
Properties of the equations of motion

By **Mike Cullen**

European Centre for Medium-Range Weather Forecasts

Abstract

Some generic properties of the nonlinear equations of fluid flow are demonstrated with simple illustrative problems. Properties of the shallow water model are then described, and the solutions shown to be close to that of a ‘balanced’ approximation to them.

In three dimensions, the generalisation of the concept of ‘balance’ leads to models from which sound waves have been filtered, in particular the incompressible and anelastic models. Properties of their solutions are described. In particular, analytic solution of the equations requires solution of an elliptic problem, which thus also has to be solved in numerical models using the equations. The hydrostatic equations can be solved without solving an elliptic problem, but it is shown that this means that the solutions break down for weak stratification. Use of the hydrostatic approximation in numerical models requires use of a numerical equivalent of a non-hydrostatic pressure to ensure stability.

Operational models are more correctly viewed as solving space-time averages of the equations. Both Eulerian and Lagrangian averaging procedures are illustrated. In particular, both suggest that the averaged variable representing the fluid trajectory is best treated as different from that representing the momentum.

Averaged equations can be related to filtered models in which all inertia-gravity waves are removed. While such models do not give a complete description of the atmosphere, since they exclude real waves, they can describe the motions that are well-resolved and predictable by operational models. Their properties are thus useful in designing models, particularly the way that the computation of the resolved flow is related to the sub-grid models which parametrise the unresolved motions.

Keywords: Nonlinear Equations Averaging Balance

Table of contents

- [1 . Introduction](#)
- [2 . Observed behaviour](#)
- [3 . Toy problems](#)
 - [3.1 Introductory remarks](#)
 - [3.2 Examples](#)
 - [3.3 Two-dimensional incompressible Euler equations](#)
- [4 . Shallow water equations](#)
 - [4.1 Basic properties](#)
 - [4.2 Properties of ‘slow’ solutions](#)
- [5 . Three dimensional equations](#)
 - [5.1 Basic equations, and filtering of fast waves](#)
 - [5.2 Solution of the inviscid incompressible equations](#)
 - [5.3 Practical implications of solution procedure](#)
 - [5.4 The hydrostatic approximation](#)



6 . Averaged equations

- 6.1 Averaged equations, and their approximation by balanced models
- 6.2 Computations of averaged solutions.
- 6.3 Balanced models
- 6.4 Structure of the solution of the semi-geostrophic model

7 . Summary

REFERENCES

1. INTRODUCTION

In operational weather forecast and climate models we have to solve the equations of fluid motion and thermodynamics, subject to boundary conditions at the Earth's surface and prescribed external forcing due to radiation. While the basic laws governing dynamics and thermodynamics are well-known, and the laws governing phase changes between water vapour, liquid water, and ice are well established, many compromises have to be made in applying these laws in an operational model. The current operational version of the ECMWF model has a resolution of about 40km in the horizontal and 60 levels in the vertical. Real processes occur on scales down to millimetres or less, and regions of the atmosphere where small scale processes are important are often highly concentrated, for instance in convective updraughts. Thus a coarse scale average of the flow may be quite misleading, an average of a localised updraught looks like a smooth wave.

These lectures concentrate on representing the equations of motion and thermodynamics. Observations show that, qualitatively, the atmosphere behaves in a similar way all the time. Weather systems move around, and the overall level of activity varies with the season. A model has to represent the evolution of individual weather systems accurately, and also to maintain the statistical 'steady state' which ensures that it always represents a possible state of the atmosphere after the limits of deterministic predictability are passed.

No model can be accurate in all respects, and compromises are necessary. In these lectures, we identify the properties of the equations which control the long-term behaviour of the solutions. These will be important aspects to treat accurately in numerical models.

2. OBSERVED BEHAVIOUR

The most useful way to gain information about the solutions of the equations of motion and thermodynamics is to study observations of the atmosphere. Satellite pictures show large scale cloud patterns, [Figure 1](#), which on closer examination contain more and more fine-scale detail, [Figure 2](#). Thus, as numerical models reach higher and higher resolution, new phenomena will appear all the time. There will also always be features that are not well resolved, and thus represented inaccurately. It may well be beneficial to exclude them from the model until they can be treated accurately. Thus the fundamental problem is to maintain accurate treatment of well-resolved features, while excluding those not treated accurately and representing their bulk effect by a parametrisation.

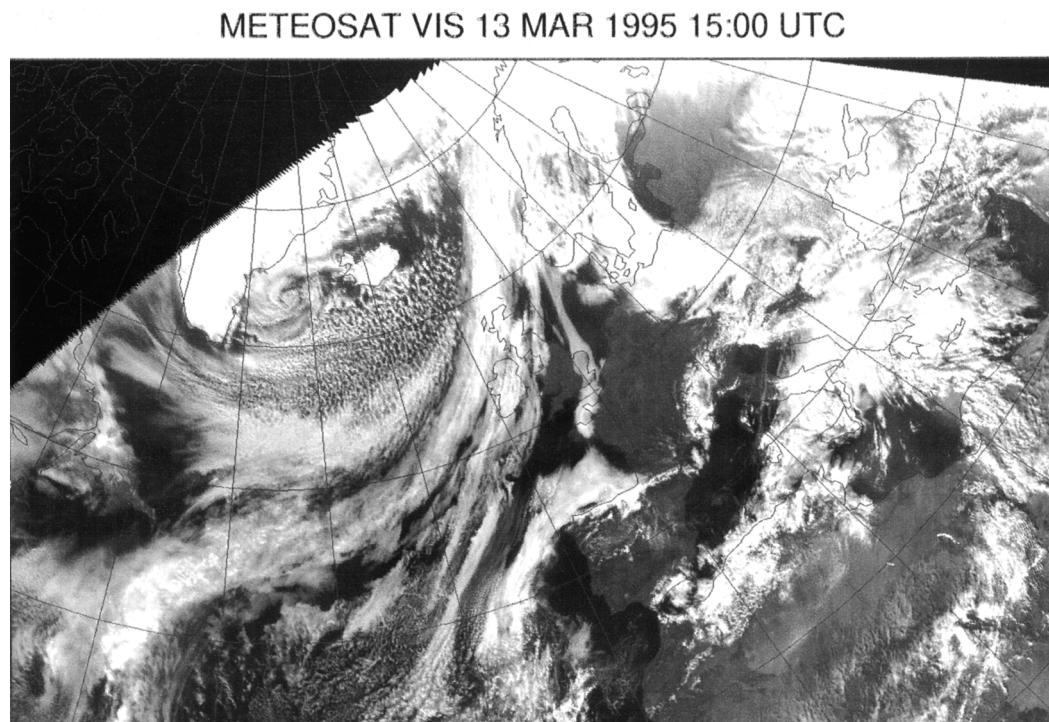


Figure 1. METEOSAT visible picture covering North Atlantic and Europe.

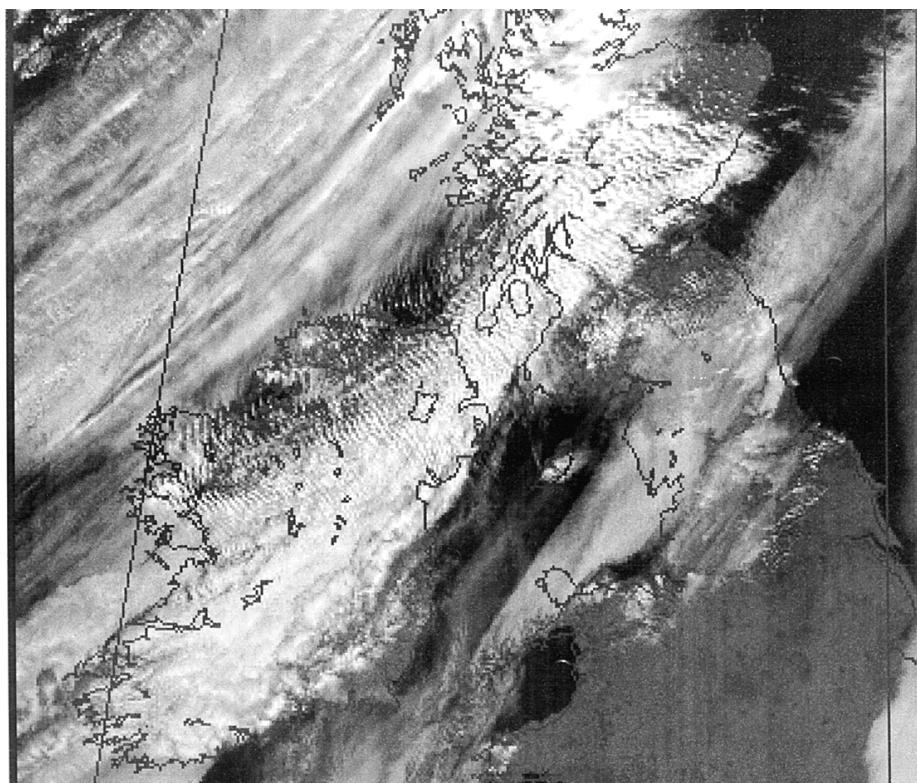


Figure 2. AVHRR visible picture covering the UK at a similar time to [Figure 1](#).

The separation problem would be easy if there was a clear difference in space and time-scale between different

phenomena. However, observations show clearly that this is not the case. [Figure 3](#) is a classical set of observations from commercial aircraft, due to [Gage](#) and Nastrom (1986, Fig.2). The energy spectrum is essentially a continuous function of horizontal scale, there is no ‘spectral gap’ which would make modelling easier. The situation is summarised in [Figure 4](#), due to [Smagorinsky](#) (1974), which shows typical space and time scales of atmospheric phenomena. Whatever resolution is chosen will cut across the active scales of some phenomena, resulting in inaccurate treatment.

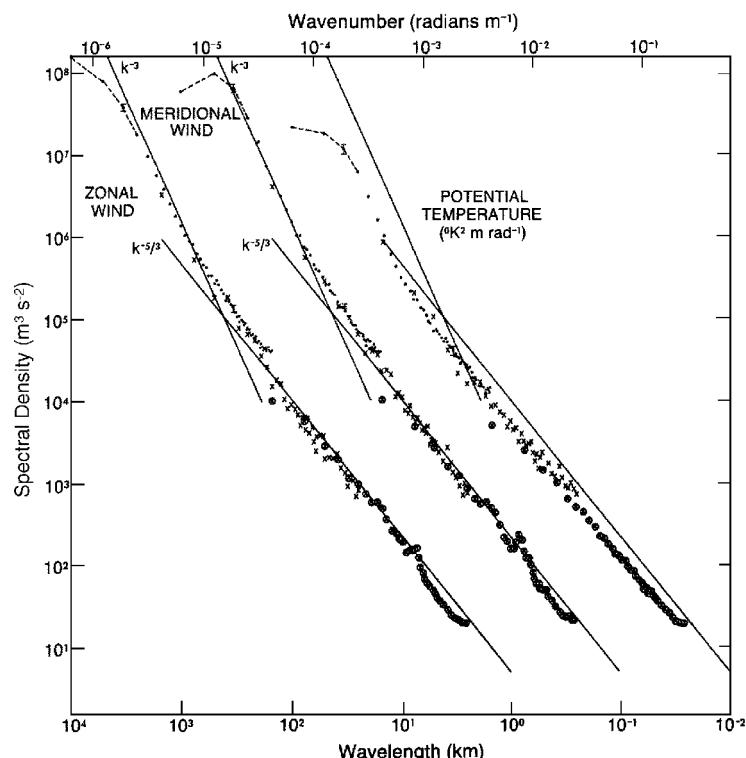


Figure 3. Wavenumber spectra of zonal and meridional velocity and potential temperature computed from 3 groups of flight segments. The meridional wind spectra are shifted one decade to the right and the potential temperature spectra two decades. From [Gage](#) and Nastrom (1986, Fig.2)

Observations also demonstrate that the equations have very non-smooth solutions. [Figure 5](#) is a balloon observation of Brunt-Vaisala frequency, essentially static stability, which shows very large variations on the smallest scale resolvable by the balloon. This illustrates that differentiated quantities, in particular, will not be robust diagnostics of atmospheric behaviour, and not be safe quantities to use in a numerical model. Another difficulty is that the small scale variations are not universal, so cannot be regarded as homogeneous turbulence in spectral space. [Figure 6](#) is an example of the wind trace from an anemometer. There are intense fast fluctuations all the time, much larger when the wind is stronger. Superposed on these are larger scale changes. Ideally, we would wish to predict the larger scale changes, together with a measure of the amplitude of the small-scale fluctuations. We would not expect, or need, to predict the small scale fluctuations deterministically.

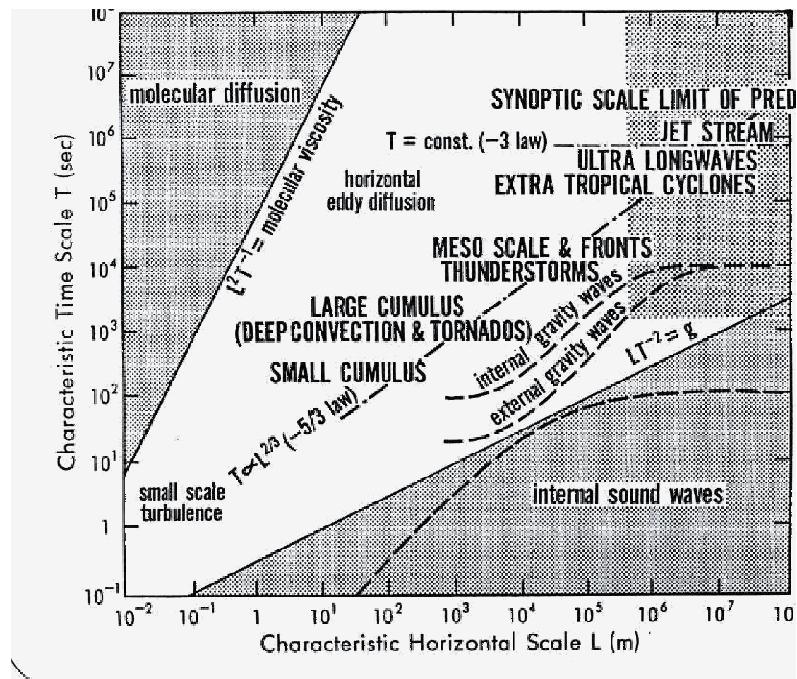


Figure 4. Characteristic horizontal and time-scales of atmospheric motions. After Smagorinsky (1974).

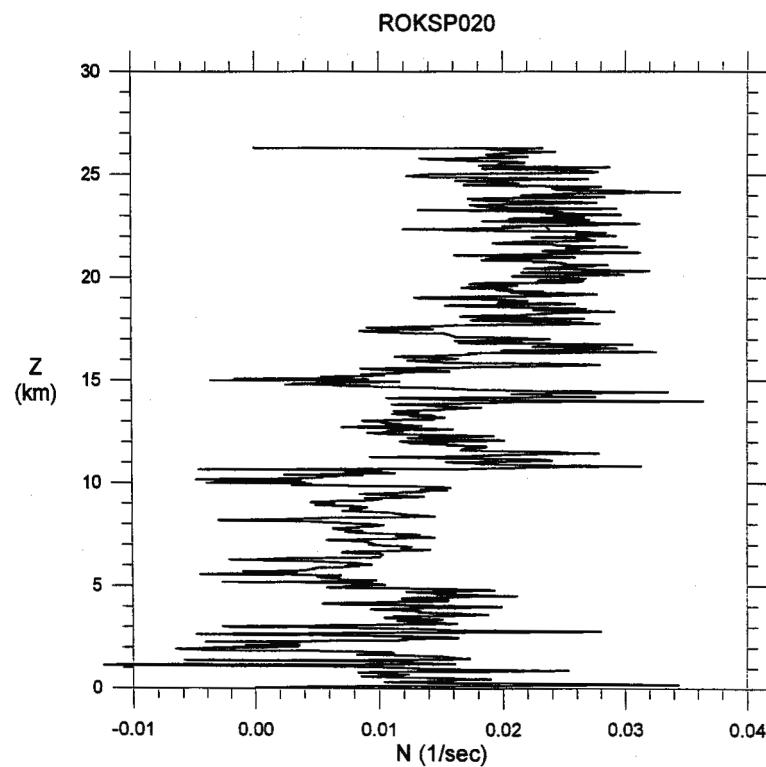


Figure 5. Plot of Brunt-Vaisala frequency against height from a balloon observation.

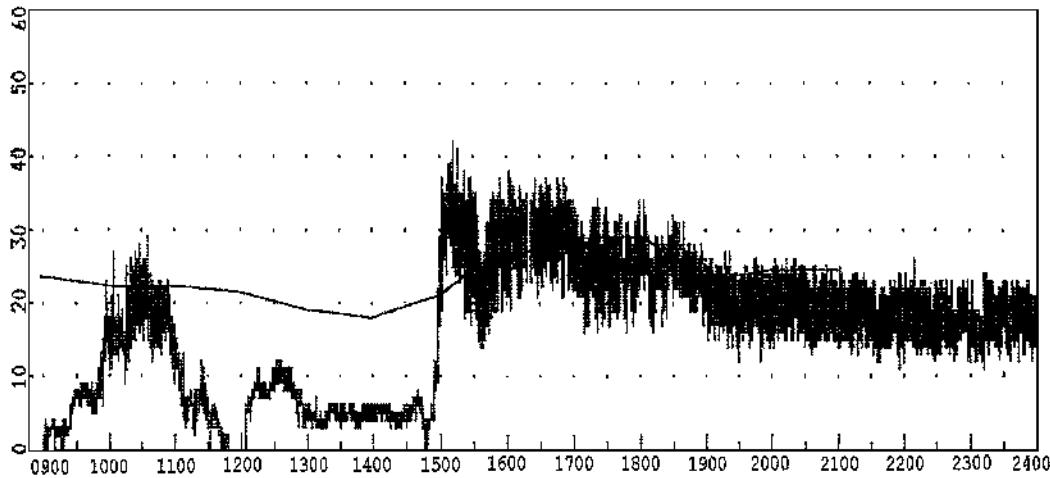


Figure 6. Anemograph trace for Bellambi Point on 26 December 1996 (wind speed in knots), taken from Batt and Leslie (1998), Fig. 7.

We can see from Figure 1 and Figure 2 that there is coherent large scale organisation in the flow, such as the cloud band with waves extending SW to NE across the Atlantic. There is also a region of regular cellular convection south of Iceland, and, in Figure 2, a regular gravity wave-train extending across Ireland and Scotland. It is well-known that many different types of organisation are possible, depending on the atmospheric state and the space and time-scales examined.

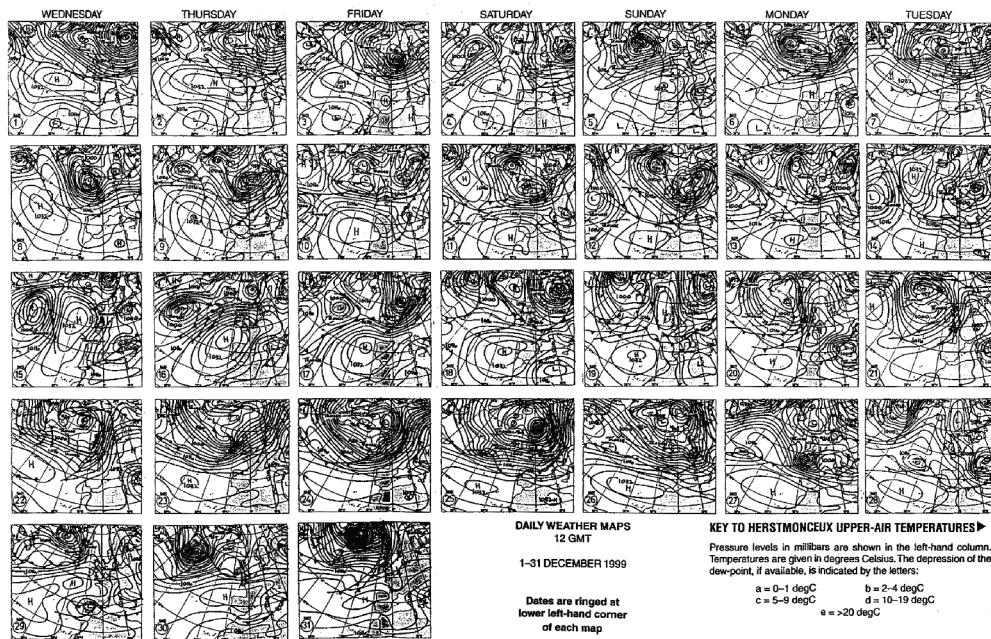


Figure 7. Daily sea-level pressure maps for December 1999, from Weather Log (Royal Meteorological Society). In operational medium-range forecasting, the prime job is to predict the coherent motions associated with weather

systems. A sequence of weather maps, such as [Figure 7](#), shows that these evolve in an irregular non-periodic way, but that statistically they are in steady state. Qualitatively, weather maps look similar from day to day. The general level of activity only varies with season, which is on a much longer time-scale than that of the individual systems.

3. TOY PROBLEMS

3.1 Introductory remarks

Very elementary models of real phenomena are often linear, so that they can be solved analytically. The solutions can be used to validate numerical calculations. Such tests were used to establish the basic rules which numerical methods have to obey to approximate solutions over large time periods. Essentially a scheme which is consistent with the analytic equation will converge to the solution provided that it is stable ([Richtmyer and Morton, 1967](#), p.45). However, realistic models have to be nonlinear, and cannot be solved analytically except in special cases. Thus numerical methods have to be used, and tend to be assumed correct if the solutions are correct in the special cases where analytic solution is possible, and do not ‘blow up’ for general data. However, the proof that the numerical methods are giving the correct answer requires knowledge that the original equation has solutions, and even then it may be very difficult to prove that a numerical method is stable. In particular, saying that a method has n th order accuracy requires the assumption that the solution has $n - 1$ continuous derivatives, allowing the Taylor expansion to be constructed. Spectral methods can only converge to solutions with infinitely many derivatives.

Equations which correctly describe real fluid flows must have solutions, since fluid properties do not diverge to infinity. We can thus expect the compressible Navier-Stokes equations to be soluble (though this has not yet been proved), since they are a very well validated model of real fluids. However, problems start when simplifying assumptions are made. For instance, there is no such thing as an inviscid fluid, but real viscosity in the atmosphere acts on scales less than millimetres except in the very high atmosphere. Thus it is negligible on practical modelling scales, and it is natural to neglect the viscosity, giving the Euler equations. However, it is not at all clear that the Euler equations have solutions. It may well be that any real flow generates small enough scales for viscosity to be important, and thus equations which omit it will ‘blow up’. This will lead to a loss of predictability, as defined by [Lorenz \(1969\)](#), since the solution will depend critically on unresolvable processes. The distinction is easiest to make at rigid boundaries. The Navier-Stokes equations will be solved with a no-slip boundary condition. However, the only boundary condition that can be set for the Euler equations is one of no normal flow, with free slip allowed along the boundary. In practice, there will always be a thin boundary layer where viscosity is important. It is therefore very dangerous to make deductions from the inviscid equations.

Another example is the use of the hydrostatic approximation. This is accurate enough for numerical models with resolution less than about 10km, and is often used at higher resolutions. However, we will see that it is very unlikely that these equations have solutions. When areas of weak stratification develop, we will show that the non-hydrostatic pressure is critical. Thus results deduced from hydrostatic equations will always be suspect.

We now illustrate these issues with some very simple problems which contain elements found in the general fluid equations.

3.2 Examples

The first example is

$$\frac{du}{dt} = u^2 \quad (1)$$



with $u = 1$ at $t = 0$. The solution is

$$u = \frac{1}{1-t} \quad (2)$$

which blows up at $t = 1$. This example illustrates the generic tendency of the solutions of nonlinear equations to collapse to singularities, as is believed to happen with vorticity in three-dimensional incompressible flow.

The second example is the one-dimensional Burger's equation

$$u = u(x, t) \quad (3)$$

$$\begin{aligned} \frac{Du}{Dt} &= 0 \\ \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} &= 0 \\ \frac{\partial u}{\partial t} + \frac{1}{2} \frac{\partial}{\partial x} u^2 &= 0 \end{aligned}$$

This is written in, successively

- (i) Lagrangian form
- (ii) Eulerian advective form
- (iii) Eulerian flux form

Equation (3) can be solved by direct integration by characteristics, giving

$$u(x, t) = u(x - u(x, t), 0) \quad (4)$$

Given periodic boundary conditions, and non-constant initial data, the solution always becomes multivalued as illustrated in Fig. 8

This is because, given $u(x_1, 0) = u_1, u(x_2, 0) = u_2, x_2 > x_1, u_1 > u_2$, the solution has values both u_1 and u_2 at $x = x_1 + u_1 t; t = (x_2 - x_1)/(u_1 - u_2)$.

Once the solution becomes multivalued, the derivatives with respect to x in eq. (3) cannot be calculated. However, the Lagrangian equation still makes sense, as it simply states that particles move at a constant speed. In order to determine how to continue the solution, it is necessary to return to the physics of the problem from which the equation was derived. There are then three options:

- (i) Say that the equation breaks down, and do not attempt to continue the solution.
- (ii) If the equation was derived from the physics of particles, and there is no reason why particles should not overtake each other, accept the solution of the Lagrangian form of eq. (3) and abandon the Eulerian forms. We will see that this is a very useful interpretation in two- and three-dimensional fluid mechanics. It is not natural for a genuinely one-dimensional problem.
- (iii) Use the Eulerian flux form, interpreted as defining momentum balance when integrated over a finite interval in x . Thus we have

$$\frac{\partial}{\partial t} \left(\int_{x_1}^{x_2} u \right) + \frac{1}{2} [u^2]_{x_1}^{x_2} = 0 \quad (5)$$

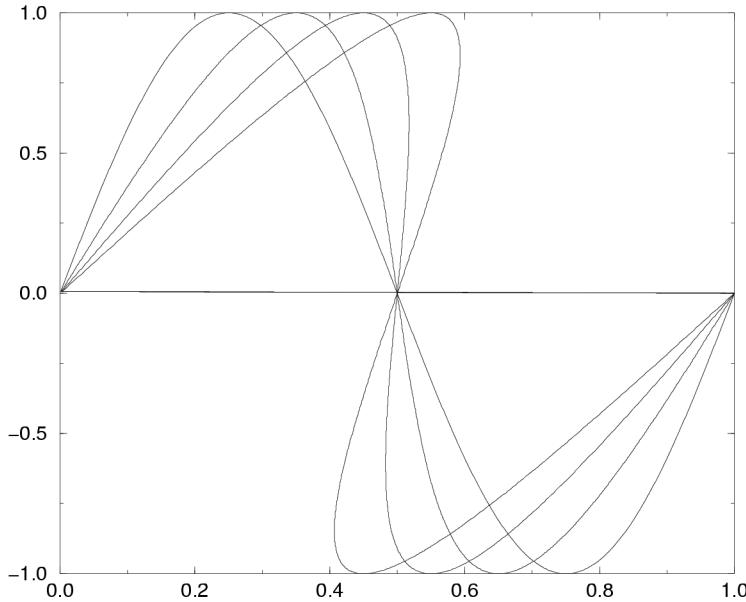


Figure 8. Solutions of eq. (3) for $u(x, 0) = \sin(2\pi x)$ at $t=0, 0.1, 0.2, 0.3$

Then, after the solution first becomes multivalued, the solution can be continued as a discontinuity propagating at a speed

$$\frac{u_R^2 - u_L^2}{2u_R - u_L} \quad (6)$$

Suffices L and R refer to values to left and right of the discontinuity. This solution can be proved to be the limit of that of a viscous version of the equation as the viscosity tends to zero. Thus if the original physical problem was a slightly viscous flow, the inviscid equation can be solved with a discontinuity, and all large scale aspects of the flow predicted without knowledge of the viscous sub-layer. Under these conditions deterministic predictability is maintained. This situation occurs in three-dimensional supersonic flow, but is not typical of subsonic flow.

The numerical method to be used also depends on the physical interpretation of the problem. In case (ii), semi-Lagrangian methods will be accurate and Eulerian methods will not be useable. In case (iii), only Eulerian flux form methods will give the correct solution. Furthermore, since the solution is discontinuous, the numerical method cannot be more than first order accurate, as a step discontinuity cannot be approximated to more than first order by a smooth function. The skill comes in blending methods which are first order accurate at discontinuities with ones which are of higher order accuracy elsewhere.

While this example suggests that advection always produces discontinuities, this is an artefact of the one-dimensional geometry. The next example shows that in two dimensions it is possible to maintain smoothness.



3.3 Two-dimensional incompressible Euler equations

These equations, for a uniformly rotating fluid, can be written

$$\begin{aligned}\frac{Du}{Dt} + \frac{\partial p}{\partial x} - fv &= 0 \\ \frac{Dv}{Dt} + \frac{\partial p}{\partial y} + fu &= 0 \\ \nabla \cdot \mathbf{u} &= 0\end{aligned}\tag{7}$$

Here, u, v are velocity components, p is the pressure and f the Coriolis parameter. We can write eq. (7) in vorticity form as

$$\begin{aligned}\frac{D(\zeta + f)}{Dt} &= 0 \\ \mathbf{u} &= \left(\frac{\partial \psi}{\partial y}, \frac{\partial \psi}{\partial x} \right) \\ \nabla^2 \psi &= \zeta\end{aligned}\tag{8}$$

Typical boundary conditions would be that $\mathbf{u} \cdot \mathbf{n} = 0$ on the boundary of a closed domain Ω .

Given smooth initial data, it can be proved that there is a unique solution which stays smooth for infinite time. Therefore it can be approximated by stable numerical methods of arbitrarily high orders of accuracy, in particular spectral methods. It is important to understand why the result is true, because this will determine what properties a numerical method will have to satisfy if it is to be nonlinearly stable and accurate over long periods of integration.

A detailed account of the proof, with references, is set out in [Gerard \(1992\)](#). A more extensive text on the subject is [Lions \(1996\)](#). The basic reason that it is possible is because the vorticity is conserved by advection, so any solution must be a rearrangement of the initial data. The key step is to write the velocity in terms of the vorticity using the Biot-Savart integral:

$$\begin{aligned}2\pi\psi &= B + \int_{\Omega} \zeta(\mathbf{x}') \ln(|\mathbf{x} - \mathbf{x}'|) d\mathbf{x}' \\ \mathbf{u} &= \left(\frac{\partial \psi}{\partial y}, \frac{\partial \psi}{\partial x} \right)\end{aligned}\tag{9}$$

B is the term from the boundary conditions. This allows the velocity gradients to be estimated in terms of the vorticity gradients by

$$\|\nabla u\| \leq C \|\zeta + f\| \ln \left(1 + \frac{\|\nabla(\zeta + f)\|}{\|\zeta + f\|} \right)\tag{10}$$

The rate of growth of the vorticity gradient can be written

$$\frac{D}{Dt} \nabla(\zeta + f) + \begin{bmatrix} -\Psi_{xy} & \Psi_{xx} \\ -\Psi_{yy} & \Psi_{yx} \end{bmatrix} \nabla(\zeta + f) = 0\tag{11}$$

Combining this with (10) gives an estimate of the form

$$\|\nabla \mathbf{u}^2\| \leq C(t) \|\mathbf{u}^2\| \quad (12)$$

where C grows exponentially in time. This growth is the symptom of the characteristic enstrophy cascade of two-dimensional turbulence (Leith 1983) and is associated with an inverse cascade of kinetic energy, and hence streamfunction, to large scales. This has been illustrated in many computations, e.g. Farge and Sadourny (1989).

There are two main implications of this proof for numerical methods. Firstly, the values of the vorticity must not be changed by advection. This favours the use of quasi-monotone schemes. Secondly, the rate of growth of the vorticity gradients must be restricted to the analytic value while they are well-resolved. This is aided in Eulerian schemes by the use of energy and enstrophy-conserving schemes, which control the mean scale of the flow (Arakawa, 1966), and in semi-Lagrangian schemes by using area-preserving trajectory mappings. The latter are an active research area, see for instance Leslie and Purser (1995). Since the exact solution generates arbitrarily small scales in parts of the flow, a dissipation mechanism is required to remove enstrophy at the smallest scales, even though the analytic solution is actually smooth without requiring viscosity.

4. SHALLOW WATER EQUATIONS

4.1 Basic properties

These equations are normally written as

$$\begin{aligned} \frac{Du}{Dt} + g \frac{\partial h}{\partial x} - fv &= 0 \\ \frac{Dv}{Dt} + g \frac{\partial h}{\partial y} + fu &= 0 \\ \frac{\partial h}{\partial t} + \nabla \cdot (h \mathbf{u}) &= 0 \end{aligned} \quad (13)$$

h is the depth of the fluid, typically chosen to have only small perturbations about a mean value h_0 . If these equations are derived as the vertical mean of the flow of an isentropic atmosphere with a free upper boundary, h has to be replaced by h^κ , $\kappa = (\gamma - 1)/\gamma$ in the pressure gradient term. If they are derived as the vertical mean of the flow of an isentropic atmosphere between rigid upper and lower boundaries, h must be replaced by $h^{\gamma-1}$ in the pressure gradient term. The equations can also be applied to a single layer with different density from the rest of the fluid. In this case g becomes a ‘reduced’ gravity, and it is likely that h can be of the same size as h_0 , and even become zero over part of the domain (as in outcropping density layers in the ocean), see Figure 9.

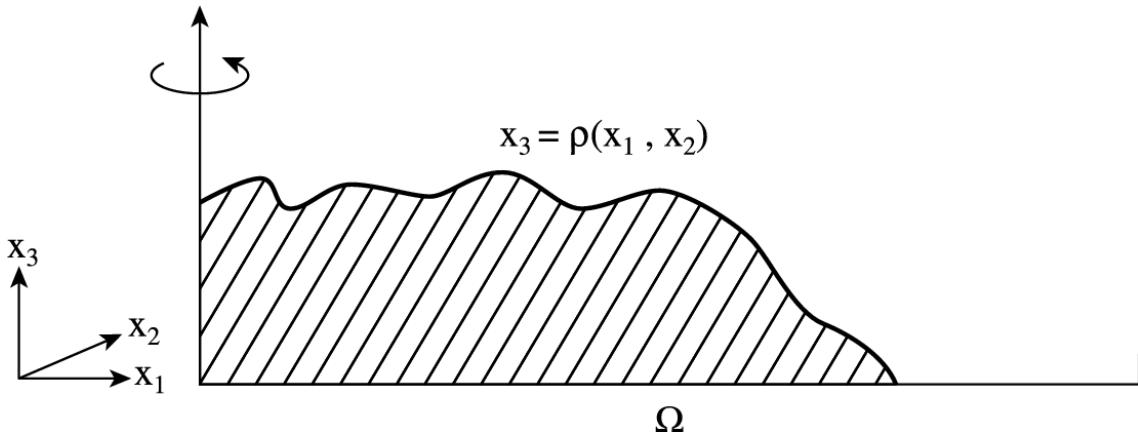


Figure 9. Schematic diagram of solution of shallow water equations with zero depth ρ over some of the domain.

If we linearise the equations about a state of rest, and assume solutions proportional to $e^{i(kx+ly+\omega t)}$, we find solutions $\omega = 0$, $\omega^2 = gh_0(k^2 + l^2) + f^2$. The propagating solutions are inertio-gravity waves, with speed

$$c = \sqrt{gh + \frac{f^2}{k^2 + l^2}} \quad (14)$$

They are pure gravity waves if $f = 0$ and pure inertial waves if $g = 0$. In the general nonlinear case, the solution corresponding to the zero value of ω becomes potential vorticity advection.

The nature of the solutions in any specific case depends on the Froude number defined by $Fr^2 = U^2/(gh_0)$, where U is a velocity scale; the Rossby number $U/(fL)$, where L is a horizontal length scale, and the Rossby radius $L_R = (\sqrt{gh_0})/f$. If h_0 is such that $Fr \ll 1$, then the gravity waves are fast compared with the advection speed, irrespective of horizontal scale. If $Ro \ll 1$ the inertio-gravity waves are fast compared with advection speeds, but this condition is only satisfied on large scales (much larger than 100km for $U=10\text{ ms}^{-1}$).

The equations conserve the energy integral

$$\frac{1}{2} \int \{ h(u^2 + v^2) + h^2 \} dx dy \quad (15)$$

and the potential vorticity $q = (\zeta + f)/h$ following fluid particles, where $\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$.

The inertio-gravity waves will always break after a time of order c/U , where U is a typical flow speed. When they break, momentum balance is preserved while energy is dissipated. Accurate numerical solution requires Eulerian flux form schemes which conserve momentum. The equations conserve the potential vorticity in a Lagrangian sense, until the waves break. Conservation is then lost.

If $\beta = \frac{\partial f}{\partial y} \neq 0$ and f_0 is a mean value of f , then Rossby waves can propagate. Their frequency is

$$\frac{-\beta k}{k^2 + l^2 + f_0^2/g h_0} \quad (16)$$

where k, l are horizontal wavenumbers. The wave speed thus increases with wavelength. On large horizontal scales the last term of the denominator dominates, and the wave speed becomes independent of wavelength.

4.2 Properties of ‘slow’ solutions

If f is constant, the solution of the linearised equations corresponding to $\omega = 0$ satisfies

$$fv = \frac{\partial h}{\partial x}, fu = -\frac{\partial h}{\partial y}, \nabla \cdot \mathbf{u} = 0 \quad (17)$$

Thus the velocity is geostrophic and non-divergent. In cases where U/c is small, we can thus approximate equations (13) by the quasi-geostrophic equations, which can be written

$$\begin{aligned} (q-f)h_0 &= \frac{gh_o}{f}\nabla^2 h - fh \\ \frac{D_g q}{Dt} &= 0 \end{aligned} \quad (18)$$

where the suffix g represents advection by the geostrophic wind and q is the quasi-geostrophic potential vorticity. Bourgeois and Beale (1994) proved that there is a solution of the shallow water equations (13) close to the solution of (18) if U/c is small. In order to study the behaviour of (18), we consider two limiting cases. If $Fr \ll 1, L \ll L_R$ then we can approximate $q-f$ in (18) by $\frac{g}{f}\nabla^2 h$. The evolution equation is then exactly equation (7) for two-dimensional incompressible flow with a streamfunction $\psi = (gh)/f$. The properties of (18) are thus determined by vorticity conservation. If $Ro \ll 1, L \gg L_R$, then (18) can be approximated by

$$\begin{aligned} (q-f)h_0 &= -fh \\ -\frac{f}{h_0} \frac{\partial h}{\partial t} + \mathbf{u} \cdot \nabla \left(\frac{g}{f} \nabla^2 h \right) &= 0 \end{aligned} \quad (19)$$

This is called the ‘equivalent barotropic’ model. The advection term in the second equation is small because $\frac{g}{f}\nabla^2 h$ is only a small part of the potential vorticity. The advection of the larger part $-fh/h_0$ is identically zero. Thus we expect the dynamics in this regime to be much less active.

We illustrate these two cases by computations with a shallow water model on a sphere with a grid of 288x193 points. The value of f is taken as uniform, so that the two limiting regimes $L \ll L_R, L \gg L_R$ can be reproduced by taking different values of h_0 , referred to as ‘deep’ and ‘shallow’. The model used a semi-Lagrangian, semi-implicit scheme with h, u, v as variables, so no conservation properties were exactly enforced. The initial data has wave-numbers between 3 and 19. The deformation radius L_R corresponds to wavenumber 3 in the ‘deep’ case and wavenumber 13 in the ‘shallow’ case. Thus the regimes computed both have L fairly close to L_R . The diagnostics illustrated are the potential and kinetic energies, the velocity gradient norm $\int h((\nabla u)^2 + (\nabla v)^2)dx dy$, the potential enstrophy $\int (hq^2)dx dy$ and the potential enstrophy gradient norm $\int h(\nabla q)^2 dx dy$. The result for the ‘deep’ case is shown in Figure 10.

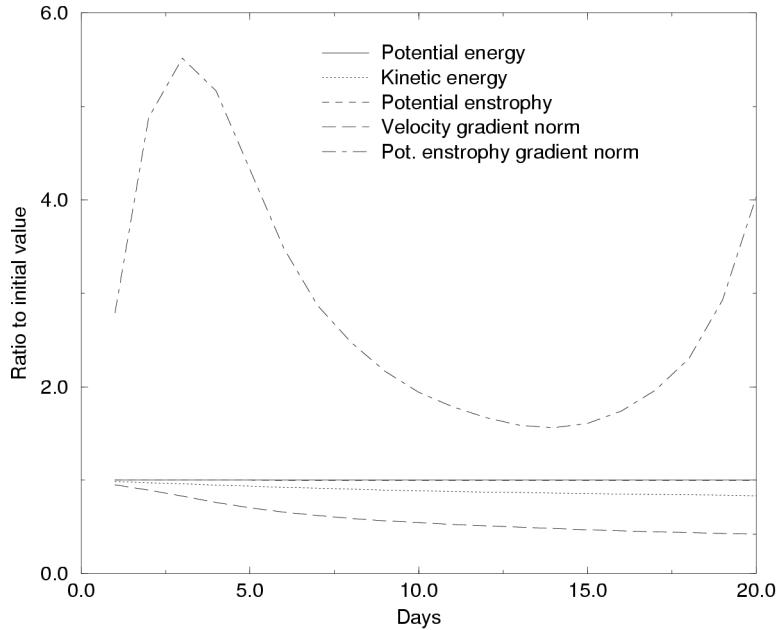


Figure 10. Diagnostics from shallow water integration with $gh_0 = 10^5$.

The results show almost exact conservation of potential enstrophy, indicating that the numerical methods, which do not use potential vorticity as a variable, are not introducing damping. They show near conservation of the kinetic energy (16% loss in 20 days) and potential energy. Since geostrophic balance is quite accurate on the scales modelled, the kinetic energy density is essentially $0.5(f^{-1}g\nabla h)^2$ and the potential energy is $0.5h^2$. Thus individual conservation of kinetic and potential energy indicates that the mean scale of the height field will be conserved. The velocity gradient norm decreases by 58% over 20 days, showing that enstrophy conservation is not sufficient to control all the velocity gradients. The enstrophy gradient norm increases rapidly over the first two days. After this time, the computations cannot resolve the filamentation of the potential vorticity field. Contour dynamics methods, [Dritschel](#) and Amabaum (1997), have to be used to follow the evolution further.

The behaviour of these computations is exactly that expected from the theory of two-dimensional incompressible flow except for the lack of an inverse cascade of energy to large scales. This is because the rate of change of potential energy is given by

$$\frac{\partial}{\partial t} \int h^2 dx dy = - \int (\mathbf{u} \cdot \nabla \left(\frac{1}{2} h^2 \right) + h^2 \nabla \cdot \mathbf{u}) dx dy \quad (20)$$

and this vanishes for non-divergent flow if there is no flow across the boundaries. The divergence is very small in shallow water integrations with U/c small, so the kinetic energy will be almost conserved and the mean scale of the height field conserved. In pure two-dimensional turbulence as described by (7), there is no potential energy; so this constraint does not operate. The inverse cascade can only operate in shallow water flow with $L \ll L_R$ so that the geostrophic constraint does not operate.

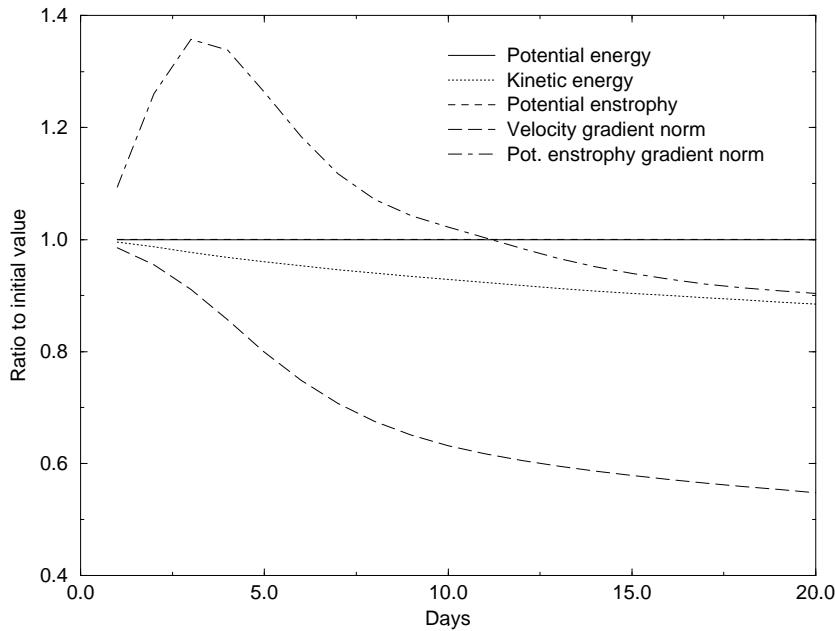


Figure 11. Diagnostics from shallow water integration with $gh_0 = 5000$.

Figure 11 illustrates the behaviour of the shallow water equations with $L > L_R$. As expected from equation (19), the effects of advection are weakened and the rate of increase of the enstrophy gradient norm is much less. Thus there is no longer a strong cascade of enstrophy to small scales. The potential vorticity distribution now looks much like the height distribution, as shown in (19). Both change only very slowly in time. The kinetic energy decrease is 12%, so the mean scale of the height field hardly changes.

These results illustrate the importance of the deformation radius L_R in determining the behaviour of ‘slow’ solutions of the shallow water equations with $U \ll c$. On scales smaller than L_R , the dynamics is essentially vortex dynamics. On scales larger than L_R , the solutions are almost stationary. Further details of these computations are given in Cullen (2002). More detailed study of these issues is given by Farge and Sadourny (1989) and Larichev and McWilliams (1991). It should also be noted that the ‘slow’ solutions of the shallow water equations can be approximated much more accurately than by the quasi-geostrophic equations. However, the qualitative distinction between flows with L greater than or less than L_R remains.

5. THREE DIMENSIONAL EQUATIONS

5.1 Basic equations, and filtering of fast waves

We start with the three-dimensional compressible Navier-Stokes equations. These are valid up to heights in the atmosphere where the continuum hypothesis breaks down. They can be written as follows:



$$\begin{aligned}\frac{D\mathbf{u}}{Dt} + C_p \theta \nabla \Pi + (-fv, fu, g) &= v \nabla^2 \mathbf{u} \\ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) &= 0 \\ \frac{D\theta}{Dt} &= \kappa \nabla^2 \theta \\ p &= R \rho \theta \Pi \\ \Pi &= (p/p_0)^\kappa\end{aligned}\tag{21}$$

Here, Π is the Exner pressure, θ the potential temperature and $\kappa = (\gamma - 1)/\gamma$, where γ is the ratio of specific heats. The rest of the notation is standard. The typical boundary conditions would be $\mathbf{u} \cdot \mathbf{n} = 0$ on the lateral and lower boundaries of a domain Ω , and $\rho, p \rightarrow 0$ as $z \rightarrow \infty$.

This system of equations has 5 independent evolution equations. Two of these are sound waves, with speed $c = \sqrt{\gamma(p/\rho)}$, two are inertio-gravity waves, and one has the time-scale of the advecting velocity. The inertio-gravity wave frequency is approximately

$$\frac{f^2 m^2 + N^2(k^2 + l^2)}{k^2 + l^2 + m^2} \tag{22}$$

where N^2 is the Brunt-Vaisala frequency $\frac{g \partial \theta}{\theta \partial z}$, k^2, l^2 are horizontal wavenumbers, and m is a vertical scale. Figure 4 shows that there is a large scale separation between sound waves and meteorologically significant motions, which tend to have the advective time-scale. There is also a large scale separation between the viscous dissipation scales and meteorological motions. However, there is not much scale separation between inertio-gravity waves and meteorological motions.

In all circumstances relevant to weather forecasting we have $|\mathbf{u}| \ll c$, so that it is appropriate to study equations from which sound waves have been filtered. Since viscous effects are also well-separated in general, we initially study inviscid equations. However, viscosity is always important near rigid boundaries. It is also likely that the local generation of small scales will make viscosity important in the interior of the atmosphere on occasions. We will see how this possibility appears naturally in the analysis of the inviscid system.

The simplest method of filtering is to impose incompressibility, as in the two-dimensional case. If viscosity is also omitted, this gives

$$\begin{aligned}\frac{D\mathbf{u}}{Dt} + C_p \theta \nabla \Pi + (-fv, fu, g) &= 0 \\ \nabla \cdot \mathbf{u} &= 0 \\ \frac{D\theta}{Dt} &= 0\end{aligned}\tag{23}$$

Π is determined implicitly by

$$\nabla \cdot (C_p \theta \nabla \Pi) = \nabla \cdot ((fv, -fu, -g) - (\mathbf{u} \cdot \nabla \mathbf{u})) \tag{24}$$

These equations are only accurate if $fL \ll c$. On larger scales, the finite propagation speed of sound waves matters.



However, they are not accurate in the presence of the large basic state vertical density gradient of the atmosphere, as $\partial\rho/\partial z$ is a large term. Useful equations with the same mathematical structure as the incompressible equations are obtained by making, instead, the anelastic approximation Lipps and Hemler (1982), which replaces (23) by:

$$\begin{aligned}\frac{D\mathbf{u}}{Dt} + C_p \theta_0 \nabla \Pi' + (-fv, fu, -g\theta'/\theta_0) &= 0 \\ \nabla \cdot (\rho_0 \mathbf{u}) &= 0 \\ \frac{D\theta}{Dt} &= 0\end{aligned}\tag{25}$$

where $\rho_0(z)$, $\theta_0(z)$, $\Pi_0(z)$ are prescribed time-independent functions satisfying the equation of state: $p_0(z) = R\rho_0(z)\theta_0(z)\Pi_0(z)$ and the hydrostatic equation $C_p\theta_0\partial\Pi_0/\partial z + g = 0$. Primes indicate perturbations from the reference profiles. Eq. (24) becomes

$$\nabla \cdot (C_p \rho_0 \theta \nabla \Pi') = \nabla \cdot (\rho_0 (fv, -fu, g\theta'/\theta_0) - \rho_0 \mathbf{u} \cdot \nabla \mathbf{u})\tag{26}$$

These equations are ‘balanced’ in the sense that the pressure is determined implicitly from the velocity field. There is a conserved potential vorticity $((\zeta + f) \cdot \nabla \theta)/\rho_0$. However, unlike the two-dimensional case, there are still 3 independent evolution equations left. The equations are thus not ‘balanced’ in the sense of Hoskins *et al.* (1985), because the potential vorticity alone does not determine all the other variables. In particular, internal gravity and inertial waves are supported.

The limitation $fL \ll c$ can be avoided in numerical models by using a semi-implicit approximation to the compressible equations, as used in the UK Met Office ‘new dynamics’, Cullen *et al.* (1997). This effectively eliminates the sound waves. However, there is still a difficulty in representing the true upper boundary condition. The best method is a current research topic. Inappropriate choices may re-introduce the condition $fL \ll c$, which would not be acceptable in global models, though adequate for local mesoscale models.

We will see in section 6 that three-dimensional equations which can be described in terms of the evolution of a single scalar can be derived by assuming that typical frequencies are slow compared with the inertial frequency, so the Rossby number $U/(fL)$ is small, or that the internal Froude number $U/(NH)$, where H is a height scale, is small. While typical values of these ratios for weather systems are about 0.1, there are significant regions where they are O(1) or greater. Thus any system of equations based on potential vorticity alone will not be very accurate, as Figure 4 suggests.

5.2 Solution of the inviscid incompressible equations

The results which were proved by the vorticity method of 4.2 no longer hold, since vorticity is not conserved. We therefore use an alternative method which is based on approximating (25) by a semi-Lagrangian, semi-implicit scheme as used in the ECMWF and many other models, and proving that the sequence of approximations converges. For simplicity, we describe the method for the special case where $\rho_0(z)$, $\theta_0(z)$, $\Pi_0(z)$ are independent of z . A similar method can be used for the general case, but is more complicated to explain.

We assume that (25) is to be solved over a finite time interval $[0, T]$. We will divide this interval into n time-steps, and prove convergence as $n \rightarrow \infty$. In order to construct an approximation to the solution over a single step δt , we initially ignore the pressure gradient terms in (25) and integrate by characteristics. With f constant, this gives a first guess solution for the positions X, Y, Z of particles initially at $x(0), y(0), z(0)$:

$$(x_c, y_c) = (x(0) + v(0)/f, y(0) - u(0)/f) \quad (27)$$

$$X(\mathbf{x}(0), \delta t) = x_c + (x(0) - x_c)\cos(f\delta t) - (y(0) - y_c)\sin(f\delta t)$$

$$Y(\mathbf{x}(0), \delta t) = y_c + (y(0) - y_c)\cos(f\delta t) + (x(0) - x_c)\sin(f\delta t)$$

$$Z(\mathbf{x}(0), \delta t) = z(0) + w(0)\delta t + \frac{1}{2}g\frac{\theta'}{\theta_0}\delta t^2$$

The mapping $x(0), y(0), z(0)$ to $X(\delta t), Y(\delta t), Z(\delta t)$ does not satisfy the continuity equation or the boundary conditions. Therefore we correct this first guess by finding $\mathbf{x}(\delta t)$ satisfying

$$\mathbf{X}(\mathbf{x}(\delta t)) = \nabla P(\mathbf{x}(\delta t)) \quad (28)$$

where P is a convex function. The convexity is because the map from $\mathbf{x}(\delta t)$ to $\mathbf{X}(\mathbf{x}(\delta t))$ is close to the identity:

$$\mathbf{X}(\mathbf{x}(\delta t)) = \mathbf{x}(\delta t) + \frac{1}{2}C_p\theta_0\nabla\Pi'(\mathbf{x}(\delta t))\delta t^2 = \nabla\left(\frac{1}{2}|\mathbf{x}(\delta t)|^2 + \frac{1}{2}C_p\theta_0\Pi'\delta t^2\right) \quad (29)$$

The difference $\mathbf{x}(\delta t) - \mathbf{X}(\mathbf{x}(\delta t))$ represents the effect of the pressure gradient term integrated over the time-step. The procedure is illustrated in Figure 12 for a one-dimensional cross-section without rotation. The solid lines represent first guess trajectories, which are straight lines, and the dashed lines represent the quadratic curves obtained by solving (28). These curves force the mapping back to the identity, which is the only possible solution in one dimension with fixed boundaries.

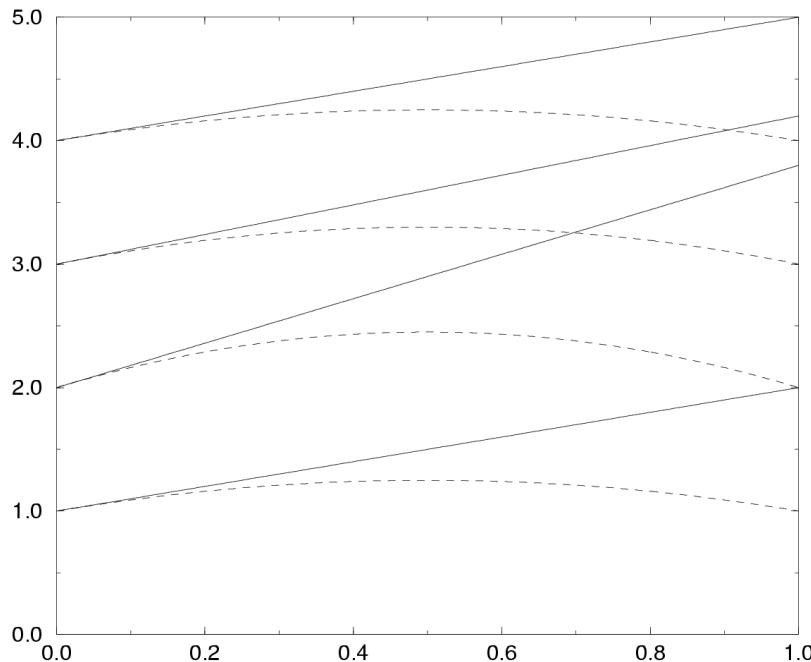


Figure 12. Illustration of the projection of a first guess trajectory onto one that satisfies the continuity equation.

The ‘polar factorisation’ theorem of Alexandrov, see [Pogorelov](#) (1973, p.475), and [Brenier](#) (1991) proves that this construction is always possible under appropriate conditions. It is equivalent to solving a Monge-Ampere equation (a nonlinear elliptic problem) for the trajectories. The right-hand side of the Monge-Ampere equation will be smooth if the velocity gradients are bounded. The result is a measure-preserving mapping $\mathbf{x}(\delta t) \rightarrow \mathbf{X}(\mathbf{x}(\delta t))$ which satisfies the boundary conditions.

If this argument is used in the two-dimensional case, the bound on the velocity gradients follows from [\(12\)](#). [Caffarelli](#) (1996) has proved in two-dimensions that, under suitable conditions, the solution of the Monge Ampere equation is smooth for a smooth right-hand side. It is then possible to prove that $\mathbf{x}(\delta t) - \mathbf{X}(\mathbf{x}(\delta t)) = O(\delta t^2)$. It should then be possible to break any finite time interval $(0, T)$ into n intervals of length δt , and construct a sequence of measure-preserving mappings $\mathbf{x}(0) \rightarrow \mathbf{x}(\delta t) \rightarrow \dots \rightarrow \mathbf{x}(n\delta t) \rightarrow \dots \rightarrow \mathbf{x}(T)$. Provided that $\mathbf{x}(\delta t) - \mathbf{X}(\mathbf{x}(\delta t)) = O(\delta t^2)$, the velocity will evolve continuously in time as $\delta t \rightarrow 0$ and the discrete sequence of mappings will converge to a measure-preserving mapping satisfying [\(25\)](#). Thus the results obtained by the vorticity argument in [3.3](#) should be reproduced.

In three dimensions, the necessary bound on the velocity gradients cannot be proved, and it is possible that they increase according to the equation

$$\frac{d}{dt} \|\nabla \mathbf{u}\| = C \|\nabla \mathbf{u}\|^2 \quad (30)$$

As we saw in [3.2](#), this has infinite solutions in finite time. At present it is a major research question as to whether singularities can be generated from smooth initial data in three-dimensional incompressible or anelastic flows. The majority opinion is that they can. If so, it may still be possible to follow this construction provided that we can prove $\mathbf{x}(\delta t) - \mathbf{X}(\mathbf{x}(\delta t)) = O(\delta t^2)$. However, the limit solution may involve particles losing their identity and ‘mixing’, thus implying the effect of a small but finite viscosity.

A numerical model that follows this procedure at finite resolution can be expected to be stable and converge as resolution increases. The key properties that must be satisfied are the accurate solution along trajectories, in particular the use of quasi-monotone schemes to ensure that new values cannot be generated by advection; and the implicit modification of the trajectories to enforce the continuity equation and boundary conditions. The possible mixing of trajectories as the resolution increases will be automatically handled by the need to average all quantities to grid-points, so no extra viscosity should be required. The requirement that the mapping from time-step to time-step is measure preserving is just as appropriate for a model solving [\(25\)](#) averaged over a finite region as for the exact solution. The difference from the vorticity method of section [3.3](#) is that the monotonicity requirement now applies to the velocity components, rather than the vorticity.

Note that, if the solutions are not smooth, and particularly if mixing of trajectories occurs, derivatives of the velocity field will become meaningless. However, quantities such as vorticity and divergence may still be meaningful in a volume-integrated sense. Thus Kelvin’s circulation theorem and local mass balance relations would still hold. However, the presence of singularities would prevent any proof that solutions were unique, so there will be a fundamental loss of predictability in the sense of [Lorenz](#) (1969). Any results deduced by manipulating or differentiating the vorticity or divergence field will also be invalid.

This method of solution can in principle be applied to the fully compressible equations [\(21\)](#) provided $u \ll c$. This is related to theoretical results, see [Lions](#) (1996), which show that, if both solutions exist, there is a solution of the compressible equations close to that of the incompressible equations, where closeness is measured in terms of u/c . The continuity equation which is enforced by the projection now expresses conservation of mass, rather than non-divergence.

We can use the solution procedure to illustrate some other properties of the solution. If the first guess pressure is

chosen to be the hydrostatic pressure p_h , then (27) becomes

$$\begin{aligned} (x_c, y_c) &= (x(0) + v(0)/f, y(0) - u(0)/f) \\ X(\mathbf{x}(0), \delta t) &= x_c - \frac{1}{2} \frac{\partial}{\partial x} (p_h) \delta t^2 + (x(0) - x_c) \cos(f\delta t) - (y(0) - y_c) \sin(f\delta t) \\ Y(\mathbf{x}(0), \delta t) &= y_c - \frac{1}{2} \frac{\partial}{\partial y} (p_h) \delta t^2 + (y(0) - y_c) \cos(f\delta t) + (x(0) - x_c) \sin(f\delta t) \\ Z(\mathbf{x}(0), \delta t) &= z(0) + w(0)\delta t \end{aligned} \quad (31)$$

The largest term that contributes to $\mathbf{X}(\mathbf{x}(0), \delta t) - \mathbf{x}(0)$ is thus removed. However, the horizontal derivatives of p_h may blow up. The non-hydrostatic pressure is still critical in order to enforce the continuity equation, even though the total pressure may be dominated by the hydrostatic part.

If the rotation is rapid, then $\mathbf{x}_c - \mathbf{x}(0)$ will be small, and the projection of $\mathbf{X}(\mathbf{x}(0), \delta t) - \mathbf{x}(0)$ onto a measure-preserving mapping will be much simpler. Thus we can expect rapid rotation to improve the behaviour of the solution.

5.3 Practical implications of solution procedure

We summarise the implications of the solution procedure set out in the previous subsection:

- (i) Quasi-monotone schemes should be used, as good behaviour of the equations integrated along particle paths (without the pressure gradient term) is required.
- (ii) In support of this, note that ECMWF only obtained satisfactory behaviour of their model at T213 resolution after quasi-monotone schemes were introduced, (Temperton et al. 2001). Now they are in use, satisfactory results at resolutions up to T799 have been obtained without further changes to the schemes. Experience with very high resolution simulations, such as of bubble convection, is also that monotone schemes are needed.
- (iii) An elliptic problem has to be solved to force the trajectories to satisfy the continuity equation and boundary conditions. Implicit recalculation of the trajectories in semi-Lagrangian methods is needed.
- (iv) Eulerian methods should be designed to be equivalent to semi-Lagrangian methods, i.e. to upwind quasi-monotone interpolation. Note that the derivatives nominally approximated by Eulerian methods may not exist, but derivations based on characteristics will be valid. Flux form schemes have the advantage of enforcing the continuity and boundary conditions exactly.

We demonstrate point (iii) by computations using an anelastic model developed by P. Smolarkiewicz. The explicit version of this model is described in Smolarkiewicz and Margolin (1997) and the implicit version by Smolarkiewicz et al. (1999). The area of integration was 2000km square. All integrations shown used 91 levels with a 300m vertical spacing and a semi-Lagrangian semi-implicit scheme with a 5 minute time-step. The lower boundary condition was free-slip, and no viscosity or turbulence model was used. The example is of flow at 10 ms⁻¹ impinging on the Scandinavian orography.

Figure 13 shows the flow at the bottom level of the model and the flow Jacobian at the bottom level. The flow Jacobian is the ratio of the volume of fluid at the beginning and end of a time-step as inferred from the departure point calculation in the semi-Lagrangian scheme. The values should be 1. The errors range from +19% to -28%, and are largest in the region where experiments with a shorter time-step and more accurate boundary condition show the numerical errors in the flow are greatest.

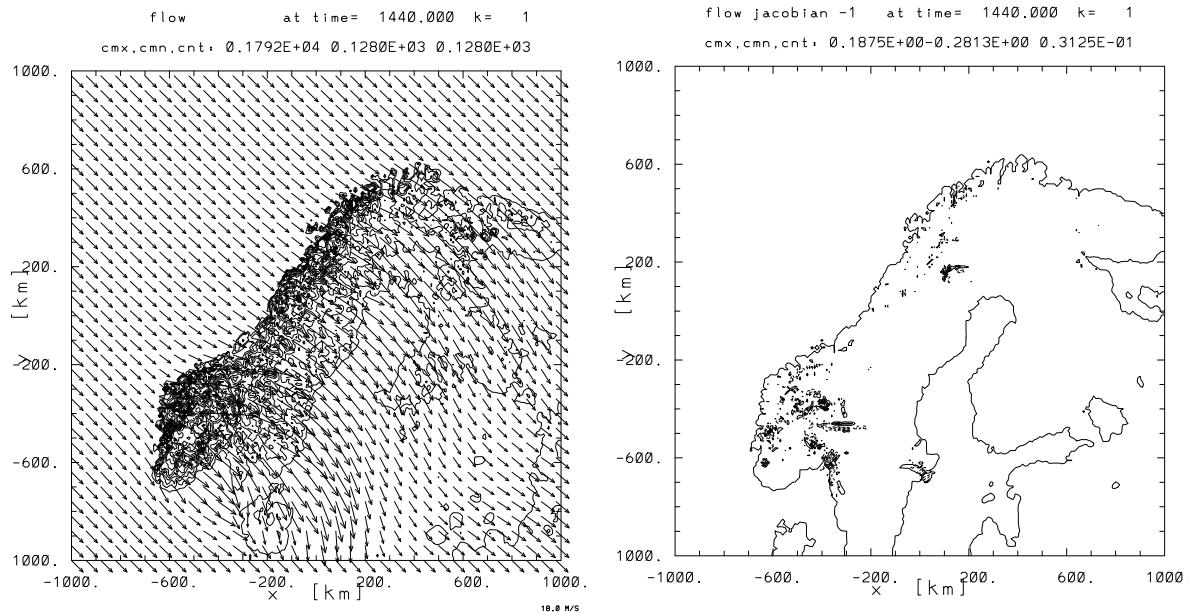


Figure 13. Low-level flow and flow Jacobian. Contour interval 3%.

Figure 14 shows a cross-section of the potential temperature, including breaking gravity waves forced by the orography. The errors in the flow Jacobian range from +12% to -19%, but are less than 3% over most of the region where the flow is smooth.

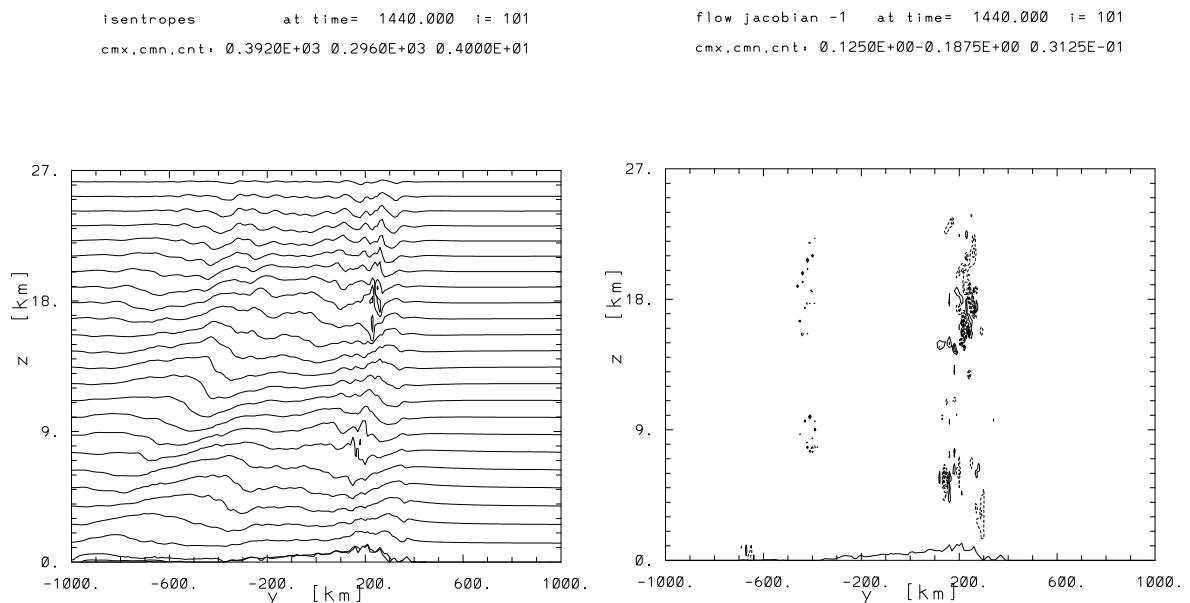


Figure 14. Cross-section of potential temperature and flow Jacobian. Contour interval 3%.

These examples show that the largest errors are where the trajectories are most distorted, or where approximations

made at the lower boundary limit the accuracy of the scheme. It should be possible to improve the simulations by correcting these errors.

5.4 The hydrostatic approximation

Though the hydrostatic equations are normally written in pressure coordinates, for uniformity we write the approximation directly into (21) as

$$\begin{aligned}
& \frac{D(u, v)}{Dt} + C_p \theta \left(\frac{\partial \Pi}{\partial x}, \frac{\partial \Pi}{\partial y} \right) + (-fv, fu) = 0 \\
& C_p \theta \frac{\partial \Pi}{\partial z} + g = 0 \\
& \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \\
& \frac{D\theta}{Dt} = 0 \\
& p = R\rho\theta\Pi \\
& \Pi = (p/p_0)^\kappa
\end{aligned} \tag{32}$$

These equations almost certainly blow up in an irrecoverable way. Suppose θ is uniform, then it will remain so for all time. Then $\frac{\partial \Pi}{\partial z}$ is independent of x and y , and $\frac{\partial \Pi}{\partial x}, \frac{\partial \Pi}{\partial y}$ are independent of z . If, initially, u and v depend on z , the pressure gradient term, which is independent of z , will be unable to prevent fluid trajectories colliding, as in the example of 3.2. The equation of state constrains ρ , in particular, if $\frac{\partial \Pi}{\partial x}, \frac{\partial \Pi}{\partial y}$ are smooth, ρ is also smooth. The continuity equation then requires $\frac{\partial w}{\partial z}$ to tend to infinity. There is no known way of resolving the difficulty in two dimensions analogous to the one-dimensional construction of 3.2. Either viscosity or non-hydrostatic pressure has to be introduced. However, if u and v are initially independent of z , they remain so for all time and two-dimensional theory can be used to show that the problem can be solved.

Scale analysis shows that the hydrostatic approximation is only valid on time-scales greater than the reciprocal of the buoyancy frequency, $N^{-1} = \theta/(g\partial\theta/\partial z)$. Typical values are 100s (a period of 600s) in the troposphere and 10s (a period of 60s) in the stratosphere. However, in moist saturated neutral regions of the atmosphere, $N^{-1} \rightarrow \infty$, assuming phase changes are instantaneous. The current ECMWF model timestep is 900s for a horizontal resolution of 40km. Thus we can expect to resolve the typical period when the hydrostatic equations break down at about a 10km resolution. However, it will break down locally in weakly stratified regions before this.

In the above example, $N^{-1} = \infty$. Thus the hydrostatic approximation is invalid, and the breakdown of the solution should be resolved by including a non-hydrostatic pressure rather than viscosity. In the current ECMWF semi-implicit scheme, a reference basic state with a large static stability is used in the implicit step. This introduces a computational non-hydrostatic pressure of $O((N - N_0)\delta t^2)$, where N_0 is the buoyancy frequency of the reference state. This is sufficient to stabilise the model. However, it may introduce errors into the simulation. Figure 15 shows the effect of changing the reference profile averaged over 12 forecasts with the ECMWF model (CY23R4) at T511 resolution.

The hydrostatic equations, unlike the anelastic equations, can be solved explicitly. However, even in the UK Met Office explicit Unified Model (Cullen and Davies (1991)), a reference profile is used in the short time-steps which compute the pressure gradient and vertical advection, so that the same stabilisation mechanism is present. It is important to remember that a hydrostatic model stabilised this way will not give correct solutions where the stratifi-

cation is weak, the solutions in these regions will be a computational artefact. It is thus important that the parametrisations maintain an adequate level of stability.

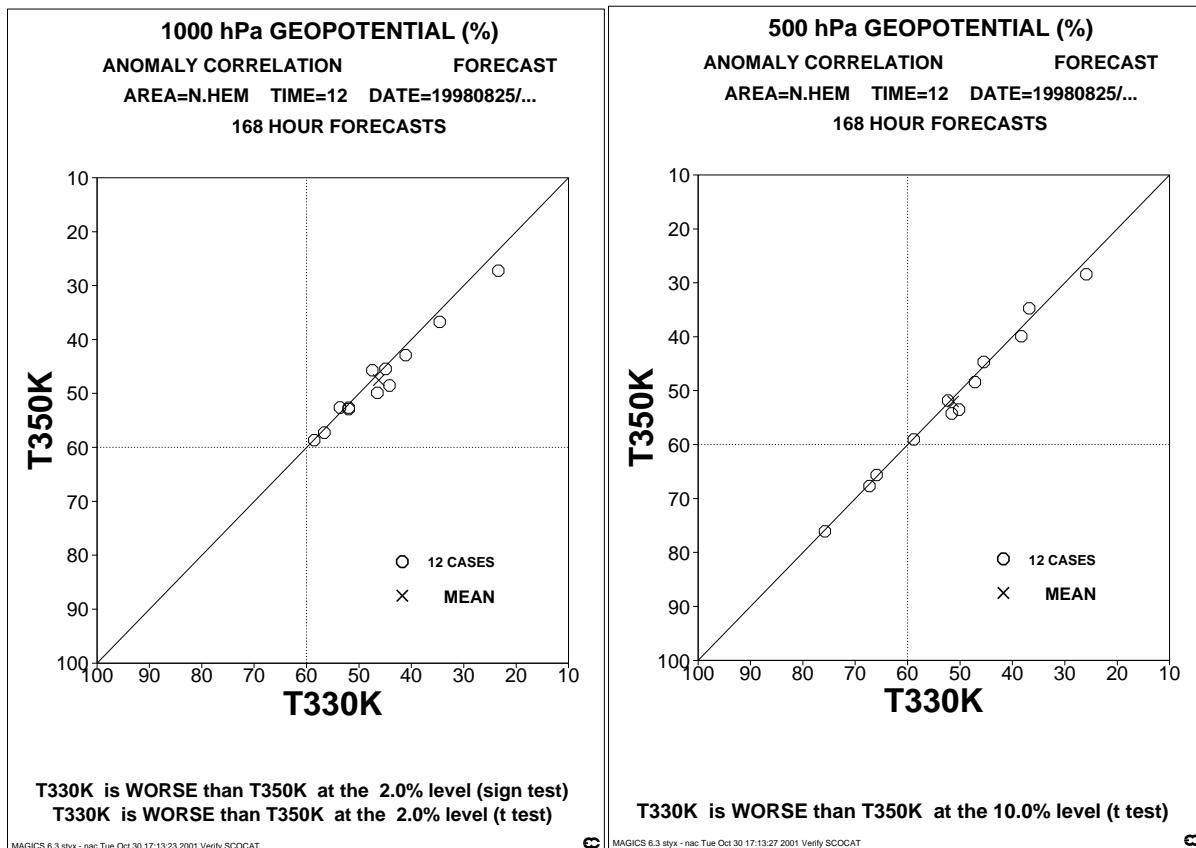


Figure 15. Anomaly correlations for 12 forecasts for periods between August 1998 and September 1999 using different reference profiles of temperature in the semi-implicit scheme.

6. AVERAGED EQUATIONS

6.1 Averaged equations, and their approximation by balanced models

The preceding sections have discussed exact equations and approximations to them which are valid for subsonic flow. However, it is clear that operational models greatly under-resolve the solutions of these equations, as illustrated by the observations reviewed in [Section 2](#). In practice, it is more realistic to view them as solving equations that have been averaged in space and time. This is well-known to large-eddy modellers, who are always disciplined in defining averaging scales in the problems they solve.

[Figure 4](#) illustrates that averaging will involve compromises. For instance, a horizontal averaging scale of 80km, compatible with the T511 ECMWF model resolution of 40km, cuts through the dominant scale of frontal zones and of internal and external gravity waves. The planned resolution of 15km achievable later in the decade is compatible with a 30km averaging scale, which is still in the most active scale for mesoscale weather systems and gravity waves. Thus accurate representation of these phenomena will not be possible.

Theory also shows that averaging involves compromises. Averaged equations include terms representing correla-

tions between resolved and unresolved motions, which can only be modelled empirically. They thus never have the fundamental validity of the original equations. Accurate deterministic prediction with averaged equations can only be expected if there is a spectral gap, which means that interactions between resolved and unresolved flows will be weak. However, [Figure 3](#) shows that there is no such gap in atmospheric data. Accurate prediction is also possible if the averaging can select resolved and unresolved processes which interact weakly for other reasons. For instance, flows which are almost entirely rotational interact weakly with flows that are almost entirely divergent. This situation is only likely if the averaging reflects a physically important scale of the flow. However, in weather forecasting, typical resolutions are finer than the Rossby radius, above which the rotational constraint dominates, but not fine enough to resolve all the key scales associated with stratification.

If the equations are averaged in space and time in a Eulerian sense, and the sub-grid model is correct, the solutions must be smooth on the averaging scale because the real flow is smooth when averaged to this scale. The analysis of section [5.2](#) suggests that this will be achieved in an incompressible non-hydrostatic incompressible model by using monotone advection schemes and implicit calculation of trajectories. This has been verified by computations using the same model as used in section [5.3](#). However, it is important that sub-grid modelling of motions which are sub-grid scale but have a major large-scale impact, such as convection, need to assume that their input data represents averaged values and be designed so that the solution of the resolved and sub-grid models together stays smooth on the averaging scale.

There is no reason why the form of the unaveraged Navier-Stokes equations has to be retained once the equations are averaged. For instance, the hydrostatic equations become appropriate when the averaging scale is greater than about 15km, as the regions of weak stability where the hydrostatic equations blow up are no longer resolved. Another possibility is discussed by [Gent](#) and McWilliams (1996), and is widely used in ocean models. Averaging the advection term $\mathbf{u} \cdot \nabla \theta$ gives

$$\overline{\mathbf{u} \cdot \nabla \theta} = \bar{\mathbf{u}} \cdot \nabla \bar{\theta} + \mathbf{u}' \cdot \nabla \theta' \quad (33)$$

assuming that perturbation quantities average to zero. The final term can be modelled in terms of mean quantities as $A\theta + B\theta'$, where A is the antisymmetric part and B the symmetric part. A can be modelled as advection by a ‘bolus’ velocity, e.g. Stokes drift or ‘entrainment velocity’. This provides a non-diffusive element to the sub-grid model, and allows smoothness to be achieved without relying entirely on diffusion.

Ideally, we would like to average without imposing smoothness, so as to capture sharp changes such as that illustrated in [Figure 6](#). One possible method is to use Lagrangian averaging, so that we average over air parcels as they move in space and time, but allow different air masses to be separated by sharp boundaries. An example of this approach is set out by [Andrews](#) and McIntyre (1978) and [Buhler](#) and McIntyre (1998). They show that the Lagrangian mean momentum equation takes the form

$$\frac{D(\mathbf{u} + \mathbf{p})}{Dt} + C_p \theta \nabla(\Pi + \pi) + (-fv, fu, g) = 0 \quad (34)$$

\mathbf{p} is the ‘pseudo-momentum’ associated with the waves removed by averaging. The trajectory is defined by the ‘Lagrangian mean’ velocity. π is a perturbation pressure. Once again we see that the trajectory is defined by a different velocity from the momentum.

A natural averaging scale is given by the inertial frequency f . The semi-geostrophic model is an accurate approximation to (34) if $(D/Dt)(\text{wind} - \text{direction}) \ll f$. It is obtained by neglecting π and assuming $\mathbf{u} + \mathbf{p}$ is in geostrophic balance. This model can be proved to have solutions, e.g. [Cullen](#) and Purser (1989), [Benamou](#) and Brenier (1998), but the condition $(D/Dt)(\text{wind} - \text{direction}) \ll f$ restricts its accuracy to scales greater than

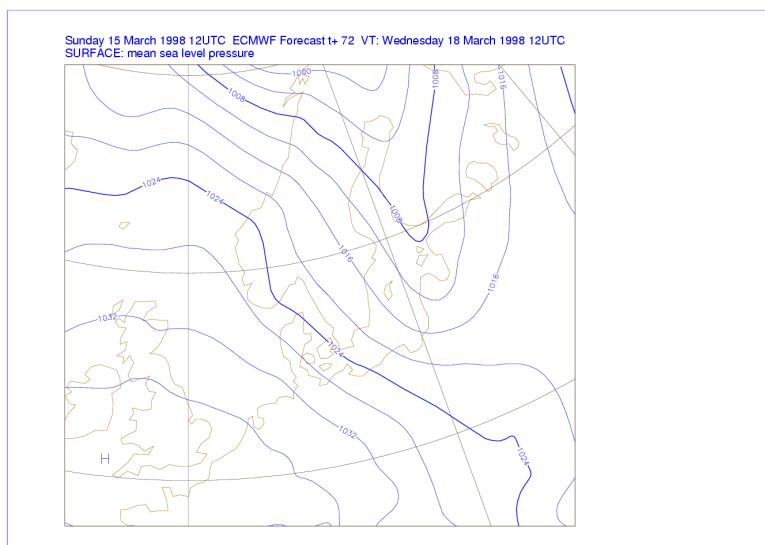
the Rossby radius (Cullen, 2000). Accuracy of semi-Lagrangian advection schemes requires that $(D/Dt)(\text{wind} - \text{direction}) \ll \delta t^{-1}$, which is a condition of the same form but much less restrictive at operational resolutions. The current operational resolution, T511, is integrated with a 15 minute time-step. The accuracy condition is thus 12 times less restrictive than the semi-geostrophic condition in middle latitudes. It thus allows a wider variety of solutions, such as inertial waves.

Another natural scale is defined by the condition that the Froude number U/NH is small. This requires internal gravity waves to propagate faster than the advection velocity. In the ECMWF model with 60 levels, only 5 vertical modes have a phase speed greater than 100ms^{-1} , and only 15 greater than 30ms^{-1} . Thus balanced models based on fast gravity wave speeds are only likely to work with limited vertical resolution.

A finer natural averaging scale would be given by the buoyancy frequency. As discussed in section 5.4, this is not resolved in operational models. Thus we either have to accept a mixture of well- and partly-resolved motions, or impose a rather restrictive averaging scale.

6.2 Computations of averaged solutions.

Since the averaging scales used in operational models do not correspond to a physical scale separation, even the correct averaged behaviour may not look like what is expected. We illustrate with an example of flow over Scandinavia shown in Figure 16. There is a north-westerly flow roughly at right angles to the Scandinavian orography. A satellite picture at the same time, Figure 17, shows possible wave activity over the mountains, but skies are clear over the Baltic. Thus we expect that the solution of the equations will contain large vertical motions over the orography, but only small vertical motions over the Baltic. A cross-section from a forecast using the then operational model at T319 resolution is shown in Figure 18. The cross-section line extends across the Baltic. However, the model orography, shown in the cross-section, has a downslope which also extends half-way across where the Baltic should be. The cross-section shows vertical motions extending right across the Baltic. However, they may well be a realistic representation of how the atmosphere would have responded to the model orography. Other experiments (not shown) show that the model waves are not numerical artefacts.



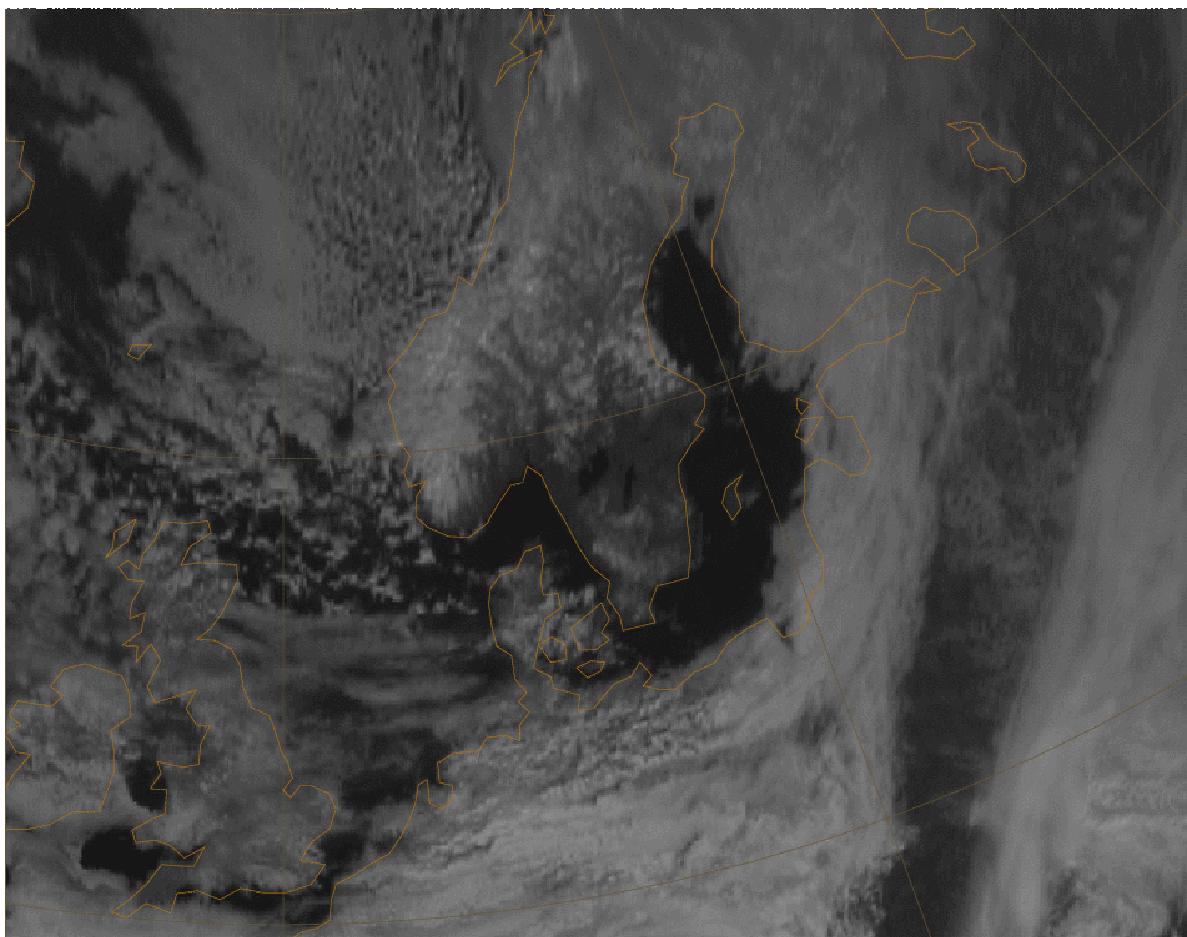


Figure 17. METEOSAT visible picture for 18 March 1998, 1200 UTC.

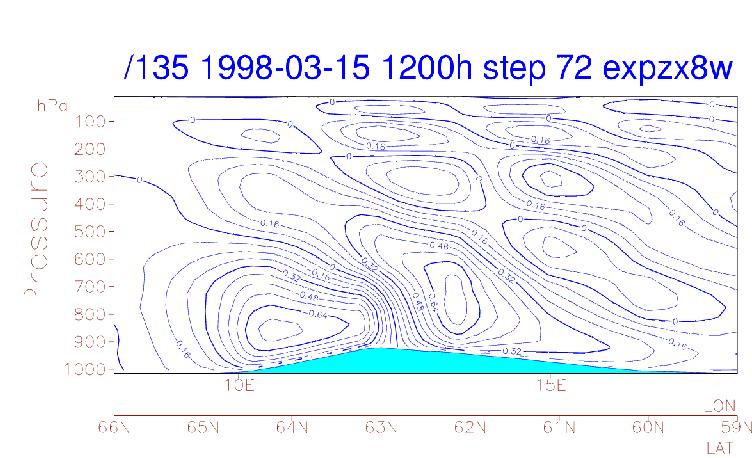


Figure 18. Cross-section of vertical velocity from operational formulation at T319 and a 20 minute time-step.

We show that the forecasts from the operational model may well be accurate forecasts of the averaged state of the atmosphere by using the research model used in section 5.3. The problem solved is exactly as in 5.3. We compare integrations with a 10km grid where the results are averaged to a 80km grid with results from the model with a 40km grid. Because of numerical errors, we can only expect results from a 40km grid to be verifiable at 80km res-

olution. The agreement is quite good, both for the low-level flow, Figure 19, and for the vertical motion, Figure 20. The operational model used to produce Figure 18 has a resolution of about 60km. The results shown in Figure 20 are sufficiently similar to this to suggest that the operational model may be giving a reasonable prediction of the averaged flow. A similar comparison was made between a 20km model and the 10km results averaged to 40km. Again there was reasonable agreement.

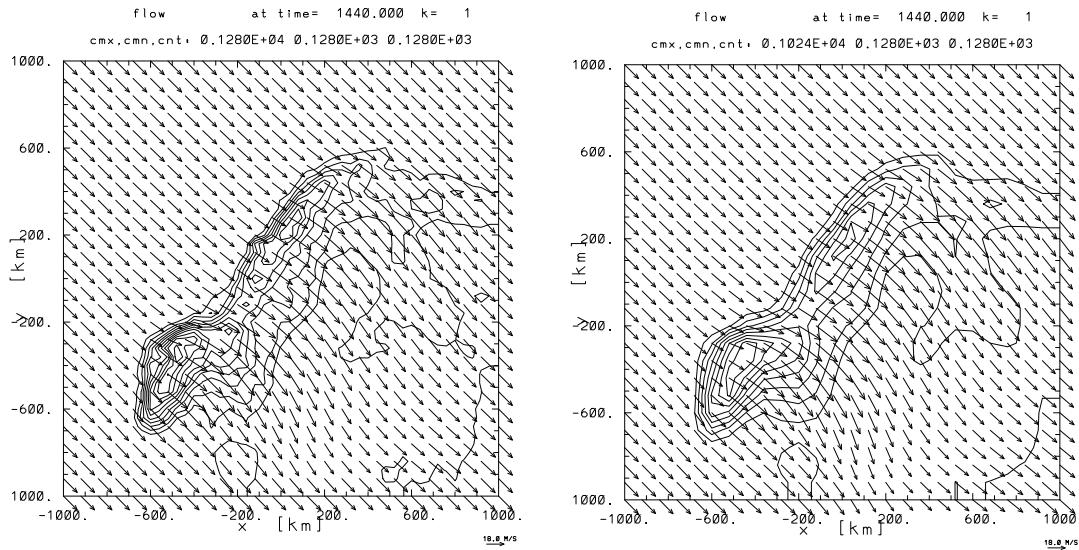


Figure 19. Comparison of low-level flow produced by 40km model (left) and 10km model averaged to 80km (right)

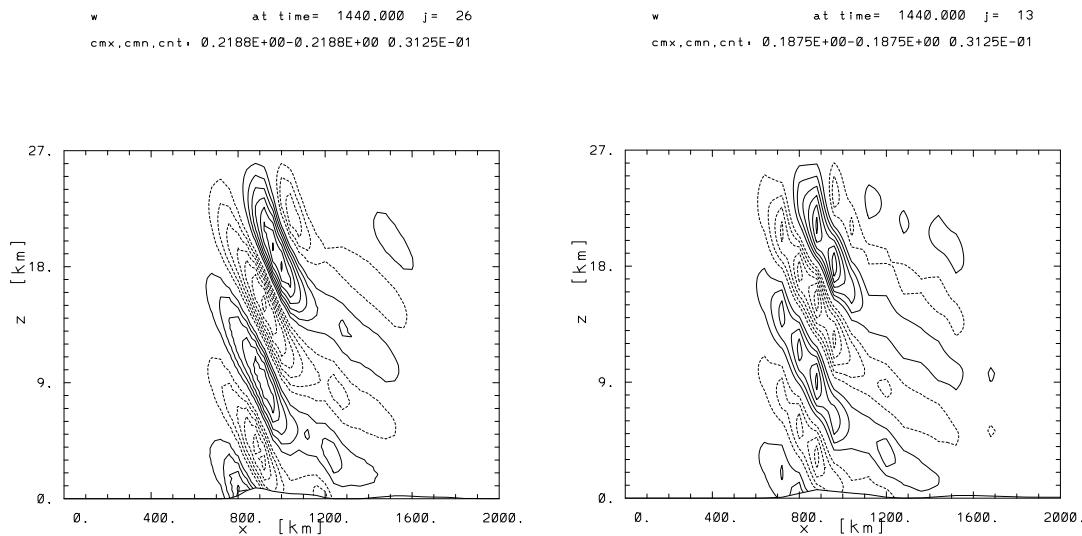


Figure 20. Comparison of vertical velocity cross-section produced by 40km model (left) and 10km model averaged to 80km (right).

6.3 Balanced models

As discussed in section 5.1, the exact equations governing the atmosphere support 5 time-like solutions. Two of these can be filtered with little loss in accuracy because the flow is very subsonic. However, we cannot further reduce the equations to give a system with one time-like variable by eliminating inertio-gravity waves and retain accuracy in all circumstances. This is because the inertio-gravity wave frequency, (22), is similar to the advection frequency under many circumstances. In order to obtain a ‘balanced model’ with a single time-like solution, we must assume that either the Rossby number U/fL , the Froude number U/NH , or the aspect ratio H/L is small. The resulting equations will have much simpler solutions than either the full equations, (21), or the anelastic approximation (25). Typically they can be described by advection of potential vorticity, together with the solution of an elliptic problem to obtain the other variables, Hoskins et al. (1985). However, we can expect the solutions to be accurate approximations to those of the full equations on large scales.

An example is shown in Figure 21. This is a sequence of ECMWF analyses of potential vorticity in the upper troposphere and accumulated precipitation. We can see the close relation between the movement and development of the potential vorticity anomaly with the main areas of precipitation. It is likely that this correspondence would be retained in a balanced model, though the small scale detail would be lost.

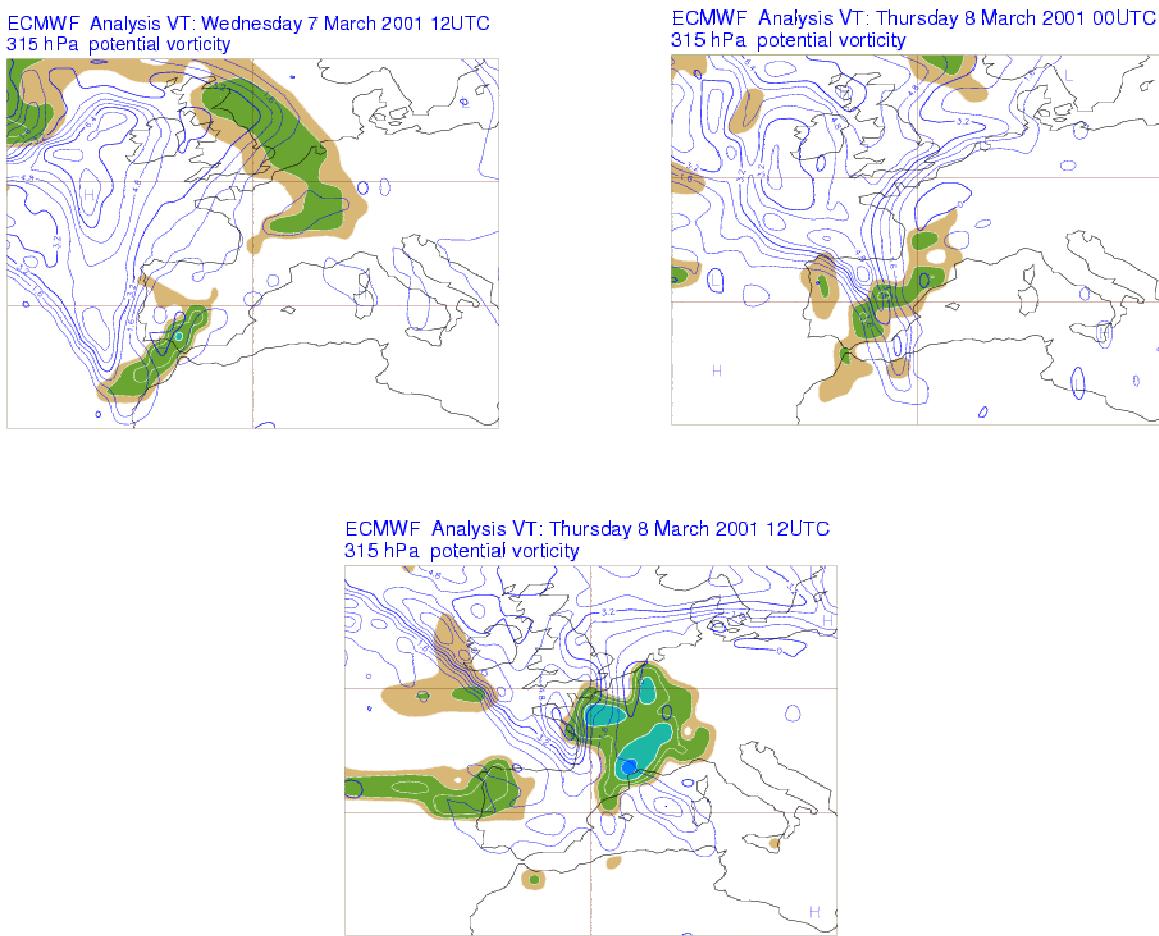


Figure 21. Analyses of potential vorticity on the 315K surface and 12 hour accumulated precipitation ending at analysis time.

There are many examples of balanced models, based on different expansions in Rossby or Froude numbers or as-

pect ratio. Two examples are the balance equations, and the semi-geostrophic model. The former is a much more accurate approximation in the context of the shallow water equations, Cullen (2000), but in three dimensions is liable to spontaneous instability McWilliams *et al.* (1998). The semi-geostrophic model can always be solved, Cullen and Purser (1989), Benamou and Brenier (1998), giving a ‘slow manifold’, but the solutions will not be as accurate. This reflects the complexity of the true solution, forcing a choice between an accurate local approximation by a simple model or a less accurate global approximation. Figure 22 is a schematic illustration of this choice. The thin line represents the exact solution of the equations of motion. The thick solid line represents the Lagrangian mean, with occasional sharp changes. The short-dashed line represents the semi-geostrophic solution, which is much simpler than the solution of the equations of motion, as many types of motion are filtered. The solution always exists but slowly loses accuracy. The long-dashed line represents the solution of the more accurate balance equations, which cannot be solved under all conditions. The breakdowns will usually correspond to places where the exact solution does not behave smoothly, so cannot be closely approximated by a simpler solution McWilliams *et al.*, 1998).

Example solution curves

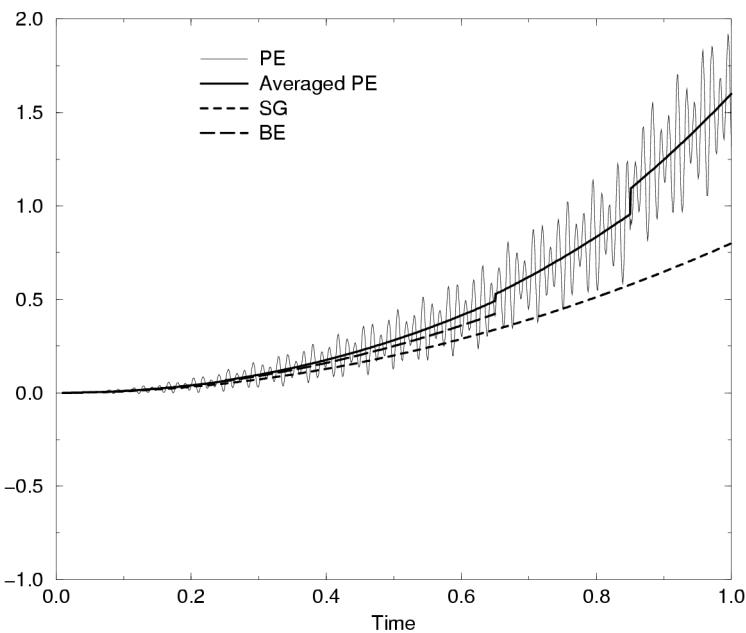


Figure 22. Schematic graph of solution curves of exact, averaged and approximate models.

A practical difficulty with the use of balance techniques in operational models is the solution of the elliptic problem to derive other variables from the potential vorticity. In order to derive the divergence, or vertical motion, or ageostrophic wind; we have to solve an ‘omega’ equation. This takes the generic form

$$L\omega = F \quad (35)$$

The accuracy of the balanced model assumes that the eigenvalues of L are large, so that a realistic estimate of the vertical motion ω is produced from the forcing F . However, as discussed in section 6.1, balance assumptions are not applicable on the smaller horizontal and vertical scales resolved by the ECMWF model, and this will be associated with small eigenvalues of L . For instance, there are 11 vertical modes in the 60 level model with internal



gravity wave speeds less than 3ms^{-1} . Straightforward application of (35) will thus lead to very large and unrealistic results. This is why, for instance, nonlinear normal mode initialisation cannot be used for more than about 5 vertical modes.

6.4 Structure of the solution of the semi-geostrophic model

As shown in section 6.1, the semi-geostrophic model is most usefully considered as an approximation to the averaged equations, so we must include the sub-grid model terms. In these notes we cannot consider more than a very limited part of the sub-grid model. A fuller discussion of this material is included in Cullen and Salmon (2002). The semi-geostrophic equations can then be written as:

$$\frac{D\mathbf{u}_b}{Dt} + C_p \theta \nabla_h \Pi + (-fv, fu) = \frac{\partial}{\partial z} \left(K_b \frac{\partial}{\partial z} (2(u_b, v_b) - (u, v)) \right) \quad (36)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$$

$$\frac{D\theta}{Dt} = S$$

$$p = R\rho\theta\Pi$$

$$\Pi = (p/p_0)^\kappa$$

$$\left(fv_b + \frac{\partial}{\partial z} \left(K_b \frac{\partial u_b}{\partial z} \right), -fu_b + \frac{\partial}{\partial z} \left(K_b \frac{\partial v_b}{\partial z} \right), -g \right) = C_p \theta \nabla \Pi$$

The boundary layer friction is represented by vertical diffusion, and S is a thermodynamic source term. The mixing coefficient K_b is computed only from the balanced variables. The last equation of (36) has been modified to include the friction terms, so that geostrophic balance has been replaced by Ekman balance, hence the suffix b . Write the last equation of (36) in the form $\nabla \Pi = \mathbf{F}(v_b, -u_b, \theta)$, where \mathbf{F} is a matrix operator. The equations can then be written, following Schubert (1985) as

$$\mathbf{Q} \begin{bmatrix} u - u_b \\ v - v_b \\ w \end{bmatrix} + \frac{\partial}{\partial t} \mathbf{F}^{-1} \nabla \Pi = \mathbf{H} \quad (37)$$

where

$$\mathbf{Q} = \begin{bmatrix} f + v_{bx} & v_{by} + \frac{\partial}{\partial z} \left(K_b \frac{\partial}{\partial z} \right) & v_{bz} \\ -u_{bx} - f \frac{\partial}{\partial z} \left(K_b \frac{\partial}{\partial z} \right) & f - u_{by} & -u_{bz} \\ \Theta_x & \Theta_y & \Theta_z \end{bmatrix} + \mathbf{Q}_E \quad (38)$$

and

$$\mathbf{H} = \begin{bmatrix} -\mathbf{u}_b \cdot \nabla v_b \\ \mathbf{u}_b \cdot \nabla u_b \\ \mathbf{u}_b \cdot \nabla \Theta + S \end{bmatrix} \quad (39)$$

We have allowed additionally for the effects of moisture, not included in (36), by using Θ to represent the potential temperature in unsaturated regions and the moist equivalent potential temperature θ_E in saturated regions, and \mathbf{Q}_E to represent a non-local matrix describing the convective plumes.

Equation (37) shows that the rate of change of the balanced pressure is driven by the forcing term \mathbf{H} . The response to the forcing, expressed as the total wind (u, v, w) , is determined by the potential vorticity matrix \mathbf{Q} . \mathbf{H} contains geostrophic advection, radiation, and vertical diffusion, all of which have slow time-scales. If \mathbf{H} contains fast time scales (compared to f^{-1}), this decomposition is inappropriate. It is equally important not to force operational models on fast time-scales, since the response is unlikely to be modelled realistically.

The PV matrix \mathbf{Q} includes the stabilising effect of boundary layer friction, and the destabilising effect of moisture. If the atmosphere is stable to moist (slantwise) ascent, the first matrix term in (38) has all positive eigenvalues. If an eigenvalue becomes zero, the smooth transport velocity (u, v, w) is replaced by mixing over the region where the zero eigenvalue exists. If there is a negative eigenvalue, the second term, which represents the non-local effects of convection, comes into play. A ‘correct’ convective parametrisation for the full equations (21) should ensure that the total \mathbf{Q} has no negative eigenvalues, or else the model will still support explicit convective instability.

It can be proved, Cullen and Purser (1989), that there is an essentially unique choice of trajectory that satisfies (37), and maintains a pressure whose implied geostrophic wind and temperature are inertially and statically stable. If the actual state is potentially unstable to moist convection, this solution will include convective plumes, as illustrated in Figure 23, and discussed in more detail by Shutts et al. (1988). Real states which are statically or inertially unstable evolve on time-scales faster than f^{-1} , so that the semi-geostrophic approximation is inappropriate. The effect of the approximation is to force such fast motions to happen instantaneously, so that there is a sharp boundary between ‘slow’ and ‘fast’ motions. As shown earlier in Figure 3, this is a simplification of the real case, where there are no sharp distinctions. In operational models of resolution similar to ECMWF, Figure 4 shows that there is not enough resolution to treat many ‘fast’ motions accurately, so they have to be parametrised. The practical situation is thus not that different from the simplified picture of equation (37).

Since the trajectory is determined implicitly by the forcing, implicit calculation of the trajectory is preferred in operational models. The boundary layer friction, which appears in the lowest order balance, and the convective mass transport, should also be determined implicitly.

Figure 24 and Figure 25 show the effect of using a predictor-corrector scheme to integrate the dynamics and physics in CY23R4 of the ECMWF model. Further details are given in Cullen and Salmon (2002). In this scheme each time-step is repeated, allowing implicit adjustment of the trajectories and the parametrised transports. Figure 25 shows that the effect is larger than that obtained by simply halving the timestep. Figure 24 shows that it is comparable, except at the end of the forecast range, to the difference made by changing the horizontal resolution from T319 to T511.

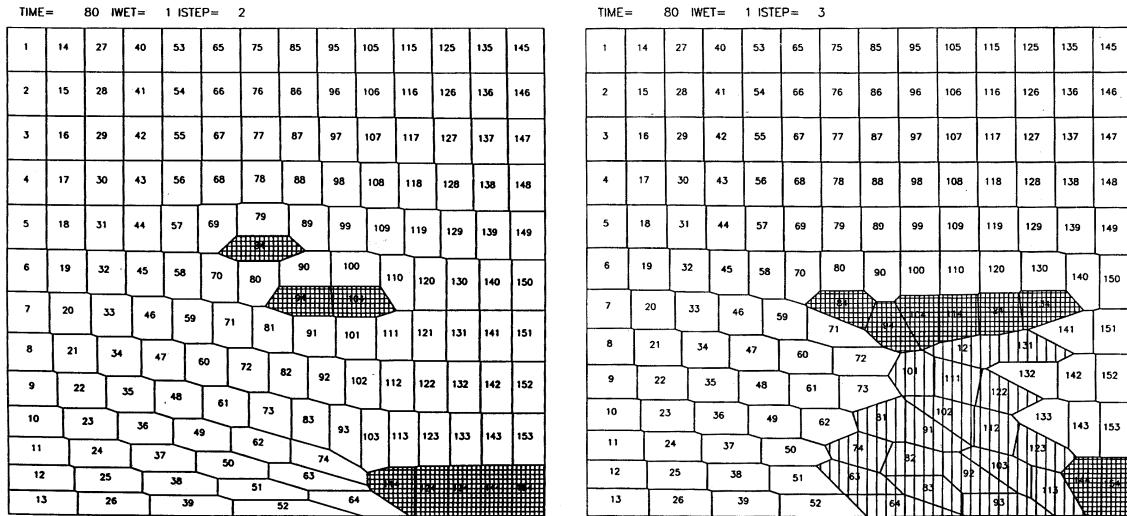


Figure 23. Left: Fluid element picture showing a vertical cross-section of a frontal zone with moisture (hatched elements). Right: the striped elements have been cooled by precipitation falling from the convecting elements.

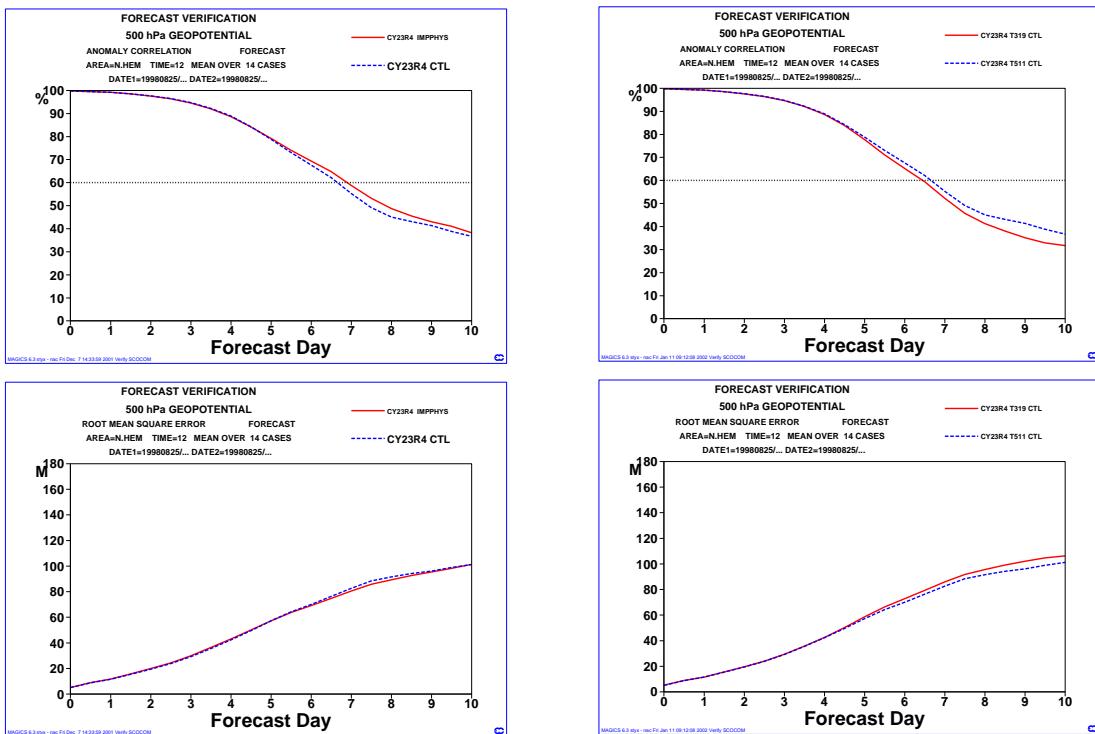


Figure 24. Left: comparison of predictor-corrector physics with control forecast, both at T511 resolution. Right: comparison of control forecasts at T511 and T319 resolutions.

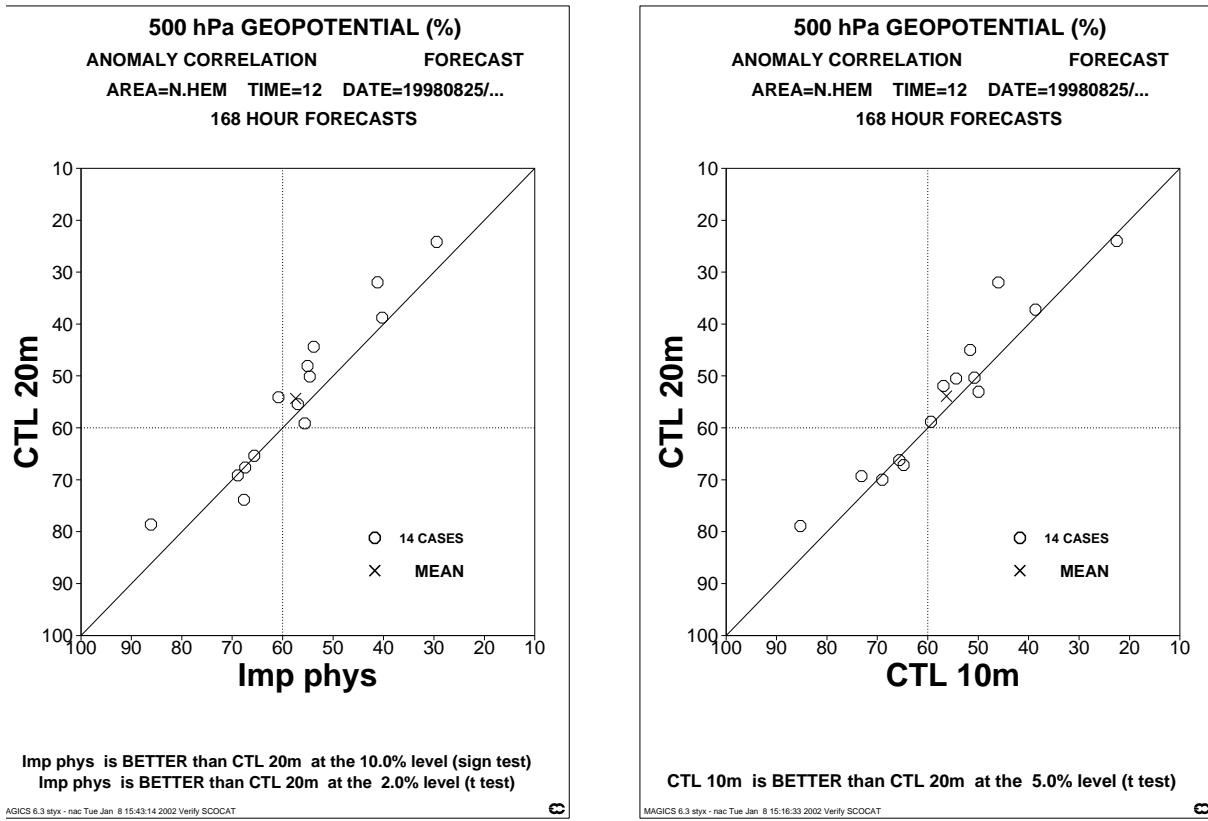


Figure 25. Left: comparison of predictor-corrector physics against control forecasts, both with 20 minute timesteps. Right: comparison between control forecasts with 10 and 20 minute timesteps.

7. SUMMARY

We have discussed the properties of some very simple models which are relevant to aspects of atmospheric flow, and then illustrated how they give information about the solution of the complete problem. The main points which are relevant to the design of operational models are summarised below.

- (i) The full solutions of the equations of motion have well-behaved solutions, but they are far too detailed to compute correctly in operational models.
- (ii) Simplifications of the full equations may or may not have well-behaved solutions. The hydrostatic equations do not have solutions for weak stratification. The semi-geostrophic equations always have solutions, but these only describe relatively large scale flow where rotation is dominant.
- (iii) Accurate solution of the complete equations for high Reynolds number subsonic flow (the relevant case) requires implicit determination of the trajectory and solution of an elliptic problem. Proper treatment of rigid boundaries is particularly important.
- (iv) Operational models are most realistically thought of as solutions of averaged equations. Only certain averaging scales correspond to physical scale separations, other choices of scale will not be so effective in practice. The average of a real atmospheric state may not look like what one would expect.

-
- (v) Balanced models are a useful guide to the large-scale behaviour of the atmosphere, in particular how to combine resolved and parametrised processes.

REFERENCES

- Andrews, D.G. and McIntyre,M.E., 1978: An exact theory of nonlinear waves on a Lagrangian-mean flow., *J. Fluid Mech.*, **89**, pp 609-646.
- Arakawa,A., 1966: Computational design for long-term numerical integrations of the equations of atmospheric motion. *J. Comput. Phys.*, **1**, pp 119-143.
- Batt,K.L. and Leslie,L.M., 1998: Verification of output from a very high resolution numerical weather prediction model: the 1996 Sydney to Hobart yacht race., *Met. Apps.*, **5**, pp 321-328.
- Benamou,J.D. and Brenier, Y., 1998: Weak existence for the semi-geostrophic equations formulated as a coupled Monge-Ampere equations/transport problem., *SIAM J. Appl. Math.*, **58** no. 5, pp 1450-1461.
- Bourgeois,A.J. and Beale,J.T., 1994: Validity of the quasi-geostrophic approximation for large-scale flow in the atmosphere and ocean., *SIAM J. Math. Anal.*, **25**, pp 1023-1068.
- Brenier,Y., 1991: Polar factorization and monotone rearrangement of vector-valued functions., *Commun. Pure Appl. Math.*, **44**, pp 375-417.
- Buhler,O. and McIntyre,M.E., 1998: On non-dissipative wave-mean interactions in the atmosphere or oceans., *J. Fluid Mech.*, **354**, pp 301-343.
- Caffarelli,L.A., 1996: Boundary regularity of maps with convex potentials II., *Annals of Math.*, **144**, pp 453-496.
- Cullen,M.J.P., 2000: On the accuracy of the semi-geostrophic approximation. *Q. J. Roy. Meteor. Soc.*, **126**, pp 1099-1116.
- Cullen,M.J.P., 2002: Large-scale non-turbulent dynamics in the atmosphere. ECMWF Tech. Memo. no.353 (submitted for publication).
- 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.
- Cullen,M.J.P. and Purser,R.J., 1989: Properties of the Lagrangian semi-geostrophic equations. *J. Atmos. Sci.*, **46**, 2684-2697.
- Cullen,M.J.P. and Salmon, D.J., 2002: On the use of a predictor-corrector scheme to couple the dynamics with the physical parametrisations in the ECMWF model, ECMWF Tech. Memo. no. 357, submitted for publication.
- Dritschel,D.G. and Ambaum,M.H.P., 1997: A contour-advection semi-Lagrangian numerical algorithm for simulating fine-scale conservative numerical fields. *Quart. J. Roy. Meteor. Soc.*, **123**, 1097-1130.
- Farge,M. and Sadourny,R., 1989: Wave-vortex dynamics in rotating shallow water., *J. Fluid Mech.*, **206**, pp 433-462.
- 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.
- Gent,P.R. and McWilliams,J.C., 1996: Eliassen-Palm fluxes and the momentum equation in non-eddy-resolving ocean circulation models. *J. Phys. Oceanogr.*, **26**, pp 2539-2546.
- Gerard,P., 1992: Resultats recents sur les fluides parfaits incompressibles bidimensionnels. *Seminaire Bourbaki*, **757**, pp 411-424.
- 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.
- Larichev, V.D. and McWilliams,J.C., 1991: Weakly decaying turbulence in an equivalent barotropic fluid., *Phys. Fluids*, **A 3(5)**, pp 938-950.
- Leith,C.E., 1983: Predictability in theory and practice., *Large-scale dynamical processes in the atmosphere*, ed. B.J.Hoskins and R.P.Pearce', Academic Press, pp 365-383.
- 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.
- Lions,P.-L., 1996: Mathematical topics in fluid mechanics. OUP, Vol.1, 237pp, Vol 2, 348 pp.
- Lipps, F. B., and R. S. Hemler, 1982: A scale analysis of deep moist convection and some related numerical calculations., *J. Atmos. Sci.*, **39**, pp 2192-2210.
- Lorenz,E.N., 1969: The predictability of a flow which contains many scales of motion., *Tellus*, **21**, pp 289-307.
- McWilliams,J.C., Yavneh, I., Cullen, M.J.P. and Gent, P.R., 1998: The breakdown of large-scale flows in rotating, stratified fluids., *Phys. Fluids*, **10**, 3178-3184.
- Pogorelov,A.V., 1973: Extrinsic geometry of convex surfaces., Israel Prog. for Scientific Translation, Jerusalem, 669pp.
- Richtmyer, R.D. and Morton, K.W., 1967: Difference methods for initial-value problems (2nd ed.), Interscience, John Wiley and Sons, New York, 405pp.
- Schubert,W.H., 1985: Semi-geostrophic theory., *J. Atmos. Sci.*, **42**, pp 1770-1772.
- 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.
- 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., Grubisic, 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., 2001: A two time-level semi-Lagrangian global spectral model., *Q. J. Roy. Meteor. Soc.*, **127**, pp 111-128.
- Weather Log Dec. 1999: in *Weather*, Feb. 2000 issue, Royal Meteorological Society.