Key numerical issues for the future development of the ECMWF model

By **Mike Cullen, Deborah Salmond and Piotr Smolarkiewicz**

European Centre for Medium-Range Weather Forecasts

Abstract

As part of its long term planning, ECMWF is assessing the capability of models with resolution of 15km for medium range forecasting. This paper discusses the key numerical issues for the optimum use of this resolution.

The evidence presented shows that both the hydrostatic approximation and the spectral method will remain competitive up to this resolution. However, the full benefit of the resolution is only likely to be obtained by a more fully implicit treatment of the model dynamics and physics, resulting in some increase in cost. It may also be necessary to filter some of the motions which can be partly resolved by the increased resolution but not treated accurately.

Forward-in-time advection schemes are now well established, and can be expected to remain appropriate since they are also successful in small-scale models. Experience with both existing high resolution operational and small-scale research models demonstrates the need to use monotone interpolation, whether within semi-Lagrangian or Eulerian schemes. The semi-Lagrangian approach is, however, the only existing monotone method which does not require definitions of local derivatives on a grid, and is thus compatible with the spectral method. Present evidence shows that semi-Lagrangian schemes as currently implemented will be adequate for the 15km resolution, but will require improvements to remain accurate enough at higher resolutions.

It is likely that further increases in resolution beyond 15km will become possible in the longer term. It is important that the research issues this involves are addressed in the next 5-10 years. In particular, it is necessary to establish the optimum integration strategy for the fully compressible equations. One important issue is the implicit solution procedure, in particular to establish whether a spectral constant coefficient problem can still be used as the inner iteration. If not, a finite difference or finite volume method is required. Research is needed both to establish the best grid arrangement and the best solver for grid-point methods on the sphere. The latitude-longitude grid becomes very expensive at resolutions below 15km, and quasi-regular grids are not yet well-established. Another important issue is to assess whether semi-Lagrangian schemes can be improved sufficiently to remain accurate. Otherwise, again, finite difference or finite volume methods have to be used.

Keywords: Strategy Resolution Implicit

1. INTRODUCTION

The approved ECMWF 10 year plan includes implementation of a deterministic model with a resolution of about 15km, or a spectral truncation of about T1280, by 2008. The current forecast model is a spectral model with resolution T319. It is expected that this will be upgraded to T511 later in 2000. Work over the last 10 years has resulted in a dramatic improvement of the computational efficiency of this model. One main contributor was the use of two-time-level semi-Lagrangian advection, allowing the use of the 'linear' transform grid and thus higher spectral resolution, Temperton et al. (1999). Another was the use of a quasi-regular transform grid on the sphere, Hortal and Simmons (1991). Since the transform grid is only used to perform local calculations, the difficulties associated with quasi-regular finite difference grids do not occur.

As the resolution is increased, there are concerns about the computational efficiency of the current method, since the Legendre transforms scale as $O(N^3)$ with increase in resolution. However, since the sophistication of the rest of the model is also increasing, this may not be a serious difficulty. We evaluate the likely costs. The cost of the trans-

forms has to be balanced against the cost of implicit solvers for finite difference models. We review these, and work presented elsewhere in these proceedings by Davies and Majewski will discuss the current status. We find that they are comparable to the cost of the transforms.

The spectral model retains stability by treating a linear constant coefficient gravity wave problem implicitly. Since the model has been used at substantially higher local resolution by Meteo-France, there is no reason to suppose that this formulation will not be stable at higher resolution. It would only become questionable if a non-hydrostatic formulation had to be used instead. Currently, most non-hydrostatic models use finite difference schemes. The stability problem becomes more difficult when two-time-level schemes are used. Golding (1992) found that the choice of reference state became very critical in this case and the UK Met Office global non-hydrostatic model, Cullen et al. (1997), therefore solves a variable coefficient implicit problem. The constant coefficient method is used in the Canadian GEM model, Cote et al. (1998), but is iterated. The non-hydrostatic version of the ECMWF model developed by Meteo-France, Geleyn and Bubnova (1997), currently requires a three-time-level scheme for stability. It is thus important both to establish at what resolution a non-hydrostatic model has to be used, and whether a spectral solver can maintain stability of such a model. Either a three-time-level scheme, or iteration of the solver within a two-time-level scheme involve increased cost. We discuss these issues in the following sections. The paper by Smolarkiewicz in these proceedings discusses the issue of solvers in more detail.

Justification of a higher resolution model at ECMWF would require high local accuracy, as well as benefit to global error measures. The efficiency of the current model involves some compromises. For instance, the 'linear' transform grid is only accurate if all processes other than advection are essentially linear. This is clearly not the case, for instance over orography. The current semi-implicit scheme only treats implicitly a linear part of the fast gravity wave terms implicitly. All other processes are explicit, apart from some parts of the physics which are implicit and use as input estimates of the current model tendencies. It is likely that at higher resolution of the order of 15km it will be necessary to ensure a stronger coupling of the different processes. We present evaluations of some of these options and show that they benefit accuracy. However, the total cost of the model increases more than by the scaling with resolution.

Forward-in-time schemes for advection, whether Eulerian or semi-Lagrangian, are now well established, Smolarkiewicz and Margolin (1998); by Eulerian, we shall mean flux-form finite volume schemes, as advective-form algorithms can be viewed as a special case of the semi-Lagrangian approach. Since they are also in almost universal use in research small-scale models, we can be confident that they will remain appropriate. Monotonicity constraints have been found to be important. In the operational ECMWF model, enforcement of monotonicity was found to remove the over-activity of the model which developed when the resolution was increased to T213. Since then, tests at T511 and T639 have shown no signs of similar problems. Experience at the UK Met Office with very high resolution bubble convection simulations also shows the need for monotone schemes. We can thus assume that this requirement will remain. However, the only currently available monotone advection scheme for spectral models is the semi-Lagrangian scheme. The present semi-Lagrangian algorithm is very efficient. We show evidence that the solutions become inaccurate near orography at resolutions of order 10km. If semi-Lagrangian schemes cannot be improved, then monotone Eulerian schemes would be required and the spectral approach would have to be abandoned. In either case, a more accurate scheme, whether a modified semi-Lagrangian scheme or a Eulerian scheme, is likely to cost more than the current one.

Higher resolution allows models to resolve more types of atmospheric motion. However, some types of flow will not be resolved well enough to be treated accurately. This problem is present at any resolution, since there is no 'spectral gap'. One cannot make a clear separation between types of motion treated accurately and others absent

from the resolved solution, with their effect treated statistically by parametrisations. We show results suggesting that current NWP models make a reasonable attempt at predicting the average solution, but that this is inevitably a very inaccurate representation of the real flow. The debate will continue on whether or how under-resolved motions should be filtered from the solutions in order not to contaminate other useful information being produced by the model.

2. THEORETICAL BACKGROUND

We assume that the atmosphere obeys the compressible Navier-Stokes equations and the laws of thermodynamics. This is sufficient except very high in the atmosphere, above any region of interest for weather forecasting. Molecular viscosity only operates on scales several orders of magnitude smaller than any practical resolution, so we consider the compressible Euler equations. The wind speed is always substantially less than the speed of sound, and usually an order of magnitude less. Thus we can identify the properties of the equations which are important for numerical simulation by using acoustically filtered models, such as the anelastic model. Such filtering may not be valid on very large horizontal scales, so we do not recommend actually using acoustically filtered models operationally, but do recognise their utility for research.

A simplified form of the anelastic equations is set out below. The notation is standard except that Π is the Exner pressure $(p/p_0)^\kappa$, θ is the potential temperature. Define $\rho_0 = \rho_0(z)$ to be a fixed reference density profile, and $\theta_0 = \theta_0(z), \Pi_0 = \Pi_0(z)$ to be potential temperature and Exner pressure profiles in hydrostatic balance with ρ_0. Define perturbation variables $\Pi' = \Pi - \Pi_0, \theta' = \theta - \theta_0$.

$$
\begin{aligned}
&\frac{Du}{Dt} + C_p\theta_0\frac{\partial}{\partial x}\Pi' - fv = 0 \\
&\frac{Dv}{Dt} + C_p\theta_0\frac{\partial}{\partial y}\Pi' + fu = 0 \\
&\frac{Dw}{Dt} + C_p\theta_0\frac{\partial}{\partial z}\Pi' - \frac{g\theta'}{\theta_0} = 0 \\
&\frac{D\theta}{Dt} = 0 \\
&\nabla\cdot(\rho_0\mathbf{u}) = 0
\end{aligned}
\qquad (1)
$$

No lateral boundary conditions are specified, since our interest is global models. The appropriate lower boundary condition for inviscid flow is $\mathbf{u}\cdot\mathbf{n} = 0$, where $\mathbf{n}$ is a vector normal to the lower boundary. An upper boundary condition for these equations is that $w = 0$ as $z \to \infty$ where $\rho_0 \to 0$.

In order to solve these equations accurately for large times, we need a solution procedure which can be proved to converge to that of the continuous equations. The key properties of the approximation used in the proof will determine the conditions that a numerical method has to satisfy to be nonlinearly stable. Most workers in the field believe that the 3-dimensional incompressible Euler equations, which have similar mathematical properties to (1), generate singular vorticity distributions in finite time from smooth initial data, see Lions (1996, chapter 4). Without such behaviour, it is hard to explain the observed properties of three-dimensional turbulence. This means, however, that no numerical method relying on computing spatial derivatives of the velocity field is likely to be nonlinearly stable and converge to the correct solution. However, integration of (1) along fluid trajectories, as in a semi-Lagrangian algorithm, would still be possible. If this is done, the solution has to be completed by a projection that constrains the trajectories to satisfy the continuity equation. Recent work by Brenier (1991) and others suggests that this projection method is well-defined for a finite timestep, though it has not been proved that the solutions converge as the time-step is reduced.

Such a projection method can also be used in the compressible subsonic case. The step which enforces the continuity equation is still elliptic in nature, even though in this case the equations appear hyperbolic. The ellipticity appears when we seek to solve the equations over finite time intervals.

It is important that numerical methods share the key properties that allow the solution procedure to work. This will ensure that they remain stable and accurate in long time integrations. Standard truncation error analysis alone only ensures accuracy for short time integrations. The lessons for numerical methods are

(i) The advection scheme should not generate values of velocity or thermodynamic variables out of the range of the initial values. Thus monotone schemes are required. This is consistent with practical experience in high resolution models, e.g. Temperton et al. (1999).

(ii) An elliptic problem has to be solved. This remains true even in the compressible subsonic case as discussed above.

(iii) The trajectories should be consistent with the continuity equation and boundary conditions. This is not accurately enforced in existing semi-Lagrangian codes, though there is much research in the area (e.g. Purser and Leslie (1995), Lin and Rood (1998)). Eulerian flux form schemes do this naturally.

We note that, if the hydrostatic approximation is made, and free surface boundary conditions are used, no elliptic problem has to be solved. However, the result of this is that the solutions are likely to collapse in finite time. Write the anelastic form of these equations in the following simplified form

$$\begin{aligned}
&\frac{Du}{Dt} + C_p\theta_0\frac{\partial}{\partial x}\Pi' - fv = 0 \\
&\frac{Dv}{Dt} + C_p\theta_0\frac{\partial}{\partial y}\Pi' + fu = 0 \\
&C_p\theta_0\frac{\partial}{\partial z}\Pi' - \frac{g\theta'}{\theta_0} = 0 \\
&\frac{D\theta}{Dt} = 0 \\
&\nabla\cdot(\rho_0\mathbf{u}) = 0
\end{aligned} \tag{2}$$

We see that, if θ' vanishes, i.e. the fluid is thermally homogeneous, then $\frac{\partial\Pi}{\partial x}$ and $\frac{\partial\Pi}{\partial y}$ are independent of z. There is thus no mechanism to deflect the trajectories of the horizontal flow, other than its vertical mean, and the solution is likely to break down. In reality, if two different particles are on a 'collision course', the non-hydrostatic pressure becomes significant and prevents a breakdown. As shown by White (2000), the hydrostatic approximation is only valid if the Lagrangian time-scale of the motion is much greater than that of buoyancy oscillations. The assumption θ constant implies zero buoyancy frequency, so the hydrostatic assumption is inappropriate anyway.

In meteorological terms, this suggests that hydrostatic models will become unstable as resolution is increased if the stratification is weak. The remedy is to provide a computational 'non-hydrostatic pressure' in these regions. This is given by the reference profile term which is used in all hydrostatic semi-implicit models, and also the explicit UK Met Office Unified Model, Cullen and Davies (1991). This has allowed the latter model to be run at resolutions as high as 5km in extreme cases such as typhoons. However, obtaining the correct solution of equation (1) requires solving an elliptic problem, so we must expect an elliptic solver (or a sufficiently accurate explicit iterative approximation to it) to be an element of any future high resolution model which has credible accuracy.

The analysis of White (2000) allows an estimate to be made of when the hydrostatic assumption is likely to break down in a widespread sense. The current ECMWF model uses a 1200s time-step with a resolution of T319. The

smallest resolvable time-scale is thus a 2400s period, the smallest accurately treatable about 5000s. A typical tropospheric buoyancy frequency is 10^{-2}, corresponding to a period of 600s. Thus we can expect just to resolve time-scales comparable to it when the resolution reaches about T1300. The current 10 year plan reaches about this resolution. Thus one can expect to be able to fulfil the 10 year plan with hydrostatic models, but one is unlikely to be able to go any further, and be sure of retaining accuracy at the scales resolved, without a non-hydrostatic model.

3. BASIC PROPERTIES OF SPECTRAL AND GRID-BASED MODELS

3.1 Properties of spectral models

Though current spectral models, such as that at ECMWF, carry out much of their work in grid-point space, the primary history-carrying variables are held as spherical harmonics and all spatial derivatives are evaluated analytically in spectral space. The only calculations done in grid-point space are multiplications and divisions. The physical parametrisations are localised in the horizontal, all the calculations involve a single column of model variables. The transforms to and from grid-point space were originally carried out in an exactly reversible manner, and were designed so that products evaluated in grid-point space were evaluated exactly without aliasing. Since the introduction of semi-Lagrangian advection, it has been found possible to reduce the grid resolution so it is sufficient that basic variables can be transformed forward and back without alteration. In a further compromise, the density of the grid is reduced in high latitudes. The latter gives a slight smoothing but allows substantial savings, Hortal and Simmons (1991).

The method inherently requires a transformation between spectral and grid-point space every time step. The cost of this has to be compared with that of elliptic solvers in finite difference methods. The spectral transforms have a fixed cost, while the cost of a finite difference solver depends on the number of iterations.

It is likely that, at higher resolution, more of the model will have to be treated implicitly. Even if the problem to be solved is nonlinear, only a linear problem can be solved at each iteration. The efficiency of the spectral method will depend on how expensive the re-evaluation of the residual is compared with the cost of the transforms. An extreme case is if all calculations have to be implicit. In that case the whole timestep calculation has to be repeated at each iteration. The relative cost of the transforms is then the same as for a non-iterated scheme. However, the total model cost increases. On the other hand, if the re-evaluation is cheap, and a finite difference method can solve the problem with a small number of iterations, the spectral method will be uncompetitive.

The present design avoids the need to compute derivatives or flux differences on a grid, and thus avoids the errors associated with many quasi-regular finite difference grids on the sphere, and also avoids the overheads associated with the pole of latitude-longitude grids. The use of a spectral method for solving the elliptic problem means that large scale residuals are eliminated efficiently. Solvers for finite difference elliptic problems have difficulty in eliminating the large scale residual. However, on small scales, where the constant-coefficient problem is unrepresentative of the full problem, the residuals will not be removed efficiently by the spectral method. Another disadvantage of the spectral method is that any computation involving derivatives requires a transformation to spectral space. The present grid resolution is liable to error in computing products, and assumes that the model is reasonably linear except for advection.

At higher resolution the local dynamics is certainly not linear. A proper co-ordinate transformation to terrain-following coordinates will be needed, e.g. Gal-Chen and Somerville (1975). The nonlinearity in the physics will become more important, and the physical parametrisations may have to include horizontal dependencies. A more sophisticated turbulence model will need many derivatives and products. We will illustrate in section 4 that more accurate enforcement of the continuity equation on the semi-Lagrangian trajectory calculation may be important.

This may require the solution of a nonlinear (Monge-Ampere) elliptic problem. Even if a linear approximation to this is good enough, calculation of the local error in this equation (a Jacobian term) needs several derivatives and products.

3.2 Properties of grid-based models

Finite difference and finite volume methods have been used extensively for hydrostatic and non-hydrostatic simulations down to very small scales. This is illustrated in section 6, which shows that they can be very accurate on scales greater than about twice the grid-length. The main difficulty is grid design on the sphere, and solution of the elliptic part of the problem. Finite difference algorithms are inherently good at removing the local residual in the elliptic problem, but not as good at removing it on large spatial scales. Good preconditioning or multigrid methods are essential to get efficient solution on all scales, see the paper by Smolarkiewicz in this workshop. Note that the existing spectral transform method could be used as a preconditioner, but this requires transforms at every iteration. This is unlikely to be competitive.

The usual latitude-longitude grid on the sphere has a proven record in maintaining accuracy. The time-step restriction resulting from the convergence of meridians is avoided in semi-implicit, semi-Lagrangian codes. However, as shown in Table 1, the cost of the elliptic solver can be high, because the grid geometry is very anisotropic, and preconditioning difficult. The minimum grid-length decreases as the square of the average grid-length. With a general grid-length of 15km, the minimum grid-length will be about 35m. With a grid-length of 1km, the minimum grid-length is 0.15m. Multigrid methods have also proved difficult to use, because the problem to be solved has to be the same on all the members of the grid hierarchy so that the residuals can be passed between resolutions. However, the problem depends on the orography, which defines the shape of the domain, and the orography changes substantially with resolution. Thus, most work has been done with preconditioned conjugate residual algorithms. Experience is summarised in the papers by Davies and Smolarkiewicz at this workshop. The latitude-longitude grid is also inherently inefficient, since the advection and physics are being computed at unnecessarily fine resolution at high latitudes. It is not clear whether this cost can be reduced by filtering without losing accuracy.

The other option is a quasi-regular grid, such as an icosahedral-based grid. This is used in the new Deutscher Wetterdienst (DWD) model, see the paper by Majewski at this workshop. Many quasi-regular grids were tested in the 1970s, but accuracy was never satisfactory. However, the use of semi-Lagrangian advection schemes may remove this problem. There are also practical difficulties in the efficient coding of these models, since the addressing of grid-points is more complicated, and there is significant difficulty in implementing semi-Lagrangian schemes, because the search algorithm for the departure point becomes more expensive. However, the preconditioning of the elliptic solver becomes easier, as illustrated in Table 1, and the inherent advantage of a quasi-regular grid for advection and physics is retained.

3.3 Costs of transforms and elliptic solvers

Table 1 illustrates the cost of a 24 hour forecast using 8 processors of a Fujitsu VPP5000 with the ECMWF model, the UK Met Office 'new dynamics', and the DWD global model. The cost of the elliptic solver or transform part of the models is also shown. All results have been scaled linearly to 60 levels (the UKMO timings were available

for 38 levels and the DWD for 31 levels) and to 8 processors where necessary.

TABLE 1. RELATIVE COSTS OF A 24 HOUR FORECAST AND SOLVERS IN EXISTING MODELS

Resolution	UKMO (New dynamics)	UKMO Solver	DWD (GME)	DWD Solver	ECMWF (IFS)	ECMWF Transforms
60km (T_L319)	678	339	679	95	248	22
40km (T_L511)			2018	282	966	122
25km (T_L799)					2661	399
20km			11829	1447		

Note that the UKMO model is non-hydrostatic, so a fully implicit scheme is used with a generalised conjugate residual solver (GCR), see the paper by Smolarkiewicz in these proceedings. The DWD model is hydrostatic, so only 5 vertical modes are treated implicitly (using successive over-relaxation). A pessimistic estimate would thus multiply the DWD solver cost by 12 for a 60 level model, but the use of vertical preconditioning as used by the UKMO would substantially reduce that. The ECMWF model has been optimised for the VPP5000, while the others have not. This might make a 20% difference in the conclusions, as might consideration of a different computer such as the CRAY T3E.

Table 1 suggests that the spectral transform method will stay competitive up to the T1200 resolution where a non-hydrostatic model would be needed. Its viability after that will depend how frequently it has to be iterated. Work to be presented by Bubnova at this workshop will provide important evidence.

4. EVALUATION OF THE ACCURACY OF SEMI-LAGRANGIAN SCHEMES

The accuracy of semi-Lagrangian advection is crucial in justifying the retention of the spectral method, since that is the only viable monotone scheme which does not require the calculation of local fluxes. We illustrate the potential problems with a test of flow over the Scandinavian orography, using a research anelastic model. The explicit version of this model is described in Smolarkiewicz and Margolin (1997), and the implicit version in Smolarkiewicz et al. (1999). The area of integration was 2000km square, and the grid-length in the examples shown was 10km. 91 levels were used, with a 300m spacing. Four methods are compared, an explicit Eulerian scheme, a semi-implicit Eulerian scheme, a semi-Lagrangian explicit scheme, and a semi-Lagrangian semi-implicit scheme. No physics is included. The lower boundary condition is free-slip with no normal flow. The first three methods use a 1 minute timestep, the semi-Lagrangian semi-implicit scheme uses a 5 minute timestep.

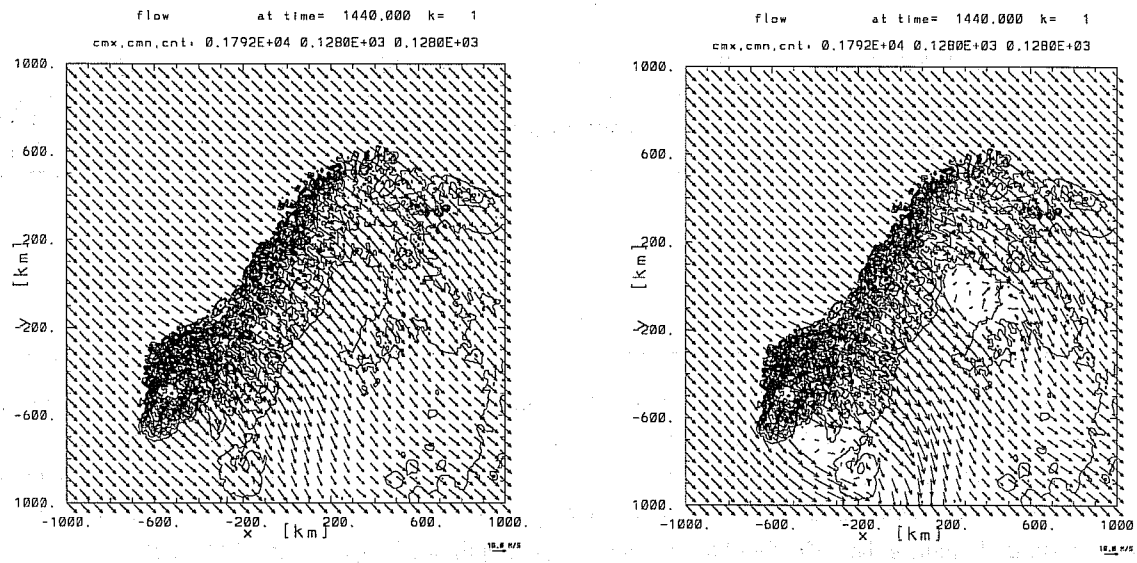

Figure 1. Comparison of solutions using explicit Eulerian (left) and semi-Lagrangian (right) schemes.

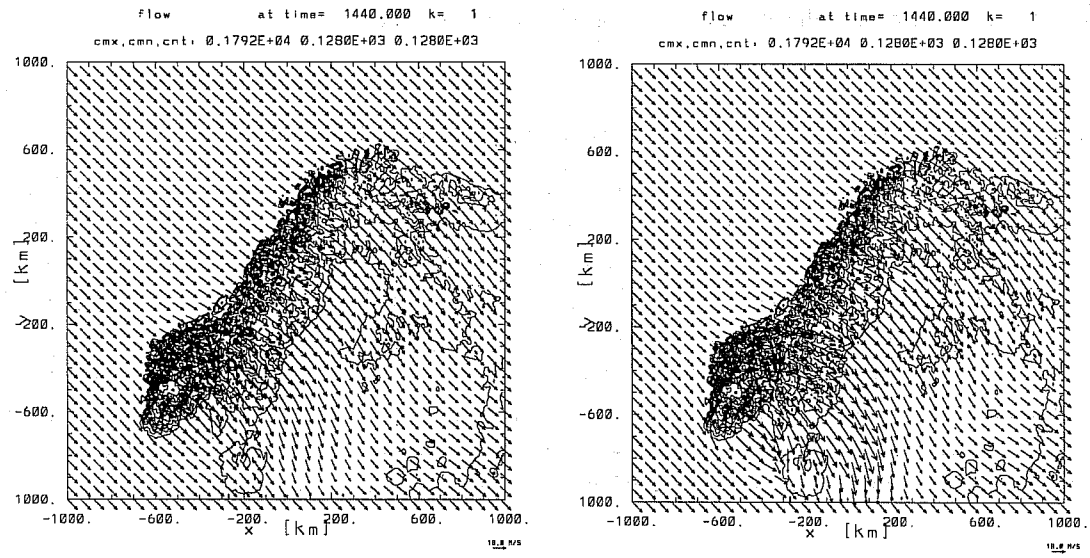

Figure 2. Comparison of semi-implicit Eulerian (left) and semi-Lagrangian (right) schemes.

Figure 1 and Figure 2 show the bottom boundary flow. The Eulerian method gives almost the same result with both methods of time integration. However, there is a large difference in the low- level downstream flow in the Skagerrak and the Gulf of Bothnia in the semi-Lagrangian results, and the latter are sensitive to the time integration. That suggests that the Eulerian integrations are more correct.

Errors in the semi-Lagrangian method must come either from interpolation or from the trajectory calculation. We can assess the trajectory calculation by diagnosing how accurately the Lagrangian continuity equation is satisfied when integrated over a timestep. This is done by calculating the density-weighted flow Jacobian which is the ratio

of the mass of a fluid element at the arrival point and the departure point. The exact solution has a unit value.

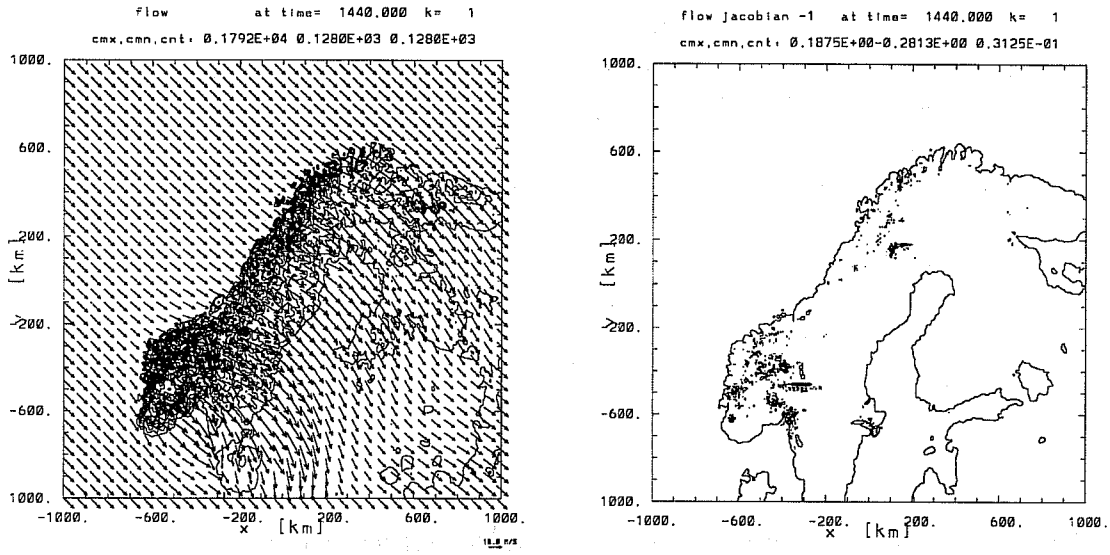

Figure 3. Low-level flow and flow Jacobian.

Figure 3 shows that the errors in the flow Jacobian range from +19% to -28%, and are largest in the region downstream of which the low-level flow is in error. Most of the time the error is less than 3%.

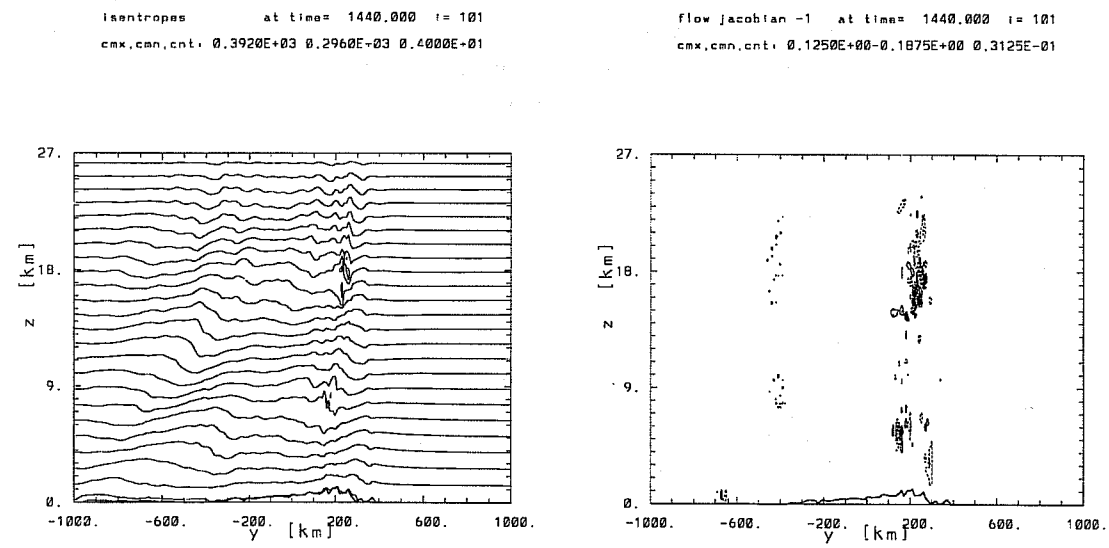

Figure 4. Cross-section of potential temperature and flow Jacobian.

Figure 4 shows that the flow Jacobian is most in error where the gravity waves break. The errors range from +12% to -19%, and are again less than 3% over most of the cross-section. The two sets of results show that the semi-Lagrangian method loses local accuracy in regions where the flow direction changes most rapidly, as is to be expected. Improvements will have to be made to the trajectory calculation, at least, if the semi-Lagrangian method is

to remain useable in very high resolution models.

Errors can also be caused by the interpolation to the departure point. Of particular concern is the lack of an obvious way to impose an exact flux boundary condition within the interpolation procedure. In effect, the simulated flow experiences a truncation-error flux of heat and momentum. The most severe errors are expected in the surface potential temperature of the explicit model, where the semi-Lagrangian transport of the total θ results in the largest amplitude of the interpolation error. In the semi-implicit model, only θ' is transported by the semi-Lagrangian scheme, whereas the reference state enters the right hand side of the equation as an Eulerian term, on which the boundary condition can be enforced. The interpolation error is proportional to the magnitude of the interpolated field, and is therefore much smaller than in the explicit model. This is well illustrated in the right panels of Figs. Figure 1 and Figure 2. It is difficult to develop a solution which is compatible with the spectral method, unless it is sufficient to use a flux-form scheme for the vertical fluxes only in the lowest layers.

5. EVALUATION OF THE IMPACT OF IMPLICIT SCHEMES

5.1 Background

The solution of a nonlinear implicit problem requires an iterative solution procedure. At each iteration a linear problem is solved, and the residual recalculated. In choosing a solution method, it is important to establish the relative cost of the re-evaluation of the residual and the solution of the linear problem. The cost of the re-evaluation depends on how much of the model has to be treated implicitly. If the most accurate method is to treat all processes implicitly, then the re-evaluation will be expensive and the cost of the solution of the linear problem is less critical. In this section we summarise results of evaluating the benefit of treating progressively more parts of the model implicitly.

The basic method is to use a predictor-corrector algorithm. This can be thought of as taking a single iteration towards the solution of an implicit problem. Both the predictor and corrector steps use the standard semi-implicit timestep, so in effect we have used the existing semi-implicit operator as a preconditioner. We assume that if benefit is to be obtained from a more implicit solution, that benefit will be visible in the first iteration.

5.2 Formulations tested

The standard semi-implicit scheme (Temperton (1997)) poses an equation for the divergence at the new time-level of the form

$$\begin{aligned} D_{t+\Delta t} &= D_t + \Delta t(\mathcal{D} - \nabla^2(\gamma_r T + \sigma_r p_{surf}) + \omega_r \varsigma)_t \\ &+ \frac{1}{4}\Delta t^2(\nabla^2(\gamma_r \tau_r + \sigma_r \nu_r)D_{t+\Delta t} + \omega_r D_{t+\Delta t} + \mathcal{G}) \\ &= D_t + \Delta t D_1 + \frac{1}{4}\Delta t^2 D_2 \end{aligned} \tag{3}$$

The Greek letters denote matrices of values at model levels, and the script letters denote known quantities. The suffix r on the matrices indicates that they are derived from a reference state.

The first experiment aims to confine the effect of the linearisation to the prediction of the divergence, and to remove its direct effect on the potential vorticity. Thus we first make an estimate D^* of $D_{t+\Delta t}$ using (3) giving:

$$D^* = D_t + \Delta t(\mathcal{D} - \nabla^2(\gamma_r T + \sigma_r p_{surf}) + \omega_r \varsigma)_t \qquad (4)$$
$$+ \frac{1}{4}\Delta t^2(\nabla^2(\gamma_r \tau_r + \sigma_r \nu_r)D^* + \omega_r D^* + \mathcal{G})$$

followed by setting

$$\varsigma^* = \varsigma_t + \Delta t(\mathcal{Z} - \omega\overline{D}) \qquad (5)$$
$$T^* = T_t + \Delta t(\mathcal{T} - \tau\overline{D})$$
$$(p_{surf})^* = (p_{surf})_t + \Delta t(\mathcal{P} - \nu\overline{D})$$

The overbar denotes an average of time t values and * values. Note that the basic state matrices are no longer used in updating the vorticity and height fields, so can have no effect on the potential vorticity. A second iteration is now carried out. Write $D' = D_{t+\Delta t} - D^*$, then solve

$$D_{t+\Delta t} = D_t + \Delta t(\mathcal{D}_t - \nabla^2(\gamma_r \overline{T} + \sigma_r \overline{p}_{surf}) + \omega_r \bar{\varsigma}) \qquad (6)$$
$$+ \frac{1}{4}\Delta t^2(\nabla^2(\gamma_r \tau_r + \sigma_r \nu_r)D' + \omega_r D')$$

for $D_{t+\Delta t}$. Finally redefine $\overline{D}$ as $\frac{1}{2}(D_t + D_{t+\Delta t})$, and calculate $\varsigma^{t+\Delta t}$, $T^{t+\Delta t}$, $(p_{surf})^{t+\Delta t}$ using (6).

This first experiment tests whether further iteration of the nonlinear terms, as carried out in the Canadian model, (Cote et al. 1998), improves accuracy, and whether allowing the reference state to influence the potential vorticity degrades accuracy. If so, the use of the hydrostatic assumption would be questionable, as the reference state is required to maintain stability in a hydrostatic model.

In the second experiment, the departure points were also recomputed, so that we set

$$\left(\frac{Dx}{Dt}\right)^* = \frac{1}{2}(u_d + u_a)^t \qquad (7)$$
$$\left(\frac{Dx}{Dt}\right)^1 = \frac{1}{2}(u_d^t + u_a^*)$$

This was combined with the iteration in (5) and (6). Work by Gospodinov (private communication) at Meteo-France suggests that this iteration converges, but that no further accuracy is gained after the first iteration.

In the third experiment, the physics was also recomputed, so that, in particular, the mixing coefficients and convective mass transport were calculated at the * time-level.

5.3 Results

Each of the three alternative versions of the ECMWF model were tested on 12 forecasts run from the 15th of every month in 1998 starting with the operational analysis. The resolution used was $T_L319L31$, which was the operational resolution for most of 1998. Figure 5 shows results for the first experiment where only the nonlinear terms were iterated, as in (5) and (6)

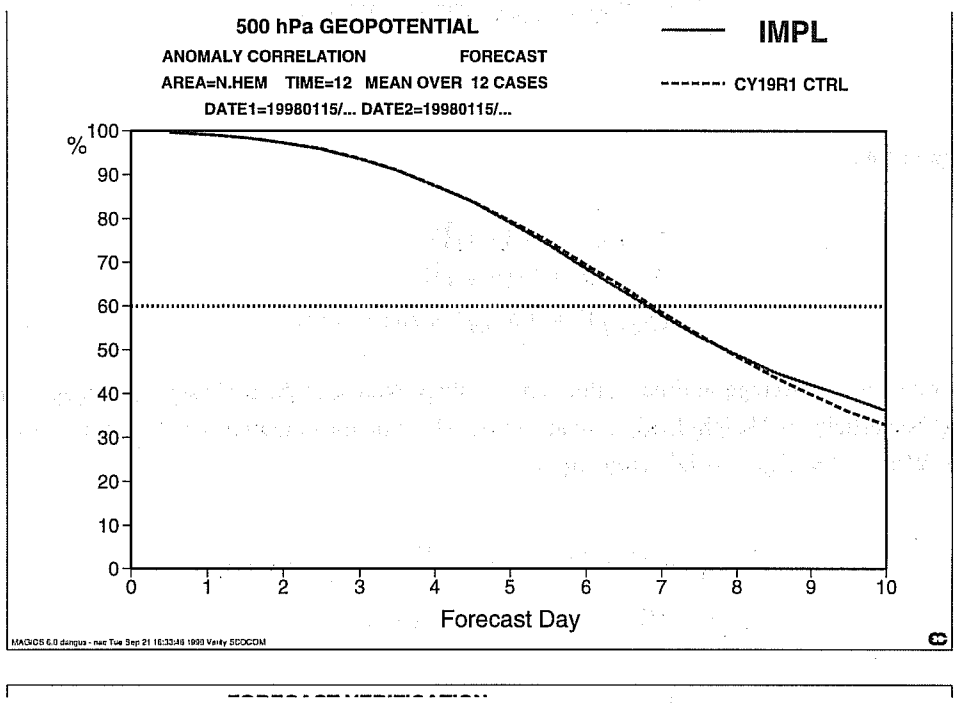

Figure 5. 500hpa scores from forecasts using iterated solver for nonlinear terms.

The differences are very small till near the end of the forecast range. Other diagnostics also suggest that there is only a small effect, for example on precipitation and vertical motion.

Figure 6 for the second experiment illustrates the effect of including an iterative recalculation of the trajectories as well.

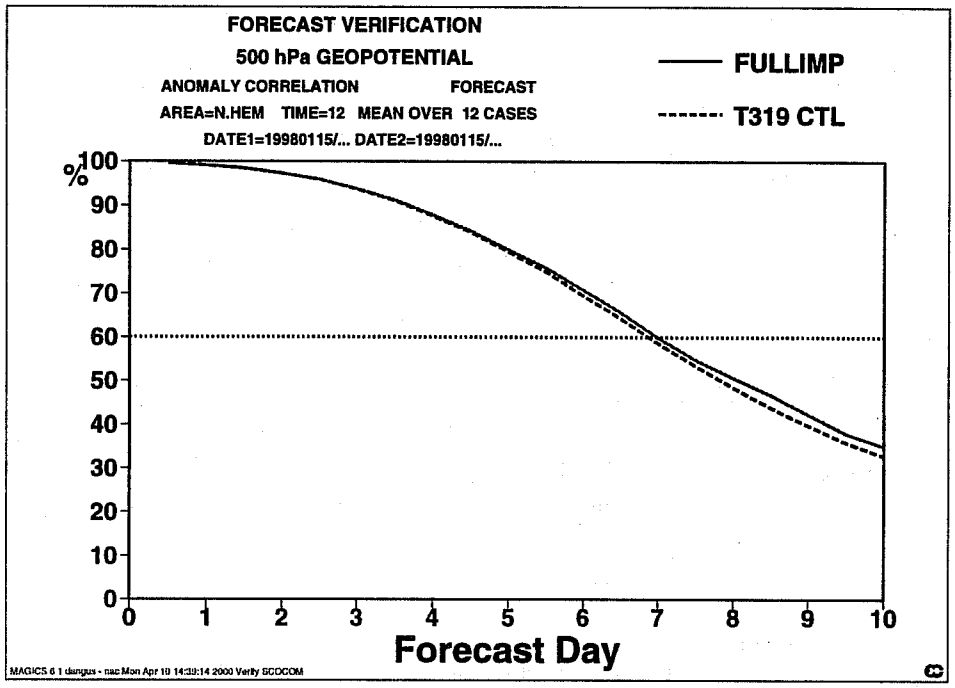

Figure 6. 500hpa forecast scores from experiment with recalculated trajectories.

Figure 6 shows that there is now a noticeable positive impact on the anomaly correlation at a time when the forecasts are still useful. The r.m.s errors are similar, and other diagnostics suggest that the level of activity of the model is not changed much.

Figure 7 for the third experiment shows the impact of also recalculating the physics. In order to be able to test the up-to-date versions of the physics, the cycle of the model used was CY21R5 and 60 levels were used in the vertical (as in the operational model from September 1999 onwards). Initial data used for the experiments was taken from assimilations carried out at T511 with T106 inner loops. 14 cases were selected for dates between August 1998 and December 1999. Figure 7 shows two experiments at T319. In the left-hand panel, the physics calculations were averaged along trajectories, as in (7). This required evaluating the coefficients at both time level n and time level *. Since the physics contains some very fast processes, which it may not be correct to average along the trajectory, a second set of experiments was done where the physics calculations, together with the vertical motion calculations, were done only at arrival points. The results are shown in the right-hand panel. Both the performance illustrated at 1000hpa in Figure 7, and other diagnostics, suggest that the impact is much greater if the physics is averaged along trajectories, giving second order accuracy in time. Figure 8 shows the results at T511, with averaging along the trajectories.

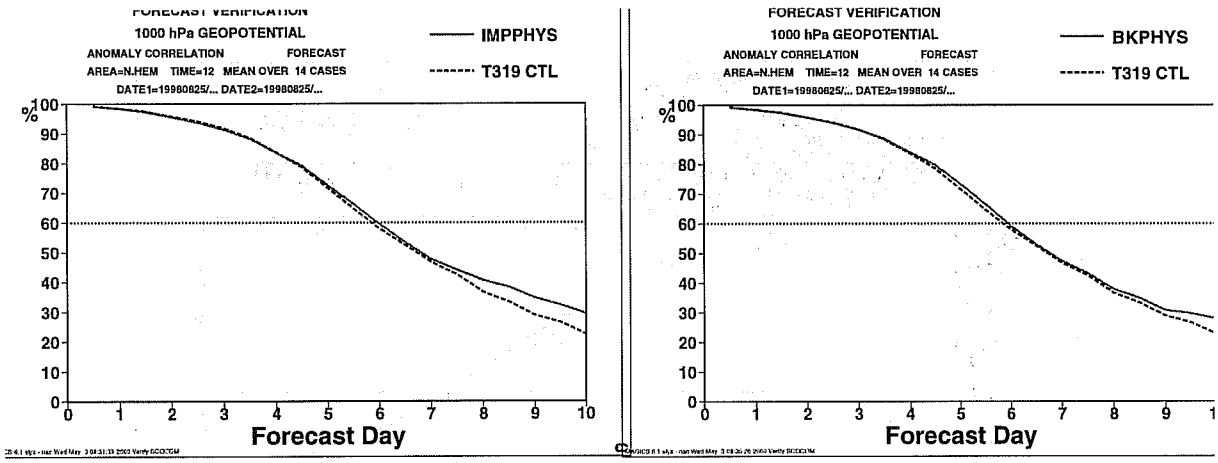

Figure 7. 1000 hpa performance of model with (left) iterated physics calculation averaged along trajectories, (right) iterated physics calculationat arrival points only .

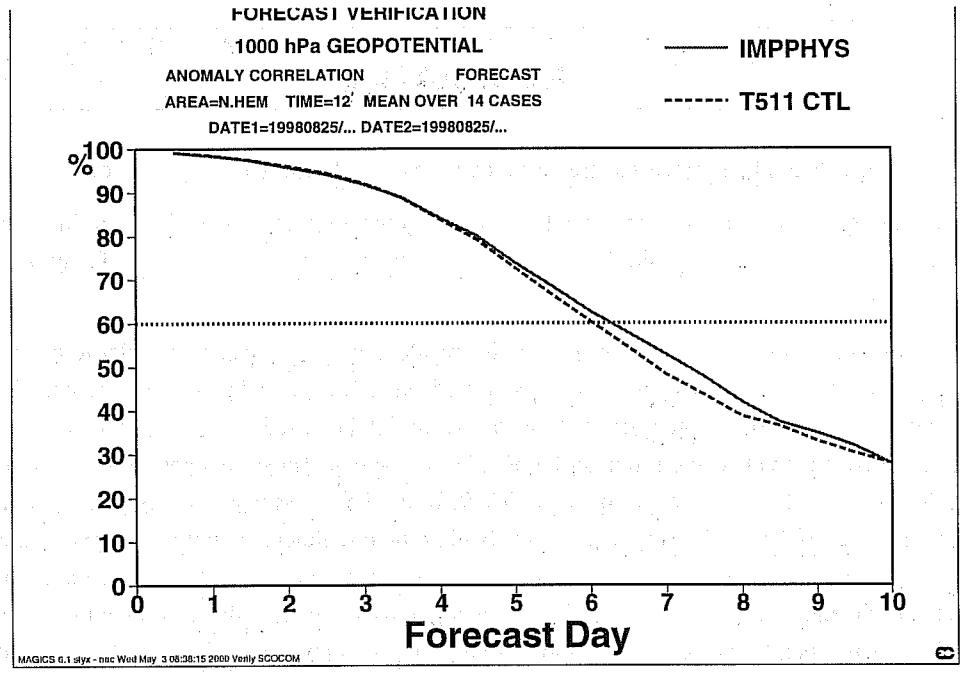

Figure 8. 1000hpa scores for experiment at T511 with physics averaged along the trajectory.

Figure 8 also shows a useful impact, a similar impact is visible in 500hpa scores. Overall, the third set of experiments suggest that there is benefit in a more implicit approach, but that the implicit treatment should extend to most of the model. Thus the cost of the transforms or solvers will not be the dominant issue, much of the cost will be in recalculating the residuals. Note also that a finite difference solver is likely to need fewer iterations after the corrector steps than in the predictor step, while the cost of transforms stays constant. Further research is needed to see if all the benefit of fully implicit schemes can be obtained by this simple predictor-corrector algorithm, or whether a more complete implicit solution would be beneficial. Experiments to date suggest that some reformulation of the physics may be needed before this question can be decided.

6. PARTLY RESOLVED MOTIONS

6.1 The problem

The atmosphere has no 'spectral gap', e.g. Gage and Nastrom (1986), so whatever resolution of model is chosen there will be a mixture of well-resolved, partly resolved and unresolved motions. In general, the flow is nonlinear, so that the presence of inaccurately treated motions may contaminate relatively well-resolved motions. Smagorinsky (1974) illustrates the main space and time-scales of different types of atmospheric motion. Both the present operational ECMWF effective minimum resolutions (120km in space and 2400s in time) and the 10-year plan (30km in space and 600s in time) cut across the most active scales of mesoscale motion and internal gravity waves.

The first issue to be resolved is whether the current ECMWF model formulation correctly predicts the averaged state of the atmosphere using an averaging scale commensurate with the model resolution. This is the problem that the model is formulated to solve. If there is significant error, the second issue will be to determine to what extent numerical errors contribute to the error. The third issue is to assess whether more useful results could be obtained in terms of practical weather forecasting by changing the averaging method so as to remove partly resolved motions. This changes the problem to be solved.

We illustrate the issues by studying a particular case of a simulation of flow over Scandinavia. Figure 9 shows the surface pressure pattern, with a north-westerly flow. Typical wind speeds in the lower troposphere were about $15ms^{-1}$. Figure 10 is a corresponding satellite picture, showing clear skies over the Baltic and cloud close to the mountains. There is some fine-scale structure in the cloud close to the limit of the resolution of the picture.

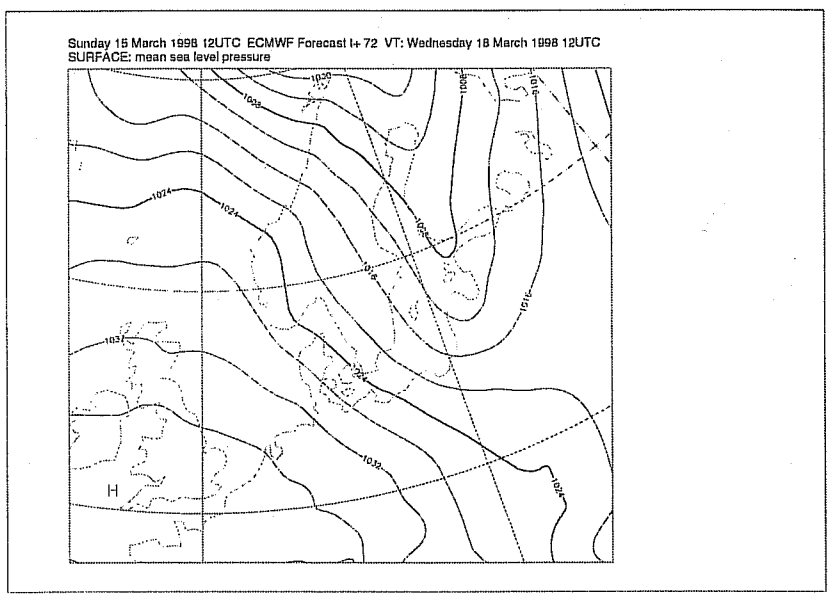

Figure 9.

Figure 10. METEOSAT visible picture for 18 March 1998, 1200UTC.

A cross-section of the vertical velocity from a T319L31 forecast using the operational model with a 20 minute timestep is shown in Figure 11. There is a mountain wave on a horizontal scale that is far larger than any cloud structures in the observations. The magnitude of the vertical velocity, up to 8 cm s-1, is consistent with the upslope wind needed for the airstream to cross the ridge. The cross-section of the orography itself, visible in Figure 11, is also quite unlike the real orography illustrated at 10km resolution in Figure 1.

We first assess whether the model solution is numerically accurate. Figure 12 shows the same cross-section from a run with a 5 minute time-step, instead of 20 minutes. The wave is similar in magnitude and extent, though somewhat less in amplitude at the upper levels plotted. Figure 13 is the result from the model with an iterated semi-Lagrangian trajectory (7), demonstrated in section 5 to be a more accurate version of the model. This is similar to Figure 11. These results suggest that the result is not a numerical artefact, but there is some time truncation error.

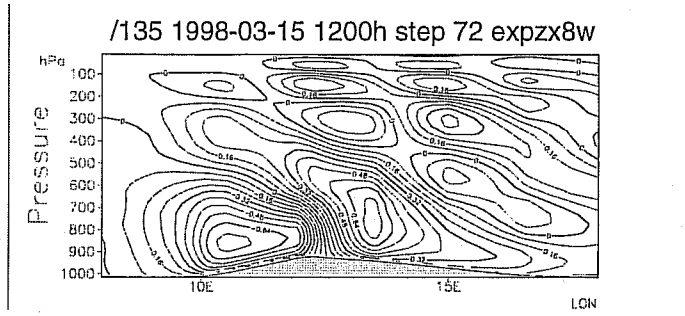

Figure 11. Cross-section of vertical velocity from control run.

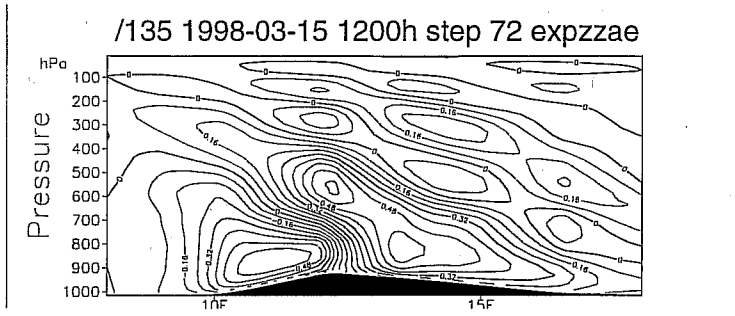

Figure 12. Cross-section of vertical velocity from integration with 5 minute timestep.

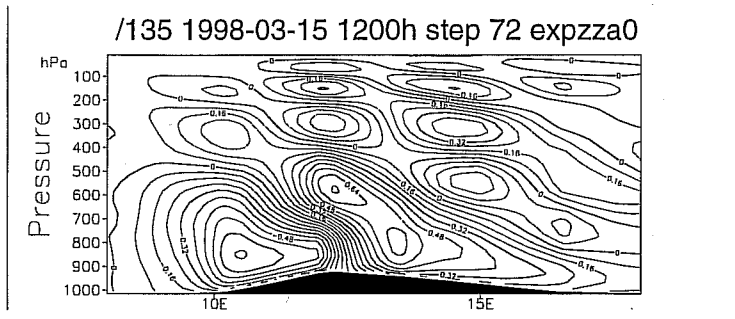

Figure 13. Cross-section of vertical velocity from integration with iterated trajectory calculation.

The next set of experiments were designed to test whether the solution obtained by the model is an accurate average of the true solution. Since it is not practical to run the ECMWF global model at much higher resolution, we use the same research model as in section 4, with an explicit Eulerian integration scheme. Bubnova will present results from the ALADIN limited area model, whose numerics are closely related to the numerics of the ECMWF model, at the workshop. The data used for the research model represents an idealisation of the real case shown in Figure 9. Figure 14 and Figure 15 compare the simulations of this case using horizontal resolutions of 10km and 40km. The results from the 10km integration have been averaged to 80km resolution.

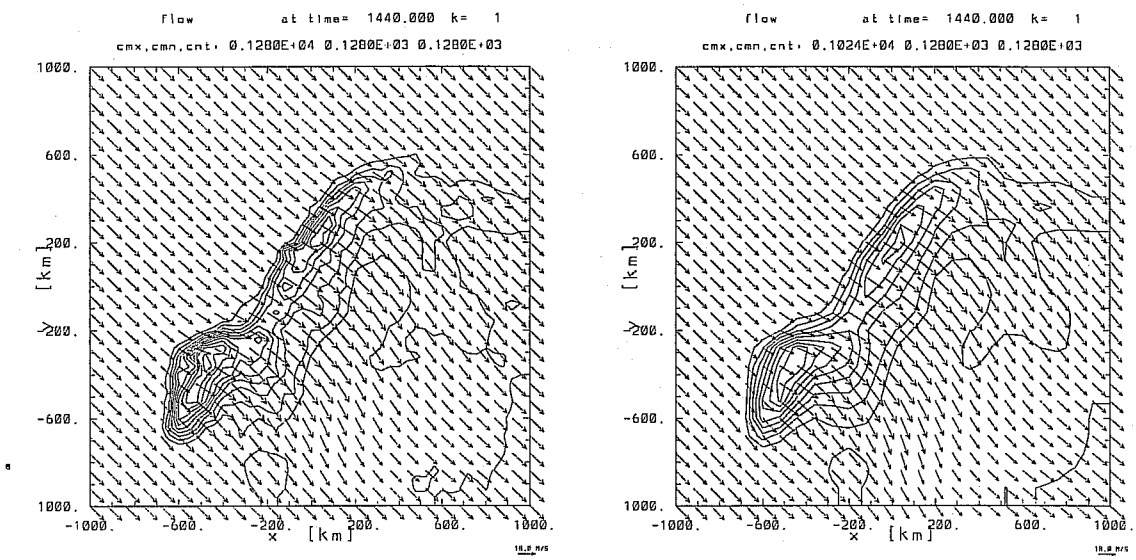

Figure 14. Comparison of low-level flow produced by 40km model (left), and 10km model averaged to 80km (right).

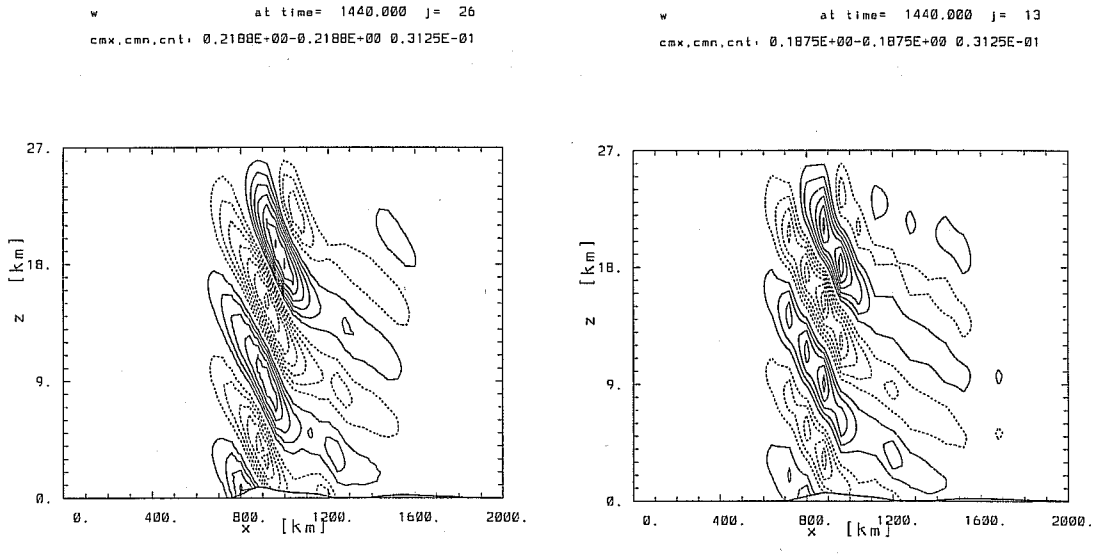

Figure 15. Comparison of vertical velocity cross-section produced by 40km model (left), and 10km model averaged to 80km (right)

We see that there is reasonable agreement between the 40km model and a 80km average of the 10km solution. A similar agreement has been obtained (not shown) between a 20km model and a 40km average of the 10km model. The values of vertical velocity in the 10km model are much more intense and localised, reaching 1 ms^{-1}, 10 times larger than in the T319 operational run. The seemingly completely wrong solution in the low resolution model is actually an average of the intense local velocities.

It is therefore likely that the solution is the best available answer to the problem posed, which is to compute the space-time average of the real flow. Note, though, that a given resolution of model should be compared with data averaged to a lower resolution to allow for numerical limitations. There is some inaccuracy because of the inability of the low resolution model to represent the shape of the orography, and evidence of time truncation error in the operational ECMWF simulations.

We now demonstrate the effect of modifying the averaging procedure to remove or reduce the gravity waves which could not be simulated accurately in the current example. We first note that mountain waves have a fast Lagrangian time scale. In particular, the vertical motion varies rapidly along the trajectory if the waves are close to the limit of resolution. These motions can be selectively filtered by decentring the semi-Lagrangian calculation along the trajectory, as is done in the Canadian model, Cote et al. (1998). Figure 16 illustrates the results when the terms associated with vertical motion are fully decentred. This are the determination of the vertical displacement of the departure point, and the nonlinear terms in the thermodynamic and surface pressure equations. The wave is somewhat damped, giving a solution is quite close to the short timestep solution illustrated in Figure 12, suggesting that the decentring is largely removing motions not treated accurately with the 20 minute timestep.

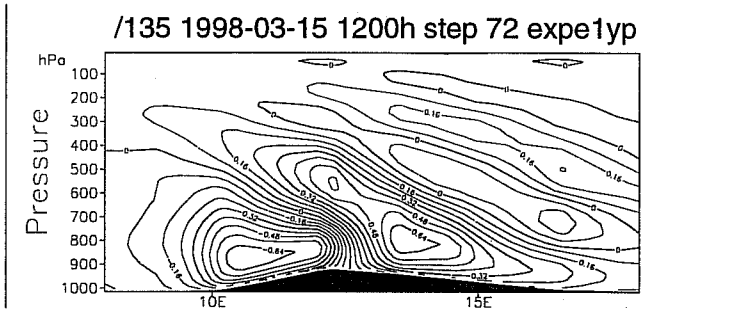

Figure 16. Cross-section of vertical velocity from experiment with decentred vertical velocity.

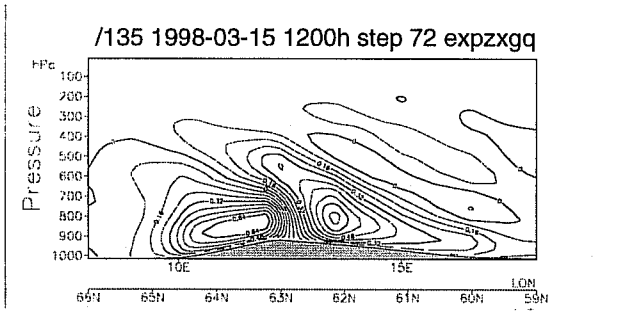

Figure 17. Cross-section of vertical velocity from experiment with all nonlinear terms decentred.

Figure 17 illustrates the results when all terms are decentred, so that all inertio-gravity oscillations will be damped. This may be justified because only a small part of the inertio-gravity wave spectrum is well-resolved. The mountain wave is now largely removed above 500hpa. The vertical motion represents only the motion necessary for the air to cross the ridge.

The overall performance of the model with decentring is illustrated in Figure 18. The experimental setup is the same as in section 5, using 12 forecasts with version CY19R1 of the ECMWF model and a resolution of T319L31. The left hand panel shows the results with only the vertical terms decentred. The impact is neutral, suggesting that

this form of damping can remove numerical errors without affecting large-scale performance.The right hand panel shows the results with all terms decentred. The anomaly correlations are improved, and the r.m.s. errors reduced. This illustrates that the more fundamental change to the averaging procedure in this experiment does affect large-scale performance.

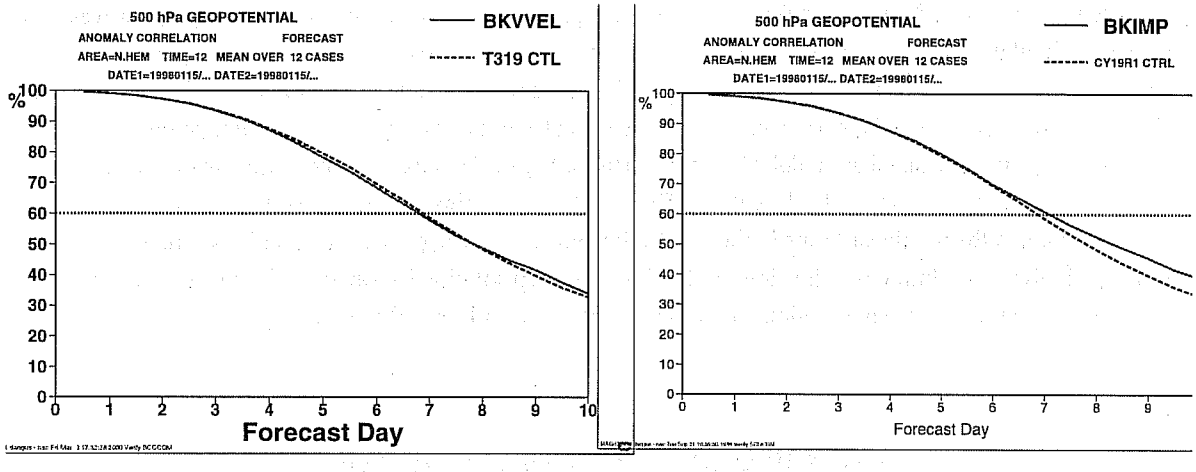

Figure 18. Performance at 500hpa of forecasts using decentring along trajectories

Table 2 shows the effect of the decentring on the vertical motion. The statistics were obtained by averaging results

TABLE 2. R.M.S. W AT 300/700 HPA (PA S^{-1}) (N. HEMISPHERE)

Experiment	Day 1	Day 3	Day 5	Day 10
Control	.184	.173	.166	.170
Imp. trajectory	.176	.168	.166	.165
All decentred	.159	.151	.148	.151
Vertical decentred	.167	.163	.159	.162

from model levels 12 and 23, about 300 and 700hpa. They should thus be representative of the troposphere. The impact of only decentring in the vertical is not very large overall, about 5% reduction in r.m.s. values. The impact of decentring all terms is larger, about 10-15% in r.m.s. values.

TABLE 3. STANDARD DEVIATION OF PMSL (HPA) (N. HEMISPHERE)

Experiment	Day 1	Day 3	Day 5	Day 10
Control	9.80	9.90	9.80	9.95
Imp. trajectory	9.60	9.76	9.60	10.08
All decentred	9.64	9.60	9.45	9.47
Vertical decentred	9.75	9.91	9.93	10.03

Table 3 shows the impact on the synoptic pattern, as represented by the standard deviation of PMSL. Decentring only in the vertical has no impact. Full decentring reduces the standard deviation by 4% by day 5, but the level of activity then stabilises. In the other experiments, the activity increases from day 5 to day 10.

These experiments illustrate that selective filtering of motions not treated accurately may be possible by using trajectory decentring, which is easy to implement and tune. There would then be no impact on the strategic development of the model. More radical methods would require using potential vorticity as a variable, as discussed in the next section.

7. METHODS TO PRESERVE LONG-TERM ACCURACY

The main aim of the ECMWF model is to predict weather systems deterministically for as long as possible, and then to predict the correct statistical behaviour of the flow for indefinitely long periods beyond that, so that ensemble-based probability forecasts have the maximum value. In terms of the dynamical formulation of the model, the predictive skill for weather systems is carried by the potential vorticity distribution, e.g. Hoskins et al. (1985). It is therefore very important to use a numerical formulation which preserves and transports the potential vorticity distribution as accurately as possible.

The papers by Budd and Dritschel in this workshop illustrate new approaches to this in simpler problems. The challenge is to blend these techniques with methods suitable to handle the rest of an operational prediction, including accurate treatment of orography and physics. The 'hybrid' approach of Dritschel and Ambaum(1997), where conventional grid-based methods are used for parts of the calculation, is the most promising. The structure would be similar to a spectral model, where many processes are treated by conventional grid-point methods. The benefits are illustrated by Mohebalhojeh and Dritschel (2000), who show that using PV as a variable even within a primitive shallow water model leads to more accurate simulation of gravity waves, while giving a very accurate prediction of the PV itself by using contour advection.

In terms of the strategy for the ECMWF infrastructure, the main considerations are not very different, since the optimum choice of grid, and the ability to solve an elliptic problem efficiently on it, are still important. However, there is a new element in the need to transfer information accurately and efficiently between a grid representation and some more idealised representation.

One particular issue in using a transformation to and from potential vorticity within an algorithm is the vertical grid structure. On small vertical scales, it is well-known that potential vorticity is determined by the mass field, Hoskins et al. (1985, p.905). However, the geostrophic or nonlinear balance relation between the mass and the horizontal winds is applied at levels where the wind is held. Thus, on small vertical scales, the potential vorticity is determined by the height or pressure at model levels. Finite difference approximations to the hydrostatic equation relate the height or pressure to a vertical mean of the temperature, but the inversion from height to temperature is ill-posed and there is thus a problem representing the correct dynamics. The effects are illustrated in Figure 19, which is a cross-section of analysed temperature increments at levels 1 to 20 (the stratosphere) of the 60 level model, derived using a PV based projection on the current ECMWF vertical grid within the data assimilation system. The solution is to use the Charney-Phillips vertical staggering, as used in the UK Met. Office 'new dynamics', Cullen et al. (1997). This makes the inversion to temperature and surface pressure well-posed.

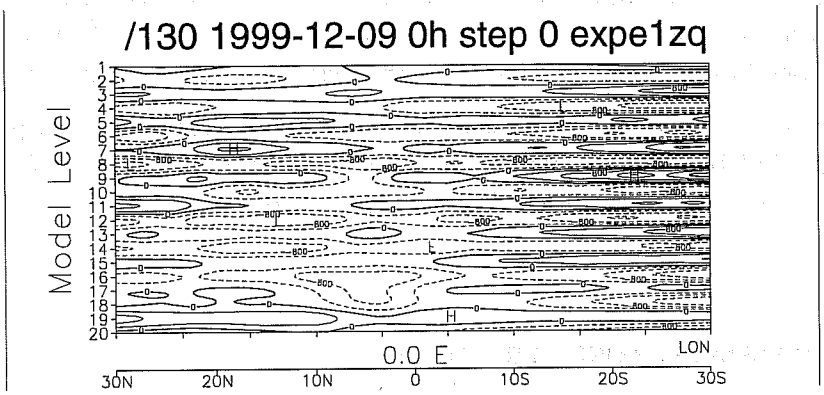

Figure 19. Cross-section of temperature increments from assimilation using PV as a control variable.

8. RECOMMENDATIONS

The recommendations based on the evidence summarised above are

- The present model structure can be retained down to resolutions of around 15km. However, the performance may well be improved by incorporating an outer iteration, by filtering solutions treated inaccurately, and by improving the semi-Lagrangian scheme to maintain conservation properties and satisfy the boundary condition more accurately.
- Research should be targeted on establishing the viability of the model at resolutions higher than 15km. The present model will only be viable if stability and accuracy of a fully compressible model can be obtained by at most 1-2 iterations of the spectral solution procedure, and if semi-Lagrangian schemes can be improved. Some aspects of the basic formulations required for fully compressible model need further research, e.g. the upper boundary condition and methods of filtering sound waves in the data assimilation. The degree of nonlinearity, and thus the resolution required of the transform grid, need to be assessed.
- There is a significant risk that, at resolutions higher than 15km, the lack of local approximations to derivatives, the inability to satisfy boundary conditions exactly, and the increased nonlinearity will become too great a handicap for spectral models. Proper terrain-following coordinate transformations will be needed (e.g. the Gal-Chen and Somerville 1975 method), and these lead to more complex equations. Research into finite difference/volume elliptic solvers, and establishing the optimum spherical grid, should be strongly encouraged.
- Given effective co-ordination with existing research programmes in the Member States, it should be possible to address most of these questions in the time available.

REFERENCES

Brenier,Y., 1991: Polar factorization and monotone rearrangement of vector-valued functions., Commun. Pure Appl. Math., 44, pp 375-417.

Cote,J., Gravel,S., Mettot,A., Patoine,A., Roch,M. and Staniforth,A. 1998: The operational CMC-MRB Global Environmental Multiscale (GEM) Model. Part I. Design consideration and formulation. *Mon. Weather Rev.*, **126**, 1373-1395.

Cullen,M.J.P. and Davies,T., 1991: A conservative split-explicit integration scheme with fourth order horizontal advection. *Q. J. Roy. Meteor. Soc.*, **117**, pp 993-1002.

Cullen,M.J.P., Davies,T., Mawson,M.H., James,J.A. and Coulter,S., 1997: An overview of numerical methods for the next generation UK NWP and climate model., in 'Numerical Methods in Atmospheric Modelling', the Andre Robert memorial volume. (C.Lin, R.Laprise, H.Ritchie eds.), Canadian Meteorological and Oceanographic Society, Ottawa, Canada, pp 425-444.

Dritschel, D.G. and Ambaum, M.H.P., 1997: A contour-advective semi-Lagrangian numerical algorithm for simulating fine-scale conservative dynamical fields., *Q. J. Roy. Meteor. Soc.*, **123**, 1097-1130.

Gal-Chen, T. and Somerville, R.C.J., 1975: On the use of a coordinate transformation for the solutions of the Navier-Stokes equations., *J. Comp. Phys.*, **17**, pp 209-228.

Golding,B.W., 1992: An efficient non-hydrostatic forecast model., *Meteor. Atmos. Phys.*, **50**, pp 89-103.

Hoskins,B.J., McIntyre,M.E. and Robertson, 1985: On the use and significance of isentropic potential vorticity maps., *Q. J. Roy. Meteor. Soc.*, **111**, pp 887-946.

Hortal,M., 1999: The development and testing of a new two-time-level semi-Lagrangian scheme (SETTLS) in the ECMWF forecast model, *Quart. J. Roy. Meteorol. Soc.*, **125**, under review.

Hortal,M. and Simmons,A.J., 1991: Use of reduced Gaussian grids in spectral models., *Mon. Weather Rev.*, **119**, pp 1057-1074.

Leslie,L.M. and Purser,R.J., 1995: Three-dimensional mass-conserving semi-Lagrangian scheme employing forward trajectories., *Mon. Weather Rev.*, **123**, pp 2551-2566.

Lin, S-J. and Rood, R.B., 1996: Multi-dimensional flux-form semi-Lagrangian transport schemes., *Mon. Weather Rev.*, **124**, 2046-2070.

Lions,P.-L., 1996: *Mathematical topics in Fluid Mechanics*, Vol. 1: Incompressible models., OUP, 237pp.

Mohebalhojeh, A.R. and Dritschel, D.G., 2000: On the representation of gravity waves in numerical models of the shallow-water equations., *Quart. J. Roy. Meteor. Soc.*, **126**, pp 669-688.

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.

Smolarkiewicz, P. K. and Margolin, L. G., 1997: On forward-in-time differencing for fluids: An Eulerian/semi-Lagrangian non-hydrostatic model for stratified flows., in 'Numerical Methods in Atmospheric Modelling', the Andre Robert memorial volume. (C.Lin, R.Laprise, H.Ritchie eds.), Canadian Meteorological and Oceanographic Society, Ottawa, Canada, pp 127-152.

Smolarkiewicz, P.K. and Margolin, L.G., 1998: MPDATA: A finite-difference solver for geophysical flows., *J. Comput. Phys.*, **140**, pp 459-480.

Smolarkiewicz, P. K., Grubivsic, V., Margolin, L. G. and A. A. Wyszogrodzki, A. A., 1999: Forward-in-time differencing for fluids: Nonhydrostatic modelling of fluid motions on a sphere., in Proc. 1998 Seminar on Recent Developments in Numerical Methods for Atmospheric Modelling, Reading, UK, ECMWF, pp 21-43.

Temperton,C., Hortal,M. and Simmons,A.J., 1999: A two-time-level semi-Lagrangian global spectral model,

Quart. J. Roy. Meteorol. Soc., **125**, under review.

Temperton, C., 1997: Treatment of the Coriolis terms in semi-Lagrangian spectral models. In "Numerical Methods in Atmospheric and Oceanic Modelling: The André J. Robert Memorial Volume" (ed. C.A. Lin, R. Laprise and H. Ritchie), pp 279-292.

White,A.A., 2000: A view of the equations of meteorological dynamics and various approximations., to appear in Proceedings of the Isaac Newton Institute on Atmosphere-Ocean Dynamics, Jul-Dec 1996, Cambridge U. P.