needed for good performance of a fully implicit scheme. Similar conclusions apply in data assimilation, where physics increments have to be perturbed. The overall performance is illustrated in Figure 9 and Figure 10 There is a positive impact in the Northern hemisphere at 500hpa, except at the very end of the forecasts where skill has been lost. This also occurs at other

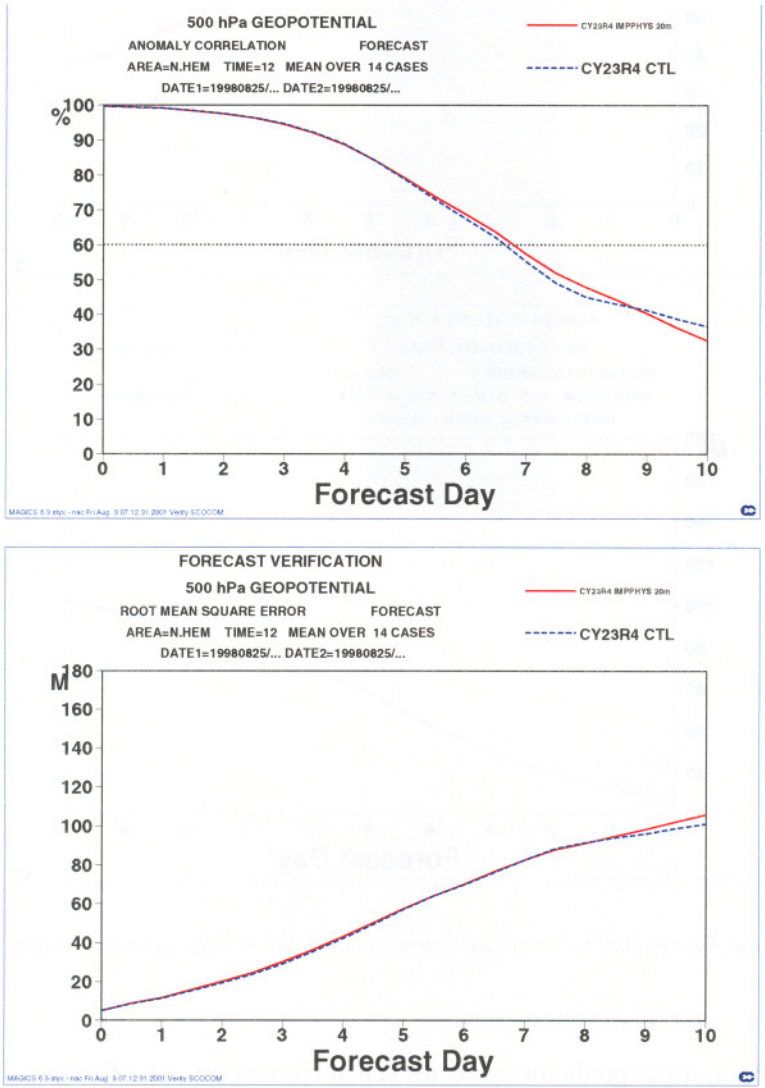

Figure 9 Anomaly correlation and r.m.s error comparing 14 control forecasts with forecasts using a longer timestep and a predictor-corrector scheme for integrating the dynamics and physics.

levels (not shown). The effect is larger than that obtained by simply using the predictor-corrector scheme for the dynamics, and the impact on individual cases is also much larger. This suggests that it is very important to establish the optimum method of integrating the combined dynamics and physics in time.The effect of the second iteration on the scores is very small, much less than the effect of using the predictor-corrector scheme itself. This suggests that most of the benefit of a fully implicit scheme could be obtained by the predictor-corrector scheme.

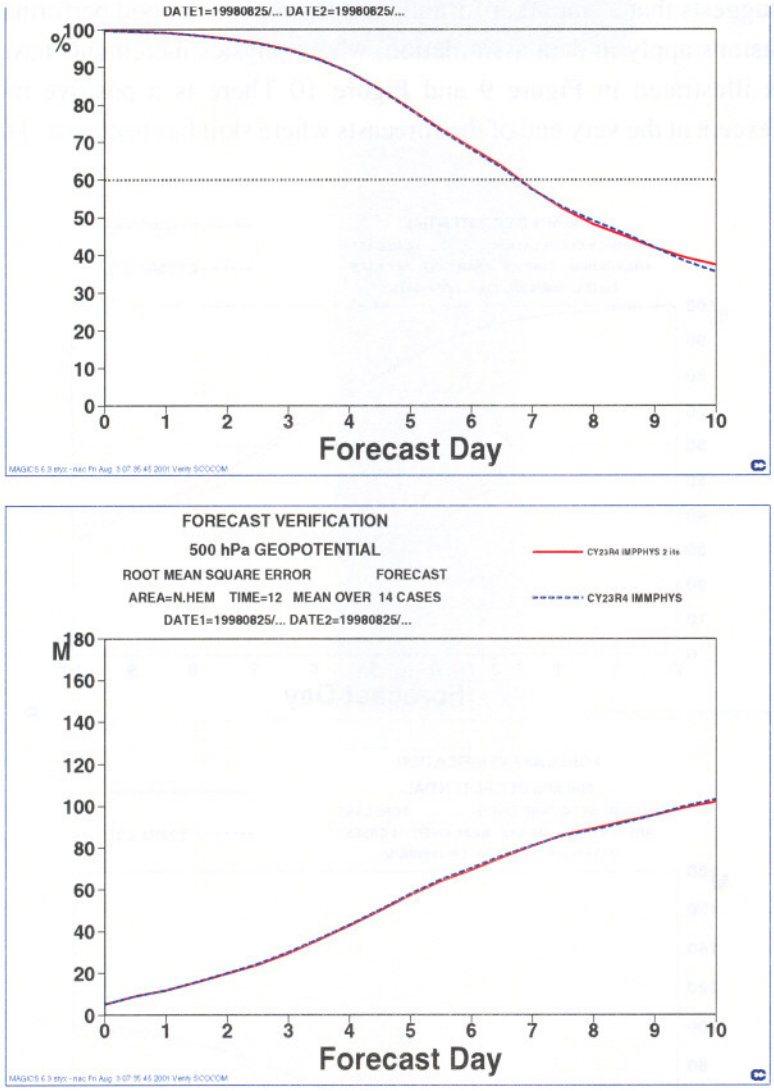

Figure 10 Verification figures for the predictor-corrector scheme with a second iteration compared with a single iteration.

We also illustrate the effect of the predictor-corrector scheme on a particular forecast of an extreme synoptic case, the French storms of December 1999. Forecasts of this case were very sensitive to analysis and forecast technique, and so the case indicates the maximum likely sensitivity of an individual forecast to this type of change. In practice, increased sensitivity would also come from using the changed formulation in the

assimilation cycle as well. Figure 11shows that the depth of the depression is significantly increased, and the position improved.

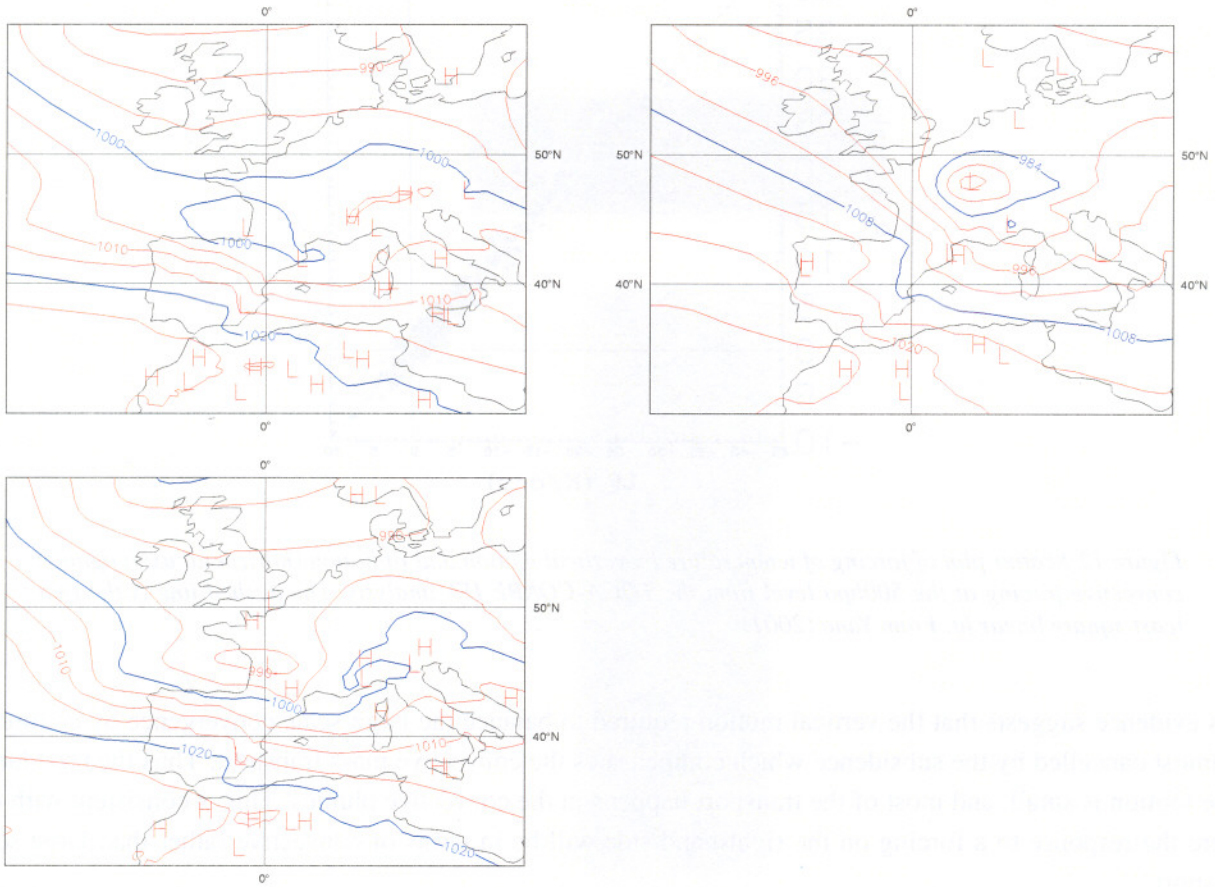

Figure 11 PMSL forecasts from 12 UTC 24 December 1999 verifying at 00UTC 28 December 1999. Top left: control forecast; bottom left: forecast using predictor-corrector scheme; top right: verifying analysis.

6. Blending of the sub-grid model into the resolved dynamics

The preceding section has demonstrated how the integration of the dynamics and physics in time can be made more consistent, and how benefit results both in reduced short-time variability and better overall performance. However, it is clear that the differing formulations used for different physical processes makes this consistency quite difficult to achieve. In this section, we illustrate how a closer coupling of the dynamics and physics can be achieved in the case of deep convection. This is motivated by observations in the tropics which show a close balance between the temperature and moisture changes due to convection and those due to large-scale vertical motion. Figure 12, after Yano (2001), shows the balance between the terms $w\frac{\partial\theta}{\partial z}-R$ (called large scale forcing) and $M_c\frac{\partial\theta}{\partial z}+E$ (called convective forcing) from (8).

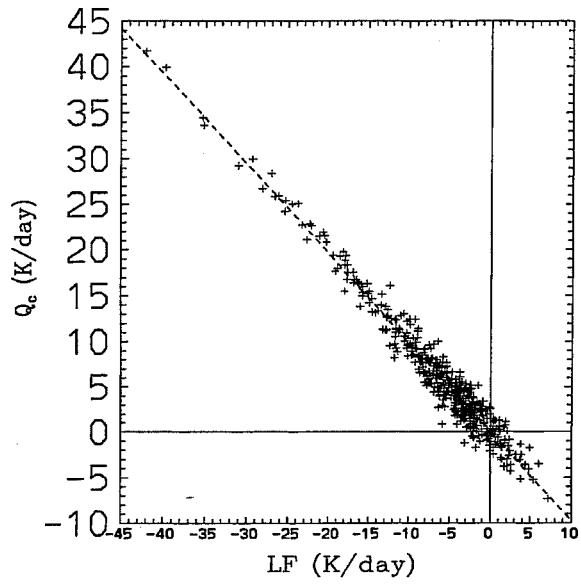

Figure 12 Scatter plot of forcing of temperature by vertical motion and radiation (horizontal axis) against convective forcing at the 500hpa level from the TOGA-COARE IFA analysis. The dashed line is the best least-square linear fit. From Yano (2001)

This evidence suggests that the vertical motion required to balance the large-scale convergence by continuity is almost cancelled by the subsidence which compensates the convective mass transport. Thus the total large-scale motion is small, and most of the transport happens in the convective plumes. This is consistent with (9), where the response to a forcing on the right-hand side will be in terms of convective rather than large-scale transport.

We can reflect this situation in the model by combining the convective transport equations with those for the large-scale dynamics. Thus equation (29) for the convective transport of horizontal momentum together with the equation for the vertical transport of horizontal momentum by the resolved dynamics becomes

$$\frac{\partial u}{\partial t} = \frac{1}{\rho}\frac{\partial}{\partial z}[M_{up}u_{up} + M_{down}u_{down} - (M_{up} + M_{down})\bar{u}] - w\frac{\partial}{\partial z}\bar{u} \tag{33}$$

We can rewrite this as

$$\frac{\partial u}{\partial t} = \frac{1}{\rho}\frac{\partial}{\partial z}[M_{up}u_{up} + M_{down}u_{down}] - \frac{1}{\rho}\bar{u}\frac{\partial}{\partial z}(M_{up} + M_{down}) \quad . \tag{34}$$
$$-\left(w + \frac{1}{\rho}(M_{up} + M_{down})\right)\frac{\partial}{\partial z}\bar{u}$$

The final term in equation (34) is then solved using semi-Lagrangian methods, with $w + \frac{1}{\rho}(M_{up} + M_{down})$ used instead of w in the departure point calculation. This has a stability advantage in the operational ECMWF scheme, because the advective part of the convection terms in (33) would be unstable if integrated with a forward timestep and centred spatial differencing, and so upwind spatial differencing has to be used. This is very diffusive. If a semi-Lagrangian method is used to solve (34), there will be no stability restriction from the

advective term and the results will be more accurate. However, there may still be a stability issue arising from the convective flux divergence.

Results are shown in Figure 13 and Figure 14. Figure 13 shows the balance between the convective mass flux and large scale vertical motion for the operational ECMWF model at T319 resolution averaged over the first 12 hours of a simulation calculated at tropical (20^{o}N to 20^{o}S) points where deep convection occurred. The results are shown for a level near 500hpa, but lower levels exhibit a similar effect. The results of a second experiment where equation (34) was used for all variables is also shown. The balance is closer, as the slope of the linear regression line is about -0.9 as against -0.7.The spread is similar.

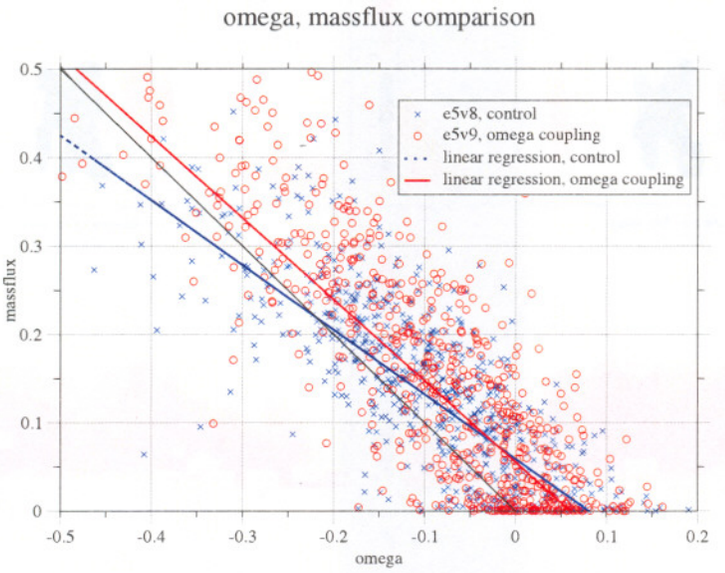

Figure 13 Balance between net convective mass flux and large scale vertical motion as computed by the ECMWF model using the operational method with convection treated separately and the 'omega coupling' method using equation (34) with semi-Lagrangian advection.

Figure 14 shows the combined results of two 3 month simulations at T63 resolution using initial data from the ECMWF reanalysis project (ERA). The periods are June to August and December to February. The integrations using (34) have rather less large scale vertical motion near the equator and a rather deeper extent of large convective mass fluxes, as in Figure 13. The balance between the two terms is not very different in the zonal means of the two experiments. This can be partly explained by the increased number of convective events found using (34).

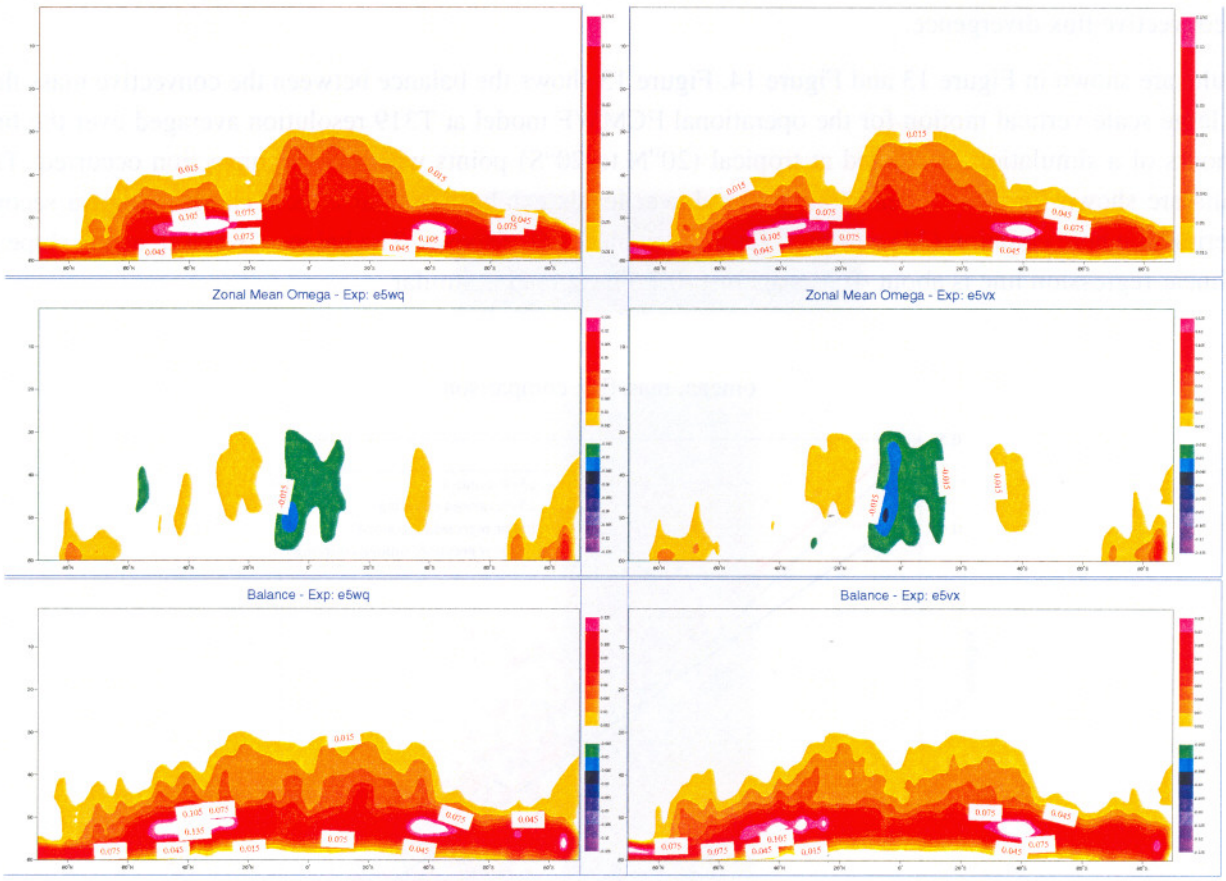

Figure 14 Comparison of zonal means for net convective mass fluxes (top), vertical motion (middle), and the balance between them (bottom). Left: 'omega coupling' using (34), right: operational scheme.

Other results (not shown) show that the large-scale forecast performance is similar to the control. Stability issues at high vertical resolution possibly caused by the loss of internal consistency within the physics are still under investigation. However, this loss of consistency does not degrade the overall conservation properties. In fact, these are improved, presumably because of the greater numerical consistency between the convection and dynamics.

7. Recommendations

We have reviewed theoretical work on the interaction of parametrised processes with large-scale atmospheric flow, and illustrated how some of the understanding thus gained can be exploited in the design of operational models. Some specific recommendations are listed below:

(i) The parametrisations should be written as far as possible as equations, so that the best way of integrating them as part of the overall model can be determined.

(ii) The integration schemes need to allow for the presence of different time-scales within each parametrised process.

(iii) There should be consistency between the way each process and the explicit dynamics is treated numerically, as many important balances involve several parametrised processes and explicit dynamics.

(iv) Implicit treatment of the physics (e.g. convective mass fluxes and vertical mixing coefficients) is desirable to allow proper treatment of those balances which involve physical processes. However, smoother formulations are needed to get the best out of an implicit scheme.

(v) The ECMWF convection scheme should be formulated in a way which is more consistent with the rest of the model.

REFERENCES

Cullen,M.J.P., 1989: On the incorporation of boundary-layer effects into a balanced model. *Quart. J. Roy. Meteor. Soc.*, *Quart. J. Roy. Meteor. Soc.*, **115**, pp 1109-1131.

Cullen,M.J.P., 2000: Alternative implementations of the semi-Lagrangian semi-implicit schemes in the ECMWF model. ECMWF Tech. Memo. no. 319.

Gage,K.S. and Nastrom,G.D., 1986: Theoretical interpretation of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft during GASP. *J. Atmos. Sci.*, **43**, pp 729-740.

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

Shutts,G.J. and Cullen,M.J.P.,1987: Parcel stability and its relation to semi-geostrophic theory. *J. Atmos. Sci.*, **46**, pp 2684-2697.

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

Smagorinsky,J., 1974: Global atmospheric modelling and the numerical simulation of climate. In 'Weather and Climate Modification', John Wiley and Sons, New York, 842pp, pp 633-686.

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

Wedi,N.P. and Cullen,M.J.P., 2001: On a stable and consistent numerical incorporation of parametrised processes into a global forecast model. Proc. ICFD Conference on Numerical Methods in Fluid Mechanics, Oxford.

Yano,J.-I., 2001: Residual cumulus parametrisation. *Quart. J. Roy. Meteor. Soc.*, **127**, pp. 1261-1276.