An assessment of the value of semi-implict schemes, semi-Lagrangian schemes and various grids for ocean dynamics

Ocean Applications Technical Note No. 25

M. J. Bell

December 2000

Summary

Semi-implict (SI) and semi-implicit semi-Lagrangian (SISL) methods have proved to be accurate and computationally efficient for weather forecasting. On the basis of an examination of the timescales and phase speeds of wave-like disturbances in the ocean, the number of iterations required to solve the elliptic equations, and the timesteps used by present ocean models, it is concluded that SI methods are likely to be 3-5 times more computationally efficient than present methods for global models of 1/3° or better horizontal resolution and limited area models of 10 km or better resolution. For forecasting of mesoscale frontal activity with a resolution of 3 km or better SISL methods are likely to be about 10 times more efficient than the methods presently used. Non-hydrostatic equations can be used with SI and SISL methods without incurring significant additional costs.

The pros and cons of various schemes for staggering the distribution of the model tracer and velocity variables in the horizontal and vertical are reviewed. The C grid is marginally better than the B grid for high resolution models and marginally worse for low resolution models. It is shown that the Lorenz grid has an analogue of potential vorticity conservation. Some of the other problems with the Lorenz grid identified by Arakawa & Moorthi (1988) may also be less severe than they suggest. The velocity errors incurred over steep bathymetry in terrain following co-ordinates are examined. An expression for the errors arising in Lin's (1997) scheme is derived and compared with similar expressions for more well established schemes. An expression for the torques on these co-ordinate cells is also derived. Finally the value of generalised horizontal orthogonal coordinates is reviewed.

Table of Contents

1. Introduction

The questions addressed in this report are:

(a) how much could be gained by using semi-implicit (SI) or semi-implicit, semi-Lagrangian (SISL) techniques in ocean modelling ?

(b) what are the pros and cons of the various methods for staggering the model variables in the horizontal and vertical and what use could be made of terrain following vertical coordinates and generalised orthogonal horizontal coordinates ?

Improvements of about a factor of 10 in the efficiency of high-resolution weather forecast models have been achieved, without significant detriment to their accuracy, by using semi-implicit semi-Lagrangian (SISL) schemes. These schemes enable the timestep used in these models to be determined by considerations of accuracy rather than by CFL type conditions for external or internal gravity waves or advection by the wind. The timesteps of our present global ocean models (at 1° and 1/3° resolutions) are limited by the CFL condition for internal gravity waves in the regions where the effective longitudinal grid-spacing is smallest (at the edge of the region where Fourier filtering is applied). Thus there are good reasons for asking how long a timestep one could take with SI or SISL schemes without significant loss of accuracy, how much more efficient these schemes would be than the present ones, and whether these schemes have other benefits or disadvantages.

Section 2.1 provides a brief overview of SI and SISL schemes. Its aim is to provide reasonably compact and simple explanations of the aspects of the schemes which are needed for the discussion in the following sections. Section 2.2 summarises the propagation speeds of the main waves in the atmosphere and ocean (see Table 1) and discusses the timesteps which could probably be used in various applications without significant loss of accuracy. Section 2.3 discusses the factors which determine the computational cost of the solution of the elliptic equations which arise in semi-implicit techniques. The range of eigenvalues of the elliptic operators, which provides a good indication of the condition number of inversion of the elliptic equations, is shown to depend on the phase speed of the waves and the dimensions of the domain in quite a simple manner. The condition numbers for the barotropic flow in NWP and ocean models are vastly different as are the condition numbers for the baroclinic and barotropic flows in ocean models.

The level of diapycnal transport in the ocean is small enough to provide a significant challenge to advection and integration schemes (in models with height or terrain following vertical coordinates). The "implicit" diffusion inherent in some schemes is larger than estimates of the actual transports. For climate integrations longer than 10 years in duration this is a major issue. The cascade of enstrophy and energy to the grid scale and the resulting "implicit" diapycnal transports are discussed in section 2.4. Finally in models with a horizontal resolution comparable with the depth of the ocean (4 km) non-hydrostatic equations provide a more accurate representation of the motion. Section 2.5 discusses the feasibility of using non-hydrostatic and compressible equations for ocean modelling.

There is more or less general agreement that the C-grid is slightly better than the B-grid when the model's grid resolves the Rossby radius. Section 3.1 examines the evidence for this by briefly reviewing the dispersion relations for gravity, Rossby and Kelvin waves on various grids and discussing the relative importance of stationary computational mode solutions and the accuracy

of the calculations. Section 3.2 discusses the relative merits of 3 schemes for staggering variables in the vertical. Particular attention is given to the representation of the normal modes and of baroclinic instabilities and to the conservation properties of the grids. Section 3.3 discusses the minimum requirements on the resolution of bathymetry needed by ocean models which use terrain-following vertical co-ordinates, such as a σ co-ordinate. An error analysis is presented for Lin's (1997) elegant formulation of the pressure force on the grid cells arising in generalised vertical co-ordinates. The results are compared with similar analyses for the standard formulations of pressure gradients in σ co-ordinates. The torques arising on the grid-cells are also discussed. Section 3.4 briefly discusses how generalised orthogonal horizontal coordinate systems could be exploited particularly in coastal modelling and global modelling.

Conclusions are presented in section 4. The main purpose of this report is to provide a factual basis on which to base decisions about the direction of ocean model development at the Met. Office. Some of the relationships and arguments laid out in this report seem not to have appeared previously in the open literature. Although these are of interest in that they may be novel, they have to be regarded as uncorroborated. As they need to be explained more fully than previously published results they will inevitably have distorted the presentation. For all these reasons section 4.1 summarises these "new" results. Section 4.2 attempts to answer the two questions posed at the start of the report emphasising areas of uncertainty where further work might help to clarify the position.

2. Time Integration schemes

2.1 Overview of semi-implicit and semi-Lagrangian schemes

The reader is referred to Staniforth & Côté (1991, 1999) for excellent introductions to this topic. As is so often the case, the main concepts are illustrated here using the shallow water equations in Cartesian geometry for a fluid of uniform resting depth H:

$$Du/Dt - fv = -g\eta_x \quad , \tag{1}$$
$$Dv/Dt + fu = -g\eta_y \quad , \tag{2}$$
$$D\eta/Dt + (H+\eta)(u_x + v_y) = 0 \quad . \tag{3}$$

Here u and v are the velocities in directions x and y, D/Dt is the Lagrangian derivative, f is the Coriolis parameter, g is gravitational acceleration and η is the free surface height. Subscripts x and y indicate partial derivatives. The fastest waves in the system (1) – (3) have phase speed c with $c^2 = gH$.

(a) Explicit integration

A simple way to step forward (1) from time step n (when u, v and η are assumed known) to time step $n+1$, time Δt later, is to write Du/Dt as $\partial u/\partial t + \mathbf{u}.\nabla u$, to evaluate fv, $-g\eta_x$ and $\mathbf{u}.\nabla u$ at timestep n and approximate $\partial u/\partial t$ by a finite difference approximation centred at time-step n, involving u at steps $n+1$ and $n-1$. Using a superscript to denote the timestep at which the variable is valid (e.g. u^n is valid at step n), (1) becomes

$$\frac{(u^{n+1} - u^{n-1})}{2\Delta t} = -(\mathbf{u}^n.\nabla)u^n + fv^n - g\eta_x^n \quad . \tag{4}$$

A similar approach can be used in (2) and (3). With this approach the time-step must be chosen so that external gravity waves do not cross a grid square within a leapfrog timestep (i.e. $c\,2\Delta t < \Delta x$).

(b) Semi-implicit integration

This restriction on the timestep is removed (as shown in (d) below) if $-g\eta_x$ in (1), $-g\eta_y$ in (2) and $(u_x + v_y)$ in (3) are calculated semi-implicitly, that is as averages of the values at timesteps $n+1$ and $n-1$. All other terms (including Du/Dt) are calculated as in (a) above. Re-organising (1) – (3) so that only the unknown values, namely u^{n+1}, v^{n+1} and η^{n+1} appear on the left hand-side of the equations one obtains:

$$u^{n+1} + g\Delta t\eta_x^{n+1} = C_1^n \quad , \tag{5}$$
$$v^{n+1} + g\Delta t\eta_y^{n+1} = C_2^n \quad , \tag{6}$$
$$\eta^{n+1} + H\Delta t(u_x^{n+1} + v_y^{n+1}) = C_3^n \quad , \tag{7}$$

where the C_i^n (for $i=1,2,3$) are determined by values of u, v and η on timesteps n and $n-1$. The horizontal divergence on step $n+1$ in (7) can be expressed in terms of η^{n+1} by taking the divergence of the horizontal momentum equations (5) and (6). This yields an elliptic Helmholtz-type equation for η^{n+1}:

$$\eta^{n+1} - gH\Delta t^2(\eta_{xx}^{n+1} + \eta_{yy}^{n+1}) = C_4^n \quad , \tag{8}$$

where C_4^n depends on C_1^n, C_2^n and C_3^n. This equation, together with its boundary conditions, is solved for η^{n+1} which is then substituted into (5) and (6) to enable u^{n+1} and v^{n+1} to be found.

Note that if only $-g\eta_x$ in (1) and $-g\eta_y$ in (2) are calculated semi-implicitly (i.e. (u_x+v_y) in (3) is calculated explicitly) a CFL condition restricting the timestep in terms of the external gravity wave speed remains, though it becomes $c\Delta t < \Delta x$, which is less stringent than the original by a factor of 2 (Brown & Campana 1978).

A similar semi-implicit scheme can be applied to the 3D hydrostatic primitive equations for u, v, θ and S (salinity). To explain which terms need to be calculated semi-implicitly in the equation for conservation of potential temperature it is useful to write it in the form

$$D\theta / Dt = \theta_t + u_h.\nabla\theta + w\theta_{0z} + w(\theta - \theta_0)_z = 0 \quad , \tag{9}$$

where a reference profile $\theta_0(z)$ which is frequently used to facilitate the computations has been introduced. This reference profile only depends on z (i.e. not on x, y or t). To obtain a scheme in which the timestep is not restricted by the internal gravity wave speed the term $w\theta_{0z}$ must be calculated semi-implicitly. The full set of equations reduces to a 3D elliptic problem for a single variable (see Ritchie et al. 1985 for the details in a specific case). Use of the reference profile makes the problem separable for simple domains – such as that of the atmosphere in terrain following coordinates. (Section 6.11 of Gill (1982) gives a clear explanation of the relationship between the shallow water equations and internal modes in hydrostatic fluids.) This device of splitting a term into a linear part which is calculated semi-implicitly (i.e. at timesteps $n-1$ and $n+1$) and the remainder which is calculated explicitly (i.e. at timestep n) is widely used and may be useful for accurate calculation of the pressure gradients (see section 3.3(e) below).

(c) Semi-Lagrangian integration

In nearly all semi-Lagrangian schemes material derivatives, such as Du/Dt and $D\theta/Dt$, for a given model gridpoint with location $\mathbf{x}$, are calculated by estimating the location at timesteps t_n and/or t_{n-1} of a particle whose trajectory passes through (or "arrives" at) $\mathbf{x}$ at timestep t_{n+1}. Consider first the case of a trajectory which is centred at time t_n and extends between times t_{n-1} and t_{n+1}. Figure 1, taken from Staniforth & Côté (1999), is a very helpful illustration of this calculation for the 1D case. The displacement of the particle from $\mathbf{x}$ at time t_n is defined to be $-\boldsymbol{\alpha}$; i.e. the particle is defined to be at point $\mathbf{x}-\boldsymbol{\alpha}$ at time t_n. $\boldsymbol{\alpha}$ is determined iteratively as the solution of the equation

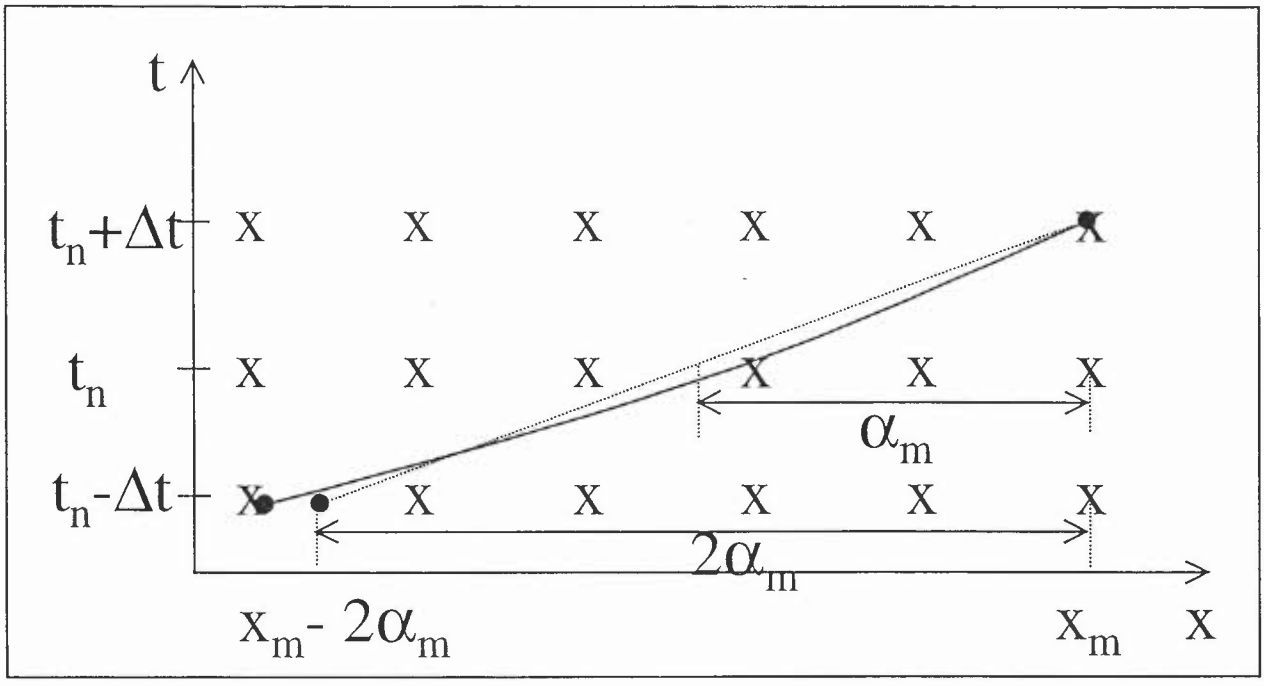

Figure 1: Schematic for semi-Lagrangian advection in 1D, illustrating the actual (solid curve) and approximate (dashed line) trajectories arriving at grid point x_m at time $t_n + \Delta t$ and the definition of α_m.

$$\boldsymbol{\alpha}_{m+1} = \Delta t \mathbf{u}^{n}(\mathbf{x} - \boldsymbol{\alpha}_m) \tag{10}$$

where $\boldsymbol{\alpha}_m$ is the estimate of $\boldsymbol{\alpha}$ on the mth iteration and evaluation of $\mathbf{u}(\mathbf{x}-\boldsymbol{\alpha})$ requires interpolation of the fields known only at the gridpoints $\mathbf{x}$ to the estimated midpoints of the trajectory at $\mathbf{x}-\boldsymbol{\alpha}$. (10) provides a value of $\boldsymbol{\alpha}$ which is 2nd order accurate (in Δt) for a trajectory whose start point at time t_{n-1} is given by $\mathbf{x}-2\boldsymbol{\alpha}$. Pudykiewicz and Staniforth (1984) show that a sufficient condition for (10) to converge to a unique solution for 2D flow is that

$$\Delta t < [\max(|u_x|, |u_y|, |v_x|, |v_y|)]^{-1} \quad . \tag{11}$$

This "deformation" Courant number restriction on the timestep is much less severe than $c\Delta t < \Delta x$ for well resolved flows. Using this trajectory one sees that $\frac{1}{2}[\mathbf{u}^{n+1}(\mathbf{x}) - \mathbf{u}^{n-1}(\mathbf{x}-2\boldsymbol{\alpha})]/\Delta t$ is a 2nd order accurate estimate of $D\mathbf{u}/Dt$ which is valid at point $\mathbf{x}-\boldsymbol{\alpha}$ on time step t_n. To use this estimate in (1) the other terms, namely fv and $g\eta_x$ need to be calculated about the same point. If the terms are calculated purely explicitly using their values at time t_n, the values must be interpolated to the points $\mathbf{x}-\boldsymbol{\alpha}$. For terms calculated semi-implicitly it is usually convenient and more accurate to use the trajectory end-points. For example $\frac{1}{2}[g\eta_x^{n+1}(\mathbf{x}) + g\eta_x^{n-1}(\mathbf{x}-2\boldsymbol{\alpha})]$ is a 2nd order accurate estimate of $g\eta_x$ at the required point. Clearly the trajectory has been chosen so that the terms on step $n+1$ are only used on the original grid. This means that after the re-organisation of terms as in (5)-(7) the form of the equations is unaltered and semi-implicit

solutions for the new timestep can be found as before. Thus with this choice, schemes which are semi-implicit and semi-Lagrangian (SISL) arise naturally.

The quality of advection by semi-Lagrangian integration schemes is largely determined by the accuracy of the interpolation of the fields. The errors in phase speeds and inherent damping in the standard schemes are summarised by McCalpin (1988) following the calculations of McDonald (1984) and discussed in section 2.4. A very wide range of advection schemes is potentially available (Smolarkiewicz & Rasch 1991, Toro 1997).

Up to this point only a 3 timestep framework has been discussed in which the calculations for the fields at t_{n+1} are centred at t_n. A 2 timestep framework in which the calculations are centred at time $\frac{1}{2}(t_n + t_{n+1})$ is now routinely used by many NWP centres. In this 2 timestep scheme all terms in (1)-(3) are calculated using values at both t_{n+1} and t_n. For these terms the "effective" timestep for the 2 timestep scheme is half that in the 3 timestep scheme or, more precisely, for the same time truncation errors twice as long a timestep can be used with the 2 timestep scheme as with the 3 timestep scheme. To calculate the trajectory, an estimate of the velocity at time $\frac{1}{2}(t_n + t_{n+1})$ is calculated by linear extrapolation of the velocities at steps t_n and t_{n-1}. Other terms calculated explicitly at step t_n in the 3 step scheme can be similarly extrapolated. This will usually give a less accurate approximation to the time integral of the term along the trajectory than obtained in the 3 step scheme. For example for a term $R(t)$ (which does not depend on x) the error in the integral for the 3 step scheme with timestep $\Delta t / 2$ is

$$E_3 \equiv \int_{\tau=-\Delta t/2}^{\tau=\Delta t/2} \{R(t_0+\tau) - R(t_0)\} d\tau = R_{tt}(t_0)\Delta t^3 / 24 \quad , \tag{12}$$

whilst the error in the 2 step scheme with timestep Δt is

$$E_2 \equiv \int_{\tau=-\Delta t/2}^{\tau=\Delta t/2} \{R(t_0+\tau) - \frac{3}{2}R(t_0 - \Delta t/2) + \frac{1}{2}R(t_0 - 3\Delta t/2)\} d\tau = \frac{11}{12}R_{tt}(t_0)\Delta t^3 \tag{13}$$

More importantly these extrapolated terms need careful treatment because they can make the integration scheme unstable (Andrew Staniforth, personal communication). They can sometimes be avoided by careful choice of model variables; for example explicit calculation of the terms $\eta(u_x + v_y)$ in (3) can be avoided by re-writing (3) as:

$$(D/Dt)\{\ln(H+\eta)\} + u_x + v_y = 0 . \tag{14}$$

(d) Stability and phase speed of plane waves

SISL schemes in which all terms are calculated semi-implicitly (as in the 2 time-level scheme) have dispersion relations for plane wave solutions which can be inferred quite simply from the dispersion relations for the continuous equations. The stability of SISL schemes and their effect on the phase speed of waves follow directly from these dispersion relations. As an example consider the solution of (1) – (3), consisting of a plane wave superimposed on a zonal flow U which is independent of x, y and t. The plane wave is of the form

$$(u, v, \eta) = \mathrm{Re}\{(\hat{u}, \hat{v}, \hat{\eta}) \exp \imath(kx + ly - \omega t)\} \tag{15}$$

in which the variables with caret superscripts are constant amplitudes (independent of x, y and t) and $\imath^2 = -1$. The total time derivative Du/Dt for the plane wave when the 3 time-level scheme is used is:

$$\frac{[u^{n+1}(x, y) - u^{n-1}(x - 2U\Delta t, y)]}{2\Delta t} = u^n(x - U\Delta t, y)\frac{-\imath \sin[(\omega - kU)\Delta t]}{\Delta t} \quad . \tag{16}$$

Similarly the semi-implicit evaluation of the Coriolis term fv is:

$$\tfrac{1}{2}[fv^{n+1}(x, y) + fv^{n-1}(x - 2U\Delta t, y)] = fv^n(x - U\Delta t, y)\cos[(\omega - kU)\Delta t] \quad . \tag{17}$$

The ratio $Du/Dt/(fv)$ for the continuous equations is $-\imath(\omega - kU)u/(fv)$ whilst for the SISL scheme it is $-\imath \tan[(\omega - kU)\Delta t]\, u/(fv\Delta t)$. The same ratio of terms holds wherever $(\omega - kU)$ appears for the continuous case in equations (1) – (3). Thus the dispersion relation for the SISL scheme can be obtained from that for the continuous case by replacing occurrences of $(\omega - kU)$ by $\tan[(\omega - kU)\Delta t]/(\Delta t)$. Similarly the dispersion relation for the explicit scheme outlined in (a) can be obtained for the case $U = 0$ by replacing occurrences of ω by $\sin[\omega\Delta t]/(\Delta t)$.

The dispersion relation for the continuous case on an f-plane is

$$(\omega - kU)^2 = f^2 + (kc)^2 \quad . \tag{18}$$

Thus the explicit scheme is unstable, for the case with $f = U = 0$, when $kc\Delta t > 1$ as then $\sin(\omega\Delta t) > 1$. The SISL scheme is stable for all values of f, k, c and U because the inverse of the tangent function provides real values for $(\omega - kU)\Delta t$ between $-\pi/2$ and $\pi/2$ whatever the value of $(f^2 + (kc)^2)\Delta t^2$. Waves of short wavelength are slowed down sufficiently by the SISL scheme to avoid breaking the CFL condition. Notice that the advection of the waves by the mean flow U is independent of wavelength. Thus when timesteps longer than the CFL condition are taken, the short waves can propagate in the model more slowly than they are advected by the mean flow when the opposite is true in reality. This can lead to false topographic resonances and the generation of noise. Rivest et al. (1994) and Ritchie & Tanguay (1996) offer techniques for avoiding excessive noise.

(e) Accuracy

Robert (1981) provides a simple comparison of the magnitudes of time and spatial truncation errors for the advection equation

$$\partial F/\partial t + c\partial F/\partial x = 0 \tag{19}$$

applied to a wave of form

$$F = A\cos k(x - ct) \quad . \tag{20}$$

Assuming that the primary errors are in the phase of the wave and using centred 2nd order differencing in both time and space he finds that the root-mean-square error, E_{rms}, intially grows linearly with time t and is given by

$$E_{rms} = \frac{\sqrt{2}}{12} k^3 cA(\Delta x^2 - c\Delta t^2)t \quad . \tag{21}$$

Writing

$$c_0 = \Delta x / \Delta t \,, \tag{22}$$

the ratio of the time truncation error and the spatial truncation error is

$$\frac{E_{rms}(\Delta t)}{E_{rms}(\Delta x)} = \frac{c^2}{c_0^2} \,. \tag{23}$$

It is computationally inefficient to use a timestep with which the ratio in (23) is very small. A ratio approaching 1 would give the most computationally efficient scheme. Ratios of 1/400 were prevalent in atmospheric models before the introduction of semi-implicit methods and are typical of present high resolution mid-latitude FOAM integrations (see section 2.2).

Clearly the above calculation is only a guide to the relative importance of spatial and time truncation errors for the quite complex waves appearing in atmosphere and ocean models. It is worth making two points relevant to its interpretation and use. Firstly (23) may not immediately convey the potential improvement of SISL schemes. Explicit schemes are limited by the worst case gridcells – where the grid spacing is shortest and/or the phase speeds largest. The waves in the regions which limit the timestep globally are generally of very little geophysical consequence. With SISL schemes there is the potential to choose the timestep suitable for the motions of interest in the region of greatest interest. Secondly a similar calculation to Robert's can be made for the propagation of Kelvin waves using (1) and (3) with $v = 0$. (23) gives the error ratio for a two timestep semi-implicit scheme whilst for a 3 timestep scheme $E(\Delta t)$ is 4 times larger.

2.2 Ocean and atmosphere timescales

Table 1 provides a summary of the phase speeds of the most important fast wave motions relevant to atmosphere and ocean models used for climate and "weather" prediction. These values are key to the discussion of this section. Most of them are easily found in Gill (1982; G hereafter in this sub-section).

The typical values for sound speeds differ greatly (see G appendix table A3.1 for water and pp 172-3 for air); the speed of sound in the atmosphere is similar to that of the external gravity waves whereas in the ocean it is more than 7 times faster. The external gravity wave speed can be estimated using $c^2 = gH$ and 10 km for the equivalent depth of the atmosphere and 4 km for the

deep ocean. On the continental shelf, the water is no more than 200 m deep and the external gravity wave speed is little more than 40 m/s.

Table 1: Phase speeds of various types of motions

Type of motion	Atmosphere (m/s)	Ocean (m/s)
Sound waves	300	1500
External gravity waves	300	200; 40 on shelf
First internal gravity wave	40	3
Max. wind/current speed	100	1□5

The phase speed of the first internal gravity wave in the ocean is a key value which is discussed on numerous occasions in G (e.g. on pages 122, 161, 437). Table 2 of Leslie & Purser (1992) gives equivalent depths for the internal modes in the atmosphere. The wind or current speeds in Table 1 are intended to be maximum values. That for the atmosphere is based on the winter-time jet level wind and is a value quoted, for example, by Robert (1981). The Gulf Stream often exceeds 1 m/s so a maximum current speed of 1□5 m/s is perhaps a useful value for stability calculations. Tidal currents on the shelf also frequently exceed 1 m/s.

Note that the maximum wind speeds in the atmosphere are about 2.5 times faster than the internal gravity waves and 3 times slower than the external gravity waves. In the ocean it is the internal gravity waves which have the intermediate velocities. They are about 60 times slower than the external gravity waves and twice as fast as the fastest currents.

For NWP the speed of the weather systems is typically 15-20 m/s. The ratio of the spatial and temporal truncation errors for systems travelling at this speed is approximately given by using $c = 20$ m/s in (23). Integration schemes in which the external gravity waves and advection are calculated explicitly (and sound waves are excluded) must limit Δt so that $c_0 = 300 + 100$ m/s. Thus the time truncation error is typically 400 times (or more) smaller than the spatial truncation error. If the external gravity waves are calculated implicitly, Δt can be increased until $c_0 = 100$ m/s. The ratio in (23) is still small (typically 1/25). Calculating the gravity waves and advection semi-implicitly, as in a SISL scheme one can reduce c_0 further to a value of about 40 m/s without significantly degrading the quality of the simulation.

In practice, of course, other processes (e.g. vertical diffusion, diurnal variations or radiative transfers) can limit the time step in an NWP model. In models with coarse resolution (e.g. 3.5°x 4.75°) , such as many coupled climate models, these other processes have limited the timestep. On a grid with (uniform) 200 km resolution SISL enables a timestep of about 90-120 minutes. SISL comes into its own for NWP in models of high resolution (10-50 km resolution).

In existing ocean models, sound waves are excluded and the barotropic flow is handled separately. This is because (see table 1) the external gravity waves move more than 60 times as fast as the internal gravity waves and other waves. Important Kelvin waves do travel along the equator (and the shelf break) at the speed of the first internal mode (3 m/s). Thus in seasonal forecasting and shelf–break modelling, and probably coupled climate models, the above arguments suggest that a value of c smaller than 3 m/s may not be particularly desirable. However an ocean model's resolution is not usually improved in order to reduce the spatial truncation error along the equator. A wave with a quarter wavelength only 4 times the grid

spacing has only a 1% phase speed error due to spatial truncation errors and Ng & Hsieh (1994) neglect this source of error completely when discussing the numerical discretisation of Kelvin waves. Other aspects of the simulation will usually limit its accuracy. So even if Kelvin waves propagating along the equator are of primary interest it may not be necessary to limit the ratio in (23) to be strictly less than 1.

For a model with 100 km resolution, demanding a ratio of ¼ in (23) gives a timestep of 4 hours. There are 2 reasons why this value is larger than the 1 hour currently employed in both the TOGA and HadCM3 models which both have 100 km east-west grid spacing at the equator and whose timesteps are limited by the internal gravity wave CFL condition. Firstly these models use a leapfrog timestep in which the timestep is effectively double that in a two-step forward method. Secondly the models do not have uniform grid resolution. In the TOGA model the north-south grid resolution is deliberately chosen to be enhanced by a factor of 3 at the equator. In the global HadCM3 model the east-west grid-spacing near the edge of the Fourier chopping region (at 75°N) is 4 times shorter than that at the equator. As discussed in section 3.3, global orthogonal coordinates can be constructed with a similar ratio of maximum to minimum grid spacing at sea points but a smaller ratio is somewhat difficult to achieve.

It is not clear that these ocean models would be stable with a 4 hour timestep. A diffusive instability limited the timestep in the 3.5° by 4.75° ocean model. Appendix A suggests however that this instability, which was first analysed by Killworth (1987), is actually specific to a particular choice of discretisation of the barotropic streamfunction so should not be as restrictive as it once was. Though FOAM attempts to resolve the diurnal cycle, neither TOGA nor HadCM3 do. Of course, a timestep of 4 hours would be completely inappropriate if the diurnal cycle was important to the simulation.

For a 1/3° global climate model a semi-implicit scheme could be expected to allow an increase in the timestep of about a factor of 4 without compromising the accuracy of the calculation. If a 2 timestep approach could be used without compromising the quality of the simulation (see section 2.4) the timestep could probably be doubled again.

In coastal ocean models the barotropic tides dominate the circulation. In the deep ocean these can travel at 200 m/s whilst in shallow water they tend to travel at 30-40 m/s. For the shallow water the phase of the tides is of crucial importance so one cannot use timesteps which would retard the barotropic wave speed. In models with relatively little deep water it might be possible to use a SI method with much longer timesteps than CFL would give in the deep water without significantly retarding the long wavelength deep water tidal fronts.

For deep ocean forecasting at mid-latitudes the phase speed of mesoscale meanders and eddies is significantly less than the speed of the currents. The present 11km FOAM model is able to use a 20-30 minute timestep. Requiring a ratio of ¼ in (23) for a phase speed c of 30 cm/s gives a timestep of about 4 hours. It thus appears possible that for very high resolution deep water FOAM (e.g. 4 km resolution) SISL could enable timesteps about 8-10 times larger than at present.

2.3 Efficiency of solution of elliptic equations

As explained in the previous sections, semi-implicit schemes are able to take longer time steps than explicit schemes. On each time-step the main task of a semi-implicit scheme is to solve an elliptic problem for a single variable. An explicit scheme traces the full time history of the evolution of the variable over this timestep whereas a semi-implicit scheme only needs to find the

final values of the variable. So it should be possible to find cheaper solutions to the ellliptic problem than a timestepping solution (Morton & Mayers 1994, p 218). This section discusses the elliptic problems which arise in semi-implicit schemes and the rate of convergence of iterative methods for their solution.

Table 8.1 of Gill (1982) provides a summary of the equations and dispersion relations for linear wave motions obtained using combinations of approximations (e.g. hydrostatic, or rapidly rotating) appropriate for various scales of motion in the atmosphere and ocean. The elliptic equations which need to be solved for a given SI or SISL scheme can be "read off" from this table. The process described in section 2.1 (b) of moving the terms calculated at the new timestep to the l.h.s. of the equations (see (5)-(7)) results in a set of equations identical in form to the linearised equations with partial derivatives for any variable (e.g. $\partial u/\partial t$) replaced by the variable divided by Δt (e.g. $u/\Delta t$). Only terms calculated semi-implicitly are included in the elliptic operator. Thus, for example, the equations for a rotating fluid in table 8.1 are relevant to the elliptic operator only if the Coriolis terms are calculated semi-implicitly.

Thus elliptic equations for the vertical velocity are easily found. For the hydrostatic case with the Coriolis terms calculated semi-implicitly

$$(\frac{1}{\Delta t^2}+f^2)w_{zz}+N^2(w_{xx}+w_{yy})=0 \quad . \tag{24}$$

For a flat bottom ocean (or if terrain following coordinates are used) good approximations to the boundary conditions for the internal modes are $w=0$ at the top and bottom of the ocean.

In NWP the standard approach to solving the 3D elliptic problem is to use a vertical mode decomposition to obtain a set of 2D Helmholtz problems. In a spectral model the 2D problems are solved directly (e.g. Ritchie et al. 1995). For semi-implicit schemes with only moderately long timesteps it is probably only necessary in a spectral model to calculate the first few vertical modes implicitly (Burridge 1975, Purser 1999b). More vertical modes need to be calculated implicitly when SISL methods and longer timesteps are used. For the ocean geometry it is not at all clear how to perform a similar decomposition without making undesirable approximations such as confining the bathymetry to the lowest model layer (as in Hurlburt and Thompson 1980) so a full 3D iterative solution for the internal modes appears to be necessary.

The rate of convergence of solutions to (8) and (24) for iterative methods of solution, such as successive overlaxation (SOR) and pre-conditioned conjugate gradient, depends largely on the range of the eigenvalues of the elliptic operators. For SOR the spectral radius of the eigenvalues determines the rate of convergence. For symmetric positive definite matrices the rate of convergence of the conjugate gradient solver can be bounded by the ratio $\kappa=|\lambda|_{\max}/|\lambda|_{\min}$ of the largest to the smallest eigenvalues (Golub & van Loan 1986 chapter 10 and p 530). The rate of convergence of the preconditioned conjugate gradient solver depends on the same ratio of the eigenvalues of the product of the matrices representing the preconditioner and elliptic operator.

Useful approximations to the eigenvalues are easily found by modelling the problems to be solved as rectangular boxes in which case simple Fourier series expansions of the variables are possible and the eigenfunctions are simple sinusoidal functions. For simplicity let the grid spacing be uniform and $\Delta x=\Delta y$, N^2 be independent of depth, I be the number of model columns, J

the number of rows, Q the number of levels and i, j, q be the column, row and level indices. (Q and q are used here to avoid conflicts with the use of k as a wavenumber, N as Brunt-Vaisala frequency and p as pressure.) To illustrate the various possibilities, the domain is taken to be periodic in x and to have sidewalls to the south at $y = 0$ and to the north.

For the external gravity waves the boundary conditions can be reduced to Neumann conditions (on the normal derivative of $\eta; \eta_n = 0$). Assuming that the usual 5 point formula is used for the horizontal Laplacian, the eigenfunctions of (8) are then

$$\eta(i, j; k, l) = \begin{matrix}\cos \\ \sin\end{matrix} k(2i\pi / I)\cos l(j\pi / J) \tag{25}$$

with eigenvalues

$$\lambda(k,l) = -2(1-\cos 2k\pi / I + 1 - \cos l\pi / J) - \frac{1}{gH}(\frac{\Delta x}{\Delta t})^2 \quad . \tag{26}$$

The eigenvalues with the largest magnitudes are obtained when k and l take their smallest values ($k = 0$ and $l = 1$ or $k = 1$ and $l = 0$) whilst those with the smallest are obtained for $k = I/2 - 1$, $l = J - 1$:

$$|\lambda|_{\max} = 4 + \frac{1}{gH}\frac{\Delta x}{\Delta t^2}^2 \quad , \tag{27}$$

$$|\lambda|_{\min} = 2[(\pi / I)^2 + (\pi / J)^2] + \frac{1}{gH}\frac{\Delta x}{\Delta t^2}^2 \quad . \tag{28}$$

The Poisson equation $\nabla^2 \eta = 0$ is ill-conditioned for large domains because the ratio of maximum to minimum eigenvalues is $[(\pi / I)^2 + (\pi / J)^2]^{-1}$. There is a limit on this ratio for the implicit free surface. This is given by the square of the number of grid points travelled at the external gravity wave speed in one timestep. This point was clearly made by Dukowicz & Smith (1994) who showed that the rate of convergence for a semi-implicit free surface solver is much better than that for a completely non-divergent barotropic streamfunction. As noted in section 2.2 the timestep in NWP models which use SISL methods is chosen so that $c_0 = \Delta x / \Delta t = 40\ ms^{-1}$ and the ratio of eigenvalues is typically about 100. In present ocean models the timestep is limited by the internal gravity wave speed ($c_0 = 3\ ms^{-1}$) and the ratio of eigenvalues is $(200/3)^2 > 4000$. In global ocean models with a regular latitude-longitude grid which have a polar island and use Fourier chopping an external gravity wave can travel a very large number of grid points round the pole in a single timestep. The rate of convergence of the free surface in such models does appear to be very slow near the pole (Malcolm Roberts personal communication). Guyon et al. (1999) appear to obtain significant improvements in the rate of convergence of the barotropic solver (on a relatively uniform grid with 2D domain decomposition) using a dual Schurr complement method.

The eigenfunctions for the internal modes are

$$w(i,j,q;k,l,m) = \begin{matrix}\cos\\ \sin\end{matrix} k(2i\pi/I)\sin l(j\pi/J)\sin m(q\pi/Q) \ , \tag{29}$$

and the eigenvalues of (24) when $f = 0$ are

$$\lambda(k,l,m) = -2(1-\cos 2k\pi/I + 1-\cos l\pi/J) - \frac{\Delta x^2}{\Delta t^2}\frac{2}{(N\Delta z)^2}(1-\cos m\pi/Q) \ . \tag{30}$$

The ranges of k, l and m determine the maximum and minimum eigenvalues. $m = 0$ gives a zero eigenfunction so $1 \le m \le Q-1$. A non-zero vertical velocity completely independent of i and j is not a solution so at least one of k and l is non-zero whilst $k \le I/2-1$ and $l \le J-1$. The maximum and minimum magnitudes of the eigenvalues are approximately:

$$|\lambda|_{\max} = 4 + \frac{\Delta x^2}{\Delta t^2}\frac{2Q^2}{(NH)^2} \ , \tag{31}$$

$$|\lambda|_{\min} = 2[(\pi/I)^2 + (\pi/J)^2] + \frac{\Delta x^2}{\Delta t^2}\frac{2\pi^2}{(NH)^2} \ . \tag{32}$$

For a timestep which is only a few times larger than that for the CFL condition for internal gravity waves the second factor will dominate in both eigenvalues and the ratio of the maximum to the minimum eigenvalues will be approximately $(Q/\pi)^2$.

Marshall et al. (1997) suggest that this internal mode problem be preconditioned by the inverse of the vertical component of the 3D Laplacian. The eigenvalues of the product of the preconditioner and the elliptic operator are:

$$\lambda(k,l,m) = 1 + \frac{(1-\cos 2k\pi/I + 1-\cos l\pi/J)}{\frac{\Delta x^2}{\Delta t^2}\frac{1}{(N\Delta z)^2}(1-\cos m\pi/Q)} \ . \tag{33}$$

The maximum and minimum magnitudes of the eigenvalues are then:

$$|\lambda|_{\max} = 1 + \frac{\Delta t^2}{\Delta x^2}\frac{(NH)^2}{\pi^2} \ , \tag{34}$$

$$|\lambda|_{\min} = 1 \ . \tag{35}$$

So the ratio of maximum to minimum eigenvalues is $(c/c_0)^2$ where c_0 is defined by (22). When $c/c_0 = 4$ (which is a likely value for climate or seasonal modelling) the resulting condition number is 16 which is rather modest. Good solutions should be obtained within 5 iterations. Of

course other pre-conditioners, including ones using multi-grid methods (Smith et al. 1996), are possible but this one is cheap to apply, easy to code and appears to be adequate. This conclusion needs to be checked using the strongly variable vertical grid spacing typical of ocean models.

2.4 Grid scale closure

As outlined in the introduction, this section discusses the diapycnal transports implicit in upwind advection schemes and the grid scale closure of the cascade of enstrophy and energy to small scales. Subsection (a) provides some motivation for the discussion and (b) derives some relationships between the diapycnal transport of density and the dissipation of potential energy by biharmonic diffusion. Subsection (c) discusses the results of Roberts & Marshall (1998) in the light of these results. Subsection (d) summarises well-known formulae for the dissipation inherent in semi-Lagrangian and upwind advection schemes (such as QUICKEST) and (e) uses results from (b) and (d) to estimate the ratio of the diapycnal transport by horizontal and vertical advection as performed by these schemes, focussing particularly on models of 1° grid spacing. Finally subsection (f) draws attention to the advantages for grid-scale closure of the variational formulations of isopycnal diffusion and the Gent-McWilliams scheme suggested by Griffies et al. (1998) and Griffies (1998).

(a) Diapycnal transport implied by horizontal biharmonic diffusion

Roberts & Marshall (1998) (hereafter RM) show that horizontal diffusion of tracers can produce a larger diapycnal flux of tracer down through the thermocline than the upward diapycnal transport associated with the upwelling "branch" of the thermohaline circulation which involves vertical velocities as small as 4 m yr^{-1}. They diagnose this from integrations using models of varying resolution in a simple box shaped mid-latitude domain which are driven by wind stresses at the surface with no heat flux across the surface. After 10 years of integration the model's density has increased considerably in the top 200 metres and the density has decreased lower down corresponding to a lowering of the thermocline by 50 metres (i.e. 5 m yr^{-1}). In particular they find that this is true not only for ocean models with a horizontal grid spacing of 100 km but also for models with grid spacings down to 10 km. They suggest that a scale selective form of the Gent & McWilliams (1990) scheme is needed to prevent diapycnal fluxes of tracer "even" in eddy-resolving ocean models.

The level of the diapycnal flux of tracers in nature is not clear. Garrett et al. (1993) (p 292) note that the effective diapycnal diffusivity κ inferred from the inflow into isolated abyssal basins is several times 10^{-4} m^2s^{-1} whilst more direct measurements in the thermocline by Gregg et al. (1996) suggest that the value is typically 10^{-5} m^2s^{-1} though it reaches 10^{-4} m^2s^{-1} in the Florida Straits and has been inferrred to be 150 10^{-4} m^2s^{-1} below 4000 m in the Romanche fracture zone (Polzin et al. 1996). There are two main sources of energy which generate mixing at small scales: firstly instabilities of the gyre-scale circulation which generate an enstrophy cascade to small scales; secondly internal gravity waves. The complex (and confusing) interactions between these motions which can lead to mixing at small scales (e.g. by wave breaking) are aptly captured by figure 1 of Thorpe (1988).

In addition to the physically realistic fluxes of tracer variance to the grid scale, numerical models contain artificial fluxes arising from the numerics of the advection scheme. Advection by a constant velocity (i.e. independent of x and y) should leave the shape of the advected tracer unchanged but in practice the phase speed of propagation depends on the wavenumber and

tracer variance is transported towards the grid scale. As Thuburn (1995 section 3a) points out particularly clearly an advection scheme must dissipate grid scale variations or give spurious results at the grid scale.

As background to the following arguments, it is perhaps worth summarising briefly the main elements of the enstrophy cascade. Quasi-geostrophic flows with no surface temperature gradients conserve both energy and enstrophy. Most of the energy is transferred to larger scales whilst enstrophy is mainly fluxed to small scales. For a statistically steady state in an inertial range of wavenumbers where dissipation is negligible, the flux of enstrophy per unit mass from one lengthscale (wavenumber) to another, ε, must be independent of the wavenumber (for enstrophy would otherwise build up or decay at particular scales). The total energy per unit mass, E, and the energy spectrum per unit mass, $E(\kappa)$, are defined by:

$$E = \frac{\int \rho(u^2 + v^2) dV}{\int \rho dV} = \int E(\kappa) d\kappa \,. \tag{36}$$

If $E(\kappa)$ is solely controlled by the enstrophy cascade, it must depend solely on ε and κ. The dimensions of $E(\kappa)$ are $m^3 s^{-2}$ whilst those of ε are s^{-3}. So

$$E(\kappa) = \gamma \varepsilon^{2/3} \kappa^{-3} \quad . \tag{37}$$

where γ is a non-dimensional constant. This distribution of energy holds approximately in the quasi-geostrophic regime at the long wavelength ends of the atmosphere synoptic scale (Charney 1971) and the ocean mesoscale. In this regime the available potential energy and kinetic energy are approximately equipartitioned. At smaller scales, quasi-geostrophy does not constrain the stretching of vortex tubes so effectively, energy cascades to small scales and the energy spectrum does not fall so steeply with the wavenumber.

The main point to be taken from the previous paragraph is that vorticity and to a lesser extent kinetic and potential energy are inevitably fluxed to the smallest scales in numerical atmosphere and ocean models. The grid scale closure of this flux cannot be circumvented. Though there are various approaches to the problem (e.g. Smagorinsky 1993 and Mason 1994) one can expect some aspects of the dissipation to be controlled by the enstrophy cascade and to be rather insensitive to the details of the numerical schemes used.

(b) Diapycnal transport of density and dissipation of APE by horizontal diffusion

Building on the arguments of RM it is possible to make an estimate of the diapycnal diffusion arising from horizontal diffusion of tracers. The horizontal fluxes of potential temperature and salinity due to biharmonic horizontal diffusion are:

$$F_h \begin{pmatrix} \theta \\ S \end{pmatrix} = -\nabla_h A_\theta \nabla_h^2 \begin{pmatrix} \theta \\ S \end{pmatrix} \quad , \tag{38}$$

where A_θ is the coefficient for biharmonic horizontal diffusion of the tracer. Assuming that the density is linearly related to the potential temperature and salinity,

$$\rho = \alpha(\theta - \theta_0) + \beta(S - S_0)\,, \tag{39}$$

and ignoring variations in α and β one obtains:

$$F_h(\rho) = -\nabla_h(A_\theta \nabla_h^2 \rho) \quad . \tag{40}$$

The slope of the isopycnal surface is $-\nabla_h \rho / \rho_z$ so the component of the horizontal flux of density across the isopycnal surface, F_d, is

$$F_d(\rho) = \frac{\nabla_h \rho . \nabla_h (A_\theta \nabla_h^2 \rho)}{\rho_z} \quad . \tag{41}$$

Neglecting horizontal variations in ρ_z (as in q-g theory) and integrating by parts with insulating boundary conditions ($\hat{\mathbf{n}}.\nabla\rho = 0$) one obtains

$$\int F_d(\rho)\, dV = \int \frac{-A_\theta (\nabla_h^2 \rho)^2}{\rho_z} dV \quad . \tag{42}$$

This flux (times g) is equal to a good approximation to the rate of decrease of available potential energy (APE) by the horizontal diffusion. From (6.10.11) of Pedlosky (1982),

$$APE = \int \frac{g}{2} \frac{\rho^2}{\rho_z} dV \quad . \tag{43}$$

The rate of change of APE due to horizontal diffusion of density, A_d, is then approximately

$$A_d = \int -\frac{g}{\rho_z} \rho \nabla_h^2 (A_\theta \nabla_h^2 \rho)\, dV = \int -\frac{g}{\rho_z} A_\theta (\nabla_h^2 \rho)^2\, dV \quad . \tag{44}$$

On the other hand, biharmonic horizontal diffusion of momentum with a viscosity coefficient A_u implies dissipation of vorticity at the rate

$$\partial \xi / \partial t = -\nabla_h^2 (A_u \nabla_h^2 \xi) \quad . \tag{45}$$

So the dissipation of enstrophy, ε, by biharmonic horizontal viscosity is approximately

$$\varepsilon = 1/2 \int (\partial \xi / \partial t)^2 dV = -\int \xi \;\; \nabla_h^2 (A_u \nabla_h^2 \xi)\, dV = -\int (A_u \nabla_h^2 \xi)^2\, dV \tag{46}$$

(where integration by parts has been performed twice and boundary conditions have been assumed negligible).

(c) Diapycnal transports in models resolving the enstrophy cascade

The dissipation of potential energy and enstrophy can be compared in various ways. Following the enstrophy and energy cascade arguments, away from the boundary layers in the enstrophy cascade regime one expects the rate of transfer of enstrophy to small scales to be almost independent of wavenumber whereas the rate of transfer of energy to small scales should scale as k^{-2}. Thus the diapycnal transport to the grid scale should decrease rapidly as the grid resolution is improved.

RM discuss the diapycnal transports in the western boundary current regions. Their arguments suggest that the contribution to the domain average diapycnal transport should decrease with the grid spacing, particularly once the resolution is high enough for the current to be inertial near the boundary.

Thus RM's finding that the diapycnal transport by the horizontal diffusion varies little with the horizontal grid spacing is not easily understood in terms of these scaling arguments. It is possible that the scaling arguments can only be applied to models with horizontal resolution which resolve the energy cascade more fully than those used by RM. It would be worth using (41) to find where the diapycnal transports are largest in their experiments. Together with the scaling with resolution of the isopycnal slopes and other variables in (41), this would illuminate RM's results.

(d) Diffusion inherent in Semi-Lagrangian and upwind flux transport schemes

The diffusion implicit (or inherent) in semi-Lagrangian schemes is dependent on the order of accuracy of the interpolation of the tracers to the upwind trajectory points. McCalpin (1988), following the analysis of McDonald (1984), provides a number of useful forms in terms of the amplification factor, A, of a wave of the form:

$$\theta(x,t) = \theta(i\Delta x, n\Delta t) = \mathrm{Re}\{A^n \exp i\, ki\Delta x\} \quad . \tag{47}$$

For **long waves** the magnitudes of the amplification factors for first order (A_1) to 4th order (A_4) integration schemes are:

$$\begin{aligned} |A_1| &= 1 - \frac{1}{2}\alpha(k\Delta x)^2 \quad , \quad |A_2| = 1 - \frac{1}{8}\alpha^2(k\Delta x)^4 \quad , \\ |A_3| &= 1 - \frac{1}{12}\alpha(k\Delta x)^4 \quad , \quad |A_4| = 1 - \frac{1}{36}\alpha^2(k\Delta x)^6 \quad , \end{aligned} \tag{48}$$

where α is the residual Courant number (which equals the Courant number $U\Delta t/\Delta x$ when the latter is less than 1). 2nd and 3rd order semi-Lagrangian advection have an inherent biharmonic viscosity proportional to the advection velocity for 3rd order advection and to its square for 2nd order advection. For **grid scale** waves (wavelength = 2 Δx) and small values of α

$$|A_1|^2 = 1-2\alpha \quad , \quad |A_2|^2 = 1-4\alpha^2 \quad ,$$
$$|A_3|^2 = 1-\frac{8}{3}\alpha \quad , \quad |A_4|^2 = 1-\frac{16}{3}\alpha^2 \quad . \tag{49}$$

These formulae show that for **small** Courant number the dissipation of the odd order schemes is much greater than that of the even order schemes and that the dissipation at the grid scale of the 2nd and 4th order schemes is rather similar. McCalpin (1988) figures 1-4 show that at the gridscale for Courant numbers larger than 0.1 the dissipation rate is **high** for all orders of interpolation (even and odd).

The upwind 3rd order flux transport schemes developed by Leonard et al. (1993, 1996) have similar dissipation rates to the 3rd order semi-Lagrangian scheme. For 1D advection of θ with small Courant number in the absence of flux correction these schemes are of the form:

$$\theta_i^{n+1} = u_{i+½}^n F_{i+½}^n - u_{i-½}^n F_{i-½}^n \quad ,$$
$$F_{i+½}^n = ½(\theta_{i+1}^n + \theta_i^n) - (1-c_{i+½}^{n\ 2})(\theta_{i'+1}^n - 2\theta_{i'}^n + \theta_{i'-1}^n)/6 \quad , \tag{50}$$
$$c_{i+½}^n = u_{i+½}^n \Delta x/\Delta t \quad and \quad i' = i + ½(1 - u_{i+½}^n / |u_{i+½}^n|) \quad .$$

In the above $F_{i+½}$ represents the flux of θ across the cell face between i and $i+1$, c is a non-dimensionalised velocity and i' is the index (i or $i+1$) of the cell upwind of the interface at $i+½$. A Taylor expansion of this scheme (as in Webb et al. 1998) shows that to leading order the effect of the scheme on tracers is:

$$\theta_t = -u\theta_x - |u|\frac{\Delta x^3}{12}\theta_{4x} \quad . \tag{51}$$

The resulting amplitude factor for long waves (for which the Taylor expansion is valid) agrees with (48) for the 3rd order semi-Lagrangian scheme.

(e) Diapycnal transports of heat by semi-Lagrangian and upwind flux transport schemes

A simple estimate of the diapycnal transport by SL and upwind schemes can be made using (51) and arguments similar to those used in (38)-(42) above. For completeness estimates will be given for both the third order accurate scheme (51) and the first order upwind scheme whose error for 1D advection is:

$$\theta_t = -u\theta_x + ½\Delta x|u|\theta_{xx} \approx -u\theta_x + \partial/\partial x\{½\Delta x|u|\theta_x\} \quad . \tag{52}$$

Here the last term in (52) has been expressed in a flux-like form, similar to (50), neglecting any terms involving spatial derivatives of u. Following (38)-(42), and treating the advection simply as three 1D terms, approximate expressions for the diapycnal transports of density for the first and third order schemes can be derived; the diapycnal components of the horizontal fluxes are similar in form to those for harmonic and biharmonic horizontal diffusion respectively. The diapycnal

transports of density by the vertical velocity, F_w, and the horizontal velocities F_u and their ratios are:

$$F_w = \tfrac{1}{2}|w|\Delta z\rho_z \quad ; \quad F_u = \tfrac{1}{2}|u|\frac{\rho_x \Delta x \rho_x}{\rho_z} \quad ; \quad F_w / F_u = \frac{\rho_z}{\rho_x}\frac{|w|}{|u|}\frac{\Delta z}{\Delta x}\frac{\rho_z}{\rho_x} \tag{53}$$

for 1st order upwind advection and

$$F_w = \frac{1}{12}|w|\Delta z^3 \rho_{zzz} \quad ; \quad F_u = \frac{1}{12}|u|\frac{\rho_{xx}\Delta x^3 \rho_{xx}}{\rho_z} \quad ; \quad F_w / F_u = \frac{\rho_z}{\rho_{xx}}\frac{|w|}{|u|}\frac{\Delta z^3}{\Delta x^3}\frac{\rho_{zzz}}{\rho_{xx}} \tag{54}$$

for 3rd order upwind advection. Furthermore the diapycnal vertical heat flux by the third order scheme is given by

$$H_w = \rho c_p |w|\Delta z^3 \theta_{zzz}/12 \ . \tag{55}$$

Useful order of magnitude estimates of the ratio F_w / F_u in (53) & (54) and the heat flux in (55) for a given model resolution can be made using representative values for the velocities and scales of motion. For our present models with 20 vertical levels and 1° horizontal resolution, $\Delta x = 10^5$ m, $\Delta z \doteq 200$ m in the thermocline, the gyre-scale current has $u \doteq 5$ cm s^{-1} and a horizontal length-scale (away from the boundary currents), $L \doteq 20\Delta x$, whilst $w \doteq 5 \cdot 10^{-4}$ cm s^{-1} and the vertical length-scale, $H \doteq 5\Delta z$. These values give $F_w / F_u \doteq 0.8$ for first order advection and $F_w / F_u \doteq 12$ for 3rd order advection. Taking the temperature difference between the surface and the deep water to be 15 K, and using $\rho = 10^3$ kg m^{-3} and $c_p = 4 \cdot 2\,10^3$ J kg^{-1} K^{-1} with the above values, the vertical heat flux, H_w in (55), is estimated to be 0.2 W m^{-2}. The estimates in this paragraph could easily be in error by a factor of 2 or 4.

Integrations of the HadCM3 model using a simple upwind tracer transport scheme produced a 1 Wm^{-2} reduction in the net outgoing top of the atmosphere flux and a 0.5 K cooling of the sea surface in the first 10 years of integration compared with a control run which used the standard 2nd order centred advection scheme (Pardaens & Gordon, private communication). A second integration in which only the horizontal advection was computed using an upwind tracer scheme gave much smaller (possibly insignificant) differences indicating that the change to the vertical advection had been of greatest significance. These results appear to be moderately consistent with the calculations presented above.

It might be thought that (55) implies that as the number of vertical levels in an ocean model are increased the diapycnal heat transport due to third order upwind advection should reduce in direct proportion with Δz^3. However the thermocline is very poorly resolved in our models at present. Thus on improving the vertical resolution smaller vertical scales will start to be represented. One should expect the diapycnal transport due to third order advection to reduce more slowly than the reciprocal of the cube of the number of levels.

Griffies et al. (2000) note that unphysical overshoots in tracers will be selectively rectified by the parametrisation of convective overturning in present ocean models. Convection will either flux density downward or leave the state unchanged. So together with a poor advection scheme it could support a spurious "unphysical" upward transport of heat. An upwind tracer transport or SL scheme will tend to diffuse density upward (a first order accurate scheme would always diffuse density upward). It would certainly be worth investigating the heat transport by the convective parametrisation in the coupled model. Convective overturning in regions of strong vertical motions such as the equator and western boundary would be unphysical and their magnitude should be assessed.

(f) Griffies et al. (1998) isopycnal diffusion scheme

In the original isopycnal diffusion scheme, as implemented by Cox (1987) and used for many years in the UM and the modular ocean model (MOM), the gradients in tracers and density were calculated as centred differences over two grid squares and were hence insensitive to some of the grid scale "noise". The Griffies et al. (1998) scheme for isopycnal diffusion of tracers is explicitly formulated to ensure dissipation of grid scale variability on isopycnal surfaces. This is a significant advantage for a scheme which needs to serve as a grid scale closure.

Gent & McWilliams (Gent et al. 1990 & 1995) formulated a parametrisation of sub-grid scale mixing in which a "bolus" velocity additional to that related to the cell-mean momentum advects the tracer field. The velocity field is derived from the slopes of the isopycnal surfaces and acts to release available potential energy. Griffies (1998) shows that the flux transported by the skew-symmetric part of the isopycnal diffusion tensor is identical (to within a vector whose divergence is zero) to the advective flux given by the Gent-McWilliams formulae. Section 5b of Griffies (1998) makes the important point that no grid scale variations are invisible to the skew diffusive flux formulation so that it will act as a reliable grid scale closure for available potential energy. This appears to be difficult to ensure with the original velocity field formulation.

Appendix B provides a concise statement of the functional used in the Griffies et al. (1998) scheme and a derivation of the expressions for the fluxes given in that paper. Griffies et al. (1998) refer to appendix C of the documentation of MOM3 (Pacanowski & Griffies 1999) for their derivation. This occupies more than 20 pages of algebra including 300 equations. A concise statement is useful as it might suggest how to formulate more satisfactory scale selective ("biharmonic") forms of the Gent & McWilliams scheme (RM section 5b). Appendix B also discusses briefly the grid scale closure of the skew diffusive formulation.

2.5 Non-hydrostatic and compressible models

Tanguay et al. (1990) were the first to perform SISL integrations in which sound waves and non-hydrostatic accelerations were calculated semi-implicitly. Marshall et al. (1997) provide a formulation for an incompressible non-hydrostatic solver in which the pressure gradient forces are calculated semi-implicitly but other terms, including the vertical advection of temperature and advection of momentum are calculated explicitly (using the Adams-Bashforth integration scheme). Marshall et al. (1997) make the point that the non-hydrostatic equations do not introduce any gravity waves with higher frequencies than the hydrostatic equations. This follows from Table 8.1 of Gill (1982) which indicates that (in the absence of rotation) in hydrostatic flows the dispersion relation is

$$\omega^2 m^2 = N^2 \kappa_h^2 \quad , \tag{56}$$

where m is the vertical wavenumber and κ_h the horizontal wavenumber whilst in non-hydrostatic flow,

$$\omega^2 (m^2 + \kappa_h^2) = N^2 \kappa_h^2 \quad , \tag{57}$$

from which it is clear that for any m and κ_h, ω is smaller for the non-hydrostatic equations. The CFL criterion for these waves turns out to be strictest for the nearly horizontal motions. The CFL criterion requires that

$$\Delta t < \Delta s / c = \kappa \Delta s / \omega \quad , \tag{58}$$

where Δs is the grid spacing in a given direction and κ is the total wavenumber. Writing the angle between the wavenumber vector and the horizontal as ϕ, (57) gives

$$\omega = N \cos\phi \quad . \tag{59}$$

On any grid the shortest waves have $\kappa \Delta s = \pi$ and using (59) in (58) one sees that the strictest stability condition is obtained when $\cos\phi = 1$, the case for horizontal travelling waves.

Killworth (personal communication) has outlined a number of reasons for wanting to avoid the Boussinesq approximation. Some aspects of it can probably be avoided in a number of ways. Clearly it is not needed when the fluid is allowed to be compressible. Tanguay et al. (1990) were able to integrate a SISL scheme for a compressible, non-hydrostatic atmosphere. As the speed of sound in the ocean but not the atmosphere is several times faster than the speed of barotropic gravity waves (see table 1), some examination is needed of the effect of allowing sound waves on the conditioning of the elliptic solver. The next paragraph suggests that allowing sound waves into the system slightly improves the conditioning of the elliptic system (just as allowing gravity waves improved the conditioning of the barotropic mode in section 2.3).

The linearised equations for a stratified, compressible, non-hydrostatic fluid are given in Gill (1982) section 6.14 – see particularly (6.14.17)-(6.14.19). Substituting $1/(\Delta t)$ for $\partial / \partial t$ in those equations, as in section 2.3, and deriving a single equation for the vertical velocity w one obtains:

$$(1 + N^2 \Delta t^2)^{-1} \left(\frac{\partial}{\partial z} + \Gamma\right)\left(\frac{\partial}{\partial z} - \Gamma\right) w + \left(\nabla_h^2 - \frac{1}{(c_s \Delta t)^2}\right) w = 0 \quad , \tag{60}$$

where

$$\Gamma = \frac{1}{2}\left(\frac{g}{c_s^2} - \frac{N^2}{g}\right) \quad . \tag{61}$$

This equation should be compared with (24) which for $f=0$ applies for the hydrostatic, incompressible case. At least for the special case when Γ is independent of height its introduction reduces the ratio between the largest and smallest eigenvalues of the vertical modes. Similarly the ratio of the maximum and minimum eigenvalues for the horizontal modes is reduced by the additional term $(c_s \Delta t)^{-2}$. The impact of the term equal to 1 supplementing $N^2 \Delta t^2$ on the range of eigenvalues is relatively subtle. It appears to slightly worsen the ratio of eigenvalues for the pre-conditioned form considered in (33)-(35) but may be beneficial for other pre-conditionings.

3. Grids

3.1 Horizontal grid staggering

As the French OPA and Princeton POM models use the C grid for their horizontal staggering of variables whereas the MOM model uses a B grid, I feel obliged to summarise the arguments in favour of each type of grid. Although most of the facts summarised here are easily established and well known the weighting to be given to them is rather unclear. For the reader's convenience the distribution of the primary variables u, v, p for grids A-D is illustrated in figure 2. The E grid is a well disguised rotation of the B grid. Randall (1995) provides an informative discussion of it. To avoid complicating the notation it is not discussed further.

(a) Dispersion relations for Rossby waves and inertia-gravity waves

The dispersion relations for inertia-gravity (Poincaré) waves and for Rossby waves for each of the grids A-D (and also for E) are stated by Dukowicz (1995) and illustrated for the cases $a = 2\Delta x$ and $a = \Delta x/2$. Here Δx is the length of a grid cell (the shortest distance between points storing the same variable) and $a = c/f$ is the Rossby radius ($c^2 = gH$ for external waves). Note that figure 3D of Dukowicz appears to be incorrect and that Dukowicz uses λ to denote the Rossby radius. The wavenumbers which are represented by the grids A-D lie in the ranges $-\pi \le k\Delta x \le \pi$, $-\pi \le l\Delta y \le \pi$. A wave with $k\Delta x = \pi$ changes its sign from one grid point to the next in the x direction (i.e. its wavelength is $2\Delta x$). It is only reasonable to expect the dispersion relations to be accurate for wavelengths longer than $4\Delta x$ (i.e. $k\Delta x < \pi/2$). Purser & Leslie (1988) argue that the solution should be viewed as only including those wavelengths.

It is well known that for inertia-gravity waves the B grid is better when the grid spacing is coarser than the Rossby radius $a < \Delta x$ and the C grid is preferable when the grid spacing is finer ($a > \Delta x$). The Rossby radius of high order vertical modes is smaller than the grid spacing even in high resolution atmosphere and ocean models so the performance of the grids for these waves in high resoltuion models is mixed.

Usually the accuracy of the Rossby wave propagation speed is viewed as more important than that of the inertia-gravity waves (this is assumed by SISL methods). Dukowicz (1995) suggests that the B grid is better than the C grid for Rossby waves at all scales whilst Wajsowicz (1986) concludes that the C grid is better when the grid resolution is better than the Rossby radius. Within the region where the Rossby wave dispersion relation can be expected to be accurate ($0 \le k\Delta x \le \pi/2,\ 0 \le l\Delta y \le \pi/2$) the relative advantages appear to be rather marginal and my opinion is that there is no really important difference between these dispersion relations for the B and C grids.

Hsieh et al. (1983) discuss boundary Kelvin waves and Ng and Hsieh (1994) equatorial Kelvin waves on the B and C grid. For poor grid resolution the C grid represents the phase speed of the waves better but the structure of the waves slightly worse than the B grid.

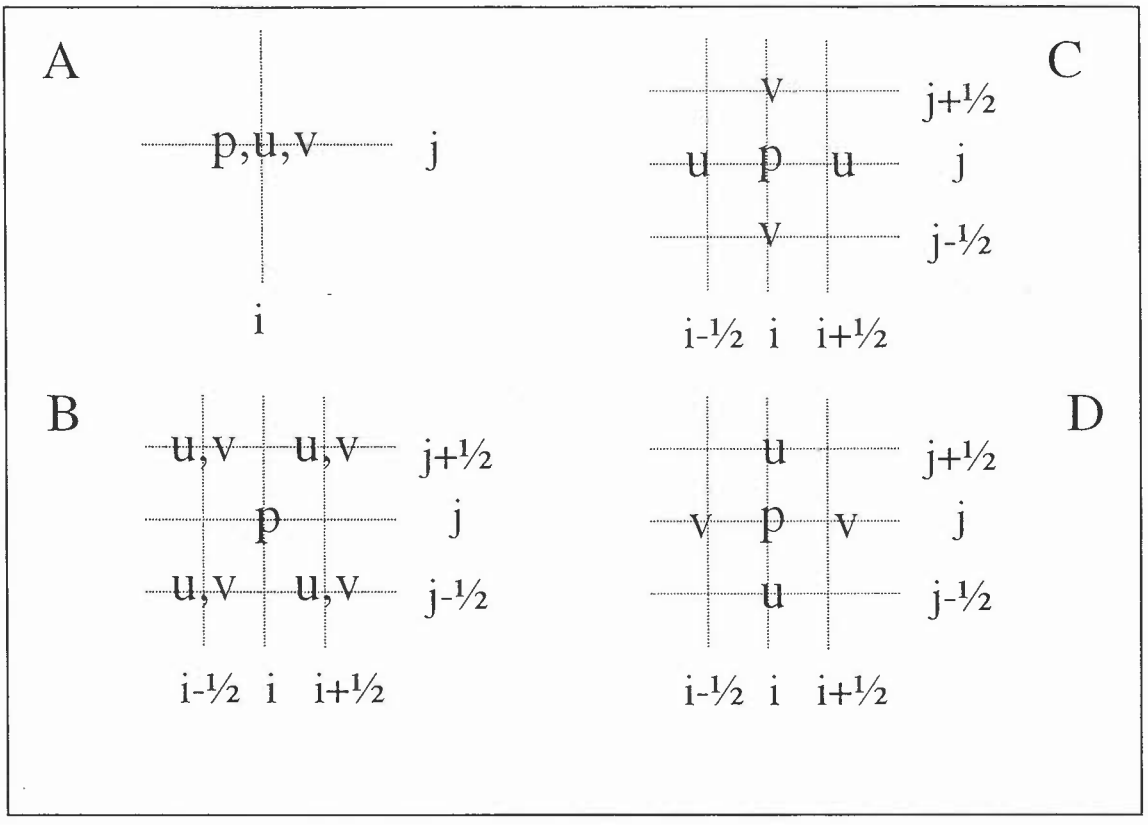

Figure 2: Distribution of primary variables for grids of type A-D.

(b) Stationary solutions

An aspect of these dispersion relations which requires some discussion is the presence of stationary grid-scale solutions (with $\omega = 0$). As the wave dispersion relations for inertia-gravity waves do not depend on the sign of k or of l it is adequate for them to consider only the first quadrant ($k, l \geq 0$). The A grid has 3 stationary solutions: at $k\Delta x = \pi, l\Delta y = 0$; $k\Delta x = 0, l\Delta y = \pi$; and $k\Delta x = \pi, l\Delta y = \pi$. The B grid has one at $k\Delta x = \pi, l\Delta y = \pi$ and the D grid has two lines of such solutions one along $k\Delta x = \pi$ and a second along $l\Delta y = \pi$. The C grid does not have any strictly stationary solutions.

Dukowicz (1995) shows that Rossby waves have false stationary solutions on the B grid for the line $k\Delta x = \pi$, for the C and D grids for two lines, namely $k\Delta x = \pi$ and $l\Delta y = \pi$ and that the A grid does not have false stationary solutions.

The significance of stationary solutions is debatable. Le Roux et al. (1998, section 3) view them as very significant, choosing their distribution of variables so as to avoid null spaces for the pressure gradient and velocity divergence operators (i.e. non-trivial grid-scale solutions of $\nabla p = 0$ and $\nabla.\mathbf{u} = 0$ respectively). Le Roux et al. argue that for solution uniqueness and convergence null spaces for these operators must be avoided – but this is not a completely compelling argument as the null solutions generally lie outside the range of wavelengths ($0 \leq k\Delta x \leq \pi/2,\ 0 \leq l\Delta y \leq \pi/2$) "resolved" by the grid. Checkerboard patterns lie in the null space of both operators on the B grid whereas the C grid operators have empty null spaces.

There are nearly stationary inertia-gravity waves on the C grid which Adcroft et al. (1999) view as a problem worth solving for coarse resolution ocean models. They suggest a formulation which has 4 velocity values within a grid cell, (the full horizontal velocity vector is stored at two points). The dispersion relation for inertia-gravity waves is improved at the expense of introducing computational modes with the inertial frequency. The dispersion relation for Rossby waves and the representation of geostrophic balance (le Roux et al. 1998, section 2 and Purser & Leslie 1988) for this arrangement do not appear to have been examined.

There is a sense in which the stationary solutions distort the surface of the dispersion relation. For example the shape of the dispersion curve for inertia-gravity waves on the B grid can be obtained from the exact dispersion curve by pulling the ω surface down to zero at the point $k\Delta x = \pi$, $l\Delta y = \pi$. The false maxima and minima in the dispersion relation result in incorrect group velocities. The main errors in the dispersion relation lie however outside the region $(0 \le k\Delta x \le \pi/2,\ 0 \le l\Delta y \le \pi/2)$ where it is reasonable to expect the dispersion relation to be accurate.

(c) Locations of vorticity and divergence tendencies

It is perhaps worth mentioning the relation between the dispersion relations and equations for the rate of change of divergence and vorticity appropriate for B and C grids (Wajsowicz 1986 appendix). Taking the divergence of the momentum equations one obtains

$$w_{zt} + f\zeta + \beta u = 1/\rho_0 \nabla_h^2 p \quad . \tag{62}$$

Taking the curl gives one:

$$\zeta_t + \beta v - f w_z = 0 \quad . \tag{63}$$

On the B grid both these equations are naturally calculated at the same point as p.

On the C grid the divergence, w_z, is naturally calculated at the same point as p but the vorticity is staggered lying half a grid square north/south of u points and east/west of v points. Thus the vortex stretching at the vorticity point must be calculated as a centred four point average of values of w_z. The dispersion relation for quasi-geostrophic flow can be obtained simply from (63) by using geostrophic velocities for ζ_t and βv. The dispersion relation for inertia-gravity waves is obtained at the divergence gridpoints. (62) involves a four point average of ζ evaluated at the vorticity points. Each of these is calculated from (63) using a four point average of fw_z. Consequently a nine-point average of fw_z is involved in (62).

(d) Accuracy

Purser & Leslie (1988) suggest that the accuracy of the dispersion relations in the range $(0 < k\Delta x < \pi/2, 0 < l\Delta y < \pi/2)$ should be the main focus of attention. They use schemes with higher order accuracies, including compact schemes, to achieve this. Given the rather poor accuracy of the 2nd order schemes, particularly for the Rossby wave dispersion relation, this appears to be a rational approach. Exploration of 4th order accuracy for the main terms in the

vorticity equation seems justified. Which of the terms need to be calculated with this accuracy implicitly and the effect on the conditioning of the elliptic solver would need consideration.

(e) Other factors

The C grid is the natural one to use to advect tracers using a flux formulation. Semi-Lagrangian schemes would appear to be more accurate at small Courant numbers when the velocity and tracer points coincide. This point is also relevant to the choice between vertical grids (see below).

The C grid is more convenient than the B for pressure solvers (Marshall et al. 1997).

If the choice between grids is significant for ocean modelling I suspect it is because of the performance of the grid in problems other than those discussed above – e.g. in modelling of western boundary layers. Certainly the C grid is convenient for specification of free-slip boundaries conditions (Adcroft & Marshall 1997).

3.2 Vertical staggering

Figure 3 illustrates the 3 most interesting grids discussed by Tokioka (1978) and Arakawa (1988) in their exhaustive coverage of the numerical discretisation of the vertical mode equations for horizontally propagating inertia-gravity and Rossby waves. Gill (1982 section 6.11) explains the derivation of the two equations and boundary conditions involved. The first equation is the advective form of the density conservation equation derived using the hydrostatic relation by linearising about a stratified background state:

$$\rho_0 N^2 w = -p_{zt} \quad . \tag{64}$$

The second equation is obtained from the horizontal divergence of the horizontal momentum equations and the incompressibility condition ($\nabla.\mathbf{u} = 0$). As shown in Gill (1982) table 8.1 this equation depends on the scales of motion represented in the model. For quasi-geostrophic flow it is

$$\rho_0 f^2 w_z = (p_{xx} + p_{yy})_t \tag{65}$$

For hydrostatic non-rotating flow it is

$$\rho_0 w_{zt} = (p_{xx} + p_{yy}) \tag{66}$$

(Gill's table 8.1 has errors in the signs of these two equations.) For a flat bottomed ocean the boundary conditions for these two equations are:

$$\begin{aligned} w &= 0 \qquad at\ z = -H\ , \\ p_t &= \rho_0 g w \quad at\ z = 0\ . \end{aligned} \tag{67}$$

For internal modes the latter is approximated well by $w = 0$ (Gill 1982, p161). Seeking separable modes of solution of (64) and (65) or (66) of the form

$$w = \hat{w}(z)\eta_t(x,y,t) \quad and \quad p = \hat{p}(z)\eta(x,y,t) \tag{68}$$

the vertical structure is determined by the two equations

$$\rho_0 N^2 \hat{w} = -\hat{p}_z \ , \tag{69}$$

$$\hat{p} = \rho_0 c_e^2 \hat{w}_z \quad . \tag{70}$$

where c_e is a separation constant with the dimensions of a velocity, and boundary conditions on $\hat{w}$ are derived from (67).

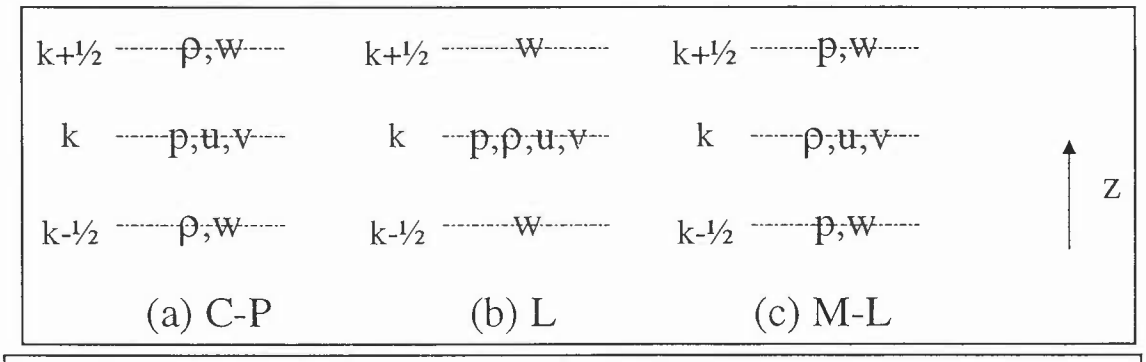

Figure 3: Distribution of basic model variables in the vertical: (a) the Charney-Phillips grid, (b) the Lorenz grid and (c) the Modified Lorenz grid.

Figure 3(a) shows the Charney-Phillips (C-P) grid which stores w and ρ at half levels and p,u and v at full levels. Note that the level number will be taken to increase with the height, z. The upper and lower boundaries are at half-levels. (64) is applied centred about the half levels (as it derives from the evolution equation for ρ) and (65) or (66) at full-levels (as they relate to the evolution equation for the horizontal divergence). This arrangement is ideal for evaluation of (64) particularly when N^2 is assumed to be known. It is also very well suited to discretisation of the vertical component of momentum in non-hydrostatic models. Figure 3(c) illustrates an alternative grid in which w and p are stored at half-levels and ρ,u and v at full levels. On this grid (64)-(66) are all evaluated at the half-levels. This arrangement is a natural one if the density is vertically advected using a flux formulation and allows a 2nd order accurate Richardson number to be easily calculated. It is not well suited to discretisation of the vertical component of the momentum equations in non-hydrostatic models. It also involves a vertical averaging of pressure in (65) and (66). Figure 3(b) illustrates the Lorenz grid which also evaluates (64)-(66) at the half levels. It differs from figure 3(c) only in that p is stored at the full levels rather than the half levels. The grid shown in figure 3(c) will be termed the Modified Lorenz grid. On the Lorenz grid

$$p_{k+1} - p_k = g\Delta z(\rho_{k+1} + \rho_k)/2 \quad . \tag{71}$$

The vertically averaged pressures at the full levels where u and v are held for the Modified Lorenz grid also satisfy (71). Thus for centred 2nd order accurate formulae the Modified Lorenz grid is equivalent to the Lorenz grid.

Arakawa and Moorthi (1988) (hereafter AM) compared the Lorenz and C-P grids and identified several apparent weaknesses of the Lorenz grid:

i. the advective term $\rho_0 N^2 w = -g\rho_z w$ in (64) is treated more accurately by the C-P grid. AM suggest that it is always poorly represented in the layers next to the upper and lower boundary on the Lorenz grid.
ii. only the C-P grid conserves an analogue of potential vorticity
iii. the Lorenz grid gives a poorer solution than the C-P grid to the Eady problem
iv. the Lorenz grid has a computational mode in which $p = 0$ at full levels, $w = 0$ at half levels and the density oscillates between levels: $\rho_{k+1}\Delta z_{k+1} = -\rho_k \Delta z_k$

Points i, ii and iv are not as clear cut as AM suggest. Algebra supporting the following statements is presented in Appendix C.

Using a flux formulation for the continuity equation on the Lorenz grid the advective term in (64) takes the form:

$$w\rho_z \rightarrow \frac{w_{k+\frac{1}{2}}(\rho_{k+1} - \rho_k)}{2\Delta z} + \frac{w_{k-\frac{1}{2}}(\rho_k - \rho_{k-1})}{2\Delta z} \qquad (72)$$

In the layer next to the wall one of the vertical velocities in (72) is zero and the advective term is half as large as the advective term for the C-P grid next to the wall. This is because the term is calculated at full levels on the Lorenz grid and half levels on the C-P grid and so is twice as far from the wall on the C-P grid as it is on the Lorenz grid. The underestimation (by a factor of 2) of the vertical advection by the Lorenz grid at mid-depth in the 2 layer problem (AM equation (5.16)) is a special case which is not typical of multi-layer models.

Section (a) of appendix C shows that a flux formulation for the continuity equation on the C-P grid can be found which gives an ideal form for (64). This formulation, which appears to be that used by Arakawa & Konor (1996) does not have the most obvious (or desirable) form for the continuity equation.

AM derive a potential vorticity equation for the Lorenz grid but conclude that it does not provide a good analogue of the quasi-geostrophic form of the potential vorticity conservation relation. It is shown in section (b) of appendix C that a small extension to the derivations of AM results in a conservation equation which though not ideal is tolerably accurate.

AM's investigation of the simulation of baroclinic instability using the quasi-geostrophic equations on the C-P and Lorenz grids shows that the Lorenz grid supports spurious short wavelength instabilities in the Eady problem. Of these only the very short waves with a steering level within a grid spacing of the boundary grow rapidly. These modes are probably very shallow and would not transport much heat (Held 1978). The Eady problem is well-known not to be robust to small changes to its formulation. AM's simulations of the Green problem (which is more robust) using the two grids look quite similar.

The computational mode obtained on the Modified Lorenz grid may be of some importance in ocean models. AM explain that it arises because the winds are driven by the pressure field on full

levels. On the C-P grid the wind shear between adjacent (full) levels is directly related to the density at the half level inbetween. In other words the thermal wind shear on the Lorenz grid involves no vertical averaging:

$$\rho_0 f (u_k - u_{k-1}) = -g \delta_y \rho_{k-\frac{1}{2}} \qquad (73)$$

By contrast on the Lorenz grid the pressure field on the full levels is the average of the pressure fields on adjoining half levels and only the pressures on the half levels are tightly coupled to the densities. As discussed briefly in section (c) of appendix C it appears that regardless of the number of levels in the model there is only one computational mode and its structure is the very specific one described in point (iv) above.

One reason why this mode may be important is that simple centred-time centred-space vertical advection schemes generate grid scale noise in the density profile which will project strongly into it. At least where the ocean bottom is flat, the computational mode is dynamically passive and stationary. Thus its amplitude is likely to increase steadily with time where the vertical advection projects strongly into it; Rossby waves and inertia-gravity waves will not act to "disperse" it. Results from 1D advection/convection experiments such as those reported by Griffies et al. (2000) are thus more relevant to 3D numerical models than would otherwise be the case.

A second reason why the mode may be significant in ocean models is that the surface pressure field and the bottom pressure field in the Modified Lorenz grid are affected by it. The mode is zero at the full pressure levels but non-zero at the surface and bottom of the model (which are at half levels). The free surface height in free surface models is thus likely to be noisy. Damping of this noise without regard for the noise remaining in the density field will generate grid scale noise in the vertical in the pressure field at full levels. It seems likely that this would drive noisy vertical velocities.

Leslie & Purser (1982) studied the accuracy of the vertical modes on an unstaggered grid and the grids shown in figure 3 using 2nd, 4th and 6th order accurate numerics. They demonstrated that using higher order schemes gives significant improvements in the representation of modes of intermediate wavenumber. The performance of the original and modified Lorenz schemes were similar for all the schemes that they used and the C-P grid had a clear edge in performance over the Lorenz grids at each order of accuracy.

3.3 Terrain-following coordinates

(a) Advantages and disadvantages

The potential advantages of vertical coordinates which allow the upper and lower boundaries of the model to smoothly follow the upper and lower boundaries of the ocean are: (a) intersection of coordinate surfaces by the boundaries or coarse representation of the lower boundary as a series of steps are avoided, (b) surface and bottom boundary layers can be resolved, (c) flows up the continental shelf break and accompanying bottom pressure torques should be smoothly represented. Ocean models using these coordinates have mainly been applied with high resolution to model the continental shelf.

Over steep bathymetry these coordinate surfaces inevitably have quite large slopes. Horizontal pressure gradients are calculated as the sum of the pressure gradients along these surfaces and

vertical pressure gradients. Each of these two gradients is typically much greater than the horizontal gradient. The angle between the coordinate and isopycnal surfaces is also typically larger than that in z coordinate models. The problems using horizontal diffusion or dissipative advection schemes discussed in section 2.4 are thus likely to be more acute in these boundary following coordinates.

(b) Errors in pressure gradient calculation

Mellor et al. (1994) present a useful example which indicates the order of magnitude of pressure gradient errors which can be expected in ocean models. The scaling of this error with grid resolution and other factors is discussed in (c) below. Their example uses a model with 20 km horizontal grid spacing and 21 layers with equal σ coordinate spacing to model flow over the continental break. The bathymetry descends (smoothly) by nearly 4000 m over a 200 km distance. A horizontally uniform density field is specified which is representative of the area-mean density field for the North Atlantic. After 90 days of integration (in which the density field is held constant) the horizontal velocity along the bathymetry (which should of course be zero) is 5 cm/s near the surface and 7 cm/s near 2000 m depth. These are substantial errors; the latter velocity is particularly undesirable as it is comparable with the velocity of deep western boundary currents. The errors depend strongly on the resolution of the bathymetry by the grid. HadCM3 and the operational global FOAM model use much rougher bathymetry than this example. Together with the scalings indicated in (c) below it appears that these coordinates are not attractive for models with worse than 20 km resolution.

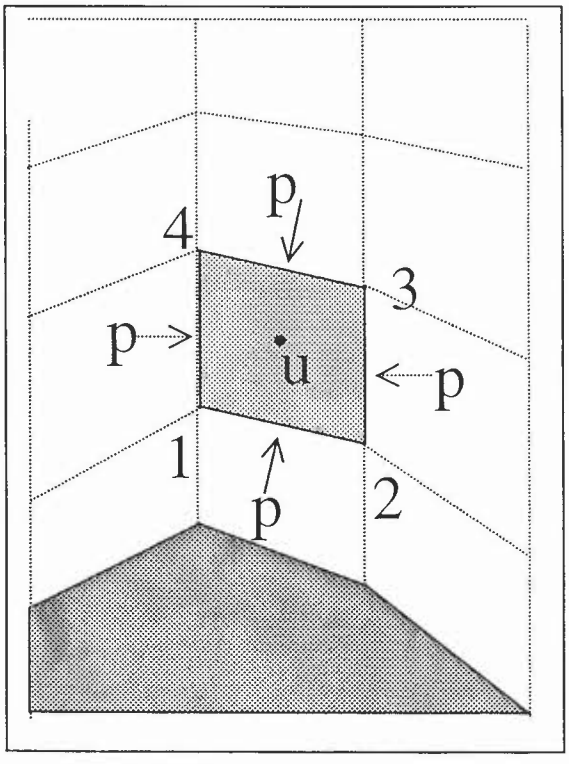

Figure 4: Pressure forces on a general vertical coordinate cell

Song (1998) provides a recent summary of methods and papers which have aimed to ameliorate this problem (and its counterpart for the atmosphere). Subtraction of the domain average density field $\rho_0(z)$ is one partially successful technique. It is not quite clear why this has not been applied using averages over much smaller domains (e.g. domains only slightly larger than the stencils required for vorticity and divergence discussed in section 3.1). A widely used approach is to ensure conservation of depth integrated angular momentum and/or analogues of the conservation of energy in conversion from kinetic to potential energy (Arakawa & Suarez 1983). Most authors also aim to obtain exact results (zero pressure gradients) for horizontal isopycnals when the density or pressure depends linearly on the chosen coordinate. McCalpin (1994) and Chu & Fan (1998) have used more accurate (4^{th} and 6^{th} order and also compact) algorithms to calculate the pressure gradients along the constant σ surfaces.

Lin (1997) and Stelling & van Kester (1994) discuss finite volume approaches in which the pressure force on each momentum cell is calculated directly as the integral over all the faces of the cell of the normal pressure forces (see figure 4). Stelling & van Kester use 1^{st} order accurate formulae where the coordinates slope particularly steeply. In view of

the steepness of ocean bathymetry this may be essential. Lin's approach is simpler and is the only one discussed here. His formulation was designed for an atmospheric model. According to his results his formulation substantially outperforms the Arakawa & Suarez (1983) formulation for an idealised test case. Sub-section (c) below derives an expression for the error in Lin's formulation ((77) below) when the isopycnals are flat and compares it with a similar expression derived by Mellor et al. (1994). Sub-section (d) discusses the pressure torques on cells which have sloping upper and lower surfaces. These torques appear to be represented well not only in Lin's formulation but in most σ coordinate formulations. Though the arguments in (c) and (d) are expressed using σ coordinates they could be applied to any generalised vertical "s" coordinate (Kasahara 1974 and Song 1998).

(c) Pressure gradient error in Lin's formulation for flat isopycnals

Figure 4, which is a copy of figure 1 from Lin (1997), illustrates the grid staggering he employs in the x- z plane. He stores p at the corner points 1-4 of the velocity cell, so the staggering is equivalent to the C grid in the horizontal and the Modified Lorenz grid in the vertical. The force in the x- direction on the cell illustrated is the projection of the pressure forces on each face along the x- axis:

$$\sum \mathbf{F}_x = -\left[\int_1^2 p\,dz + \int_2^3 p\,dz + \int_3^4 p\,dz + \int_4^1 p\,dz\right] \tag{74}$$

It is clear from this equation that if pressure is strictly a function of z each grid cell pressure force will be identically zero. Thus "only" errors in the numerical approximation of the integrals of the pressure give rise to errors in the net acceleration. Lin suggests calculating each integral as a simple average of the pressures at each end of the domain of integration. When the density is purely a function of z (flat isopycnals) the error in the first two integrals obtained using the simple average formula advocated by Lin is:

$$\begin{aligned}
\Delta I_{12} &\equiv \Delta\int_1^2 p\,dz = \int_1^2 p\,dz - 1/2\,(p(z_1)+p(z_2))H_x\Delta x = -\frac{p_{zz}\sigma^3 H_x^3 \Delta x^3}{12} + h.o.t. \\
\Delta I_{23} &\equiv \int_2^3 p\,dz = \int_2^3 p\,dz - 1/2\,(p(z_2)+p(z_3))H\Delta\sigma = -\frac{p_{zz}H^3\Delta\sigma^3}{12} + h.o.t.
\end{aligned} \tag{75}$$

where Δx and $\Delta\sigma$ are the dimensions of the cell, p_{zz} is evaluated at the central points of the line integrals, *h.o.t.* stands for higher order terms, and following Mellor et al (1994), $z=\sigma H$ with $\sigma=-1$ at $z=-H$ and $\sigma=0$ at $z=0$. Expressing the sum of ΔI_{12} and ΔI_{34} in terms of derivatives at the centre of the cell, and combining ΔI_{23} and ΔI_{41} in a similar manner, one obtains

$$\begin{aligned}
\Delta I_{12} + \Delta I_{34} &= \frac{\Delta\sigma\,\Delta x^3}{12} H_x^3 (3\sigma^2 p_{zz} + H\sigma^3 p_{zzz}) + h.o.t. \quad , \\
\Delta I_{23} + \Delta I_{41} &= -\frac{\Delta\sigma^3\,\Delta x H^2 H_x}{12} (3\sigma^2 p_{zz} + H\sigma^3 p_{zzz}) + h.o.t. \quad ,
\end{aligned} \tag{76}$$

where p_{zz} and p_{zzz} are calculated at the centre of the cell. Finally, adding these two terms together and re-arranging to aid comparison with Mellor et al. (1994) the total error in the sum of forces is:

$$\frac{\Delta\sum \mathbf{F}_x}{H\Delta\sigma\Delta x} = \frac{HH_x}{4}(p_{zz} + \frac{\sigma H}{3} p_{zzz})(\Delta\sigma^2 - \sigma^2 \frac{H_x^2 \Delta x^2}{H^2}) + h.o.t. \quad . \tag{77}$$

$H\Delta x\Delta\sigma$ is the volume of the cell, thus (77) is equivalent to an error in p_x.

To explain the relationship of (77) to (7) of Mellor et al. (1994) it is necessary to note that Mellor et al. discuss two expressions for the pressure gradient in σ coordinates, namely

$$p_x|_z = p_x|_\sigma - \frac{\sigma}{H} H_x p_\sigma \quad \text{, and} \tag{78}$$

$$p_x|_z = p_{0x}|_\sigma + \rho_0 H \int_\sigma^0 b'_x|_z \, d\sigma' \quad ; \quad b_x|_z = b_x|_\sigma - \frac{\sigma}{H} H_x b_\sigma \quad . \tag{79}$$

In (79) p_0 is the surface pressure, ρ_0 is a constant reference density and $b = \rho g / \rho_0$ is the buoyancy. As Mellor et al. (1994) note, (78) and (79) are equivalent analytically but may differ after approximation by finite difference algebra. Mellor et al.'s equation (7) is an expression for the error in the second half of expression (79) when the isopycnals are flat. Their expression when applied to (78) yields (77).

There are four further points relating to (77) – (79) that are worth making. Firstly the Arakawa & Suarez (1983) formulation is based on (78), so it remains to be explained why Lin found his formula more accurate than theirs for his flat isopycnal calculation. Secondly (79) is exact when the density is a linear function of the chosen coordinates whereas (78) is only exact when the pressure is a linear function of the coordinates. Song (1998) views this as a significant advantage of formulation (79). As the vertical density gradients are nearly always smallest near the top and bottom of the ocean quantitative calculations for realistic test cases are needed to establish whether one formulation or other does have a significant practical advantage. Thirdly the derivation of (77) for Lin's formulation seems to be more intuitive than that from (78). Starting from (78) it seems that brute force algebra is needed to establish (77). Finally Mellor et al. (1994) note that for the example quoted in subsection (b) above the maximum value of $|(\sigma H_x \Delta x)/H| = 13/20$. The square of this quantity appears as the 2nd half of the last term in (77). Thus the resolution used in the example is the coarsest at which concepts like 2nd order accuracy make any sense and the points made by Stelling & van Kester (1994) appear to be very relevant.

(d) Pressure torques on grid cells

Lin's construction should ensure that a property corresponding to curl grad p = 0 is satisfied for σ and s coordinates. The transformed pressure gradient in σ coordinates does not satisfy this property directly because the the depth of the grid cells, h, varies spatially. Applying Stokes' theorem to a σ coordinate grid cell (as illustrated in figure 4) and considering only the acceleration due to the pressure gradient one obtains:

$$\oint h\rho\mathbf{u}_t.d\mathbf{l} = \int \nabla \wedge (h\rho\mathbf{u}_t).d\mathbf{n} = -\int \nabla \wedge (h\nabla p).d\mathbf{n} = -\int (\nabla h \wedge \nabla p).d\mathbf{n} \quad . \qquad (80)$$

The latter expression can be related to the curl of the pressure forces on the top and bottom of the sigma coordinate cell as follows. The (area weighted) normal to the surface $z = h_1(x, y)$ is given by $(h_{1x}, h_{1y}, -1)$. So the vertical component of the curl of the pressure force on the surface $z = h_1$ is:

$$\hat{\mathbf{z}}.\int curl(p(h_{1x}, h_{1y}, -1))dxdy = \int (ph_{1y})_x - (ph_{1x})_y dxdy = \int p_x h_{1y} - p_x h_{1y} dxdy \quad . \qquad (81)$$

In other words the depth integrated rate of change of circulation around the cell equals the difference between the pressure torques on the top and bottom of the cell. The standard transformed coordinate formulae (e.g. (78)) give the same expressions. This is an advantage for s co-ordinates as it implies that the pressure torque on the bottom of the ocean should be represented faithfully in s coordinates. Song & Wright (1998) indeed confirm that this is the case for Song's (1998) finite difference formulation. Use of the modified Lorenz grid which has the pressure held at the upper and lower surfaces of the domain gives a clean representation of the bottom pressure torque.

(e) Semi-implicit calculations

It is clear from Lin's formulation how to obtain more accurate calculations of the pressure gradient term (assuming the grid resolution is sufficient). In ocean modelling the interaction of the time mean flow and the bathymetry is very important. To avoid complicating solution of the elliptic equation if semi-implicit methods are used, the more accurate calculations could be used only on the central timestep and simpler formulations used to approximate the 2nd order time derivative of the pressure gradient term (see Ritchie et al. 1995).

3.3 Generalised curvilinear orthogonal horizontal coordinates

There are two quite different ways in which these more general co-ordinates could be useful.

(a) Orientation of a coordinate along the shelf break

This idea has been widely used by workers with the Princeton Ocean Model (e.g. Ezer & Mellor 1994). It could enable the resolution perpendicular to the shelf break to be significantly enhanced so as to reduce the errors in the calculation of the pressure gradients (see previous section).

A second possibility when a small region is the main focus of attention is to choose the grid resolution to be finest in this region and to increase the grid spacing in both orthogonal directions as the coordinate leaves the area of interest. Staniforth et al.(1999) discuss the application of this idea to variable resolution NWP models.

For coastal modelling it is probably desirable also to enable the resolution on the shelf to be finer in both x and y than in the deep water – because the Rossby radius is smaller on the shelf. This cannot be achieved using an orthogonal grid – it would require an unstructured grid.

(b) More uniform resolution global models

The timestep in our current global models is limited by a CFL type condition on the propagation of internal gravity waves in the region of convergence of the grid longitudes. The longitude spacing at 75°N (beyond which Fourier filtering is used) is 4 times smaller than at the equator.

Madec & Imbard (1996) and Bentsen et al. (1999) have proposed that the singularity in the latitude-longitude grid at the North geographical pole be moved onto a continent (such as Asia) using conformal mappings. These mappings (Kreyszig, 1979, chapter 13) preserve angles (and hence provide orthogonal grids) and are locally isotropic (the grid stretching is uniform in all directions at a given point). Moving the pole well away from sea points enables an ocean model to be integrated without Fourier filtering. The ratio between the maximum and minimum longitude spacing which can be achieved by this means appears to be not much smaller than 4. An isotropic grid can be obtained if the "latitude" spacing is made proportional to the cosine of latitude as in the standard Mercator grid.

Several authors including Coward et al. (1994) have proposed that the majority of the ocean be represented using a standard latitude-longitude grid but that north of the equator the Atlantic and Arctic Oceans be covered by a second grid with its north pole on the geographical equator at 90°E. This gives very uniform coverage of the Atlantic and Arctic but the ratio of the minimum to the maximum longitude spacing near the Antarctic is again not much smaller than 4. (Also the grid-spacing perpendicular to the geographical equator in the 2 grids cannot match up exactly.) Murray (1996) reviews the projections discussed above and suggests some others.

The Pacific ocean occupies almost an entire hemisphere. The stereographic projection is generally recommended as the "best" conformal mapping of a hemisphere onto a plane. The ratio of the maximum to minimum grid resolution of this projection is 2. Thus this value appears to be a lower limit on the variation in the grid spacing of ocean points of any set of orthogonal coordinates representing the Earth on a plane surface. More uniform mappings of the sphere can only be achieved by projecting it onto other surfaces, like the surfaces of a cube centred at the centre of the sphere (Purser 1999a, Ronchi et al. 1996).

Semi-implicit methods enable the model's timestep to be much less sensitive to the minimum grid spacing than explicit methods. Thus the incentive to find uniform meshes is weakened if semi-implicit methods are used. Different grid resolutions parallel and perpendicular to the equator have been widely used in ocean models for seasonal forecasting, so uniform meshes have not found universal favour even with explicit methods. Some incentive to control the uniformity of the grid remains in that enhanced resolution in areas where the solution does not require it degrades computational efficiency. A simple estimate of the number of gridpoints required by the grids considered in the first paragraph relative to the number required by a uniform distribution with the same gridspacing at the equator can be obtained as follows. For these grids with displaced poles the latitude range where there are sea points is approximately $\pm 5\pi/12$ (i.e. 75°S to 75°N) and the density of gridpoints in a latitude-longitude interval is proportional to $\cos^{-n}\phi$ where ϕ is the "latitude" and $n=0$ for a uniform distribution, $n=1$ for a standard lat-long distribution and $n=2$ for a lat-long grid with Mercator latitude spacing (and isotropic grid spacing). Thus the relative number of gridpoints, G_n, for these grids is:

$$G_n = \int_{-5\pi/12}^{5\pi/12} (\cos\phi)^{-n} \cos\phi \, d\phi \quad and$$
$$G_0 = 2\sin(5\pi/12) \simeq 1{\cdot}93 \quad ; \quad G_1 = 5\pi/6 \simeq 2{\cdot}6 \quad ; \quad G_2 = 2\ln\tan(11\pi/24) \simeq 4 \quad . \tag{82}$$

About 25% fewer points are required by a truly uniform grid spacing than the normal lat/long grid and the Mercator grid spacing requires 50% more points than a normal lat/long grid.

4. Summary and Conclusions

4.1 Summary of new arguments and results

There appears to be very little discussion in the literature about the timestep which semi-implicit or semi-implicit, semi-Lagrangian ocean models could use without significant loss of accuracy. Explicit calculations of the expected range of eigenvalues of the elliptic operator for the internal modes, though straightforward, also seem to be lacking. Thus much of sections 2.2 and 2.3 must be viewed as uncorroborated.

The relationship between dissipation of available potential energy and diapycnal transports by horizontal diffusion appears not to have been noted previously. The estimate of the ratio of diapycnal transports by horizontal and vertical advection when semi-Lagrangian or upwind flux transport schemes are used also appears to be a new result. The approximations used in these derivations need to be assessed more carefully.

The points made in the discussion of non-hydrostatic and compressible motions are probably implicit in Tanguay et al. (1990) or Marshall et al. (1997) but are stated more explicitly here.

No new points are made about horizontal grid staggering. The analogue for the quasi-geostrophic form of the conservation of potential vorticity on the "modified" Lorenz grid does not appear to have been noticed previously. The Charney-Phillips grid is likely to be superior to the Lorenz grid in a non-hydrostatic model, but in a hydrostatic model its advantages appear to be marginal. The calculation in section 3.3 (c) of the error in Lin's formula when applied to a fluid with horizontal isopycnals is new. The discussion of the torque on the grid-cells used with generalised vertical coordinates can hardly be described as new but may help to dispel some misconceptions.

The onset of viscous "overstability" in the barotropic streamfunction was shown by Killworth (1987) to be precipitated by steep bathymetric gradients. Appendix A gives reasons for thinking that this problem was dependent on the details of the discretisation used. Appendix B gives a more concise presentation than has previously appeared of the numerical discretisation used in the Griffies et al. (1998) isopycnal diffusion.

4.2 Conclusions

The U.S. Navy National Research Laboratory (NRL) routinely use high resolution semi-implicit ocean models in which the timestep is limited by the Courant number for advection by the ocean currents (e.g. Hurlburt & Metzger 1998). For global models of 1/3° or better horizontal resolution, and limited area models of 10 km resolution or better, semi-implicit methods would certainly enable timesteps at least 3-4 times longer than those employed in ocean models using explicit schemes. Whether models of coarser resolution would benefit similarly is not quite clear as other considerations might limit the timestep. Models with finer north-south than east-west resolution near the equator would also benefit from semi-implicit techniques. A further factor of 3 increase in timestep could probably be gained for very high resolution forecasting of mesoscale fronts (and eddies) by using SISL methods.

It appears to be relatively easy to pre-condition the solution of the internal modes to enable them to be solved sufficiently cheaply. When semi-implicit methods are used it appears that non-

hydrostatic and even compressible equations could be used for ocean modelling without significant computational penalties.

Grid scale closure has been discussed in detail because it is relevant to the maintenance of the thermocline in present climate integrations and also because for computational efficiency and accuracy it is desirable to use 2 timestep integration schemes either with semi-Lagrangian or upwind flux formulations of advection but the results of Pardaens and Gordon (private communication) suggest that this might result in a diapycnal flux of heat from the surface into the thermocline which would be unacceptable for climate simulations. The diapycnal transports by vertical velocities and convective overturning and the vertical resolution required by climate simulations all appear to be worth further study.

Although there is a shared body of knowledge relating to the choice of staggering of variables in the horizontal, the weighting given to the evidence varies widely between modellers. For low resolution models (e.g. 1° grid spacing) the B grid is to be preferred to the C grid, but for high resolution models (e.g. 10 km resolution) the C grid appears to be marginally preferable. To obtain the best solutions for a given computational cost the option of calculating at least some terms to higher order accuracy should definitely be kept in mind.

The errors obtained using terrain-following vertical coordinates over steep bathymetry appear too large for them to be competitive in simulations with horizontal grid spacing coarser than 20 km. In climate simulations the smoothing of bathymetry required at this resolution would still be undesirable because of its impact on the pathways by which tracers can be advected (Roberts & Wood 1997). Terrain-following coordinates are used very extensively for coastal modelling. For models including the shelf-break it is desirable to concentrate the model resolution across the break. Generalised orthogonal horizontal coordinates enable one to do this and are used for this purpose in some applications of the Princeton Ocean Model (POM). They also enable one to prevent the singularities (i.e. poles) of the coordinates from lying close to sea points in global models.

Acknowledgements

Andy White and Andrew Staniforth provided very constructive criticisms of and corrections to the first draft of this report.

Appendices

Appendix A: Dependence of a viscous instability on the calculation of reciprocal depth

Killworth (1987) will be referred to as K87 in this appendix.

K87 provides an explanation for an instability in the barotropic streamfunction which is driven by viscous dissipation and arises only over steep bathymetry. He performs a partial stability analysis for the special case when the bathymetry is of depth H_1 for all velocity points $j < J$ and of depth H_2 for all points $j \geq J$. He considers a disturbance to the barotropic streamfunction, ψ, which is periodic in x and **localised** close to the bathymetric step in y and looks for growing mode solutions of the form

$$\psi^n_{i,j} = A^n \phi_j \exp(ikx_i) \quad , \tag{A.1}$$

where n is the timestep (K87 uses ζ to denote the amplitude A).

The evolution of the streamfunction is taken to be determined by the depth averaged vorticity equation driven solely by viscous dissipation:

$$\delta_x[\frac{1}{H}\delta_x\delta_t\psi] + \delta_y[\frac{1}{H}\delta_y\delta_t\psi] = \delta_x G^v - \delta_y G^u \quad , \tag{A.2}$$

where δ_x, δ_y are the finite difference versions of the partial derivatives with respect to x and y usual on a B grid, δ_t , is calculated by the leapfrog scheme and G^u and G^v are the viscous dissipation terms:

$$G^u = \nu(\delta_x^2 + \delta_y^2)u \quad and \quad G^v = \nu(\delta_x^2 + \delta_y^2)v \quad , \tag{A.3}$$

which are calculated on the backward timestep.

The main point of this appendix is that the averaging of $1/H$ used in (A.2) is crucial to the instability. Midway between points with depths H_1 and H_2, K87 uses

$$1/H = 2\,(H_1 + H_2)^{-1} \quad . \tag{A.4}$$

An alternative advocated in the MOM manual by Charles Goldberg (Pacanowski et al. 1999) to increase consistency with the Coriolis terms, is to use

$$1/H = ½(1/H_1 + 1/H_2) \quad . \tag{A.5}$$

K87 assumes that ϕ_j is zero except at the bathymetric step (i.e. $\phi_j = \delta_{jJ}$) and calculates (A.2) for $j = J$ to obtain an approximate stability condition. For the 2nd term on the l.h.s. of (A.2) he obtains

$$[\delta_y(\frac{1}{H}\delta_y\delta_t\psi)]^n_{i,J} = -\frac{(A^2-1)}{A}(1/H_1+1/H_2)\frac{1-s^2}{\Delta y^2}\psi^n_{i,J} \quad , \tag{A.6}$$

where

$$s \equiv \sin ½k\Delta x \quad . \tag{A.7}$$

The first term on the l.h.s. of (A.2) depends on whether (A.4) or (A.5) is used to calculate $1/H$ between points at differing depths. For both cases it may be expressed as:

$$[\delta_x(\frac{1}{H}\delta_x\delta_t\psi)]^n_{i,J} = -\frac{(A^2-1)}{A}(1/H)\frac{2s^2}{\Delta x^2}\psi^n_{i,J} \quad , \tag{A.8}$$

where $1/H$ is given by (A.4) or (A.5) depending on which formulation is used. The term on the r.h.s. of (A.2) for $j = J$ is given by:

$$[\delta_x G^v - \delta_y G^u]^n_{i,J} = \nu\frac{\psi^n_{i,J}}{A}\left\{\frac{1}{H_1}+\frac{1}{H_2}\right\}\left\{\frac{4s^4}{\Delta x^4}+\frac{s^2}{\Delta x^2\Delta y^2}+(1-s^2)\left(\frac{4s^4}{\Delta x^2\Delta y^2}+\frac{3}{\Delta y^4}\right)\right\} . \tag{A.9}$$

The stability condition is obtained by substituing (A.6), (A.8) and (A.9) into (A.2). When (A.5) is used in (A.8) all the terms in (A.2) are directly proportional to $1/H_1+1/H_2$ and it is clear that the stability is independent of H_1 and H_2. When (A.4) is used in (A.8) the stability does depend on the ratio H_1/H_2 and as K87 shows stability requires

$$\nu\Delta t/\Delta x^2 < 2H_1/H_2 \quad , \tag{A.10}$$

which, when the ratio H_1/H_2 is very small, is much more restrictive than the usual condition to avoid diffusive instability.

Appendix B: Discretisation of isopycnal diffusion used by Griffies et al. (1998)

In this appendix Griffies et al. (1998) will be referred to as G and (G.N) will refer to equation N in that paper. The continuous functional used in the Griffies scheme for the isopycnal flux with the small-slope approximation is given by (G.25) which is:

$$\mathcal{J} = -½\int A_I\{(T_x+S^xT_z)^2+(T_y+S^yT_z)^2\}d\mathbf{x} \quad . \tag{B.1}$$

Here $S^x = -\rho_x/\rho_z$ and $S^y = -\rho_y/\rho_z$ denote the isopycnal slopes. To aid clarity of presentation, this derivation will limit consideration to the x-z plane and hence the first term in (B.1). G introduce a discrete analogue of this functional in which each term $T_x+S^xT_z$ is calculated using 3 points; namely a central point $T_{i,k}$, a point to its east or west $T_{i\pm1,k}$ and a point above or below it,

$T_{i,k\pm1}$. The discrete functional consists of the sum over all points in the domain of four terms at each point, one for each of the four possible combinations. Denoting the term for the upper-right quadrant with the subscript 1 – see figure B – the first of the four terms in the functional is:

$$\mathfrak{I}_1 = \sum_{i,k} \mathfrak{I}_{1;i,k} = -\tfrac{1}{2} \sum_{i,k} A_{1;i,k} \left(\frac{T_{i+1,k} - T_{i,k}}{2\Delta x_{1;i,k}} + S^x_{1;i,k} \frac{T_{i,k+1} - T_{i,k}}{2\Delta z_{1;i,k}} \right)^2 \Delta x_{1,i,k} \Delta z_{1;i,k} \ . \qquad \text{(B.2)}$$

Here $\Delta x_{1;i,k}$ and $\Delta z_{1;i,k}$ are the dimensions of the upper-right quadrant's cell whose edges bisect the tracer points (see figure B), and $A_{1;i,k}$ and $S^x_{1;i,k}$ are calculated specifically for this cell (see G for details).

The aim of this appendix is to express the derivative of $\mathfrak{I}$ with respect to $T_{i,k}$ as the sum of a flux $\mathbf{F} = (F_x, F_z)$ through the faces (indicated by dashed lines in figure B) of the $T_{i,k}$ tracer cell:

$$-\partial \mathfrak{I} / \partial T_{i,k} = (F_{i+\frac{1}{2},k} - F_{i-\frac{1}{2},k}) \Delta z_{i,k} + (F_{i,k+\frac{1}{2}} - F_{i,k-\frac{1}{2}}) \Delta x_{i,k} \quad , \qquad \text{(B.3)}$$

and to show that $\mathbf{F}$ is given by equations (G.30) and (G.33).

Consider first the contribution to $\partial \mathfrak{I} / \partial T_{i,k}$ from $\mathfrak{I}_{1;i,k}$ (i.e. the contribution to $\mathfrak{I}_1$ from the (i,k)th box). From (B.2)

$$\partial \mathfrak{I}_{1;i,k} / \partial T_{i,k} = A_{1;i,k} \left(\frac{T_{i+1,k} - T_{i,k}}{2\Delta x_{1;i,k}} + S^x_{1;i,k} \frac{T_{i,k+1} - T_{i,k}}{2\Delta z_{1;i,k}} \right) \left(\frac{\Delta z_{1;i,k}}{2} + \frac{S^x_{1;i,k} \Delta x_{1;i,k}}{2} \right) \ . \qquad \text{(B.4)}$$

Comparison of (B.3) and (B.4) suggests that the term proportional to $\Delta z_{1;i,k}$ be viewed as a contribution to $F_{i+\frac{1}{2},k}$ whilst the term proportional to $S^x_{1;i,k} \Delta x_{1;i,k}$ be viewed as a contribution to $F_{i,k+\frac{1}{2}}$. Comparison of (B.3) and (B.4) with (G.30) shows that the term proportional to $\Delta z_{1;i,k}$ is one of the four to contribute to the eastern face flux $F_{i+\frac{1}{2},k}$. Similarly the term proportional to $S^x_{1;i,k} \Delta x_{1;i,k}$ is one of the four contributions to $F_{i,k+\frac{1}{2}}$ given in (G.33).

The previous paragraph showed that $\partial \mathfrak{I}_{1;i,k} / \partial T_{i,k}$ provides one of the four contributions to both $F_{i+\frac{1}{2},k} \Delta z_{i,k}$ and $F_{i,k+\frac{1}{2}} \Delta x_{i,k}$. Similarly the contribution to $\mathfrak{I}$ from the lower right quadrant (see figure B.1) is given by

$$\mathfrak{I}_2 = \sum_{i,k} \mathfrak{I}_{2;i,k} = -\tfrac{1}{2} \sum_{i,k} A_{2;i,k} \left(\frac{T_{i+1,k} - T_{i,k}}{2\Delta x_{2;i,k}} + S^x_{2;i,k} \frac{T_{i,k-1} - T_{i,k}}{2\Delta z_{2;i,k}} \right)^2 \Delta x_{2;i,k} \Delta z_{2;i,k} \quad , \qquad \text{(B.5)}$$

and $\partial \mathfrak{I}_{2;i,k} / \partial T_{i,k}$ provides one of the contributions to each of $F_{i+\frac{1}{2},k} \Delta z_{i,k}$ and $F_{i,k-\frac{1}{2}} \Delta x_{i,k}$. $\partial \mathfrak{I}_{3;i,k} / \partial T_{i,k}$ and $\partial \mathfrak{I}_{4;i,k} / \partial T_{i,k}$ each provide 2 further contributions so that in total we have now accounted for 2 contributions to (B.3) for each of the 4 faces.

(B.4) gives the contribution to $\partial \Im_1 / \partial T_{i,k}$ from the (i,k)th box. There are two other contributions to $\partial \Im_1 / \partial T_{i,k}$: one from the $(i+1,k)$ th box (i.e. from $\Im_{1;i+1,k}$) and one from the $(i,k+1)$ th box (i.e. from $\Im_{1;i,k+1}$). For each contribution to $\partial \Im / \partial T_{i,k}$ from the (i,k)th box there is a second contribution to the flux through the same face from an adjacent box. The four contributions to the flux through a face are the ones indicated by figure 9 of G and the sign of each contribution is in agreement with (G.30) and (G.33).

The flux through the west face of the $(i+1,k)$ th box includes a contribution from

$$\partial \Im_{1;i,k} / \partial T_{i+1,k} = -A_{1;i,k} \left(\frac{T_{i+1,k} - T_{i,k}}{2\Delta x_{1;i,k}} + S^x_{1;i,k} \frac{T_{i,k+1} - T_{i,k}}{2\Delta z_{1;i,k}} \right) \frac{\Delta z_{1;i,k}}{2} \quad . \tag{B.6}$$

This is **identical but opposite** to the corresponding contribution to the flux through the east face from the (i,k)th box. Thus the functional derivative does represent a true conservative flux.

For isopycnal diffusion, the choice of triads in the calculation of each term together with the fact that all terms in the sum comprising the functional are positive ensures that grid scale noise will not lie in the kernel of ("be invisible to") the functional. For the skew-flux formulation only terms linearly proportional to $S^x_{m;i,k}$ contribute (here $1 \le m \le 4$). $S^x_{m;i,k}$ is calculated using the same triad of points as for the temperatures. If the potential density is taken to be a local linear function of θ and S and the model is stably stratified, all the terms contributing to the flux of **density** at any given face will have the same sign. Thus the skew-flux formulation will also have an empty kernel, at least for the potential density field.

Appendix C: Properties of the Lorenz and Charney-Phillips grids

This appendix provides algebra to support the assertions made in section 3.2.

(a) Continuity equation: advective and flux forms

On the Lorenz grids the flux form of the continuity equation and the incompressibility condition are:

$$\rho_{kt} + \nabla.(\mathbf{u}_k \rho_k) + \frac{w_{k+½}(\rho_{k+1} + \rho_k)}{2\Delta z} - \frac{w_{k-½}(\rho_k + \rho_{k-1})}{2\Delta z} = 0 \quad , \tag{C.1}$$

$$\nabla.\mathbf{u}_k + \left(w_{k+½} - w_{k-½} \right) / \Delta z = 0 \quad . \tag{C.2}$$

Subtracting ρ_k times (C.2) from (C.1) one finds the advective form of the continuity equation to be:

$$\rho_{kt} + \mathbf{u}_k.\nabla \rho_k + \frac{w_{k+½}(\rho_{k+1} - \rho_k)}{2\Delta z} + \frac{w_{k-½}(\rho_k - \rho_{k-1})}{2\Delta z} = 0 \quad , \tag{C.3}$$

as anticipated in (72).

On the C-P grid, (C.2) still holds but the natural form for the vertical component of the flux of ρ, F^v, and the desirable form for the vertical component of the advective flux of ρ, A^v, are:

$$F^v_{k+½} = \frac{(w_{k+3/2} + w_{k+½})(\rho_{k+3/2} + \rho_{k+½})}{4\Delta z} - \frac{(w_{k+½} + w_{k-½})(\rho_{k+½} + \rho_{k-½})}{4\Delta z} \quad , \tag{C.4}$$

$$A^v_{k+½} = \frac{w_{k+½}(\rho_{k+3/2} - \rho_{k-½})}{2\Delta z} \quad . \tag{C.5}$$

Subtracting (C.5) from (C.4) and re-arranging one finds that :

$$F^v_{k+½} - A^v_{k+½} = \frac{(w_{k+3/2} - w_{k+½})(\rho_{k+3/2} + \rho_{k+½})}{4\Delta z} + \frac{(w_{k+½} - w_{k-½})(\rho_{k+½} + \rho_{k-½})}{4\Delta z} \quad . \tag{C.6}$$

Thus the simplest advective form for the continuity equation on the C-P grid consistent with the flux formulation (C.4) and (C.2) appears to be:

$$\rho_{k+½t} + \frac{\mathbf{u}_k}{2}.\nabla\left(\frac{\rho_{k+½} + \rho_{k-½}}{2}\right) + \frac{\mathbf{u}_{k+1}}{2}.\nabla\left(\frac{\rho_{k+3/2} + \rho_{k+½}}{2}\right) + \frac{w_{k+½}(\rho_{k+3/2} - \rho_{k+½})}{2\Delta z} = 0 \quad . \tag{C.7}$$

This is the averaging used by Arakawa and Konor (1996) in their equation (3.6). It is to be contrasted with the more desirable form that is usually assumed, namely

$$\rho_{k+½t} + \frac{\mathbf{u}_k + \mathbf{u}_{k+1}}{2}.\nabla\rho_{k+½} + \frac{w_{k+½}(\rho_{k+3/2} - \rho_{k+½})}{2\Delta z} = 0 \quad . \tag{C.8}$$

(b) Potential vorticity conservation on the Lorenz grid

AM show that the C-P grid has an analogue of the quasi-geostrophic form of "generalised" potential vorticity conservation when (C.8) is the advective form used for the continuity equation. The principal approximations of quasi-geostrophic theory used in the derivation are:

that the vertical advection of potential density is approximated by vertical advection of a "background" stratified state, $\rho_0(z) \equiv \Pi(z)$, which depends only on z and that horizontal advection by the ageostrophic flow is neglected so that $(\mathbf{u}.\nabla)\varphi = J(\psi, \varphi)$ for any field φ where $p = \Pi(0) f_0 \psi$. (Π has been introduced here simply to avoid confusion of subscripts.) The conservation equation (on the C-P grid) is of the form:

$$q_{kt} + \nabla.(\mathbf{u}_k q_k) = 0 \quad for\ k = 1, 2, ...K \tag{C.9}$$

where K is the number of interior full levels and

$$q_k = \nabla^2_h \psi_k + f + \frac{f_0^2}{\Delta z_k^2}\left[\frac{\psi_{k+1} - \psi_k}{N^2_{k+½}} - \frac{\psi_k - \psi_{k-1}}{N^2_{k-½}}\right] \quad for\ k = 2, 3, ..K-1 \ , \tag{C.10}$$

$$N^2_{k+½} = g(\Pi_{k-½} - \Pi_{k+3/2})/(2\Pi_{k+½}\Delta z_k) \ . \qquad \text{(C.11)}$$

See AM equations (4.6) and (4.8) for expressions for q_1 and q_K.

AM derive an expression for the potential vorticity equation on the Lorenz grid which can be written in the form:

$$a_k \frac{D_k}{Dt}(\nabla^2_h \psi_k + f) + b_{k+1}\frac{D_{k+1}}{Dt}(\nabla^2_h \psi_{k+1} + f) = \frac{2f_0}{\Delta z_k}\left\{ c_k \frac{D_k \rho_k}{Dt} - c_{k+1}\frac{D_{k+1}\rho_{k+1}}{Dt} \right\} \quad for\ k = 2,..K-1 \ , \qquad \text{(C.12)}$$

where

$$D_k\varphi / Dt \equiv \varphi_t + J(\psi_k, \varphi) \quad , \qquad \text{(C.13)}$$

$$a_k = \frac{\Pi_{k-1} - \Pi_k}{\Pi_{k-1} - \Pi_{k+1}} \quad ; \quad b_{k+1} = \frac{\Delta z_{k+1}}{\Delta z_k}\frac{\Pi_{k+1} - \Pi_{k+2}}{\Pi_k - \Pi_{k+2}} \quad ; \quad c_k = \frac{\Delta z_k}{\Pi_{k-1} - \Pi_{k+1}} \ . \qquad \text{(C.14)}$$

See AM equations (5.17) - (5.21) for expressions for the equations for levels 1,2, K-1 and K.

The l.h.s. of (C.12) involves a weighted vertical average over two layers of the horizontal advection of the vertical component of the vorticity whereas (C.9) and (C.10) involve advection of the vorticity at a single level. This difference between the Lorenz and C-P grids derives from the difference between the continuity equations (C.3) and (C.8). The potential voticity equation is derived from the vorticity equation,

$$\frac{D}{Dt}\left(\nabla^2_h \psi + f\right) - f w_z = 0 \ , \qquad \text{(C.15)}$$

and these continuity equations. On the C-P grid the vertical velocities can be eliminated directly whereas on the Lorenz grid the vertical velocities can only be eliminated by using a unique combination (weighted average) of the vorticity equation at two levels and (C.3). As AM remark, the potential vorticity on the Lorenz grid should be considered to lie at half levels rather than full levels. Recognising this point one easily appreciates the need for some form of vertical averaging of the advection of vorticity in the potential vorticity equation. The form in (C.12) is not ideal (as noted when discussing (C.7)) but is acceptable.

A more important issue is whether the r.h.s. of (C.12), the vortex stretching term, can be re-organised into an advective form. It is easiest to show this using the pressures of the Modified Lorenz grid which lie at the model half levels and are related to ψ, p and ρ on the full levels of the Lorenz grid by:

$$p_k = \Pi(0) f_0 \psi_k = \frac{p_{k+½} + p_{k-½}}{2} \quad ; \quad g\rho_k \Delta z_k = p_{k-½} - p_{k+½} \quad . \qquad \text{(C.16)}$$

Using (C.16)

$$gJ(\psi_k,\rho_k)\Delta z_k = J\left(\frac{\psi_{k+½}+\psi_{k-½}}{2}, p_{k-½}-p_{k+½}\right) = J(\psi_{k+½}, p_{k-½}) = gJ(\psi_{k+½},\rho_k)\Delta z_k \quad . \qquad \text{(C.17)}$$

Similarly one can show that

$$J(\psi_{k+1},\rho_{k+1}) = J(\psi_{k+½},\rho_{k+1}) \quad . \qquad \text{(C.18)}$$

Thus the r.h.s. of (C.14) can be re-expressed as:

$$\begin{aligned}\frac{2f_0}{\Delta z_k}\left\{c_{k+1}\frac{D_{k+1}\rho_{k+1}}{Dt} - c_k\frac{D_k\rho_k}{Dt}\right\} &= -\frac{2f_0}{\Delta z_k}\frac{D_{k+½}}{Dt}\left\{\frac{\Delta z_{k+1}\rho_{k+1}}{\Pi_k-\Pi_{k+2}} - \frac{\Delta z_k\rho_k}{\Pi_{k-1}-\Pi_{k+1}}\right\} \\ &= -\frac{f_0^2}{\Delta z_k}\frac{D_{k+½}}{Dt}\left\{\frac{\psi_{k+3/2}-\psi_{k+½}}{N^2_{k+1}\Delta z_{k+1}} - \frac{\psi_{k+½}-\psi_{k-½}}{N^2_k\Delta z_k}\right\}\end{aligned} \qquad \text{(C.19)}$$

in which

$$N_k^2 = g(\Pi_{k-1}-\Pi_{k+1})/(2\Pi_0\Delta z_k) \quad . \qquad \text{(C.20)}$$

(c) The Computational Mode of the Lorenz grid

The vertical normal mode problem with rigid lid boundary conditions involves only the interior vertical velocities. On both the C-P and Modified Lorenz grids these are stored at the half-levels. So the number of non-zero vertical normal modes equals the number of internal half-levels on both grids. On the C-P grid the density is also stored at the half-levels so one can expect any density field, and hence any initial conditions for density, to be expressible in terms of the density modes paired with the vertical velocity modes. On the Lorenz grids the densities are stored at full levels so there is an additional level at which the density is stored. It is clear that none of the non-zero vertical velocity modes has an accompanying density structure which varies rapidly enough in the vertical to represent a density field which changes sign with grid level. Thus there must be an additional normal mode with this density structure whose vertical velocity is zero at all levels. From (64) it is clear that this mode is stationary. Some care is needed to determine the rest of the vertical structure; for example if both (69) and (70) were applied one would conclude that the mode had $p_k = \rho_{k+½} = 0$. Using (66) with suitable homogeneous boundary conditions for p_k one could conclude that the p_k are zero, but this cannot be established directly from (65). The usual argument is that the mode has $p_k = u_k = v_k = 0$ as this is clearly a solution. Any initial density field should be expressible using this mode together with the standard normal modes.

References

Adcroft, A., and D. Marshall, 1997: How slippery are piecewise-constant coastlines in numerical ocean models ? *Tellus*, **50A**, 95-108.

Adcroft, A. J., C. N. Hill, and J. C. Marshall, 1999: A new treatment of the Coriolis terms in C-grid models at both high and low resolutions. *Mon. Wea. Rev.*, **127**, 1928-1936.

Arakawa, A., 1988: Finite-difference methods in climate modelling. In M. E. Schlesinger (ed.), *Physically-based modelling and simulation of climate and climatic change* – Part I, 79-168. Kluwer Academic Publishers.

Arakawa, A., and M. J. Suarez, 1983: Vertical differencing of the primitive equations in sigma coordiantes. *Mon. Wea. Rev.*, **111**, 34-45.

Arakawa, A., and S. Moorthi, 1988: Baroclinic instability in vertically discrete systems. *J. Atmos. Sci.*, **45**, 1688-1707.

Arakawa, A., and C. S. Konor, 1996: Vertical differencing of the primitive equations based on the Charney-Phillips grid in hybrid σ-p vertical coordinates. *Mon. Wea. Rev.* , **124**, 511-528.

Bentsen, M., G. Evensen, H. Drange, and A. D. Jenkins, 1999: Coordinate transformation on a sphere using a conformal mapping. *Mon. Wea. Rev.*, **127**, 2733-2740.

Brown, J. A. and K. A. Campana, 1978: An economical time-differencing system for numerical weather prediction. *Mon. Wea. Rev.*, **106**, 1125-1136.

Burridge, D. M., 1975: A split semi-implicit reformulation of the Bushby-Timpson 10 level model. *Quart. J. Roy. Meteor. Soc.*, **101**, 777-792.

Charney, J. G., 1971: Geostrophic turbulence. *J. Atmos. Sci.*, **28**, 1087-1095.

Chu, P. C., and C. Fan, 1998: High-order compact difference schemes for coastal ocean models. Second conference on coastal atmospheric and oceanic prediction and processes, 11-16 January, 1998, 68-75, American Meteor. Soc.

Coward, A. C., P. D. Killworth, and J. R. Blundell, 1994: Test of a two-grid world ocean model. *J. Geophys. Res.*, **99**, 22725-22735.

Cox, M. D., 1987: Isopycnal diffusion in a z-coordinate ocean model. *Ocean Modelling*, **74**, 1-5.

Dukowicz, J. K., 1995: Mesh effects for Rossby waves. *J. Comp. Phys.*, **119**, 188-194.

Dukowicz, J. K., and R. D. Smith, 1994: Implicit free-surface method for the Bryan-Cox-Semtner ocean model. *J. Geophys. Res.*, **99**, C4, 7991-8014.

Ezer, T., and G. L. Mellor, 1994. Diagnostic and prognostic calculations of the North Atlantic circulation and sea level using a sigma coordinate ocean model. *J. Geophys. Res.*, **99**, C7, 14,159-14,171.

Garrett, C., P. MacCready, and P. Rhines, 1993: Boundary layer mixing and arrested Ekman layers: rotating stratfied flow near a sloping boundary. *Ann. Rev. Fluid Mech.*, **25**, 291-323.

Gent, P. R., and J. C. McWilliams, 1990: Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.*, **20**, 150-155.

Gent, P. R., J. Willebrand, T. J. McDougall, and J. C. McWilliams, 1995: Parametrizing eddy-induced tracer transports in ocean circulation models. *J. Phys. Oceanogr.*, **25**, 463-474.

Gill, A. E., 1982: *Atmosphere-ocean dynamics*. International Geophysics Series, vol 30, Academic Press. 662 pp.

Golub, G. H., and C. F. van Loan, 1986: *Matrix Computations*, 2nd edition, North Oxford Academic, Oxford.

Gregg, M. C., D. P. Winkel, T. B. Sanford, and H. Peters, 1996: Turbulence produced by internal waves in the oceanic thermocline at mid and low latitudes. *Dyn. Atmos. Oceans*, **24**, 1-14.

Griffies, S. M., 1998: The Gent-McWilliams skew flux. *J. Phys. Oceanogr.*, **28**, 831-841.

Griffies, S. M., A. Gnanadesikan, R. C. Pacanowski, V. D. Larichev, J. K. Dukowicz, and R. D. Smith, 1998: Isoneutral diffusion in a z-coordinate ocean model. *J. Phys. Oceanogr.*, **28**, 805-830.

Griffies, S. M., R. C. Pacanowski, and R. W. Hallberg, 2000: Spurious diapycnal mixing asssociated with advection in a z-coordinate ocean model. *Mon. Wea. Rev.*, **128**, 538-564.

Guyon, M., G. Madec, F-X Roux, C. Herbaut, M. Imbard, and P. Fraunie, 1999 Domain decomposition as a nutshell for massively parallel ocean modelling with the OPA model. *IPSL Notes du Pôle de modélisation*, **12**, January 1999, LOYDC, Paris.

Held, I. M., 1978: The vertical scale of an unstable baroclinic wave and its importance for eddy heat flux parametrizations. *J. Atmos. Sci.*, **35**, 572-576.

Hortal, M., 1999: Aspects of the numerics of the ECMWF model. *ECMWF Seminar Proceedings on Recent developments in numerical methods for atmospheric modelling*, 127-143, 7-11 September 1998. July 1999.

Hsieh, W. W., M. K. Davey, and R. C. Wajsowicz, 1983: The free Kelvin waves in finite-difference numerical models. *J. Phys. Oceanogr.*, **13**, 1383-1397.

Hurlburt, H. E., and J. D. Thompson, 1980: A numerical study of loop current intrusions and eddy shedding. *J. Phys. Oceanogr.*, **10**, 1611-1651.

Hurlburt, H. E., and E. J. Metzger, 1998. Bifurcation of the Kuroshio extension at the Shatsky Rise. *J. Geophys. Res.*, 103, C4, 7549-7566.

Kasahara, A., 1974: Various vertical coordinate systems used for numerical weather prediction. *Mon. Wea. Rev.*, **102**, 509-522.

Killworth, P. D., 1987: Topographic instabilities in level model OGCMs. *Ocean Modelling*, **75**, 9-12.

Kreyszig, E., 1979: *Advanced Engineering Mathematics*. Wiley, New York, 939 pp.

Le Roux, D., A. Staniforth, and C. A. Lin, 1998: Finite elements for shallow-water equation ocean models. *Mon. Wea. Rev.*, **126**, 1931-1951.

Leonard, B. P., M. K. MacVean, and A. P. Lock, 1993: Positivity preserving numerical schemes for multidimensional advection. *NASA Tech Memo* 106055.

Leonard, B. P., A. P. Lock, and M. K. MacVean, 1996: Conservative explicit unrestricted-time-step multidimensional constancy-preserving advection schemes. *Mon. Wea. Rev.*, **124**, 2588-2606.

Leslie, L. M., and R. J. Purser, 1992: A comparative study of the performance of various vertical discretisation schemes. *Met. Atm. Phys.*, **50**, 61-73.

Lin, S.-J., 1997: A finite volume integration method for computing pressure gradient force in general vertical coordinates. *Quart. J. Roy. Meteor. Soc.*, **123**, 1749-1762.

Madec, G., and M. Imbard, 1996: A global ocean mesh to overcome the north pole singularity. *Climate Dyn.*, **12**, 381-388.

Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey, 1997: A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *J. Geophys. Res.*, **102**, C3, 5753-5766.

Mason, P. J., 1994: Large-eddy simulation: a critical review of the technique. *Quart. J. Roy. Met. Soc.*, **120**, 1-26.

McCalpin, J. D., 1994: A comparison of second-order and fourth-order pressure gradient algorithms in a σ-coordinate ocean model. *Int. J. Num. Meth. Fluids*, **18**, 361-383.

McCalpin, J. D., 1988: A quantitative analysis of the dissipation inherent in semi-Lagrangian advection. *Mon. Wea. Rev.*, **116**, 2330-2336.

McDonald, A., 1984: Accuracy of multiply-upstream, semi-Lagrangian advective schemes. *Mon. Wea. Rev.*, **112**, 1267-1275.

Mellor, G. L., T. Ezer, and L.-Y. Oey, 1994: The pressure gradient conundrum of sigma coordinate ocean models. *J. Atmos. Ocean. Tech.*, **11**, 1126-1134.

Morton, K. W., and D. F. Mayers, 1994 *Numerical solution of partial differential equations.* Cambridge University Press, Cambridge, 227 pp.

Murray, R. J., 1996: Explicit generation of orthogonal grids for ocean models. *J. Comp. Phys.*, **126**, 251-273.

Ng, M. K. F., and W. W. Hsieh, 1994: The equatorial Kelvin wave in finite difference models. *J. Geophys. Res.*, **99**, C7, 14,173-14,185.

Pacanowski, R. C., and S. M. Griffies, 1999: *MOM Manual 3.0.* Draft version. Geophysical Fluid Dynamics Laboratory/NOAA. [Available from GFDL, Princeton University, Princeton, NJ 08542.]

Pedlosky, J., 1982: *Geophysical fluid dynamics.* Springer-Verlag. 624 pp.

Polzin, K. L., K. G. Speer, J. M. Toole, R. W. Schmitt, 1996: Intense mixing of Antarctic bottom water in the equatorial Atlantic Ocean. *Nature*, **380**, 54-57.

Pudykiewicz, J., and A. Staniforth, 1984: Some properties and comparative performance of the semi-Lagrangian method of Robert in the solution of the advection-diffusion equation. *Atmos. Ocean*, **22**, 283-303.

Purser, R. J., 1999a: Non standard grids. *ECMWF Seminar Proceedings on Recent developments in numerical methods for atmospheric modelling*, 44-72, 7-11 September 1998. July 1999.

Purser, R. J., 1999b: Efficient high-order semi-Lagrangian methods. , *ECMWF Seminar Proceedings on Recent developments in numerical methods for atmospheric modelling*, 73-94, 7-11 September 1998. July 1999.

Purser, R. J., and L. M. Leslie, 1988: A semi-implicit, semi-Lagrangian finite-difference scheme using high-order spatial differencing on a nonstaggered grid. *Mon. Wea. Rev*, **116**, 2069-2080.

Randall, D. A., 1994: Geostrophic adjustment and the finite-difference shallow-water equations. *Mon. Wea. Rev.*, **122**, 1371-1377.

Ritchie, H., C. Temperton, A. Simmons, M. Hortal, T. Davies, D. Dent, and M. Hamrud, 1995: Implelementation of the semi-Lagrangian method in a high-resolution version of the ECMWF forecast model. *Mon. Wea. Rev.*, **123**, 489-514.

Ritchie, H., and M. Tanguay, 1996: A comparison of spatially averaged Eulerian and semi-Lagrangian treatment of mountains. *Mon. Wea. Rev.*, **124**, 167-181.

Rivest, C., A. Staniforth, and A. Robert, 1994: Spurious resonant response of semi-Lagrangian discretisations to orographic forcing: diagnosis and solution. *Mon. Wea. Rev.*, **122**, 366-376.

Robert, A., 1981: A stable numerical integration scheme for the primitive meteorological equations. *Atmos. Ocean*, **19**, 35-46.

Roberts, M. J., and R. A. Wood, 1997: Topographic sensitivity studies with a Bryan-Cox type ocean model. *J. Phys. Oceanogr.*, **27**, 823-836.

Roberts, M., and D. Marshall, 1998: Do we require adiabatic dissipation schemes in eddy-resolving ocean models *? J. Phys. Ocean.*, **28**, 2050-2063.

Ronchi, C., R. Iacono, and P. S. Paolucci, 1996: The 'cubed' sphere: A new method for the solution of primitive equations in spherical geometry. *J. Comp. Phys.*, **124**, 93-114.

Smagorinsky, J., 1993: Some historical remarks on the use of nonlinear viscosities. *In Large eddy simulation of complex engineering and geophysical flows*, 3-36, B. Galperin and S. A. Orszag (eds.) Cambridge Univ. Press.

Smith, B., P. Bjorstad, and W. Gropp *1996 Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge Univ. Press.

Smolarkiewicz, P. K. and P. Rasch, 1991: Monotone advection on the sphere: an Eulerian versus a semi-Lagrangian approach. *J. Atmos. Sci.*, **48**, 793-810.

Song, Y. T., 1998: A general pressure gradient formulation for ocean models. Part I: Scheme design and diagnostic analysis. *Mon. Wea. Rev.*, **126**, 3213-3230.

Song, Y. T., and D. G. Wright, 1998: A general pressure gradient formulation for ocean models. Part II: Energy, momentum and bottom torque consistency. *Mon. Wea. Rev.*, **126**, 3231-3247.

Staniforth, A., and J. Côté, 1991: Semi-Lagrangian integration schemes for atmospheric models. *Mon. Wea. Rev.*, **119**, 2206-2223.

Staniforth, A., J. Côté, S. Gravel, A. Methot, A. Patoine, and M. Roch, 1999: Environment Canada's GEM (global environmental multiscale) model. *ECMWF Seminar Proceedings on Recent developments in numerical methods for atmospheric modelling*, 144-171. 7-11 September 1998. July 1999.

Staniforth, A., and J. Côté, 1999: Semi-Lagrangian methods. *ECMWF Seminar Proceedings on Recent developments in numerical methods for atmospheric modelling*, 95-111, 7-11 September 1998. July 1999.

Stelling, G. S., and J. A. T. M. van Kester, 1994: On the approximation of horizontal gradients in sigma co-ordinates for bathymetry with steep slopes. *Int. J. Num. Meth. Fluids*, **18**, 915-935.

Tanguay, M., A. Robert, and R. Laprise, 1990: A semi-implicit semi-Lagrangian fully compressible regional forecast model. *Mon. Wea. Rev.*, **118**, 1970-1980.

Thorpe, S. A., 1988: An overview of the 19th international Liege colloquium on ocean hydrodynamics: "Small-scale turbulence and mixing in the ocean". In *Small-scale turbulence and mixing in the ocean, Proceedings of the 19th international Liege colloquium on ocean hydrodynamics.* J. C. J. Nihoul and B. M. Jamart (eds.) Elsevier Oceanography Series, vol 46, 533-541.

Thuburn, J., 1995: Dissipation and cascades to small scales in numerical models using a shape-preserving advection scheme. *Mon. Wea. Rev.*, **123**, 1888-1903.

Tokioka, T., 1978: Some considerations on vertical differencing. *J. Met. Soc. Jap.*, **56**, 98-111.

Toro, E. F., 1997: *Riemann solvers and numerical methods for fluid dynamics: a practical introduction.* Springer-Verlag.

Wajsowicz, R. C., 1986: Free planetary waves in finite-difference numerical models. *J. Phys. Oceanogr.*, **16**, 773-789.

Webb, D. J., B. A. de Cuevas, and C. S. Richmond, 1998: Improved advection schemes for ocean models. *J. Atm. Ocean. Tech.*, **15**, 1171-1187.