JOURNAL OF COMPUTATIONAL PHYSICS 43, 79–94 (1981)

Solutions of the Shallow Water Equations on a Sphere

RONALD L. GILLILAND

High Altitude Observatory and Advanced Study Program, National Center for Atmospheric Research, P. O. Box 3000, Boulder, Colorado 80307*

Received December 2, 1980

Various numerical solutions of the shallow water equations in two dimensions are studied in an effort to develop a computational technique applicable to hydrodynamics in spherical geometry. The equations are first cast in a form which allows periodic boundary conditions in both angular coordinates. Explicit numerical solutions using leap-frog centered differencing in time and either second, fourth, or compact fourth order centered spatial differencing are studied. The fourth order compact differencing is found to be easily adapted to spherical geometry and is superior to the second order technique. We also consider an alternating-direction implicit (ADI) scheme in an attempt to increase computational efficiency by taking larger time steps. Both analytically steady state and time dependent solutions are examined to investigate stability properties and discretization errors in time and space. Implicit methods require more computation per time step than explicit methods for solution of the shallow water equations. However, the total time for a simulation can be less with the implicit method. The ADI formalism also has advantages of importance for more physically complex problems.

INTRODUCTION

Multidimensional problems in spherical geometry naturally arise in attempting global solutions with a central force field, e.g., the meteorological problem of modeling atmospheric circulation or the astrophysical problem of modeling the solar convection zone and interior. Many numerical methods have been developed for solving multidimensional hydrodynamics problems in Cartesian coordinates, e.g., the Navier–Stokes equations modeling flows over airfoils. Numerical solutions of systems of partial differential equations in spherical geometry have generally been limited to explicit methods. Explicit methods for multidimensional problems in spherical geometry have been used for two reasons: (1) Sound waves are usually the highest frequency waves which are present in the system. The computational burden of an explicit scheme when these waves are present is excessive. Sound waves are usually excluded by some additional assumption, e.g., the hydrostatic approximation. Once these waves are removed an explicit scheme becomes more attractive. (2) The

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

0021-9991/81/090079–16$02.00/0

computational and programming efforts required for implicit calculations have generally been much greater than the effort required for explicit solutions.

The noniterative Douglas–Gunn [1] formalism for generating alternating-direction implicit schemes, as developed by Briley and McDonald [2], is an efficient implicit method which is straightforward to program. The unconditional stability associated with the implicit technique allows longer time steps which offset the increased effort per time step relative to explicit methods. The use of compact, Padé (first and second order spatial derivatives are approximated using only three adjacent points) fourth order differences [3, 4] allows a significant reduction in the number of grid points to obtain a specified accuracy.

In the astrophysical context the model equations are often solved over a domain which includes a very broad range of physical properties. For example, a global model of the sun must contend with sound speeds of greater than $500\,\mathrm{km\,sec^{-1}}$ near the core and less than $10\,\mathrm{km\,sec^{-1}}$ near the surface. The simultaneous presence of such discrepant conditions in a single model leads to the necessity of very short time steps for an explicit approximation if the core is adequately resolved spatially. In a reasonably zoned multidimensional model of the sun the Courant–Friedrichs–Lewy (CFL) condition ($\Delta t \leqslant \Delta x/(|u| + c)$, where Δx is the mesh size, u is the flow velocity, and c is the sound speed) restricts explicit time steps to $\Delta t < 1$ sec. Since the time scales on which model changes of interest occur are many orders of magnitude greater than 1 sec the CFL coondition makes realistic computer simulations impractical. With an unconditionally stable ADI scheme the unreasonable time step restriction of an explicit scheme is avoided. Furthermore the ADI formalism allows for a straightforward inclusion of complex auxiliary physics which would be quite cumbersome in an explicit algorithm.

An alternative approach to circumventing the severe CFL restriction due to acoustic waves is to filter out the sound waves by dropping the time derivative of density in the continuity equation. This anelastic approximation allows an explicit solution of the model equations in which sound waves have been suppressed. However, this approximation is valid only for cases in which the resulting flow velocities remain small compared to the sound speed. In the outer layers of the solar convection zone the convective velocities approach the sound speed, thus invalidating use of the anelastic approximation. Thus to model the outer, directly observable, convective layers of the sun a more general approach which is capable of following advective flows with Mach numbers approaching unity is desirable.

As a first step in the development of a three-dimensional code for compressible hydrodynamics (e.g., as required for detailed study of the solar convection zone) various two-dimensional numerical solutions of the shallow water equations on a sphere are studied. Explicit time differencing with second order spatial accuracy is developed as a base case. The technique of primary interest is a spatially fourth order approximation over the two angular coordinates on a sphere with second order implicit temporal accuracy.

Shallow Water Equations

Throughout this paper spherical polar coordinates with $\theta =$ colatitude, $\varphi =$ longitude are used as a basis for the differential and finite difference forms of the governing equations. The primitive barotropic model equations written in advection form are (e.g., see [5])

$$\frac{\partial u}{\partial t} = \frac{-u}{a \sin^2 \theta} \frac{\partial u}{\partial \varphi} - \frac{v}{a \sin \theta} \frac{\partial u}{\partial \theta} - \frac{g}{a} \frac{\partial h}{\partial \varphi} - vf, \tag{1a}$$

$$\frac{\partial v}{\partial t} = \frac{-u}{a \sin^2 \theta} \frac{\partial v}{\partial \varphi} - \frac{v}{a \sin \theta} \frac{\partial v}{\partial \theta} - \frac{g \sin \theta}{a} \frac{\partial h}{\partial \theta} + \frac{u^2 + v^2}{a \sin^2 \theta} \cos \theta + uf, \tag{1b}$$

$$\frac{\partial h}{\partial t} = \frac{-u}{a \sin^2 \theta} \frac{\partial h}{\partial \varphi} - \frac{v}{a \sin \theta} \frac{\partial h}{\partial \theta} - \frac{h}{a \sin^2 \theta} \frac{\partial u}{\partial \varphi} - \frac{h}{a \sin \theta} \frac{\partial v}{\partial \theta}, \tag{1c}$$

where u is longitudinal velocity, v is latitudinal velocity, h is height of the free surface, g is gravity, a is the radius of the earth, and the physical Coriolis parameter is $f = 2\Omega \cos \theta$ (Ω is angular rotation frequency). In order to simplify treatment at the poles the above equations are the result of transforming the standard pseudo-vectors $u_\varphi = a\dot{\varphi} \sin \theta$ and $u_\theta = a\dot{\theta}$ by $u = \sin \theta u_\varphi$, $v = \sin \theta u_\theta$. For continuous flow across a pole the vector velocity components u_φ and u_θ are discontinuous. The scalar components u, v are continuous across a pole. By introducing the above transformation, which is used in spherical harmonic expansions of u_φ and u_θ [6] to maintain uniform convergence at the poles, complicated special treatments [5] of the difference equations near the poles can be avoided.

The primitive shallow water equations (1a)–(1c) have an analytic steady state solution of zonal geostrophic flow. Introducing a rotated coordinate system [5] and the Coriolis term, $f = 2\Omega \sin \theta \cos \varphi$, the solution with flow across poles is

$$\begin{aligned} u &= -u_0 \cos \varphi \cos \theta \sin \theta, \\ v &= -u_0 \sin \varphi \sin \theta, \\ h &= h_0 - \frac{u_0}{g} \left(a\Omega + \frac{u_0}{2} \right) \sin^2 \theta \cos^2 \varphi. \end{aligned} \tag{2}$$

The constants which roughly correspond to parameters for the earth are

$$\begin{aligned} u_0 &= 500.0 \text{ cm sec}^{-1}, \\ g &= 980.0 \text{ cm sec}^{-2}, \\ h_0 &= 3.0 \times 10^5 \text{ cm}, \\ \Omega &= 7.3 \times 10^{-5} \text{ sec}^{-1}, \\ a &= 6.4 \times 10^8 \text{ cm}. \end{aligned} \tag{3}$$

The above equations will be used to test for both spatial and temporal discretization errors [7, 8] and stability limits of the CFL type. A time dependent solution with balanced initial conditions [9] to be used for energy conservation checks is

$$
\begin{aligned}
s &= \sin\theta \\
A(\theta) &= \tfrac{1}{2}\omega(2\Omega+\omega)s^2 + \tfrac{1}{4}\omega^2 s^{2R}[(R+1)s^2 + (2R^2 - R - 2) - 2R^2 s^{-2}], \\
B(\theta) &= \frac{2(\Omega+\omega)\omega}{(R+1)(R+2)}\, s^R[(R^2+2R+2) - (R+1)^2 s^2], \\
C(\theta) &= \tfrac{1}{4}\omega^2 s^{2R}[(R+1)s^2 - (R+2)] \qquad (4)\\
u &= a\omega s^2 + a\omega s^R[R\cos^2\theta - s^2]\cos R\varphi, \\
v &= a\omega R s^R \cos\theta \sin R\varphi, \\
h &= h_0 + \frac{a^2}{g}\,(A(\theta) + B(\theta)\cos R\varphi + C(\theta)\cos 2R\varphi),
\end{aligned}
$$

with $h_0 = 8.0 \times 10^5$ cm, $\omega = 7.821 \times 10^{-6}\,\mathrm{sec}^{-1}$, and $R = 4.0$. Total energy (kinetic + potential) should be conserved during an integration of (1a)–(1c) with initial conditions (4). This represents a solution with wavenumber four which rotates around the globe in longitude with frequency $\approx\Omega/30.0$.

COMPUTATIONAL GRID

The grid adopted for this study uses equally spaced intervals in both θ and φ. It is desirable to have points on the equator, but not to have a point at the poles where φ derivatives are indefinite. To facilitate taking derivatives across the poles the θ spacing across the pole of points on common longitude circles should be equal to the general θ spacing. The above requirements can be met by restricting $\Delta\theta = \Delta\varphi = 180°/(2n+1)$, where n is an integer. This places latitude circles $1/2\Delta\theta$ from each pole and at the equator. For this study values of $n = 7$ and 22 have been selected which yield integral angular spacings of $\Delta\theta = 12$ and $4°$, respectively. The choice of a uniform grid produces the well known pole problem—a crowding of points in longitude by the $\sin\theta$ factor at high latitudes. The crowding near the poles leads to a very restrictive CFL condition for explicit time steps, but this should not be a significant problem with the ADI methods. The simplicity of having equal increments in θ and in φ (but not necessarily $\Delta\theta = \Delta\varphi$) over the full grid is desirable for efficient implementation of fourth order and alternating-direction implicit schemes.

Explicit Solutions

Anticipating the need for a compact notation the shallow water equations may be written as (following notation of [2])

$$\partial H(x)/\partial t = D(x) + S(x), \tag{5}$$

where x is the column vector, $x^T = (u, v, h)$ (superscript T implies transpose), H and S are column vector functions of x, and D is a column vector of differential operations. In the present context

$$H = x, \tag{6}$$

$$S^T(x) = \left(-vf, \frac{u^2+v^2}{a\sin^2\theta}\cos\theta + uf, 0\right), \tag{7}$$

$$D_\theta^T(x) = \left(\frac{-v}{a\sin\theta}\frac{\partial u}{\partial\theta}, \frac{-v}{a\sin\theta}\frac{\partial v}{\partial\theta} - \frac{g\sin\theta}{a}\frac{\partial h}{\partial\theta}, \frac{-v}{a\sin\theta}\frac{\partial h}{\partial\theta} - \frac{h}{a\sin\theta}\frac{\partial v}{\partial\theta}\right), \tag{8}$$

$$D_\varphi^T(x) = \left(\frac{-u}{a\sin^2\theta}\frac{\partial u}{\partial\varphi} - \frac{g}{a}\frac{\partial h}{\partial\varphi}, \frac{-u}{a\sin^2\theta}\frac{\partial v}{\partial\varphi}, \frac{-u}{a\sin^2\theta}\frac{\partial h}{\partial\varphi} - \frac{h}{a\sin^2\theta}\frac{\partial u}{\partial\varphi}\right), \tag{9}$$

where $D(x) = D_\theta(x) + D_\varphi(x)$. Using the notation of Eq. (5) the explicit solutions are obtained using leap-frog differencing as

$$H^{n+1}(x) = H^{n-1}(x) + 2\Delta t[D^n(x) + S^n(x)], \tag{10}$$

where the superscript n implies evaluation of the grid function at time $t = n\,\Delta t$. This yields a second order accurate scheme in time.

Second order spatial derivatives are evaluated by the standard central difference form, e.g., $(\partial/\partial\theta)(x_i) = (1/2\Delta\theta)(x_{i+1} - x_{i-1})$. Cyclic boundary conditions are simply applied by making the identification $x_{N+1} = x_1$ and $x_0 = x_{N-1}$, where N is the number of grid points over a full latitude or longitude circle. The algorithm (10) can be trivially coded and represents a very fast solution per time step. As expected, integrations of (10) with Δt exceeding that allowed by the CFL condition lead to instabilities. See Table I for a summary of test runs of the analytically steady state solution and Table II of the time variable solution. Solutions at different grid resolutions confirm second order accuracy.

Fourth order spatial derivatives are evaluated by solving a set of linear equations coupling three adjacent derivatives Q_i with the standard second order central differences as source terms. This system is [3]

$$\frac{1}{6}Q_{i+1} + \frac{2}{3}Q_i + \frac{1}{6}Q_{i-1} = \frac{1}{2\Delta\theta}(x_{i+1} - x_{i-1}), \tag{11}$$

where the Q_i are fourth order derivatives. Note that this is a globally dependent solution for the spatial derivatives. The above system generates a cyclic, symmetric matrix equation which can be efficiently solved by a simplified version of the Ahlberg–Nilson–Walsh algorithm [10] for cyclic tridiagonal matrices (a generalized version for handling cyclic block tridiagonal matrices is presented below). The solution of (10) then follows with the Q_i used in evaluating $D^n(x)$. In applying cyclic boundary conditions in both φ and θ it is useful to note that a mapping exists to transform a discrete array $x(\theta = 0$ to π, $\varphi = 0$ to $2\pi)$ into an equivalent array $x'(\varphi = 0$ to $\pi, \theta = 0$ to $2\pi)$. Using this mapping cyclic boundary conditions can be trivially applied by following great circles in θ as well as φ.

For smooth solutions the fourth order compact differencing works very well in comparison with the second order scheme. An accuracy increase of more than two orders of magnitude is obtained at an increase of less than a factor of 2 in computer time (see Table I). As noted in [3] the CFL condition is reduced for the fourth order scheme, but this is more than offset by the possibility of using fewer grid points to obtain a given discretization accuracy and less restrictive CFL limit. As would be expected by the improvement in accuracy energy conservation is also better with the fourth order scheme (Table II).

A comparison of the accuracy of compact, Padé differences with a central five point formula, e.g., $(\partial/\partial\theta)(x_i) = (1/12\Delta\theta)(-x_{i+2} + 8x_{i+1} - 8x_{i-1} + x_{i-2})$, also shows the superiority of compact differences. The compact differences have a coefficient in the truncation error which is smaller by a factor of 6 with respect to the above five point formula [3]. The computations listed in Table I show a factor of 5.9 gain in accuracy with compact differences with a loss of only $\sim$10% in computational speed. Thus a grid coarser by $(5.9)^{1/4} \approx 1.56$ can be used with the compact differencing to obtain the same accuracy as that obtained with the five point formula. A factor of 1.56 in each of three dimensions translates to an overall reduction of about 3.8 in the required number of grid points.

Alternating-Direction Implicit Solutions

For a recent and thorough review of linearized block implicit schemes for the numerical solution of systems of nonlinear multidimensional partial differential equations see Briley and McDonald [11]. As with most implicit methods the present one involves a Taylor series expansion in time. In standard approaches the expansion is done about the time level to be solved for resulting in a system of nonlinear algebraic equations which must be solved by an iterative technique, e.g., the Newton–Raphson method. The method to be used here utilizes an expansion about the known time level, coupled with use of the Douglas–Gunn [1] technique for deriving ADI schemes, to generate a system of linear difference equations for the dependent variables at the new time level. This has the obvious advantage that the solution is obtained in one direct step without iteration.

The ADI solutions obtained with second order spatial differences follow the method of [2] with the exception of cyclic boundary conditions. This linearization technique is conceptually equivalent to that used by Beam and Warming [4]. The temporal discretization of (5) may be expressed as

$$H^{n+1}(x) = H^n(x) + \beta\,\Delta t[D^{n+1}(x) + S^{n+1}(x)] + (1-\beta)\,\Delta t[D^n(x) + S^n(x)], \tag{12}$$

where β is a centering parameter, e.g., $\beta = 0.5$ yields Crank–Nicolson (second order) and $\beta = 1.0$ yields a backward (first order) difference form. For the shallow water equations the two step Douglas–Gunn ADI solution of (12) can be written

$$\left[\mathbf{I} - \beta\,\Delta t\frac{\partial S^n}{\partial x}\right]\psi^* - \beta\,\Delta t\left[F^T_{\varphi q}\frac{\partial}{\partial\varphi} + \frac{\partial F_{\varphi q}}{\partial x}\frac{\partial}{\partial\varphi}x^n\right]\psi^*$$

$$= \Delta t\left[F^T_{\varphi q}\frac{\partial}{\partial\varphi}x^n + F^T_{\theta q}\frac{\partial}{\partial\theta}x^n + S^n\right], \tag{13a}$$

$$\left[\mathbf{I} - \beta\,\Delta t\frac{\partial S^n}{\partial x}\right]\psi^{n+1} - \beta\,\Delta t\left[F^T_{\theta q}\frac{\partial}{\partial\theta} + \frac{\partial F_{\theta q}}{\partial x}\frac{\partial}{\partial\theta}x^n\right]\psi^{n+1}$$

$$= \left[\mathbf{I} - \beta\,\Delta t\frac{\partial S^n}{\partial x}\right]\psi^*, \tag{13b}$$

where the $F_{\varphi q}$, $G_{\varphi q}$, $F_{\theta q}$, and $G_{\theta q}$ (see Appendix) are defined as in [2]:

$$F^T_{\varphi q}(x)\frac{\partial}{\partial\varphi}G_{\varphi q}(x) = D_{\varphi q}(x), \tag{14a}$$

$$F^T_{\theta q}(x)\frac{\partial}{\partial\theta}G_{\theta q}(x) = D_{\theta q}(x). \tag{14b}$$

Differencing Eq. (13a) leads to the system of equations

$$a^n_i\psi^*_{i-1} + b^n_i\psi^*_i + c^n_i\psi^*_{i+1} = d^n_i. \tag{15}$$

Differencing Eq. (13b) leads to a similar set of equations. The second order spatial difference solution explicitly follows the formalism of [2] in generating the a_i, b_i, c_i and d_i for (15); therefore further details are deferred to the Appendix. Systems (15) for successive ADI steps determine the consistent intermediate and final approximations to $\psi^{n+1} = x^{n+1} - x^n$. The above solutions are often referred to as the "delta" form, where one solves for changes in given quantities, rather than for the full quantities. The delta form is presumably better from the standpoint of roundoff error accumulation. An examination of Tables I and II show that the ADI solution compares favorably with explicit solutions regarding accuracy. The maximum time step allowed for the explicit scheme can be exceeded by more than an order of magnitude, confirming unconditional stability.

The claim [2] that the computational effort per time step for the ADI solutions is only about twice that of explicit methods is not supported by this study. The computational timing ratio for this very simple test case is approximately 80–90. The high observed ratio in implicit to explicit run time may be highly dependent on the vector processing capability of the CRAY-1 computer used for this study. The explicit scheme is easily and fully vectorized, while the implicit computational scheme is only partly vectorized due to the increased complexity of block inversion algorithms. A rough operation count [2] for the periodic block-tridiagonal eliminations $[(3N-2)(m^3+m^2)$ plus a factor of ~ 1.4 for periodic boundary conditions$]$ yields $\sim 300N^2$ operations for the two ADI sweeps, where N^2 is the total number of grid points and $m=3$ the block size. Evaluating Eq. (1) for either the explicit or implicit schemes requires $\sim 50N^2$ operations. The operation counts suggest a basic computational labor ratio of order 350/50 or 7. It should be noted that the ratio of computational time for the ADI to explicit solution would drop significantly as the overhead of more complex physics is added to a problem. In a computation where evaluation of an equation of state is the dominate time element the ADI to explicit ratio would approach unity. For typical problems of astrophysical interest the ADI computational burden could be expected to be less than an order of magnitude more per time step than the corresponding explicit solution.

The cyclic boundary conditions inherent in Eq. (15) for spherical polar coordinates introduce blocks in the antidiagonal corners of a block tridiagonal matrix. The resulting matrix equations take the form

$$\begin{bmatrix} b_1 & c_1 & \cdot & \cdot & \cdot & a_1 \\ a_2 & b_2 & c_2 & \cdot & \cdot & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdot & \cdot & \cdot & a_{n-1} & b_{n-1} & c_{n-1} \\ c_n & \cdot & \cdot & \cdot & a_n & b_n \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{n-1} \\ \psi_n \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{n-1} \\ d_n \end{bmatrix}, \tag{16}$$

where the a_i, b_i, c_i are $m \times m$ square matrices (3×3 for shallow water equations), and ψ_i and d_i are column vectors.

The generalized Ahlberg–Nilson–Walsh algorithm [10] (see also Appendix B of Schnack and Killeen [12]) for solving the above system is as follows:

For $i=1$ solve the augmented system

$$[b_1 \mid -c_1, -a_1, d_1];$$

this is a set of linear equations with $2m+1$ right hand sides.

Place solution vectors in

$$[b_1, c_1, d_1].$$

For $i=2$ to $n-1$ solve the augmented systems

$$[a_i b_{i-1} + b_i \mid -c_i, -a_i c_{i-1}, d_i - a_i d_{i-1}]$$

and place solution vectors in

$$[b_i, c_i, d_i]. \tag{17}$$

Now do a backward sweep:

For $i = n-1$, $a_{n-1} = c_{n-1} + b_{n-1}$.

For $i = n-2$ to 1 set

$$a_i = b_i a_{i+1} + c_i,$$
$$d_i = b_i d_{i+1} + d_i.$$

For the solution at $i = n$ solve the augmented system

$$[b_n + c_n a_1 + a_n a_{n-1} \mid d_n - c_n d_1 - a_n d_{n-1}]$$

and place the solution vector in d_n.

A final sweep generates the remaining solution vectors.

For $i = 1$ to $n-1$ set $d_i = d_i + a_i d_n$.

The above is a very efficient algorithm both in terms of execution speed, when properly programmed, and in terms of storage requirements. The execution time for the cyclic algorithm is only ~40% greater than that for the simpler algorithm for block tridiagonal systems. (All computations have been done on the NCAR CRAY-1A computer.)

The fourth order compact differencing scheme leads to a cyclic block tridiagonal matrix structure as in the second order ADI case (see [4] for fourth order solutions in Cartesian coordinates and [13] for the introduction of periodic Cartesian boundary conditions). However, for a given ADI sweep the dimension of the blocks increases since the spatial derivatives are now evaluated implicitly and simultaneously with the dependent variables. As before, cyclic boundary conditions are easily applied. The structure of the blocks is significantly altered in the fourth order scheme. Equations (13a), (13b) become, in the fourth order scheme,

$$\begin{aligned}
&\left[\mathbf{I} - \beta\,\Delta t\left(\frac{\partial S_q^n}{\partial x} + P_\varphi^T \frac{\partial F_{\varphi q}}{\partial x}\right)\right]\psi_i^* - \beta\,\Delta t\left[F_{\varphi q}^T + (x^n)^T \frac{\partial F_{\varphi q}}{\partial x}\right] Q_{\varphi,i}^* \\
&\qquad = \Delta t[F_{\varphi q}^T P_\varphi + F_{\theta q}^T P_\theta + S_q^n], \\
&\frac{3}{\Delta\varphi}\psi_{i-1}^* - \frac{3}{\Delta\varphi}\psi_{i+1}^* + Q_{\varphi,i-1}^* + 4Q_{\varphi,i}^* + Q_{\varphi,i+1}^* = 0,
\end{aligned} \tag{18a}$$

$$\begin{aligned}
&\left[\mathbf{I} - \beta\,\Delta t\left(\frac{\partial S_q^n}{\partial x} + P_\theta^T \frac{\partial F_{\theta q}}{\partial x}\right)\right]\psi_i^{n+1} - \beta\,\Delta t\left[F_{\theta q}^T + (x^n)^T \frac{\partial F_{\theta q}}{\partial x}\right] Q_{\theta,i}^{n+1} \\
&\qquad = \left[\mathbf{I} - \beta\,\Delta t \frac{\partial S_q^n}{\partial x}\right]\psi_i^*, \\
&\frac{3}{\Delta\theta}\psi_{i-1}^{n+1} - \frac{3}{\Delta\theta}\psi_{i+1}^{n+1} + Q_{\theta,i-1}^{n+1} + 4Q_{\theta,i}^{n+1} + Q_{\theta,i+1}^{n+1} = 0.
\end{aligned} \tag{18b}$$

Equation (18a) can be rewritten as (with a similar representation for (18b))

$$[a_i^n]\begin{bmatrix}\psi_{i-1}^* \\ Q_{\varphi,i-1}\end{bmatrix}+[b_i^n]\begin{bmatrix}\psi_i^* \\ Q_{\varphi,i}\end{bmatrix}+[c_i^n]\begin{bmatrix}\psi_{i+1}^* \\ Q_{\varphi,i+1}\end{bmatrix}=[d_i^n]. \tag{19}$$

The $P_\varphi=(\partial/\partial\varphi)(x)$ and $P_\theta=(\partial/\partial\theta)(x)$ are column vectors of fourth order differences evaluated explicitly at time level n by solution of a symmetric, cyclic tridiagonal matrix (Eq. (11)) for each component derivative. The blocks a_i^n, b_i^n, c_i^n and vector d_i^n are now of twice the dimension of those in Eq. (15); however, the fourth order blocks are sparser. The a_i^n, etc., written as partitioned matrices where the partitions are of rank 3, become

$$[a_i^n]=\left[\begin{array}{c|c}0 & 0\\ \hline \dfrac{3}{\Delta\varphi}\mathbf{I} & \mathbf{I}\end{array}\right]\qquad [c_i^n]=\left[\begin{array}{c|c}0 & 0\\ \hline \dfrac{-3}{\Delta\varphi}\mathbf{I} & \mathbf{I}\end{array}\right],$$

$$[b_i^n]=\left[\begin{array}{c|c}\mathbf{I}-\beta\,\Delta t\left(\dfrac{\partial S_q^n}{\partial x}+P_\varphi^T\dfrac{\partial F_{\varphi q}}{\partial x}\right) & -\beta\,\Delta t\left(F_{\varphi q}^T+(x^n)^T\dfrac{\partial F_{\varphi q}}{\partial x}\right)\\ \hline 0 & 4\mathbf{I}\end{array}\right], \tag{20}$$

$$[d_i^n]=\left[\begin{array}{c}\Delta t(F_{\varphi q}^T P_\varphi+F_{\theta q}^T P_\theta+S_q^n)\\ \hline 0\end{array}\right].$$

Although inclusion of fourth order differencing doubles the rank of the submatrices, which should increase by $\times 7$ the number of required operations, the submatrices have simple properties. The solution algorithm (17) can be significantly optimized to take account of the simple structures of a_i^n, c_i^n, and d_i^n. With the simplified matrix multiplies the compact fourth order solution requires only 80% more computer time per step than the second order solution. The computed accuracy is fourth order as expected. As with the second order ADI scheme the CFL time step can be exceeded by more than an order of magnitude.

The use of five point central differences to yield fourth order spatial accuracy would result in a block pentadiagonal system analogous to (13a)–(13b) and (15) for the ADI solution. The blocks would be only $m\times m$ dimensional, but not sparse as with the $2m\times 2m$ blocks (20) of the compact scheme. The solution of a dense $m\times m$ pentadiagonal system (with a total of six antidiagonal blocks for the cyclic case) would presumably take more computer time than that of the sparse $2m\times 2m$ tridiagonal system (19). Thus for the ADI solution of the shallow water equations (1) the Padé differences are superior in terms of both computational speed and accuracy. However, for a case with mixed order derivatives the block size for the Padé operators must be increased, while the five point formulas are completely general for second order systems.

COMPUTATIONAL ACCURACY AND STABILITY

As argued above and demonstrated by the solutions of (2) given in Table I the fourth order compact differencing schemes show substantial advantages over the second order schemes. The increase in computer time (less than a factor of 2) is more than offset by the gain in accuracy. The results for energy conservation as given in Table II are also favorable for the ADI schemes. A comparison of the explicit and ADI solutions at the same spatial resolution and time step show very little difference, with the ADI having ~10% smaller total energy changes. At 12° grid spacing the fourth order schemes conserve energy better by more than an order of magnitude with respect to the second order schemes. At time steps greater than 10 times the CFL condition the total energy change is still only 0.26% for the fourth order ADI

TABLE I

Computational Speed and Accuracy for Steady State Solution

Method–Order–Grid spacing (degrees)	Time step (sec)	Error[a]	Computer time/step (sec)	Courant limit (sec)	β
Explicit 2nd 12	50	7.53–3[b]	8.50–4	842.	
Explicit 2nd 12	100	7.54–3	8.50–4	842.	
Explicit 2nd 12	900	Diverged	8.50–4	842.	
Explicit 2nd 4	50	8.30–4	4.97–3	94.	
Explicit 2nd 4	100	Diverged	4.97–3	94.	
Explicit 4th 12	100	4.37–5	1.46–3	345.	
Explicit 4th 12	500	Diverged	1.46–3	345.	
Explicit 4th 4	25	5.40–7	1.28–2	38.	
Explicit 4th(5) 12	100	2.58–4	1.35–3	614.	
Implicit 2nd 12	100	7.50–3	7.24–2	842.	0.5
Implicit 2nd 12	900	7.49–3	7.24–2	842.	0.5
Implicit 2nd 12	900	6.91–3	7.24–2	842.	1.0
Implicit 2nd 12	4000	1.54–2	7.24–2	842.	0.5
Implicit 2nd 12	10000	Diverged	7.24–2	842.	0.5
Implicit 2nd 4	100	8.28–4	6.52–1	94.	0.5
Implicit 4th 12	100	4.38–5	1.30–1	345.	0.5
Implicit 4th 12	4000	7.51–5	1.30–1	345.	0.5
Implicit 4th 12	4000	7.54–5	1.30–1	345.	0.51
Implicit 4th 12	4000	8.20–5	1.30–1	345.	0.75
Implicit 4th 12	8000	2.31–2	1.30–1	345.	0.5001
Implicit 4th 12	8000	4.28–2	1.30–1	345.	0.51
Implicit 4th 12	8000	Diverged	1.30–1	345.	0.55

[a] Root mean square deviation of discretized variables from analytic solution.
[b] Signed integer indicates power of 10 to multiply the mantissa.

solution over 7 days. The second order ADI solution at large Δt encounters a nonlinear instability after ~6 days; the increased resolution of the fourth order scheme allows evolution through 7 days with no problems. The solutions at 12° in second order develop high frequency gravity waves at an amplitude of ~1% of the total geopotential height, fourth order solutions reduce the initial imbalance, resulting in amplitudes of the gravity waves of ~0.1%.

As can be seen from an examination of Tables I and II variation of β (time centering) for the implicit solutions has generally small effects on the solution. The accuracy of a real steady state solution should not have any relation to the accuracy of the temporal differencing. The solutions of (2) are steady state only in an analytic sense. The solutions given in Table I are time dependent numerical solutions (with spatial truncation errors) which approximate the analytically time independent, steady state solution [5, 7, 8]. For the steady state solutions of Table I accuracy is not sensitive to the temporal differencing (see below also). Stability is not sensitive to variation of β for time steps much greater than the CFL condition. With respect to energy conservation (Table II) the error increases slightly for larger β.

The numerical simulations (see Table I) of the analytically steady state solution (2) are independent of the time step for steps not greatly in excess of the Courant limit. The numerical solutions for time steps less than or about the Courant limit show deviations from the analytic solution which vary by a factor of about 3 over

TABLE II

Energy Conservation with Time Dependent Solution

Method–Order–Grid spacing (degrees)	Time step (sec)	ΔE (%)[a] 7 days	$\Delta E/E$	Courant limit (sec)	β	Computer time (sec) at 5 days
Explicit 2nd 4	45.	+2.5–3	4.3–5	57.3		56.1
Explicit 2nd 12	45.	−1.3–2	2.0–3	516.		11.4
Explicit 2nd 12	180.	−1.2–2	2.0–3	516.		2.9
Explicit 4th 12	180.	−1.3–2	1.9–4	211.4		4.9
Implicit 2nd 12	180.	−4.9–2	1.9–3	516.	0.5	243.
Implicit 2nd 12	1800.	Diverged		516.	0.5	24.3
Implicit 2nd 12	1800.	Diverged		516.	0.51	24.3
Implicit 4th 12	180.	9.0–3	1.7–4	211.4	0.5	437.
Implicit 4th 12	1800.	2.1–1	2.2–3	211.4	0.5	43.7
Implicit 4th 12	1800.	2.6–1	2.8–3	211.4	0.75	43.7
Implicit 4th 12	2400.	2.5–1	2.6–3	211.4	0.5	29.1
Implicit 4th 12	3600.	Diverged		211.4	0.5	

[a] Change in sum of potential and kinetic energies with respect to initial conditions. $\Delta E/E$ refers to the difference of maximum and minimum total energies relative to the energy over the full 7-days integration.

0.25 days (~200 steps at $\Delta t = 100$ sec). This variation is steady in time with the maximum error for several extended test runs not exceeding that listed in Table I. The explicit solutions experience exponential error growth for time steps in excess of the Courant limit. The ADI simulations of the analytically steady-state solution (2) are time dependent for time steps much greater than the Courant limit. For the fourth order solutions at 12° resolution and $\Delta t = 4000$ sec (~12 × Courant limit) maximum deviations from the analytical solution grow slowly over the 10-day test integration.

Table I gives results for analytically steady state solutions (2 and 3) of the shallow water equations (1a)–(1c). All errors are maximum change over integrations of 0.25 days, or 200 time steps, whichever is larger. Five point difference scheme is denoted in first column by 4th(5).

Table II gives results for time dependent solutions (4) of the shallow water equations (1a)–(1c).

Alternate Time Discretizations and Gravity Waves

The implicit solutions considered above have been based on the two time level discretization of Eq. (12) in which all terms in the equations are centered at the same time. The most rapid wave solutions to the shallow water equations are generated by only the h_φ, h_θ, and $h(u_\varphi + v_\theta)$ terms. Although not unconditionally stable a semi-implicit scheme which evaluates only these terms implicitly will then allow the use of time steps limited only by the advection terms [14]. Thus the fast mode gravity waves, which are usually not of interest for solutions of the shallow water equations, will not limit the allowed time step. Semi-implicit schemes for simple systems can be developed which are nearly as efficient computationally per time step as an explicit scheme.

In order to compare results of the implicit discretization (12) with previously adopted [14] semi-implicit schemes the following time discretization of (5) has been implemented within the general ADI framework

$$H^{n+1} = H^{n-1} + \Delta t[D^{n+1} + D^{n-1}] + 2\Delta t S^n. \tag{21}$$

Here only the terms responsible for the gravity waves are included in $D(x)$, with all other terms being put in the explicitly evaluated $S(x)$. The solution vector of Eq. (21) becomes $\psi^{n+1} = x^{n+1} - x^{n-1}$, which requires only minor modifications of the ADI Eqs. (18a), (18b). At long time steps ($\Delta t = 1800$ sec) numerical solutions using Eq. (21) are similar to those obtained with Eq. (12) through 3–4 days with numerical instability developing in the semi-implicit solution past day 4.

In order to examine the treatment of the gravity wave modes initial solutions (4) were used with quadratic terms in ω removed. This initial imbalance resulted in fast waves of amplitude ~1% in the height field for the fourth order 12° resolution approximation. With short ($\Delta t = 180$ sec) time steps numerical integrations using the semi-implicit (21) and explicit (10) algorithms yielded nearly identical results with the gravity waves maintaining the same amplitude through 7 days. A short time step integration using the implicit (12) scheme with $\beta = 0.5$ shows a damping of the fast mode by a factor of 2 over each 3-day interval. The implicit scheme thus has a small numerical dissipation not present in the other schemes. The small numerical dissipation can explain the superior stability of the implicit scheme in relation to the dissipation free semi-implicit and explicit schemes. At short time steps the implicit scheme is slightly superior to the other schemes as regards conservation of total energy. Thus for problems in which the slow changes or advection terms are of primary interest the implicit scheme (12) is superior in terms of accuracy and stability.

Accurate solutions of the fast waves are not possible with implicit or semi-implicit schemes with the time step chosen significantly longer than that allowed by the CFL condition [15]. Since time steps shorter than the CFL limit are generally required to resolve the fast waves, the more efficient explicit methods should be used in this case. Clearly the implicit scheme which damps the fast modes would not be useful for their study. Therefore the choice of an approximation scheme must entail a careful consideration of the solution to be followed in the study.

Discussion

Various methods for the solution of the shallow water equations on a sphere have been outlined. Variables have been used which permit periodic boundary conditions in both angular coordinates. Fourth order compact spatial differencing has been found to be superior to both second order differencing and a five point fourth order method. Normally boundary conditions pose significant complications for fourth order methods; however, cyclic conditions on a sphere are quite easily set in the fourth order scheme. The application of alternating-direction implicit techniques has been accomplished in two-dimensional spherical geometry. The equations studied above provide a worst case for comparison of ADI solution speed to explicit solutions. Since the shallow water equations are closed without any auxiliary relations, e.g., an equation of state, essentially no overhead exists in addition to the mechanics of applying the time stepping in the explicit case. Thus the factor of 80–90 speed loss of ADI versus explicit methods may not be relevant for more complex problems. In the solution of physically complex problems the speed loss of ADI to explicit methods would drop to less than 10. Since time steps more than an order of magnitude in excess of those allowed in explicit calculations are allowed, the ADI solution should be more efficient computationally for complex problems. The ADI formalism of [2] holds one final advantage over explicit techniques. Through the generality of (5) it would be possible to use variables which do not explicitly appear in the equations, but whose dependence through $H(x)$, $D(x)$, and $S(x)$ is explicitly known. For example, in the construction of stellar models with realistic physics it is useful to parameterize the state variables (e.g., density and pressure) in terms of a "degeneracy parameter" and temperature [16]. The degeneracy parameter appears in the equations only implicitly through the auxiliary variables which depend upon it. To use such an equation of state in the Briley and McDonald ADI formalism would be straightforward—maintaining the desired non-iterative property. (To maintain second order temporal accuracy will require adoption of a three-level difference scheme for the general case [17]). However, inclusion of such an equation of state in an explicit or semi-implicit algorithm would require a very wasteful inverse iteration on the equation of state.

The extension of the above two-dimensional ADI scheme over spherical polar coordinates θ and φ to a full three-dimensional scheme over r, θ, and φ should be straightforward using the Douglas and Gunn [1] formalism. Assuming fourth order compact differencing in θ and φ, $\Delta\theta = \Delta\varphi$, number of variables $NEQU$, and N_r radial zones with second order central differencing the computation time per time step on a CRAY-1 computer will be $\sim 5.4(N_r/30)(12°/\Delta\theta)^2(NEQU/3)^{1.6}$ sec. The above estimate is based on timings for the solutions of Eqs. (18a), (18b) and solutions of algorithm (17) with variable NEQU. The inclusion of realistic physics would be expected to increase the above by <50%.

Appendix

The terms of Eq. (13a), (13b) which represent the shallow water equations (1) in the ADI formalism are

$$F_{\varphi q} = \begin{bmatrix} \dfrac{-u}{a \sin^2\theta} & 0 & \dfrac{-h}{a \sin^2\theta} \\ 0 & \dfrac{-u}{a \sin^2\theta} & 0 \\ \dfrac{-g}{a} & 0 & \dfrac{-u}{a \sin^2\theta} \end{bmatrix},$$

$$F_{\theta q} = \begin{bmatrix} \dfrac{-v}{a \sin\theta} & 0 & 0 \\ 0 & \dfrac{-v}{a \sin\theta} & \dfrac{-h}{a \sin\theta} \\ 0 & \dfrac{-g \sin\theta}{a} & \dfrac{-v}{a \sin\theta} \end{bmatrix}, \tag{A1}$$

where q corresponds to separate columns in the matrices.

$$G_\theta = G_\varphi = H = x = (u, v, h)^T. \tag{A2}$$

The qth rows of matrices a_i, b_i, c_i and vector d_i of Eq. (15) are

$$\begin{aligned}
a_i^n &= \left[F_{\varphi q}^T\big|_i + \frac{\partial F_{\varphi q}}{\partial x}\bigg|_i x_{i-1} \right], \\
b_i^n &= \frac{2\,\Delta\varphi}{\beta\,\Delta t}\left[\mathbf{I} - \beta\,\Delta t \frac{\partial S_q}{\partial x} \right], \\
c_i^n &= -\left[F_{\varphi q}^T\big|_i + \frac{\partial F_{\varphi q}}{\partial x}\bigg|_i x_{i+1} \right], \\
d_i^n &= \frac{2\,\Delta\varphi}{\beta}\left[F_{\varphi q}^T\big|_i \frac{\partial x}{\partial \varphi}\bigg|_i + F_{\theta q}^T\big|_i \frac{\partial x}{\partial \theta}\bigg|_i + S_q\big|_i \right].
\end{aligned} \tag{A3}$$

Similar matrices exist for the second ADI step as given by Eq. (13b).

ACKNOWLEDGMENTS

I thank Dr. Gerald Browning for pointing out the limitations of implicit and semi-implicit schemes with respect to treatment of fast wave solutions, and for several general comments on the manuscript. I thank anonymous reviewers for suggesting several clarifications to the original presentation.

REFERENCES

1. J. DOUGLAS AND J. E. GUNN, *Numer. Math.* **6** (1964), 428.
2. W. R. BRILEY AND H. MCDONALD, *J. Comput. Phys.* **24** (1977), 372.
3. R. S. HIRSH, *J. Comput. Phys.* **19** (1975), 90.
4. R. A. BEAM AND R. F. WARMING, *J. Comput. Phys.* **22** (1976), 87.
5. D. L. WILLIAMSON AND G. L. BROWNING, *J. Appl. Meteor.* **12** (1973), 264.
6. S. A. ORSZAG, *Mon. Weather Rev.* **102** (1974), 56.
7. L. UMSCHEID, JR., AND M. SANKAR–RAO, *Mon. Weather Rev.* **99** (1977), 686.
8. C. H. DEY, *Mon. Weather Rev.* **97** (1969), 597.
9. N. A. PHILLIPS, *Mon. Weather Rev.* **87**, No. 9 (1959), 333.
10. H. H. AHLBERG, E. N. NILSON, AND J. L. WALSH, "The Theory of Splines and Their Applications," p. 14, Academic Press, New York, 1967.
11. W. R. BRILEY AND H. MCDONALD, *J. Comput. Phys.* **34** (1980), 54.
12. D. SCHNACK AND J. KILLEEN, *J. Comput. Phys.* **35** (1980), 110.
13. R. M. BEAM AND R. F. WARMING, *AIAA J.* **16** (1978), 393.
14. T. ELVIUS AND A. SUNDSTRÖM, *Tellus* **25** (1973), 132.
15. B. GUSTAFSSON, *J. Comput. Phys.* **36** (1980), 209.
16. P. P. EGGELTON, J. FAULKNER, AND B. P. FLANNERY, *Astronom. and Astrophys.* **23** (1973), 325.
17. H. MCDONALD AND W. R. BRILEY, *J. Comput. Phys.* **19** (1975), 150.