INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2000; **34**: 527–558

An unconditionally stable semi-Lagrangian method for the spherical atmospherical shallow water equations

M. F. Carfora*,[1]

Istituto per Applicazioni della Matematica CNR, Napoli, Italy

SUMMARY

A semi-implicit, semi-Lagrangian, mixed finite difference–finite volume model for the shallow water equations on a rotating sphere is introduced and discussed. Its main features are the vectorial treatment of the momentum equation and the finite volume approach for the continuity equation. Pressure and Coriolis terms in the momentum equation and velocity in the continuity equation are treated semi-implicitly. Moreover, a splitting technique is introduced to preserve symmetry of the numerical scheme. An alternative asymmetric scheme (without splitting) is also introduced and the efficiency of both is discussed. The model is shown to be conservative in geopotential height and unconditionally stable for $0.5 \leq \theta \leq 1$. Numerical experiments on two standard test problems confirm the performance of the model. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: finite difference; finite volume; rotating sphere; semi-implicit; semi-Lagrangian; shallow water equations

1. INTRODUCTION

The semi-Lagrangian treatment of advection is currently used in several global and regional models for numerical weather prediction (NWP); among them we cite here just the Australian Global Assimilation and Prediction system, the UK MetOffice Unified Model, the National Center for Atmospheric Research (NCAR)/Penn State mesoscale MM5 model and the European Centre for Medium-Range Weather Forecast (ECMWF) spectral model. Advantages of the semi-Lagrangian approach have been exhaustively presented in the review by Staniforth and Côté [1] and more recently discussed in Bartello and Thomas [2].

In particular, application of the semi-Lagrangian scheme to global grid point models is described in McDonald and Bates [3] for the shallow water equations and in Bates *et al.* [4] for primitive equations.

* Correspondence to: Istituto per Applicazioni della Matematica CNR, Via Pietro Castellino 111, 80131 Napoli, Italy.
[1] E-mail: carfora@iamna.iam.na.cnr.it

Received March 1999
Revised May 1999

Shallow water equations have been often used in NWP to test new numerical methods. In this paper we introduce a numerical model of the inviscid shallow water equations for atmospherical circulation. It can be an attractive alternative over existing models owing to its specific features, which are not yet considered in the meteorological literature

- the hybrid approach we use (finite volume–finite difference) guarantees fluid mass conservation while retaining ease of implementation;
- the problems arising from the spherical geometry, summarized in the expression 'pole problem', i.e. the polar convergence of the meridians and the consequent singularity of velocity components at the poles, are faced by discretization of the momentum equations in their vectorial form, as in Bates *et al.* [5], but resulting in simpler corrective terms;
- the semi-implicit differencing of the Coriolis and pressure terms in the momentum equations guarantees unconditional stability to the numerical scheme; we also introduce a splitting technique to preserve symmetry of the system to be solved;
- the semi-Lagrangian treatment of advection is combined with an accurate reconstruction of the trajectories, due to a sub-stepping procedure that arises from a paper by Casulli [6], where a semi-Lagrangian numerical model of the shallow water equations for oceanic circulation is devised.

The organization of the paper is as follows: Section 2 is devoted to the analysis of the numerical method, including discretization of the equations, splitting of the momentum equation and accurate reconstruction of characteristics curves for the evaluation of semi-Lagrangian terms. Section 3 contains a linear stability analysis of the method. In Section 4, the numerical solution of the system is analysed and the conditioning of the resulting linear system evaluated. Finally, Section 5 shows the performance of the method on some test problems.

2. THE NUMERICAL METHOD

Let us consider the inviscid shallow water equations in spherical components

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{V}}{\mathrm{d}t} = -f\mathbf{k}\times\mathbf{V} - \nabla_h\Phi \\ \dfrac{\partial\Phi}{\partial t} + \nabla_h\cdot((\Phi-\Phi^s)\mathbf{V}) = 0 \end{cases} \tag{1}$$

Here $\mathbf{V}\equiv(u, v)$ is the wind field, with curvilinear components toward the east and the north respectively; it means that we use curvilinear co-ordinates x and y, with

$$\mathrm{d}x = R\cos\varphi\,\mathrm{d}\lambda, \qquad \mathrm{d}y = R\,\mathrm{d}\varphi$$

λ, φ are longitude and latitude respectively and R is Earth's radius; f is the Coriolis parameter, $\Phi(x, y, t)$ is the geopotential height field, $\mathrm{d}\Phi = g\,\mathrm{d}Z$, where Z is the height field and g is gravity

acceleration; $\Phi^s(x, y)$ is the orography field, $\Phi^s = gZ^s$, with Z^s being height of mountains. Finally, ∇_h is the horizontal gradient operator

$$\nabla_h = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = \left(\frac{1}{R\cos\varphi}\frac{\partial}{\partial\lambda}, \frac{1}{R}\frac{\partial}{\partial\varphi}\right)$$

We shall operate in a mixed finite difference–finite volume environment. The grid will be uniform in (λ, φ) co-ordinates and staggers the variables using an Arakawa C-grid: with respect to the Φ grid, the u grid is shifted half a length eastwards and the v grid is shifted half a length northwards.

2.1. Discretization of the geopotential height conservation equation

We solve the equation for geopotential height in conservative form. It is important to develop algorithms that preserve this property, also from the numerical standpoint. Consider the continuity equation in the following integral form:

$$\int_D \frac{\partial\Phi}{\partial t}\,\mathrm{d}x\,\mathrm{d}y + \int_D \nabla_h\cdot((\Phi - \Phi^s)\mathbf{V})\,\mathrm{d}x\,\mathrm{d}y = 0 \tag{2}$$

where integrals are taken over any horizontal domain D.

By considering $D_{i,j}$ as the grid cell centred around the variable $\Phi_{i,j}$, whose corners are clockwise labelled A, B, C, D, and using Green's formula, Equation (2) results in

$$\int_{D_{i,j}} \frac{\partial\Phi}{\partial t}\,\mathrm{d}x\,\mathrm{d}y + \int_B^A (\Phi - \Phi^s)u\,\mathrm{d}y - \int_D^C (\Phi - \Phi^s)u\,\mathrm{d}y + \int_A^D (\Phi - \Phi^s)v\,\mathrm{d}x - \int_C^B (\Phi - \Phi^s)v\,\mathrm{d}x = 0 \tag{3}$$

where orthogonality of cell boundaries with velocity components has been exploited. We recall that, due to our definition of the grid, the latitudinal grid length Δy (i.e. length of CD and AB) is constant, whereas Δx_j (the longitudinal diameter of a cell) varies with latitude. We indicate length AD by $\Delta x_{j+1/2}$ and length CB by $\Delta x_{j-1/2}$.

The following finite volume approximations will now be considered in Equation (3): the transient term

$$\int_{D_{i,j}} \frac{\partial\Phi}{\partial t}\,\mathrm{d}x\,\mathrm{d}y \approx \frac{\Phi_{i,j}^{n+1} - \Phi_{i,j}^n}{\Delta t} A_j$$

is obtained by first-order finite difference in time and assuming Φ is constant over the cell. A_j is the area of the grid cell:

$$A_j = 2R\sin\frac{\Delta\varphi}{2}\Delta x_j = \Delta x_j \Delta y \frac{\sin(\Delta\varphi/2)}{\Delta\varphi/2} \tag{4}$$

The divergence terms are discretized implicitly in time for stability reasons. We introduce an implicitness parameter $\theta \in [0, 1]$ and put

$$\int_B^A (\Phi - \Phi^s)u \, \mathrm{d}y \approx H_{i+1/2,j}\Delta y(\theta u^{n+1}_{i+1/2,j} + (1-\theta)u^n_{i+1/2,j})$$

$$\int_D^C (\Phi - \Phi^s)u \, \mathrm{d}y \approx H_{i-1/2,j}\Delta y(\theta u^{n+1}_{i-1/2,j} + (1-\theta)u^n_{i-1/2,j})$$

$$\int_A^D (\Phi - \Phi^s)v \, \mathrm{d}x \approx H_{i,j+1/2}\Delta x_{j+1/2}(\theta v^{n+1}_{i,j+1/2} + (1-\theta)v^n_{i,j+1/2})$$

$$\int_C^B (\Phi - \Phi^s)v \, \mathrm{d}x \approx H_{i,j-1/2}\Delta x_{j-1/2}(\theta v^{n+1}_{i,j-1/2} + (1-\theta)v^n_{i,j-1/2})$$

where $H_{i,j} = \Phi^n_{i,j} - \Phi^s_{i,j}$. Approximations are obtained by supposing geopotential height and velocity field constant over the boundaries of the cell; the values of geopotential height on the boundaries can be obtained by averaging the corresponding adjacent values. The following approximation of Equation (3) comes out for quadrangular cells:

$$\begin{aligned}
&\frac{\Phi^{n+1}_{i,j} - \Phi^n_{i,j}}{\Delta t} A_j + H_{i+1/2,j}\Delta y(\theta u^{n+1}_{i+1/2,j} + (1-\theta)u^n_{i+1/2,j}) \\
&- H_{i-1/2,j}\Delta y(\theta u^{n+1}_{i-1/2,j} + (1-\theta)u^n_{i-1/2,j}) + H_{i,j+1/2}\Delta x_{j+1/2}(\theta v^{n+1}_{i,j+1/2} + (1-\theta)v^n_{i,j+1/2}) \\
&- H_{i,j-1/2}\Delta x_{j-1/2}(\theta v^{n+1}_{i,j-1/2} + (1-\theta)v^n_{i,j-1/2}) = 0
\end{aligned} \tag{5}$$

The approximation still holds for triangular cells ($j = 1$ and $j = N_y$), where we have $\Delta x_{1/2} = 0$ and $\Delta x_{N_y+1/2} = 0$ and so the value of the variable v at only one location is involved.

As is well known, the lack of conservation is a common problem for semi-Lagrangian methods (see, for example, Staniforth and Côté [1]). Even the best schemes do not possess this property, although they may show interesting features in terms of accuracy. It is quite easy to prove that the adopted discretization procedure ensures conservation of total geopotential height. Indeed, if we sum Equation (5) over i and j, we obtain after cancellation of opposite terms

$$\begin{aligned}
\sum_{i,j} A_j \Phi^{n+1}_{i,j} = \sum_{i,j} A_j \Phi^n_{i,j} &- \theta \Delta t \Delta y \left[\sum_j H_{N_x+1/2,j} u^{n+1}_{N_x+1/2,j} - H_{1/2,j} u^{n+1}_{1/2,j} \right] \\
&- (1-\theta)\Delta t \Delta y \left[\sum_j H_{N_x+1/2,j} u^n_{N_x+1/2,j} - H_{1/2,j} u^n_{1/2,j} \right]
\end{aligned}$$

and since for the periodicity in longitude of the grid at every time level k it is $u^k_{N_x+1/2,j} = u^k_{1/2,j}$, we finally have

$$\sum_{i,j} A_j \Phi_{i,j}^{n+1} = \sum_{i,j} A_j \Phi_{i,j}^{n}$$

This result ensures that, in the discrete model, the total geopotential height is conserved during the numerical integration. Numerical experiments confirmed this exact conservation within the machine rounding.

2.2. Discretization of the momentum equation

We discretize the momentum equation starting from its vectorial form. This is an attractive choice for the semi-Lagrangian approach in spherical co-ordinates as we use a co-ordinate system that rotates with the Earth. Thus, unit vectors in the horizontal $(\mathbf{i}, \mathbf{j})$, $(\mathbf{i}_*, \mathbf{j}_*)$ at the arrival and departure point of characteristic lines respectively, in general do not coincide. Derivation of the scalar equations for velocity components will be now described in detail. This procedure is similar to the approach of Bates *et al.* [5] but with different (and simpler) projection coefficients for the three-dimensional vectors on the Earth's surface (Equations (7)). If $\theta \in [0, 1]$ is again the implicitness parameter, we have

$$\mathbf{V}^{n+1} = \mathbf{V}_*^n - \theta \Delta t [f\mathbf{k} \times \mathbf{V} + \nabla_h \Phi]^{n+1} - (1-\theta)\Delta t [f\mathbf{k} \times \mathbf{V} + \nabla_h \Phi]_*^n \tag{6}$$

where symbols $(\cdot)_*^n$ refer to values of the corresponding variables at departure points of the characteristic curves ending at time t^{n+1} in grid points. Since these departure points will generally not coincide with grid points, corresponding values of the variables must be found by interpolation.

To resolve Equation (6) into components we need to express $(\mathbf{i}_*, \mathbf{j}_*, \mathbf{k}_*)$ in terms of the unit vector triad $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ at time t^{n+1}; it means that in the following the term 'horizontal' refers to the arrival point of the trajectory.

Now, if $A(\lambda, \varphi)$ and $B(\lambda_*, \varphi_*)$ are arrival and departure points of the trajectory respectively, the relationship between the related unit vectors is

$$\begin{pmatrix} \mathbf{i} \\ \mathbf{j} \\ \mathbf{k} \end{pmatrix} = R_{A,B} \begin{pmatrix} \mathbf{i}_* \\ \mathbf{j}_* \\ \mathbf{k}_* \end{pmatrix}$$

where the rotation $R_{A,B}$ is given by

$$\begin{pmatrix} \cos\delta\lambda & -\sin\varphi_*\sin\delta\lambda & \cos\varphi_*\sin\delta\lambda \\ \sin\varphi\sin\delta\lambda & \cos\varphi_*\cos\varphi + \sin\varphi_*\sin\varphi\cos\delta\lambda & \sin\varphi_*\cos\varphi - \cos\varphi_*\sin\varphi\cos\delta\lambda \\ -\cos\varphi\sin\delta\lambda & \cos\varphi_*\sin\varphi - \sin\varphi_*\cos\varphi\cos\delta\lambda & \sin\varphi_*\sin\varphi + \cos\varphi_*\cos\varphi\cos\delta\lambda \end{pmatrix}$$

with $\delta\lambda$ denoting $(\lambda_* - \lambda)$.

As we are interested only in the horizontal part of the transform, we need only consider the sub-matrix obtained from the first two rows and two columns of $R_{A,B}$, whose elements we

indicate by $r_{i,j}$. They depend on the co-ordinates (λ, φ), (λ_*, φ_*) of arrival and departure points of the trajectory; in the same notation of Equation (6), they are

$$r_{1,1} = \cos \delta\lambda$$

$$r_{1,2} = -\sin \delta\lambda \sin \varphi$$

$$r_{2,1} = \sin \delta\lambda \sin \varphi_*$$

$$r_{2,2} = \cos \delta\lambda \sin \varphi \sin \varphi_* + \cos \varphi \cos \varphi_* \tag{7}$$

If now we indicate by $(X_u)^{n+1}$, $(X_v)^{n+1}$ the two horizontal components of the left-hand side of Equation (6) and by $(Y_u)^n_*$, $(Y_v)^n_*$ the corresponding components of the right-hand side

$$X_u^{n+1} = \left[u - f\theta v \Delta t + \theta \Delta t \frac{\partial \Phi}{\partial x} \right]^{n+1}$$

$$X_v^{n+1} = \left[v + f\theta u \Delta t + \theta \Delta t \frac{\partial \Phi}{\partial y} \right]^{n+1}$$

$$(Y_u)^n_* = \left[u + f(1-\theta) v \Delta t - (1-\theta)\Delta t \frac{\partial \Phi}{\partial x} \right]^n_*$$

$$(Y_v)^n_* = \left[v - f(1-\theta) u \Delta t - (1-\theta)\Delta t \frac{\partial \Phi}{\partial y} \right]^n_*$$

we have

$$\begin{pmatrix} X_u \\ X_v \end{pmatrix}^{n+1} = \begin{pmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{pmatrix} \begin{pmatrix} Y_u \\ Y_v \end{pmatrix}^n_* \tag{8}$$

From system (8) one can formally obtain new values u^{n+1}, v^{n+1} in all the grid points, depending on values of Φ^{n+1} in surrounding points. Hence, we find

$$\begin{pmatrix} u \\ v \end{pmatrix}^{n+1} = -\frac{\theta \Delta t}{1 + f^2 \theta^2 \Delta t^2} \begin{pmatrix} 1 & f\theta\Delta t \\ -f\theta\Delta t & 1 \end{pmatrix} \begin{pmatrix} \Phi_x \\ \Phi_y \end{pmatrix}^{n+1} + \begin{pmatrix} \mathscr{L}^u \\ \mathscr{L}^v \end{pmatrix} \tag{9}$$

where the first scalar equation is collocated on the u grid and the second one on the v grid. To simplify the notation we indicated by $\mathscr{L}$ the terms to be evaluated at departure points of the characteristic lines

$$\begin{pmatrix}\mathscr{L}^u\\ \mathscr{L}^v\end{pmatrix}=\frac{1}{1+f^2\theta^2\Delta t^2}\begin{pmatrix}1 & f\theta\Delta t\\ -f\theta\Delta t & 1\end{pmatrix}\begin{pmatrix}r_{1,1} & r_{1,2}\\ r_{2,1} & r_{2,2}\end{pmatrix}\begin{pmatrix}u+f(1-\theta)v\Delta t-(1-\theta)\Delta t\dfrac{\partial\Phi}{\partial x}\\ v-f(1-\theta)u\Delta t-(1-\theta)\Delta t\dfrac{\partial\Phi}{\partial y}\end{pmatrix}^n_* \tag{10}$$

Now we discretize the spatial derivatives of Φ with centred finite differences

$$\left(\frac{\partial\Phi}{\partial x}\right)_{i+1/2,j}=\frac{\Phi_{i+1,j}-\Phi_{i,j}}{\Delta x_j}$$

$$\left(\frac{\partial\Phi}{\partial y}\right)_{i+1/2,j}=\frac{\Phi_{i+1/2,j+1/2}-\Phi_{i+1/2,j-1/2}}{\Delta y}$$

$$\left(\frac{\partial\Phi}{\partial x}\right)_{i,j+1/2}=\frac{\Phi_{i+1/2,j+1/2}-\Phi_{i-1/2,j+1/2}}{\Delta x_{j+1/2}}$$

$$\left(\frac{\partial\Phi}{\partial y}\right)_{i,j+1/2}=\frac{\Phi_{i,j+1}-\Phi_{i,j}}{\Delta y}$$

Once we have chosen a suitable interpolation formula for values of geopotential Φ at half-indexes, we finally obtain the expressions for u^{n+1}, v^{n+1}. For simplicity's sake we use linear interpolation, i.e. the mean of the four surrounding values; in this case, the resulting values of $u^{n+1}_{i+1/2,j}$, $v^{n+1}_{i,j+1/2}$ each depend on the unknown values of Φ^{n+1} in the six neighbouring grid points. Then, substitution of u^{n+1} and v^{n+1} in the approximated continuity equation (5) leads to a nine-point scheme for Φ^{n+1}

$$\begin{aligned}
&[A_j+\theta^2\Delta t^2(C_{i+1/2,j}+C_{i-1/2,j}+D_{i,j+1/2}+D_{i,j-1/2})]\boldsymbol{\Phi}^{n+1}_{\mathbf{i},\mathbf{j}}\\
&\quad-\theta^2\Delta t^2(C_{i+1/2,j}-\theta\Delta t(G_{i,j+1/2}-G_{i,j-1/2}))\boldsymbol{\Phi}^{n+1}_{\mathbf{i}+1,\mathbf{j}}\\
&\quad-\theta^2\Delta t^2(C_{i-1/2,j}+\theta\Delta t(G_{i,j+1/2}-G_{i,j-1/2}))\boldsymbol{\Phi}^{n+1}_{\mathbf{i}-1,\mathbf{j}}\\
&\quad-\theta^2\Delta t^2(D_{i,j+1/2}+\theta\Delta t(E_{i+1/2,j}-E_{i-1/2,j}))\boldsymbol{\Phi}^{n+1}_{\mathbf{i},\mathbf{j}+1}\\
&\quad-\theta^2\Delta t^2(D_{i,j-1/2}-\theta\Delta t(E_{i+1/2,j}-E_{i-1/2,j}))\boldsymbol{\Phi}^{n+1}_{\mathbf{i},\mathbf{j}-1}-\theta^3\Delta t^3(E_{i+1/2,j}-G_{i,j+1/2})\boldsymbol{\Phi}^{n+1}_{\mathbf{i}+1,\mathbf{j}+1}\\
&\quad-\theta^3\Delta t^3(-E_{i+1/2,j}+G_{i,j-1/2})\boldsymbol{\Phi}^{n+1}_{\mathbf{i}+1,\mathbf{j}-1}-\theta^3\Delta t^3(-E_{i-1/2,j}+G_{i,j+1/2})\boldsymbol{\Phi}^{n+1}_{\mathbf{i}-1,\mathbf{j}+1}\\
&\quad-\theta^3\Delta t^3(E_{i-1/2,j}-G_{i,j-1/2})\boldsymbol{\Phi}^{n+1}_{\mathbf{i}-1,\mathbf{j}-1}\\
&=A_j\Phi^n_{i,j}-(1-\theta)\Delta t[\Delta y(H_{i+1/2,j}u^n_{i+1/2,j}-H_{i-1/2,j}u^n_{i-1/2,j})+\Delta x_{j+1/2}H_{i,j+1/2}v^n_{i,j+1/2}\\
&\quad-\Delta x_{j-1/2}H_{i,j-1/2}v^n_{i,j-1/2}]-\theta\Delta t[\Delta y(H_{i+1/2,j}\mathscr{L}^u_{i+1/2,j}-H_{i-1/2,j}\mathscr{L}^u_{i-1/2,j})\\
&\quad+\Delta x_{j+1/2}H_{i,j+1/2}\mathscr{L}^v_{i,j+1/2}-\Delta x_{j-1/2}H_{i,j-1/2}\mathscr{L}^v_{i,j-1/2}]
\end{aligned} \tag{11}$$

where

$$C_{i+1/2,j} = \frac{H_{i+1/2,j}}{1+f_j^2\theta^2\Delta t^2}\frac{\Delta y}{\Delta x_j}$$

$$D_{i,j+1/2} = \frac{H_{i,j+1/2}}{1+f_{j+1/2}^2\theta^2\Delta t^2}\frac{\Delta x_{j+1/2}}{\Delta y}$$

$$E_{i+1/2,j} = \frac{1}{4}\frac{f_j}{1+f_j^2\theta^2\Delta t^2}H_{i+1/2,j}$$

$$G_{i,j+1/2} = \frac{1}{4}\frac{f_{j+1/2}}{1+f_{j+1/2}^2\theta^2\Delta t^2}H_{i,j+1/2} \tag{12}$$

2.3. Splitting of the momentum equation

The choice of discretizing implicitly in time Coriolis and pressure terms in the momentum equation leads to the presence of asymmetric extra-diagonal terms in system (11). These terms, and the presence of 'mixed' derivatives of the geopotential, Φ_y in the u-equation and Φ_x in the v-equation, increase the computational effort for the solution of the system. So it is interesting to consider a simplifying approach (splitting technique), and to evaluate the balance between reduced computational cost and loss of accuracy. The basic idea is to split the momentum equation in two parts by considering the total derivative of velocity as the sum of two terms: first (step 1) we take into account only the contribution due to Coriolis terms; then (step 2) the contribution due to pressure terms. Formally, we solve

$$\mathbf{V}_{\text{split}}^{n+1} = \mathbf{V}_*^n - \theta_1\Delta t[f\mathbf{k}\times\mathbf{V}]_{\text{split}}^{n+1} - (1-\theta_1)\Delta t[f\mathbf{k}\times\mathbf{V}]_*^n \tag{13}$$

and then

$$\mathbf{V}^{n+1} = \mathbf{V}_{\text{split}}^{n+1} - \theta\Delta t\nabla_h\Phi^{n+1} - (1-\theta)\Delta t\nabla_h\Phi^n \tag{14}$$

coupled with Equation (5). We stress that we introduced two different implicitness parameters (θ and θ_1) for the two steps; this choice will be explained by stability considerations in the next section.

The resulting two-step scheme is clearly first-order in time. It is possible, however, to generalize it to a second-order scheme. Our experience in simplified tests [7] suggests that the introduction of a second-order in time scheme should not improve the overall performance of the method; comparison between first- and second-order schemes for characteristics reconstruction (presented in the section devoted to numerical experiments) confirms this suggestion. However, we intend to investigate further this kind of scheme in a forthcoming paper.

The introduction of the splitting technique leads to some simplifications in system (11). In fact, Equation (9) reduces to

$$\begin{pmatrix} u \\ v \end{pmatrix}^{n+1} + \theta\Delta t\begin{pmatrix} \Phi_x \\ \Phi_y \end{pmatrix}^{n+1} = \begin{pmatrix} \mathscr{L}^u \\ \mathscr{L}^v \end{pmatrix} - (1-\theta)\Delta t\begin{pmatrix} \Phi_x \\ \Phi_y \end{pmatrix}_*^n \tag{15}$$

where now the expressions $\mathscr{L}$ contain only Coriolis terms

$$\begin{pmatrix}\mathscr{L}^u\\ \mathscr{L}^v\end{pmatrix}=\frac{1}{1+f^2\theta_1^2\Delta t^2}\begin{pmatrix}1 & f\theta_1\Delta t\\ -f\theta_1\Delta t & 1\end{pmatrix}\begin{pmatrix}r_{1,1} & r_{1,2}\\ r_{2,1} & r_{2,2}\end{pmatrix}\begin{pmatrix}u+f(1-\theta_1)v\Delta t\\ v-f(1-\theta_1)u\Delta t\end{pmatrix}_*^n \tag{16}$$

It follows that system (11) reduces to

$$\begin{aligned}&[A_j+\theta^2\Delta t^2(C_{i+1/2,j}+C_{i-1/2,j}+D_{i,j+1/2}+D_{i,j-1/2})]\mathbf{\Phi}_{\mathbf{i},\mathbf{j}}^{n+1}-[\theta^2\Delta t^2C_{i+1/2,j}]\mathbf{\Phi}_{\mathbf{i}+1,\mathbf{j}}^{n+1}\\ &\quad-[\theta^2\Delta t^2C_{i-1/2,j}]\mathbf{\Phi}_{\mathbf{i}-1,\mathbf{j}}^{n+1}-[\theta^2\Delta t^2D_{i,j+1/2}]\mathbf{\Phi}_{\mathbf{i},\mathbf{j}+1}^{n+1}-[\theta^2\Delta t^2D_{i,j-1/2}]\mathbf{\Phi}_{\mathbf{i},\mathbf{j}-1}^{n+1}\\ &=A_j\Phi_{i,j}^n-(1-\theta)\Delta t[\Delta y(H_{i+1/2,j}u_{i+1/2,j}^n-H_{i-1/2,j}u_{i-1/2,j}^n)\\ &\quad+\Delta x_{j+1/2}H_{i,j+1/2}v_{i,j+1/2}^n-\Delta x_{j-1/2}H_{i,j-1/2}v_{i,j-1/2}^n]-\theta\Delta t[\Delta y(H_{i+1/2,j}\mathscr{L}_{i+1/2,j}^u\\ &\quad-H_{i-1/2,j}\mathscr{L}_{i-1/2,j}^u)+\Delta x_{j+1/2}H_{i,j+1/2}\mathscr{L}_{i,j+1/2}^v-\Delta x_{j-1/2}H_{i,j-1/2}\mathscr{L}_{i,j-1/2}^v]\end{aligned} \tag{17}$$

where the same notations of system (11) have been used, except for the terms $\mathscr{L}$, defined by Equation (16).

2.4. Semi-Lagrangian terms

This section is devoted to the evaluation of 'semi-Lagrangian terms' $\mathscr{L}$ defined by Equations (10) and (16), requiring velocity components u_*^n and v_*^n at departure points of characteristic lines. This Lagrangian procedure goes through two steps: determination of the departure points of characteristic lines and evaluation of variables in these points. Then we firstly need to integrate (backward in time) the following system from time t^{n+1} to t^n at any vertex of the u grid and v grid in order to determine departure points in both u and v grids:

$$\begin{cases}R\dfrac{\mathrm{d}}{\mathrm{d}t}\lambda\cos\varphi=u(\lambda,\varphi)\\ R\dfrac{\mathrm{d}\varphi}{\mathrm{d}t}=v(\lambda,\varphi)\end{cases} \tag{18}$$

The choice of integrating the system in variables λ, φ rather than in x, y is suggested by the differences in the geometry of the grid cells in the two planes (λ,φ) and (x,y) and it is much more accurate in practice. Since the overall numerical scheme is $O(\Delta t)$ accurate, a simple first-order algorithm will suffice for the integration of the system (18); however, we consider also a more accurate numerical algorithm, based on a second-order Runge–Kutta method.

In both procedures we fix the tracking sub-time steps $\tau^{(k)}$ by the conditions

$$\tau^{(k)}=\frac{\Delta t}{N_\tau},\quad N_\tau=1+\left[\Delta t\max_{i,j}\left(\frac{u_{i+1/2,j}^n}{\Delta x_j},\frac{u_{i-1/2,j}^n}{\Delta x_j},\frac{v_{i,j+1/2}^n}{\Delta y},\frac{v_{i,j-1/2}^n}{\Delta y}\right)\right]$$

which ensures that, in any sub-step, Courant numbers do not become greater than one.

This sub-stepping procedure is not a common choice. In many atmospheric circulation models that use a semi-Lagrangian approach [1], the characteristic lines are approximated with straight lines, whose direction is determined iteratively. The method we use was introduced by Casulli [6] for the numerical solution of shallow water equations in plane geometry and was modified by Amato and Carfora [7] to adapt to spherical geometry, with particular attention to the treatment of singularity of the grid.

Now it is possible to see that, in the case of Courant numbers not greater than one, the vectorial correction of the Lagrangian terms does not introduce significant improvements in the numerical solution. Indeed, if in Equation (16) we do not perform the correction, matrix $(r)_{i,j}$ must be substituted by the identity matrix. In this case, the error is $\|(r)_{i,j} - I_2\|$ and it is clear that

$$|r_{1,1} - 1| \leq \|(r)_{i,j} - I_2\|_2 \leq \|R_{A,B} - I_3\|_2$$

Now, the eigenvalues of $R_{A,B}$ are given by $(1, \alpha \pm I\beta)$, with $\alpha^2 + \beta^2 = 1$ and $2\alpha = Tr(R_{A,B}) - 1$ (Tr stands for the trace of a matrix). Thus, if we define, as usual, the spectral radius ρ of a matrix as the maximum modulus of its eigenvalues, we obtain

$$\rho^2(R_{A,B} - I_3) = (\alpha - 1)^2 + \beta^2 = 3 - Tr(R_{A,B}) = 4 - (1 + \cos \delta\lambda)(1 + \cos \delta\varphi)$$

and, recalling the expression of $r_{1,1}$ and the fact that $R_{A,B} - I_3$ is a normal matrix, so its spectral radius coincides with its L_2-norm, we finally obtain

$$1 - \cos \delta\lambda \leq \|(r)_{i,j} - I_2\|_2 \leq (4 - (1 + \cos \delta\lambda)(1 + \cos \delta\varphi))^{1/2}$$

which means, for $\delta\lambda$ and $\delta\varphi$ tending to zero

$$O(\delta\lambda)^2 \leq \|(r)_{i,j} - I_2\|_2 \leq O(\delta\lambda) + O(\delta\varphi)$$

As a consequence, in the sub-stepping procedure, where Courant numbers are not greater than one, the error introduced by the absence of vectorial correction is not greater than the error $(\Delta\lambda, \Delta\varphi)$ due to spatial discretization.

In order to evaluate u, v at the foot of characteristic lines, an interpolation procedure has to be invoked, because points of the characteristic curves generally do not coincide with grid points. Two interpolation schemes have been considered: bilinear and bicubic (Lagrange type). In both schemes we use the surrounding points; for grid cells near to the Poles this means that we resort to points located across the Pole. In this case, the scheme for u, v differs from the one for Φ, as crossing the Pole we have to take into account the change of the conventional direction in the velocity components. We have shown by extensive tests presented in a previous work [7] on the simpler problem of the semi-Lagrangian treatment of advection on the sphere that, in general, the regions of higher error level do not coincide with the numerical poles, but rather with medium (numerical) latitudes. This confirms the good approximation reached in pole regions, mainly in the case of cubic interpolation. The introduction of accurate spatial and

temporal approximation dramatically reduces the error: in the same paper we compared the bilinear scheme with the bicubic one and found a growth of accuracy of about three significant digits.

3. STABILITY ANALYSIS

We perform the von Neumann stability analysis of the model equations (5) and (15) for the case of a resting basic state on a tangent plane ($f = f_0 = \text{constant}$). Discretization of the linearized shallow water equations leads to the system

$$\begin{cases} u^{n+1}_{i+1/2,j} + \theta K_j \Delta t(\Psi^{n+1}_{i+1,j} - \Psi^{n+1}_{i,j}) = \mathscr{L}^u_{i+1/2,j} - (1-\theta)K_j \Delta t(\Psi^n_{i+1,j} - \Psi^n_{i,j}) \\ v^{n+1}_{i,j+1/2} + \theta K \Delta t(\Psi^{n+1}_{i,j+1} - \Psi^{n+1}_{i,j}) = \mathscr{L}^v_{i,j+1/2} - (1-\theta)K \Delta t(\Psi^n_{i,j+1} - \Psi^n_{i,j}) \\ \Psi^{n+1}_{i,j} + \theta K_j \Delta t(u^{n+1}_{i+1/2,j} - u^{n+1}_{i-1/2,j}) + \theta K \Delta t(v^{n+1}_{i,j+1/2} - v^{n+1}_{i,j-1/2}) \\ \quad = \Psi^n_{i,j} - (1-\theta)K_j \Delta t(u^n_{i+1/2,j} - u^n_{i-1/2,j}) - (1-\theta)K \Delta t(v^n_{i,j+1/2} - v^n_{i,j-1/2}) \end{cases} \tag{19}$$

where $K_j = \sqrt{\bar{H}}/\Delta x_j$; $K = \sqrt{\bar{H}}/\Delta y$; $\Psi = \Phi/\sqrt{\bar{H}}$ and $\bar{H}$ is the mean geopotential height. The terms $\mathscr{L}^u$, $\mathscr{L}^v$ are to be specified according to the order of the interpolation used in the semi-Lagrangian phase. We have also assumed that $A_j = \Delta x_j \Delta y$ and that $\Delta x_{j+1/2} = \Delta x_j = \Delta x_{j-1/2}$.

Then we introduce a Fourier mode for the dependent variables u, v and Ψ and carry out a stability analysis on the corresponding amplitude functions. We write the equations for a single mode $\bar{w}^n \, e^{Iix} \, e^{Ijy}$, where $\bar{w}^n$ is the amplitude of the variable w (w standing for u, v, Ψ) at the time level n. After some simplification, the system

$$B\bar{\mathbf{w}}^{n+1} = C\bar{\mathbf{w}}^n$$

where at the time level k

$$\bar{w}^k = (\bar{u}^k, \bar{v}^k, \bar{\Psi}^k)$$

$$B = \begin{pmatrix} 1 & 0 & 2I\theta \Delta t K_j \sin(x/2) \\ 0 & 1 & 2I\theta \Delta t K \sin(y/2) \\ 2I\theta \Delta t K_j \sin(x/2) & 2I\theta \Delta t K \sin(y/2) & 1 \end{pmatrix}$$

and

$$C = \begin{pmatrix} c_1 & c_2 & -2I(1-\theta)\Delta t K_j \sin(x/2) \\ c_3 & c_4 & -2I(1-\theta)\Delta t K \sin(y/2) \\ -2I(1-\theta)\Delta t K_j \sin(x/2) & -2I(1-\theta)\Delta t K \sin(y/2) & 1 \end{pmatrix}$$

where coefficients c_i depend on the first part of the splitting procedure as described by Equation (16). For completeness, we recall here their expression

$$\begin{pmatrix} c_1 & c_2 \\ c_3 & c_4 \end{pmatrix} = E \frac{1}{1+f^2\theta_1^2\Delta t^2} \begin{pmatrix} 1 & f\theta_1\Delta t \\ -f\theta_1\Delta t & 1 \end{pmatrix} \cdot \begin{pmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{pmatrix} \begin{pmatrix} 1 & f(1-\theta_1)\Delta t \\ -f(1-\theta_1)\Delta t & 1 \end{pmatrix} \tag{20}$$

where E is the amplification factor of the interpolation procedure (in our case $|E| = 1$).

To prove the stability of the numerical method we have to show that the L_2-norm of the amplification matrix $B^{-1}C$ is not greater than one.

First of all, we decompose matrix C as $C = C_1 + C_2$, where

$$C_1 = \begin{pmatrix} c_1 & c_2 & 0 \\ c_3 & c_4 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$C_2 = \begin{pmatrix} 0 & 0 & -2I(1-\theta)\Delta t K_j \sin(x/2) \\ 0 & 0 & -2I(1-\theta)\Delta t K \sin(y/2) \\ -2I(1-\theta)\Delta t K_j \sin(x/2) & -2I(1-\theta)\Delta t K \sin(y/2) & 0 \end{pmatrix}$$

In order to prove that the considered numerical method is unconditionally stable with a suitable choice of the implicitness parameters θ and θ_1 and to give a sharp estimate of the instability due to Coriolis terms, we need some preliminary results.

Lemma 1

$$\|B^{-1}\|_2 = 1$$

Proof

Since the real part of matrix B is the identity matrix and its imaginary part is symmetric, B is a normal matrix and so is B^{-1}. Then, their L_2-norms coincide with their spectral radius and it suffices to evaluate their eigenvalues to prove the lemma.

It is easy to show that the eigenvalues of B (v_1, v_2, v_3) are all outside the unit open circle, since they are given by

$$v_1 = 1, \qquad v_{2,3} = 1 \pm 2I\theta\Delta t \sqrt{K_j^2 \sin^2(x/2) + K^2 \sin^2(y/2)}$$

Since the eigenvalues of B^{-1} are the reciprocals of the eigenvalues of B, the spectral radius of B^{-1} is equal to one. □

Lemma 2

$$\|C_1\|_2 \leq \max\left(1, \sqrt{\frac{1+f^2(1-\theta_1)^2\Delta t^2}{1+f^2\theta_1^2\Delta t^2}}\right)$$

Proof
Since

$$\|C_1\|_1 = \max_j \sum_i |c_{i,j}| = \max(\|C_{3,3}\|_1, 1)$$

$$\|C_1\|_\infty = \max_i \sum_j |c_{i,j}| = \max(\|C_{3,3}\|_\infty, 1)$$

where $C_{3,3}$ denotes the sub-matrix determined by the first two rows and columns of C_1. By the equivalence of norms it is also

$$\|C_1\|_2 = \max(\|C_{3,3}\|_2, 1)$$

Then, from Equation (20), we have

$$\|C_{3,3}\|_2 \leq \frac{\sqrt{(1+f^2\theta_1^2\Delta t^2)(1+f^2(1-\theta_1)^2\Delta t^2)}}{1+f^2\theta_1^2\Delta t^2}\|r_{i,j}\|_2 \leq \sqrt{\frac{1+f^2(1-\theta_1)^2\Delta t^2}{1+f^2\theta_1^2\Delta t^2}}$$

since $\|r_{i,j}\|_2 \leq \|R_{AB}\|_2 = 1$. □
This completely proves the lemma.

Lemma 3

$$\|C_2\|_2 \leq 2\left[\bar{H}\left(\frac{1}{\Delta x_1^2}+\frac{1}{\Delta y^2}\right)\right]^{1/2}(1-\theta)\Delta t$$

Proof
It is easy to see that C_2 is also a normal matrix. Its eigenvalues are

$$v_1 = 0, \qquad v_{2,3} = \pm 2I(1-\theta)\Delta t\sqrt{K_j^2 \sin^2(x/2) + K^2 \sin^2(y/2)}$$

Then we have

$$\|C_2\|_2 = \rho(C_2) \leq 2(K_1^2+K^2)^{1/2}(1-\theta)\Delta t = 2\left[\bar{H}\left(\frac{1}{\Delta x_1^2}+\frac{1}{\Delta y^2}\right)\right]^{1/2}(1-\theta)\Delta t \qquad \square$$

The main result easily follows from these lemmas. Indeed, the following theorem holds:

Theorem 1
It is

$$1 \leq \|B^{-1}C\|_2 \leq \max\left(1, \sqrt{\frac{1+f^2(1-\theta_1)^2\Delta t^2}{1+f^2\theta_1^2\Delta t^2}}\right) + 2\left[\bar{H}\left(\frac{1}{\Delta x_1^2}+\frac{1}{\Delta y^2}\right)\right]^{1/2}(1-\theta)\Delta t \tag{21}$$

Then, the considered numerical method is unconditionally stable for $\theta = 1$ provided that $\theta_1 \in [0.5, 1]$. Moreover, for $\theta_1 \in [0, 0.5)$ we have

$$\|B^{-1}C\|_2 \approx 1 + f^2\Delta t^2 \tag{22}$$

Proof
To prove the lower bound in Equation (21), consider the particular case of the zero mode in Fourier analysis ($x = y = 0$). In this case, the amplification matrix $B^{-1}C$ reduces to matrix C_1, whose norm, from Lemma 2, is not less than one.

By application of the usual norm inequalities, and by Lemma 1, it is

$$\|B^{-1}C\|_2 \leq \|B^{-1}\|_2 \cdot \|C\|_2 \leq \|C\|_2 \leq \|C_1\|_2 + \|C_2\|_2$$

Then, by application of Lemmas 2 and 3, the upper bound in Equation (21) also holds. Now, in the case $\theta = 1$, it is $C_2 = 0$ and

$$\|B^{-1}C\|_2 \leq \|C_1\|_2 \leq \max\left(1, \sqrt{\frac{1+f^2(1-\theta_1)^2\Delta t^2}{1+f^2\theta_1^2\Delta t^2}}\right)$$

The maximum of this quantity is achieved for $\theta_1 = 0$ when we have (by Taylor expansion)

$$\|B^{-1}C\|_2 \leq \sqrt{1+f^2\Delta t^2} = 1 + \frac{1}{2}f^2\Delta t^2 + O(\Delta t^4)$$

whereas, for $\theta_1 \in [0.5, 1]$, it is

$$\|B^{-1}C\|_2 \leq 1$$

and then, using both the bounds, $\|B^{-1}C\|_2 = 1$, i.e. unconditional stability. □

4. COMPUTATIONAL ISSUES

4.1. Poles treatment

We include here some considerations on the numerical treatment of cells containing poles (triangular cells). Owing to the definition of the grid, $\Delta x_j \neq 0$, $j = 1, \ldots, N_y$. Let us consider the South Pole: a cell containing the Pole ($j = 1$) physically degenerates to a triangular cell, whose lower vertex is given by the Pole. Integral representation (3) still holds and the integral over the cell lower boundary becomes null, so the flux on that boundary need not be considered. Similar arguments hold for the North Pole. It comes out that all the involved coefficients must vanish for $j = \frac{1}{2}$ and for $j = N_y + \frac{1}{2}$ (values corresponding to the South and North Pole respectively) so that Equation (11) and correspondingly Equation (17) hold for all cells, including upper and lower ones, provided that these positions are set. An important consequence is that geopotential height and, in particular, the y component of velocity, are not required on the poles, totally avoiding the difficulties due to the singularity of the transform to polar co-ordinates. Then no 'special treatment' of the Poles, such as the introduction of an auxiliary (rotated) co-ordinate system for trajectories near the Poles as in Bates *et al.* [5], is required.

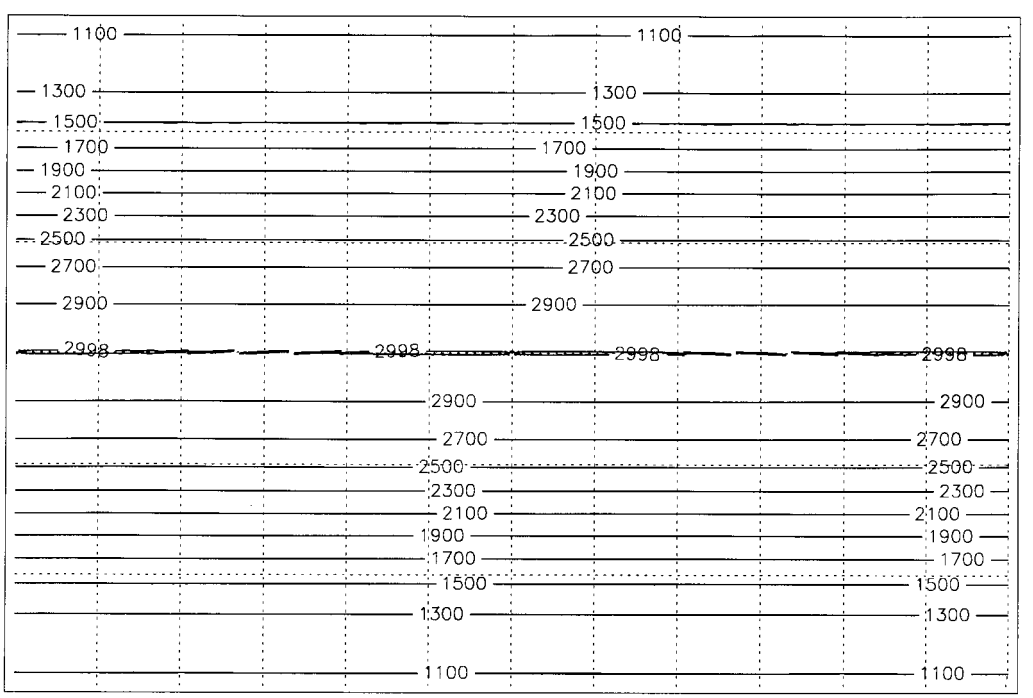

Figure 1. Contour lines of the true geopotential height field of the test problem.

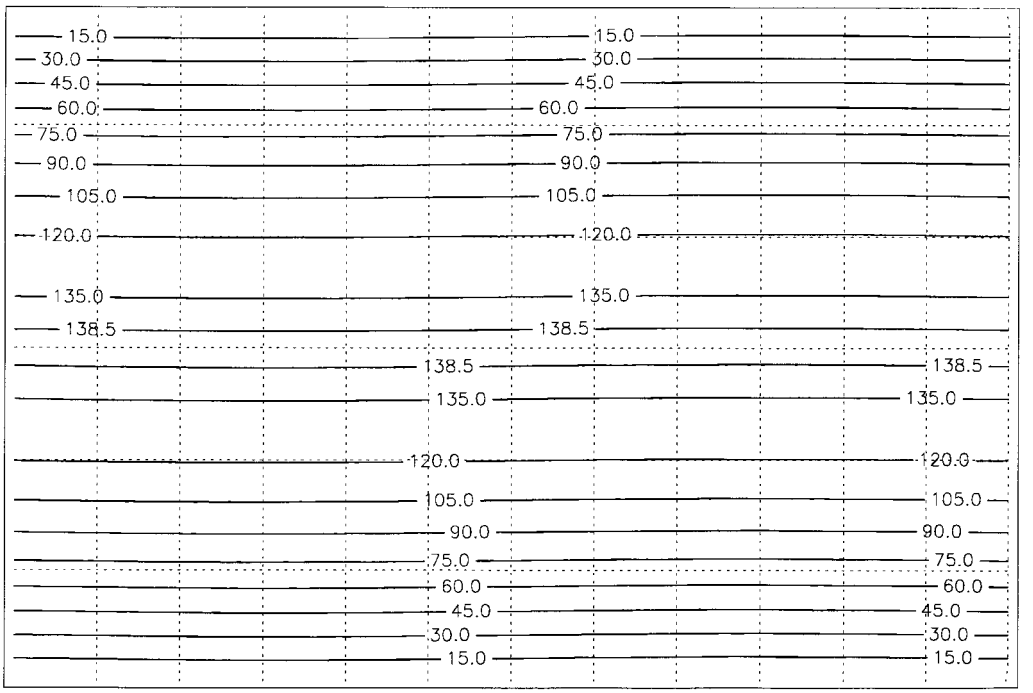

Figure 2. Contour lines of the true (zonal) velocity field of the test problem.

4.2. Analysis of the algebraic system

To obtain the numerical solution of system (17) we have to order the $N_x \cdot N_y$ equations for the unknowns $\Phi_{i,j}$ and to solve a linear system.

Therefore, we define the matrix A of the coefficients of the system by the following notations:

- row $\mathbf{a_k}$ contains the coefficients of all the unknowns that appear in equation (i, j), with $k = (i-1)N_y + j$;
- column $\mathbf{a_h}$ contains the coefficients of the unknown (i, j), with $h = (i-1)N_y + j$, in all the equations where it appears.

In each row of A there are at most five non-zero entries; they could be obtained directly from system (17); however, we write them out here for completeness, referring to Equations (12) for the detailed description of coefficients $C_{i,j}$ and $D_{i,j}$

$$a_{h,h} = A_j + \theta_1^2 \Delta t^2 (C_{i+1/2,j} + C_{i-1/2,j} + D_{i,j+1/2} + D_{i,j-1/2})$$

$$a_{h,h+N_y} = -\theta_1^2 \Delta t^2 C_{i-1/2,j} = a_{h+N_y,h}$$

$$a_{h,h+1} = -\theta_1^2 \Delta t^2 D_{i,j-1/2} = a_{h+1,h}$$

Then, matrix A is

- symmetric;
- diagonally dominant, as the quantity $a_{i,i} - \Sigma_{j \neq i} |a_{i,j}|$ representing the area A_k of a grid cell (with $k = i/N_y + 1$), is obviously positive;
- sparse and band-structured, as there are only seven diagonals with non-zero entries (for periodicity reasons, due to the spherical geometry, the five non-zero entries on each row spread out in seven diagonals instead of five).

Therefore, matrix A is positive-definite. For such a matrix, a standard iterative solver for large, sparse and symmetric linear systems, such as the conjugate gradient method, does not meet with serious difficulties.

Indeed, a convergence analysis of the conjugate gradient algorithm (Golub and van Loan [8]) shows that its time complexity is $O(N\sqrt{\mu})$, where $N = N_x \cdot N_y$ is the size of A and μ is the condition number of A, defined as $\|A\|_2 \|A^{-1}\|_2$. For our scheme we are able to prove the following result on the conditioning of the linear system.

Theorem 2
The condition number of matrix A is of polynomial type

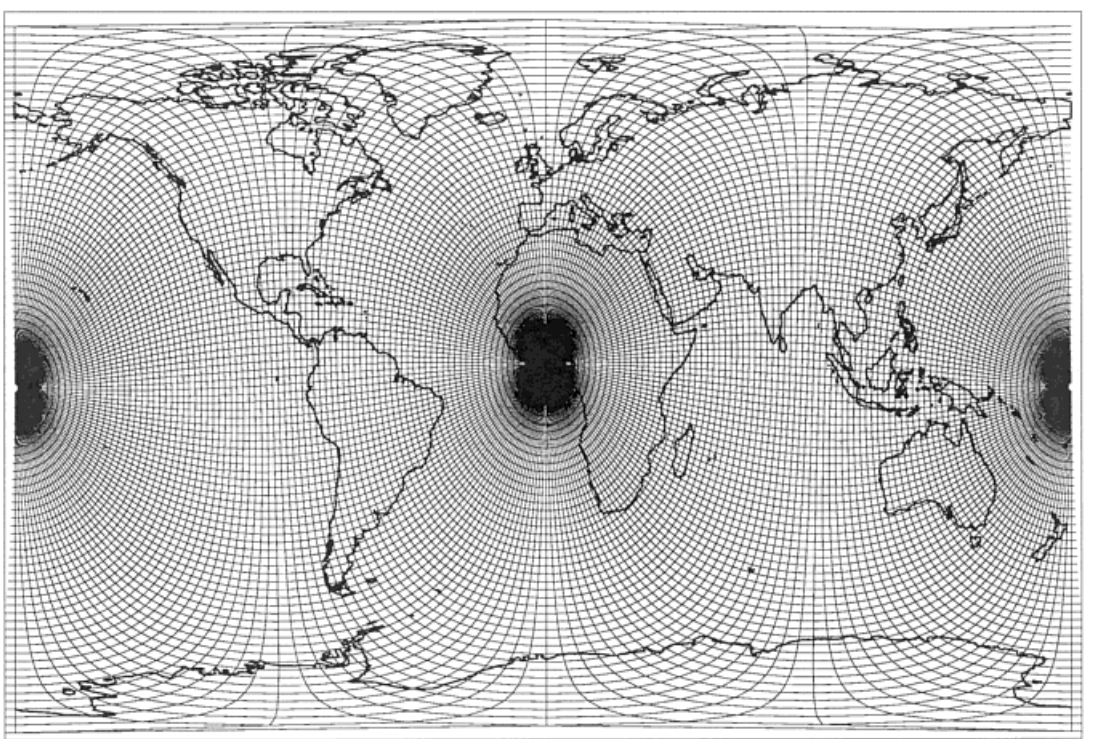

Figure 3. Fully rotated grid with resolution 180×90 grid points, NP latitude 0.05.

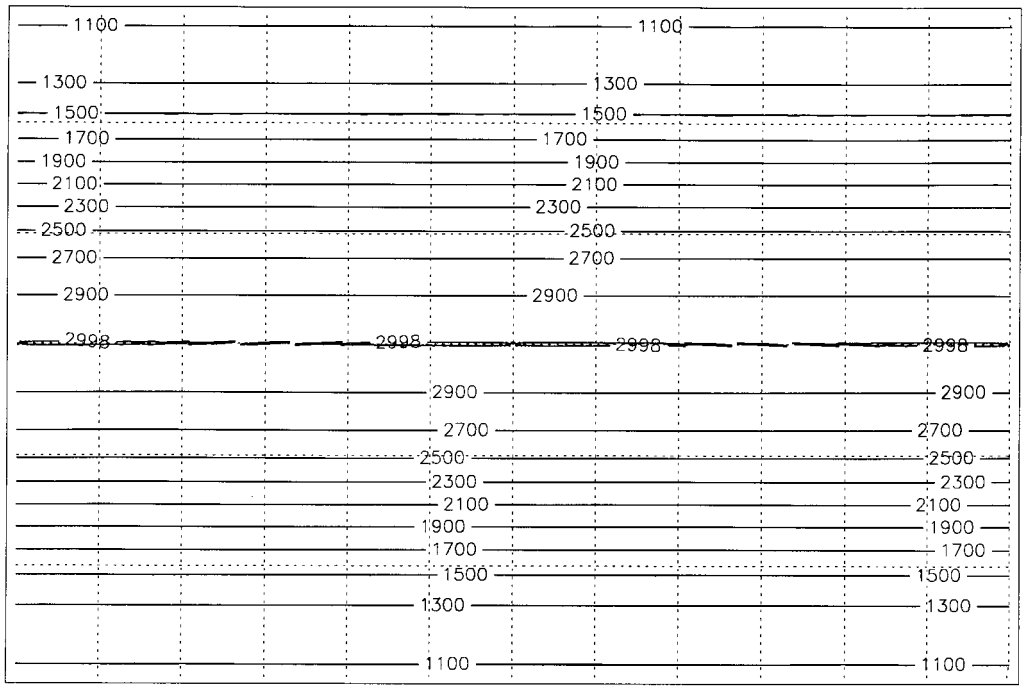

Figure 4. Contour lines of the error on the retrieved geopotential height after 5 days of simulation; slightly rotated grid 360×180, $\Delta t = 10$ min, Courant max ≈ 1.4.

Table I. Grid 180×90—NP latitude $\pi/2 - 0.05$ rad; $T_{\max} = 24$ h. First- and second-order scheme with cubic interpolation (Schemes 1 and 2): comparison of the error indicators for Φ.

θ	Δt	Scheme 1			Scheme 2		
		$l_1(\Phi)$	$l_2(\Phi)$	$l_\infty(\Phi)$	$l_1(\Phi)$	$l_2(\Phi)$	$l_\infty(\Phi)$
0	10	0.899e−3	0.978e−3	0.140e−2	0.936e−3	0.102e−2	0.141e−2
	30	0.282e−2	0.305e−2	0.422e−2	0.293e−2	0.316e−2	0.429e−2
	60	0.602e−2	0.652e−2	0.942e−2	0.623e−2	0.673e−2	0.960e−2
0.5	10	0.114e−1	0.125e−1	0.179e−1	0.114e−1	0.125e−1	0.179e−1
	30	0.326e−1	0.358e−1	0.541e−1	0.326e−1	0.358e−1	0.541e−1
	60	0.601e−1	0.666e−1	0.109	0.601e−1	0.666e−1	0.109
1	10	0.212e−1	0.233e−1	0.345e−1	0.212e−1	0.233e−1	0.345e−1
	30	0.572e−1	0.636e−1	0.106	0.573e−1	0.636e−1	0.106
	60	0.974e−1	0.109	0.199	0.975e−1	0.110	0.199

Table II. Grid 180×90—NP latitude $\pi/2 - 0.05$ rad; $T_{\max} = 24$ h. First- and second-order scheme with cubic interpolation (Schemes 1 and 2): comparison of the error indicators for **v**.

θ	Δt	Scheme 1			Scheme 2		
		$l_1(\mathbf{v})$	$l_2(\mathbf{v})$	$l_\infty(\mathbf{v})$	$l_1(\mathbf{v})$	$l_2(\mathbf{v})$	$l_\infty(\mathbf{v})$
0	10	0.307e−2	0.352e−2	0.452e−2	0.353e−2	0.380e−2	0.467e−2
	30	0.876e−2	0.982e−2	0.125e−1	0.100e−1	0.107e−1	0.131e−1
	60	0.145e−1	0.175e−1	0.242e−1	0.174e−1	0.196e−1	0.259e−1
0.5	10	0.368e−1	0.418e−1	0.532e−1	0.368e−1	0.418e−1	0.532e−1
	30	0.103	0.115	0.144	0.103	0.115	0.144
	60	0.190	0.209	0.255	0.190	0.209	0.255
1	10	0.675e−1	0.771e−1	0.976e−1	0.679e−1	0.773e−1	0.977e−1
	30	0.181	0.202	0.247	0.182	0.202	0.247
	60	0.315	0.344	0.408	0.317	0.345	0.408

Table III. Grid 180×90—$T_{\max} = 120$ h. Scheme 1: error indicators for Φ and **v** for different values of the implicitness parameter.

θ	Δt	Courant	$l_1(\Phi)$	$l_2(\Phi)$	$l_\infty(\Phi)$	$l_1(\mathbf{v})$	$l_2(\mathbf{v})$	$l_{1\infty}(\mathbf{v})$
NP latitude $\pi/2 - 0.05$								
0	10	0.4	0.446e−2	0.488e−2	0.724e−2	0.140e−1	0.163e−1	0.212e−1
	30	1.2	0.128e−1	0.141e−1	0.214e−1	0.414e−1	0.476e−1	0.609e−1
	60	2.5	0.268e−1	0.282e−1	0.377e−1	0.831e−1	0.939e−1	0.224
0.5	10	0.4	0.505e−1	0.562e−1	0.926e−1	0.134	0.183	0.226
	30	1.2	0.116	0.131	0.244	0.398	0.428	0.493
	60	2.5	0.165	0.187	0.351	0.595	0.627	0.695
1	10	0.4	0.859−1	0.965−1	0.173	0.284	0.312	0.371
	30	1.2	0.164	0.185	0.349	0.581	0.618	0.688
	60	2.5	0.195	0.221	0.407	0.748	0.783	0.825
NP latitude 0.05								
0	10	6.0	0.443e−2	0.485e−2	0.715e−2	0.139e−1	0.162e−1	0.213e−1
	30	18.0	0.127e−1	0.140e−1	0.213e−1	0.413e−1	0.475e−1	0.609e−1
	60	36.0	0.273e−1	0.287e−1	0.368e−1	0.939e−1	0.984e−1	0.194
0.5	10	6.0	0.505e−1	0.561e−1	0.926	0.163	0.182	0.226
	30	18.0	0.116	0.131	0.244	0.397	0.427	0.493
	60	36.0	0.165	0.186	0.351	0.593	0.627	0.698
1	10	6.0	0.859e−1	0.964e−1	0.173	0.283	0.311	0.372
	30	18.0	0.164	0.185	0.349	0.584	0.618	0.688
	60	36.0	0.195	0.221	0.407	0.750	0.784	0.855

Table IV. Grid 360×180 — $T_{\max} = 120$ h. Scheme 1: error indicators for Φ and $\mathbf{v}$ for different values of the implicitness parameter.

θ	Δt	Courant	$l_1(\Phi)$	$l_2(\Phi)$	$l_\infty(\Phi)$	$l_1(\mathbf{v})$	$l_2(\mathbf{v})$	$l_\infty(\mathbf{v})$
NP latitude $\pi/2-0.05$								
0	10	1.4	0.466e−2	0.508e−2	0.724e−2	0.161e−1	0.175e−1	0.218e−1
	30	4.2	0.133e−1	0.146e−1	0.215e−1	0.477e−1	0.514e−1	0.623e−1
	60	8.5	0.269e−1	0.284e−1	0.332e−1	0.996e−1	0.104	0.174
0.5	10	1.4	0.566e−1	0.562e−1	0.927e−1	0.164	0.183	0.266
	30	4.2	0.116	0.131	0.244	0.398	0.428	0.493
	60	8.5	0.165	0.187	0.267	0.595	0.628	0.695
1	10	1.4	0.861e−1	0.966e−1	0.173	0.285	0.312	0.372
	30	4.2	0.164	0.185	0.349	0.585	0.619	0.688
	60	8.5	0.195	0.221	0.407	0.753	0.784	0.853
NP latitude 0.05								
0	10	24	0.457e−2	0.500e−2	0.722e−2	0.155e−1	0.170e−1	0.216e−1
	30	73	0.131e−1	0.144e−1	0.215e−1	0.464e−1	0.505e−1	0.229
	60	146	0.273e−1	0.287e−1	0.354e−1	0.972e−1	0.101	0.929
0.5	10	24	0.505e−1	0.561e−1	0.926e−1	0.163	0.182	0.226
	30	73	0.116	0.131	0.244	0.397	0.428	0.494
	60	146	0.165	0.186	0.351	0.597	0.627	0.970
1	10	24	0.860e−1	0.965e−1	0.173	0.285	0.312	0.372
	30	73	0.164	0.185	0.349	0.584	0.619	0.689
	60	146	0.195	0.221	0.407	0.754	0.785	0.965

$$\mu(A) = O(N^2)$$

where N is the dimension of A.

As a consequence of this statement, the number of iterations required in the solution of the linear system by conjugate gradients is $O(N)$.

The proof of this theorem is quite tedious and is left to Appendix A.

A suitable conjugate gradient (CG) solver has been developed, optimized for the structure of the matrix at hand. Some preconditioners were also considered to speed up convergence, but they did not prove effective, as expected from the presented analysis of the condition number of the matrix. Owing to the large dimension of A, we just preprocessed the elements of the matrix, in order to increase the computational accuracy. Then we introduced a simple diagonal preconditioner, $D = \mathrm{diag}(a_{i,i}^{-1/2})$ and we define

$$\hat{A} = DAD$$

Matrix $\hat{A}$ is still symmetric and positive definite; moreover $\hat{a}_{i,i} = 1$ for each i.

5. NUMERICAL EXPERIMENTS

In this section we show experiments performed with the new numerical scheme devised based on two test problems proposed by Williamson *et al.* [9]. In the last few years, this set had been used for the validation of several shallow waters numerical models of the atmosphere (e.g. Swarztrauber [10]). Our first test is a steady state solution to the non-linear shallow water equations, i.e. a solid body rotation or zonal flow with the corresponding geostrophic height field. It involves the complete set of equations and has been chosen also to provide a benchmark for computer timing. The test is comprised of an initial height profile (for simplicity, a cosine bell), which rotates with constant angular velocity Ω around the Earth axis (through the Poles) and we consider this rotation in a spherical co-ordinate system (λ, φ), having its North Pole at point P (not coinciding with the physical North Pole (NP) in general). If $(0, \varphi_0)$ are NP co-ordinates in this system, the analytical solution to this test problem is given by

$$\Phi = \Phi_0 - \left[\Omega R u_0 + \frac{u_0^2}{2}\right](\cos\lambda\cos\varphi\sin\varphi_0 + \sin\varphi\cos\varphi_0)^2 \tag{23}$$

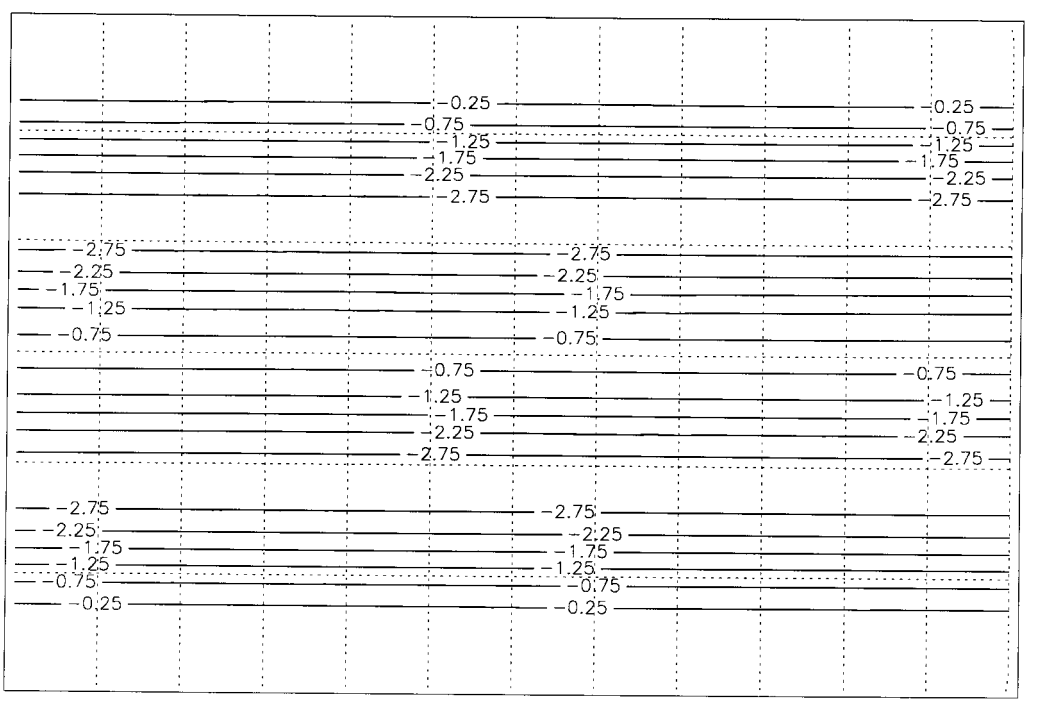

Figure 5. Contour lines of the error on the retrieved velocity (u component) after 5 days of simulation for the case of Figure 4.

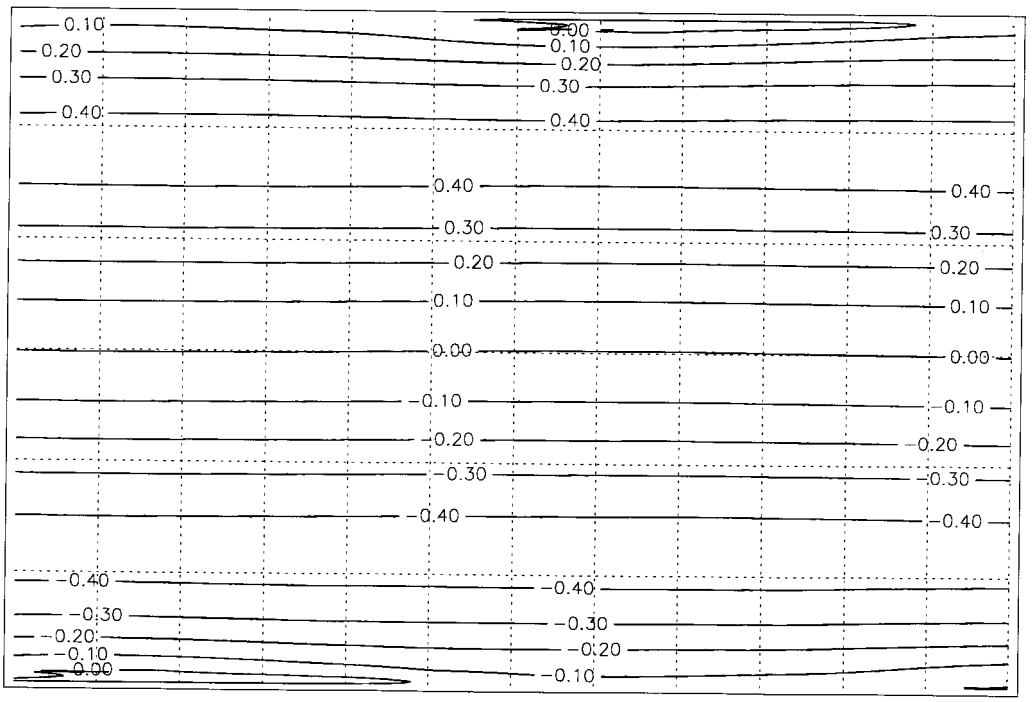

Figure 6. Contour lines of the absolute error on the retrieved velocity (v component) after 5 days of simulation for the case of Figure 4.

whereas wind components are

$$u = u_0[\cos\varphi\ \cos\varphi_0 + \cos\lambda\ \sin\varphi\ \sin\varphi_0]$$

$$v = -u_0 \sin\lambda\ \sin\varphi_0 \tag{24}$$

and where $u_0 = 2\pi R/12$ days and $\Phi_0 = 2.94 \times 10^4\ \mathrm{m^2\,s^{-2}}$.

Figures 1–3 show contour lines of the analytical solution for Φ and u in the absence of grid rotation.

We tested our model on a slightly rotated grid ($\varphi_0 = \pi/2 - 0.05$ rad) and on a full rotated grid ($\varphi_0 = 0.05$ rad). This is Problem 2 described in Williamson *et al.* [9].

In both cases, executions were made for several values of Δt and for two grid resolutions: 180×90 grid points (corresponding to 2° resolution in both directions) and 360×180 grid points (1° resolution). Figure 4 shows an example of rotated grid with 2° resolution.

The global relative errors on the retrieved fields in L_1-, L_2- and L_∞-norms (indexes l_1, l_2, l_∞) have been calculated after 5 days of simulation. We recall here their definition for the case of Φ

$$l_1 = \frac{\sum_{i,j} |\Phi_{i,j} - \Phi_{i,j}^{\text{true}}|}{\sum_{i,j} |\Phi_{i,j}^{\text{true}}|}$$

$$l_2 = \left(\frac{\sum_{i,j} (\Phi_{i,j} - \Phi_{i,j}^{\text{true}})^2}{\sum_{i,j} (\Phi_{i,j}^{\text{true}})^2} \right)^{1/2}$$

$$l_\infty = \max_{i,j} \frac{|\Phi_{i,j} - \Phi_{i,j}^{\text{true}}|}{|\Phi_{i,j}^{\text{true}}|}$$

The same error measures indicators $l_1(v)$, $l_2(v)$, $l_\infty(v)$ were evaluated for the velocity field.

Tables I and II compare the accuracy of the two schemes for the evaluation of the Lagrangian terms: runs were made for several time steps values and implicitness parameters and confirm that the second-order in time scheme does not introduce significant improvements, as expected for this stationary test case. As a result, in the subsequent tests we

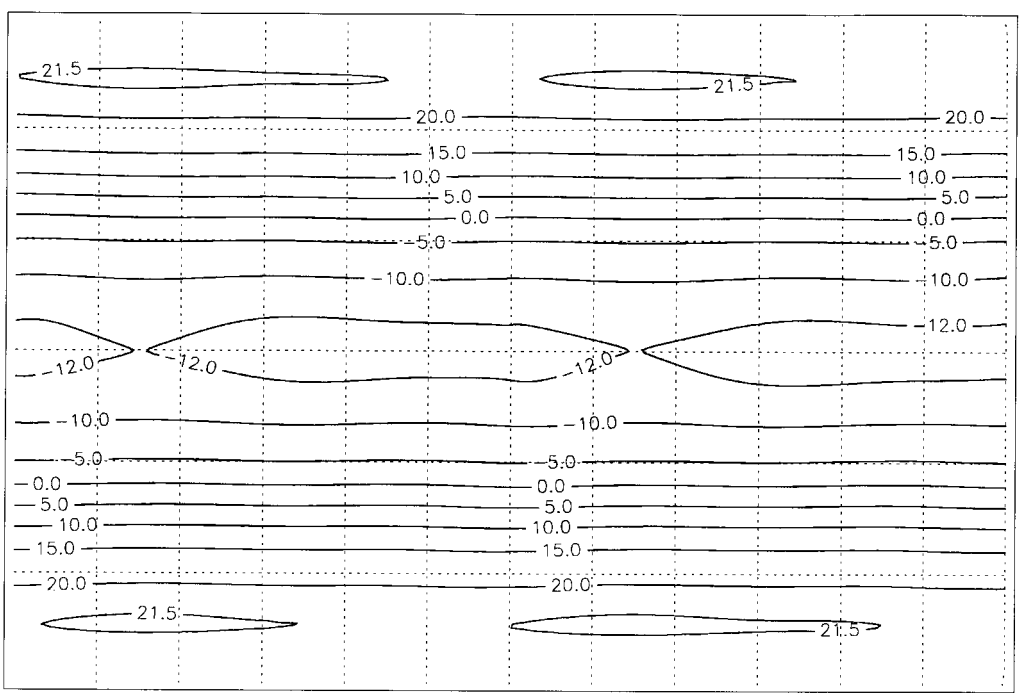

Figure 7. Contour lines of the error on the retrieved geopotential height after 5 days of simulation; fully rotated grid 360×180, $\Delta t = 10$ min, Courant max ≈ 24.

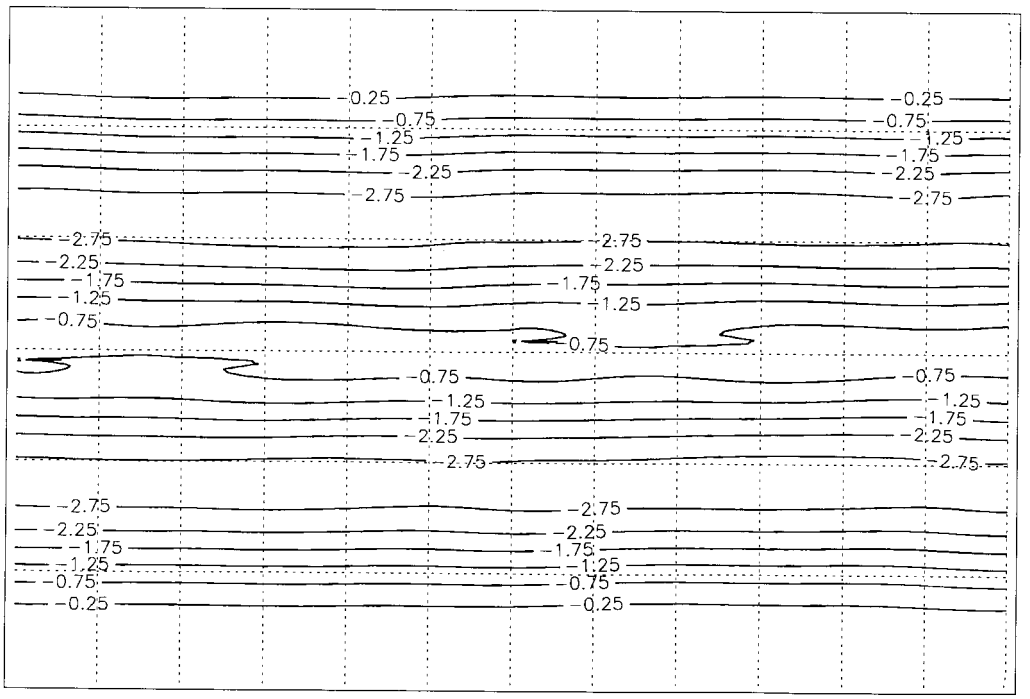

Figure 8. Contour lines of the error on the retrieved velocity (u component) after 5 days of simulation for the case of Figure 7.

considered only the first-order scheme. Tables III and IV show that the accuracy of the method does not depend on the rotation of the numerical grid (as in the tests shown in Amato and Carfora [7]). The case of a full-rotated grid, with higher Courant numbers, gives the same error levels of a slightly rotated grid. The same tables also compare the effects of the choice of the implicitness parameter: in partial disagreement with the theoretical stability analysis, the best performance is for $\theta_1 = 0$, i.e. for explicit Coriolis terms. Indeed, the instability due to Coriolis terms leads to very small effects, as proved by the estimate of the amplification matrix given in Theorem 1 (22), due to the order of the Coriolis parameter: $f = O(10^{-5})$. Only in experiments where the angular velocity of Earth rotation was increased by two or three orders of magnitude we have seen the effects of this instability.

Figures 4–6 show the error fields (difference between retrieved and true values) for the retrieved geopotential height and wind components in the case of a slightly rotated grid (NP latitude $\pi/2 - 0.05$ rad); Figures 7–9 show the corresponding error fields in the case of a fully rotated grid (NP latitude 0.05 rad).

It is possible to see that the main indicators (maximum, minimum and location of the error) are quite the same, not depending on the grid rotation: this also means that the error is not concentrated near the representational poles. In particular, for the geopotential height field,

the retrieved solution is smoother than the true one, with higher minima and lower maxima; for the u component of the wind field there is in general a small underestimation, while for the v component the errors are quite evenly distributed. The main effect of the grid rotation, apart from some perturbation near the representational poles (see the centre of Figures 8 and 9, corresponding to the numerical NP), is a modification in the shape of the error fields, particularly in the v component of wind velocity; however, such a modification does not affect the level of the error, always between -0.4 and $+0.4$ m s^{-1}.

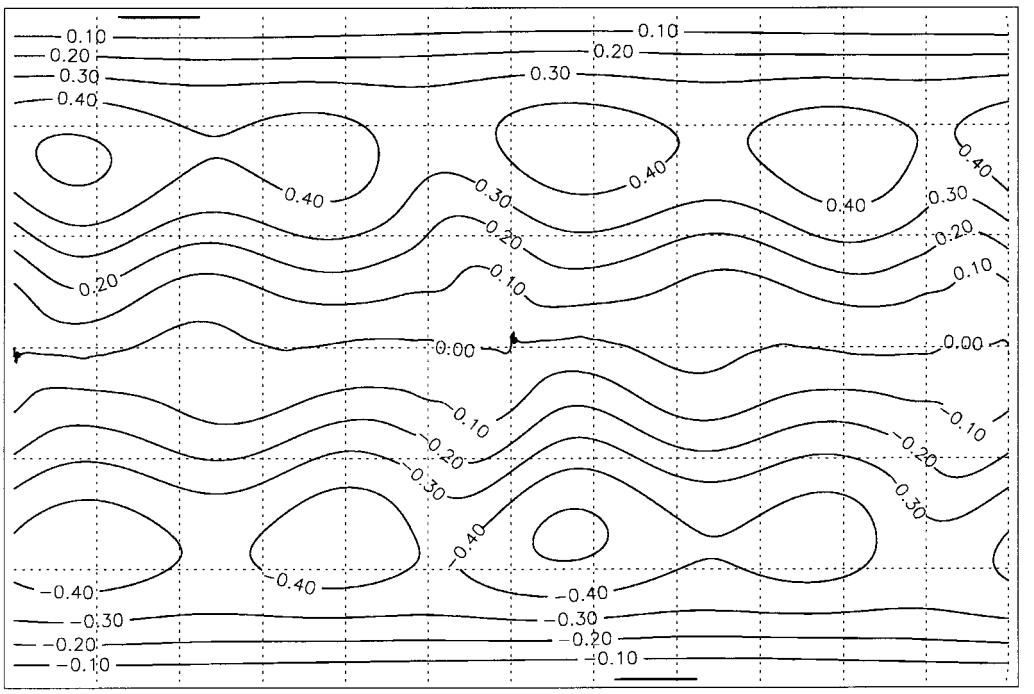

Figure 9. Contour lines of the absolute error on the retrieved velocity (v component) after 5 days of simulation for the case of Figure 7.

Table V. Execution time depending on Δt. Grid resolution is 180×90, NP latitude $\pi/2 - 0.05$.

Δt (min)	Total time (s)	Lagrangian phase (s)	System solving (s)
5	750	306	388
10	560	156	373
20	450	80	356
40	350	62	278

Table VI. Execution time depending on Δt. Grid resolution is 180×90, NP latitude 0.05.

Δt (min)	Total time (s)	Lagrangian phase (s)	System solving (s)
5	1278	657	565
10	1188	596	563
20	1101	558	528
40	928	544	376

Table VII. Execution time depending on grid resolution. Time step is 10 min, NP latitude $\pi/2 - 0.05$.

Grid points	Total time (s)	Lagrangian phase (s)	System solving (s)
90×45	80	37	37
180×90	560	156	373
360×180	5060	990	3915

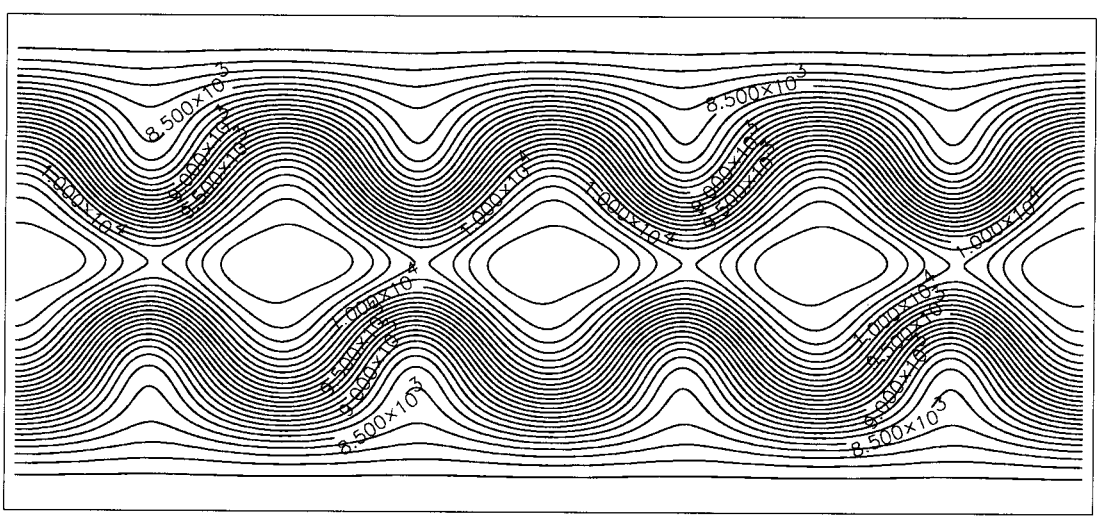

Figure 10. Contour lines of the initial solution for the Rossby–Haurwitz test problem.

To conclude with this test case, let us present some results on the computational cost of the method. It is not possible to simply define a 'cost per cycle' since in every cycle of the numerical integration n sub-cycles are performed, where n is approximately equal to the maximum value of Courant numbers on all grid points. Then it depends on the chosen time

step, grid resolution and also on grid rotation. To evaluate the cost of the method, we performed three set of runs: in the first set only the time step varies, while the grid resolution is constant (180×90 grid points) and NP latitude is $\pi/2 - 0.05$ rad; in the second set again the

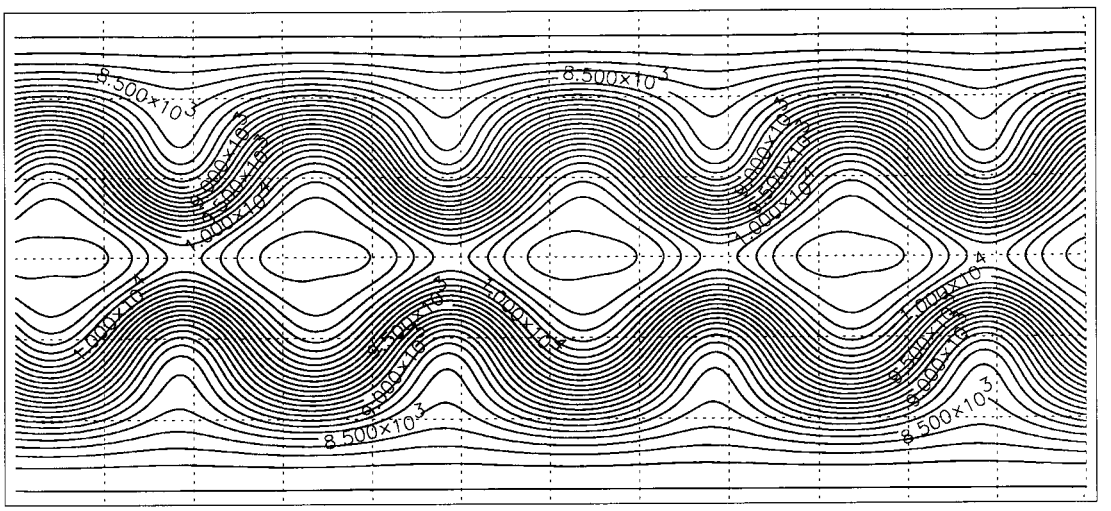

Figure 11. Contour lines of the retrieved solution after 1 day of integration.

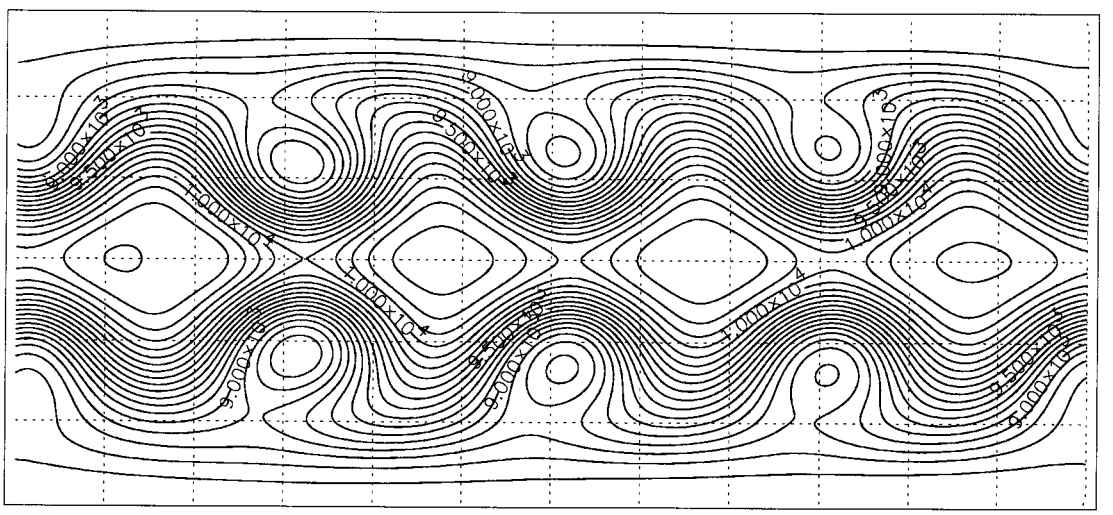

Figure 12. Contour lines of the retrieved solution after 7 days of integration.

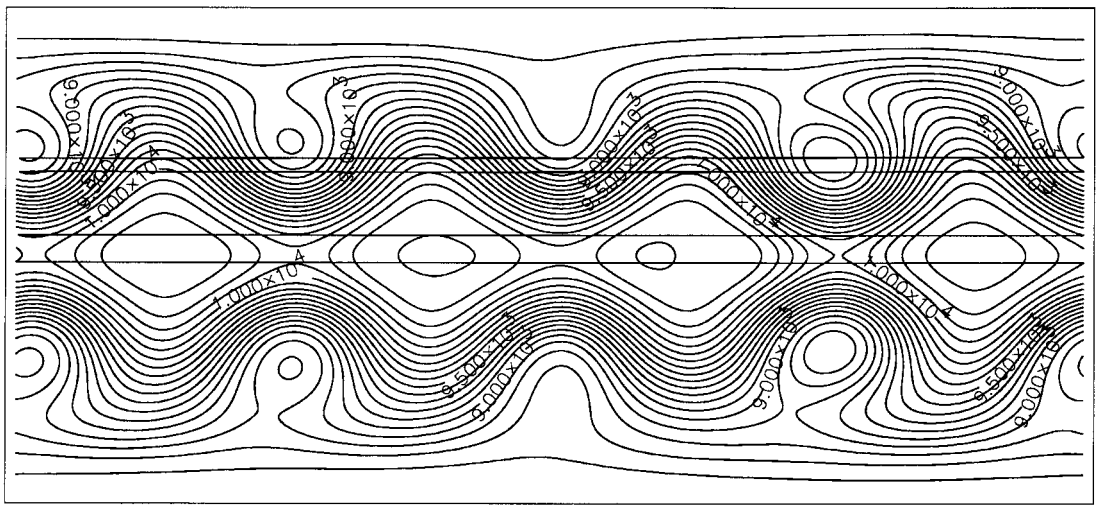

Figure 13. Contour lines of the retrieved solution after 10 days of integration.

time step varies, but NP latitude is 0.05 rad to show the effects of high Courant numbers; in the third set only the grid resolution varies, while time step (10′) and NP latitude (0.05) are fixed.

Tables V, VI and VII show the execution time for the complete simulation ($T_{\max} = 120$ h) for the three set of runs. Tests were performed on an Alpha Server 2100/250. This machine, if used as single processor, is rated at about 120 Mflop/s in the LINPACK benchmark.

It is possible to see (Table V) that, when Courant numbers are small, the computational cost strongly decreases with the increasing time step: indeed, the evaluation of the Lagrangian terms in the method does not require the sub-stepping procedure and the cost of this evaluation depends (almost linearly) just on the performed cycles; while the cost for the solution of the linear system is quite not dependent on the time step. This trend is completely confirmed by the first three rows in Table V; the case $\Delta t = 40'$ is a little bit different: since in this case the Courant numbers are somewhere greater than one, the use of the sub-stepping procedure slightly increases the cost of the run.

Table VI presents the corresponding test for a different NP latitude: the same set of runs, when Courant numbers are higher, shows a different trend: when the sub-stepping procedure is widely used, the cost for the evaluation of the Lagrangian terms is no more linear decreasing with the time step. However, it must be stressed that, also in this test where Table VII shows the cost variations with increasing grid resolution: as could be expected, since the number of grid points grows by a factor four, and the Courant numbers doubles, the cost approximately increases by a factor eight.

The second group of results we present refer to a more complicated test problem, a planetary scale Rossby–Haurwitz wave dominated by wave number four (Test no. 6 in

Williamson *et al.* [9]). This is a frequently used meteorological test, but no analytical solution is known. As the solution evolves, the flow field does not maintain its initial structure, due to the impact of the horizontal resolution. Then, as suggested in Reference [9], we qualitatively compare the results of the numerical integration with the reference solutions presented in the literature. Figure 10 shows the initial state.

The time step for our integration is 180 s and the grid resolution is 1° (360×180 grid points). Figures 11–13 show the retrieved solution for day 1, day 7 and day 10. As can be seen in these maps, the phase of the wave structure is quite well represented, although there is a tendency for some sharpening in the mid-latitude. The error field structure is very similar to the literature tests, with a general erosion of zonal and meridional gradients.

With regard to conservation properties, the runs confirm the theoretical results: there is an almost exact conservation (within the machine rounding) of the total geopotential height during the iteration.

6. CONCLUSIONS

The present paper introduces a numerical method for the shallow water equations in the atmosphere and evaluates the impact of its specific features on the overall performance of the model. The numerical method is based on a mixed finite difference–finite volume approximation of the equations, which results in conservation of the total geopotential height. The momentum equation is discretized in its vectorial form. The advective terms are treated in a semi-Lagrangian way and pressure and Coriolis terms in a semi-implicit way. A splitting procedure is invoked to preserve symmetry of the algebraic system of equations. Finally, the introduction of a sub-stepping procedure for the evaluation of the Lagrangian terms allows us to increase the spatial and temporal accuracy. The resulting method is stable for any choice of the time step provided that the implicitness parameter is in [0.5, 1]. The resulting matrix is sparse and structured, symmetric and positive definite, which makes iterative methods for solving the system of equations very attractive.

As far as the different improvements are concerned, our conclusions are that the splitting procedure is considerably useful to simplify the numerical scheme, whereas the vectorial correction in the momentum equation has been shown to be of minor importance, due to the sub-stepping procedure. For the reconstruction of the characteristic lines, the improvement due to accurate spatial interpolation is evident; this is not the case for the second-order in time algorithm. These results completely confirm those obtained in the study of semi-Lagrangian treatment of advection on the sphere [7].

ACKNOWLEDGMENTS

The author wants to thank Dr Amato for many useful discussions and two anonymous referees, whose comments were very helpful in improving a previous version of this paper.

APPENDIX A

Proof of Theorem 2
Since A is positive-definite, if we indicate its eigenvalues, using v, it is

$$\mu(A) = \frac{v_{\max}}{v_{\min}}$$

Then, to locate the eigenvalues of A we consider the Gerschgorin circles

$$\mathscr{C}_j = \left\{ z \in C:\ |z - a_{j,j}| \leq \sum_{k \neq j} |a_{j,k}| \right\}$$

For simplicity's sake, we assume that H and F are constants in Equation (12); therefore, all the coefficients will depend only on index j. Since the spectrum of A is contained in the union of all the $\mathscr{C}_j$, it is straightforward to see, from the definition of $a_{j,j}$ and $a_{j,k}$, that

$$\mu(A) \leq \frac{\max_j (A_j + 2\theta^2 \Delta t^2 \alpha_j)}{\min_j A_j}$$

where we have posed

$$\alpha_j = C_{1+1/2,\,j} + C_{i-1/2,\,j} + D_{i,\,j+1/2} + D_{i,\,j-1/2} = 2\frac{H}{F}\,\Delta x_j \Delta y \left(\frac{1}{\Delta x_j^2} + \frac{1}{\Delta y^2} \cos\frac{\Delta\varphi}{2} \right)$$

Now, since A_j is the area of a cell centred at latitude φ_j, as evaluated in Equation (4), it is clear that

$$\min_j A_j = A_1 = 2R \sin\frac{\Delta\varphi}{2}\,\Delta x_1$$

Moreover, since it is

$$A_j + 2\theta^2 \Delta t^2 \alpha_j = \left[2R \sin\frac{\Delta\varphi}{2} + 4\frac{H}{F}\,\theta^2 \frac{\Delta t^2}{\Delta y} \cos\frac{\Delta\varphi}{2} \right] \Delta x_j + 4\frac{H}{F}\,\theta^2 \frac{\Delta t^2 \Delta y}{\Delta x_j} \tag{25}$$

this quantity is a convex function of Δx_j, which attains its absolute maximum at one of the extrema $j = 1$ or $j = N_y/2$.

Therefore, we have just to compare $A_1 + 2\theta^2 \Delta t^2 \alpha_1$ with $A_{N_y/2} + 2\theta^2 \Delta t^2 \alpha_{N_y/2}$.

Recalling that $\Delta x_j = R\Delta\lambda \cos\varphi_j$ we have

$$\Delta x_1 = R\Delta\lambda \sin\frac{\Delta\varphi}{2}, \qquad \Delta x_{N_y/2} = R\Delta\lambda \cos\frac{\Delta\varphi}{2}$$

and it is easy to prove that the maximum of Equation (25) is reached for $j = 1$. Indeed, after some manipulation, one can see that, in the case $\theta \neq 0$, the inequality

$$A_1 + 2\theta^2 \Delta t^2 \alpha_1 \geq A_{N_y/2} + 2\theta^2 \Delta t^2 \alpha_{N_y/2}$$

is equivalent to the identity

$$\frac{R^2 F \Delta\varphi}{2H\theta^2 \Delta t^2} \sin^2 \frac{\Delta\varphi}{2} \left(1 - \cos \frac{\Delta\varphi}{2}\right) + \cos^2 \frac{\Delta\varphi}{2} \left(1 - \sin \frac{\Delta\varphi}{2}\right) \geq 0$$

where the only simplification we made is $\Delta\lambda = \Delta\varphi$.

Thus, it is

$$\mu(A) \leq \frac{A_1 + 2\theta^2 \Delta t^2 \alpha_1}{A_1}$$

and finally, using that $\Delta\varphi = \pi/N_y$ and making the usual approximations for increasing N_y

$$\sin \frac{\Delta\varphi}{2} \approx \frac{\Delta\varphi}{2}, \qquad \cos \frac{\Delta\varphi}{2} \approx 1 - \left(\frac{\Delta\varphi}{2}\right)^2$$

we obtain

$$A_1 = 4R^2 \left(\frac{\pi}{2N_y}\right)^3$$

$$\alpha_1 = 2 \frac{H}{F} \left[\frac{2N_y}{\pi} + \left(\frac{\pi}{2N_y}\right)\left(1 - \left(\frac{\pi}{2N_y}\right)^2\right)\right]$$

so that

$$\mu(A) = O\left(\frac{N_y + N_y^{-1} + N_y^{-3}}{N_y^{-3}}\right) = O(N_y^4) = O(N^2)$$

REFERENCES

1. Staniforth A, Côté J. Semi-Lagrangian Integration Schemes for atmospheric models—a review. *Monthly Weather Review* 1991; **119**: 2206–2223.
2. Bartello P, Thomas SJ. The cost-effectiveness of semi-Lagrangian advection. *Monthly Weather Review* 1996; **124**: 2883–2897.
3. McDonald A, Bates JR. Semi-Lagrangian integration of a gridpoint shallow water model on the sphere. *Monthly Weather Review* 1989; **117**: 130–137.
4. Bates JR, Moorthi S, Higgins RW. A global multilevel atmospheric model using a vector semi-Lagrangian finite-difference scheme. Part I: adiabatic formulation. *Monthly Weather Review* 1993; **121**: 244–263.

5. Bates JR, Semazzi FHM, Higgins RW, Barros SRM. Integration of the shallow water equation on the sphere using a vector semi-Lagrangian scheme with a multigrid solver. *Monthly Weather Review* 1990; **118**: 1615–1627.
6. Casulli V. Semi-implicit finite difference methods for the two-dimensional shallow water equations. *Journal of Computational Physics* 1990; **86**: 56–74.
7. Amato U, Carfora MF. Semi-Lagrangian treatment of advection on the sphere with accurate spatial and temporal approximations. *Mathematical and Computer Modelling* 2000 (in press).
8. Golub GH, van Loan CF. *Matrix Computations* (2nd edn). Johns Hopkins University Press: Baltimore, MD, 1989.
9. Williamson DL, Drake JB, Hack JJ, Jakob R, Swarztrauber PN. A standard test set for numerical approximations to the shallow water equations in spherical geometry. *Journal of Computational Physics* 1992; **102**: 211–224.
10. Swarztrauber PN. Spectral transform methods for solving the shallow water equations on the sphere. *Monthly Weather Review* 1996; **124**: 730–744.

