

On Simulating Flows with Multiple Time Scales Using a Method of Averages¹

B.T. Nadiga, M.W. Hecht, and L.G. Margolin

Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A.

P.K. Smolarkiewicz

National Center for Atmospheric Research, Boulder, CO 80307, U.S.A.

Communicated by H.J.S. Fernando

Received 25 September 1996 and accepted 11 April 1997

Abstract. We present a new method, based on averaging, to simulate certain systems with multiple time scales efficiently and demonstrate its utility in the context of the shallow-water equations. We first develop the method in a simple linear setting and analytically prove its stability. This is followed by an extension to the full equations and a presentation of a computational model for it. In this preliminary study, we find that the new method produces results that are very close to a fully explicit (spatially and temporally) second-order accurate scheme and much better than a fully explicit (spatially and temporally) first-order accurate scheme, while costing less than the first-order accurate scheme.

1. Introduction

The simulation of physical processes with multiple time scales provides a severe challenge to the numerical modeler. Fully explicit methods must employ the most restrictive time step associated with ensuring the stability of the fastest waves, and so may be prohibitively expensive. Thus, in addition to implicit and semi-implicit methods, a variety of alternate strategies have been developed, each more specially adapted for particular flows. The purpose of this paper is to describe a new method for treating flows with multiple time scales. We have chosen to illustrate the methodology in the context of shallow fluid approximations of ocean-basin circulations. Nevertheless, the method itself does not depend on the details of such flows and is likely to be more generally applicable to other problems with multiple time scales.

The circulations in the atmosphere and the oceans, as approximated by even the highly simplified shallow-water equations, are two examples of flows with multiple time scales. In both cases the fastest waves are the gravity waves, with speeds up to a few hundred meters per second. In the atmosphere, material motions tend

¹ The support of the Computer Hardware, Advanced Mathematics and Model Physics (CHAMMP) research program (within the U.S. Global Change Research Program), and the Institute for Geophysics and Planetary Physics (IGPP) at Los Alamos National Laboratory is acknowledged. MH was supported through the Global Change Distinguished Postdoctoral Fellowship program, sponsored by the U.S. Department of Energy (DOE), Office of Health and Environmental Research, administered by the Oak Ridge Institute for Science and Education. The Los Alamos National Laboratory is supported by the U.S. Department of Energy and the National Center for Atmospheric Research is sponsored by the National Science Foundation.

to be slower by a factor of about four, whereas material motions in the ocean are much slower. This has led to different algorithmic strategies—atmospheric models typically use split-explicit methods whereas ocean models depend on modal decomposition.

Split-explicit methods take advantage of the separation between the fast modes and slow modes by splitting the exact solution operator into fast and slow parts and using small time steps only on the fast part. Such a method then requires less work than using small time steps on the full unsplit problem provided the accuracy is not too adversely affected by the splitting errors (LeVeque and Olinger, 1983). These methods have a long lineage and are based on variations of schemes originally proposed by Marchuk (1974) and Strang (1968). Of particular interest to the meteorological community have been the splitting schemes of Klemp and Wilhelmson (1978; KW) and Mesinger *et al.* (1988). Analyses of the destabilizing effects of the splitting errors and the reduction in the order of accuracy due to a consequent use of filters, along with efficiency issues of these methods have been considered by many workers and the interested reader is referred to LeVeque and Olinger (1983), Skamarock and Klemp (1994), Browning and Kreiss (1994), and references therein.

The underlying idea of modal decomposition is the separation of the flow into a fast and a slow part (Bryan, 1969). Based on physical reasoning, the fast motions are identified with the column-averaged motion. Because this part of the flow, termed the barotropic component, is two-dimensional, either implicit methods coupled with efficient elliptic solvers (Dukowicz *et al.*, 1993) or explicit subcycling methods (Bleck and Smith, 1990) can be used to solve the approximate equations governing these motions. The remaining part of the flow, termed the baroclinic component, is then integrated explicitly at a time step nearly 20 times larger. Both strategies lead to significant improvements in computational efficiency over purely explicit calculations.

The identification of the fast motions with the column-averaged motions is based on a eigenmode analysis of the linearized equations. In reality, the barotropic and baroclinic motions are coupled by the nonlinearity of the equations, the bottom topography, and the free surface boundary conditions. This coupling raises the possibility of “bleeding” of the fast motions into the equations governing the slow motions. For example, the theoretical and computational studies of Higdon and Bennett (1996) point to such a bleeding as a possible reason for the large amount of time filtering that is necessary to ensure stability of computations in the Miami Isopycnic Coordinate Ocean Model (MICOM).

In this paper we propose an alternate computational approach that is *not* based on modal decomposition. As discussed by Browning and Kreiss (1987), when the time step is restricted by the fast waves, errors in geophysical fluid dynamic simulations are dominated by spatial errors. Thus, our method begins with a *low-order* integration of a set of equations which allows for both the fast modes and the primary balances of the original system. At this stage the fast modes of the system are explicitly resolved and the evolution is continued for \mathcal{M} steps, where \mathcal{M} is approximately the ratio of the slow time scale to the fast time scale. This provisional low-order solution is then averaged in time over the \mathcal{M} steps to provide a representative slow time scale solution. This solution in turn is used in the high-order accurate evolution of the slow time scale variables over the (single) long time step. While the choice of the set of approximate equations that is used in the initial step is important and will determine the computational efficiency of the overall scheme, in this preliminary study we focus on demonstrating the realizability of the strategy. We use the full equation unless otherwise specified. The choice of more computationally efficient equations for the initial pass will be the subject of future research. We note that Madala (1981) considered a splitting scheme in the context of leapfrog time discretization that has similarities to the present scheme, but that an analysis of its stability and accuracy properties led to the conclusion that there were no advantages over the KW splitting (Skamarock and Klemp, 1992). (This work was brought to our attention only after initial computational verifications of our scheme.) The very different nature of the truncation errors and stability properties of a three-time-level scheme—the framework in which Madala constructed his splitting scheme—and the present framework of a two-time-level scheme render a direct comparison inconclusive.

Although the methodology is likely to be of greater generality, we illustrate it here in the context of an idealized ocean problem, which is described in the next section. In Section 3 we develop the method first in the highly simplified context of one-dimensional gravity wave propagation and demonstrate its linear stability. This is followed by a generalization of the method to the full nonlinear shallow-water equations on a β -plane. Results of simulations employing the method of averaging are presented and compared with the fully explicit solutions in Section 4. A few concluding remarks complete the paper in Section 5.

2. Monopole Evolution in a Closed β -Plane Ocean Basin

Milliff and McWilliams (1994) recently considered the evolution of a monopole vortex in a closed rectangular β -plane ocean basin in their study elucidating the boundary pressure adjustment process on an ocean-basin scale. This problem embodies the prototypical interactions that exist in the ocean between the fast, boundary-trapped signals—the primarily gravitational Kelvin waves—and the slower, open-ocean bulk modes—the Rossby waves. It is important that a model, either in which the physics of the gravity wave modes are not fully treated or in which the gravity wave modes are computationally distorted, allows for the above physical interactions between the fast and slow modes. Hence our choice of problem. We note, however, that the primary goal of the paper is to demonstrate the new methodology, and not to illustrate any advantage in accuracy of this method over alternate approaches.

2.1. The Single-Layer Reduced-Gravity Approximation

We consider the problem in the simplified context of the single-layer reduced-gravity shallow-water equations on a β -plane:

$$\frac{\partial \eta}{\partial t} + \nabla \cdot \eta \mathbf{u} = 0, \quad (1a)$$

$$\frac{\partial \mathbf{q}}{\partial t} + \nabla \cdot \mathbf{q} \mathbf{u} = -g' \eta \nabla \eta - f \eta \hat{\mathbf{z}} \times \mathbf{u} - \nu_4 \eta \nabla^4 \mathbf{u}. \quad (1b)$$

In (1), η is the depth of the homogeneous fluid layer, \mathbf{q} is the depth-averaged two-dimensional horizontal momentum vector, $\mathbf{u} = \mathbf{q}/\eta$, ∇ is the two-dimensional horizontal gradient operator ($\partial/\partial x, \partial/\partial y$) with x pointing east and y pointing north, g' is reduced gravity, $f = f_0 + \beta y$ is twice the vertical component of the Earth's rotation at the north–south location y , and ν_4 is the coefficient of biharmonic viscosity. The implied operation between the two vectors \mathbf{q} and \mathbf{u} is the outer or tensor product $q_i u_j$. Note that (1b) is written in conservative form. The corresponding advective form is

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -g' \nabla \eta - f \hat{\mathbf{z}} \times \mathbf{u} - \nu_4 \nabla^4 \mathbf{u}. \quad (1b')$$

For the closed rectangular domain, the boundary conditions we use are: no flow through the rigid boundaries

$$\mathbf{u} \cdot \hat{\mathbf{n}} = 0, \quad (2a)$$

where $\hat{\mathbf{n}}$ is the unit normal, and free slip

$$\frac{\partial}{\partial \hat{\mathbf{n}}} (\mathbf{u} \cdot \hat{\mathbf{s}}) = 0, \quad (2b)$$

where $\hat{\mathbf{s}}$ is the unit tangent. The preservation of the global integral relation for potential enstrophy given a dissipation of the biharmonic form requires

$$\nabla^2 (\mathbf{u} \cdot \hat{\mathbf{n}}) = 0, \quad (2c)$$

$$\frac{\partial}{\partial \hat{\mathbf{n}}} \nabla^2 (\mathbf{u} \cdot \hat{\mathbf{s}}) = 0, \quad (2d)$$

at the boundaries (see Milliff and McWilliams, 1994).

The setup of the problem is identical to that in Milliff and McWilliams (1994): The model domain is 3600 km in the east–west direction and 2800 km north–south, with a grid spacing of 20 km in each direction. The Coriolis parameters are such that the midbasin latitude is about 38.2 N, and the coefficient of biharmonic viscosity ν_4 is $9.3 \times 10^9 \text{ m}^4 \text{ s}^{-1}$. The value of reduced gravity is 0.081 m s^{-2} and η_0 , the undisturbed depth of the layer is 1 km, giving a midbasin radius of deformation r of 100 km.

The initial condition consists of a cyclostrophically balanced (see below) monopole positioned at a distance $9r$ offshore of the western boundary at the midbasin latitude. The monopole has a Gaussian profile with a peak amplitude of the perturbation dynamic pressure of $3.2 \text{ m}^2 \text{ s}^{-2}$ and an e -folding radius of $2r$. The cyclostrophically balanced velocity field is obtained as a solution to the quadratic equation

$$-\frac{u_\theta^2}{a} = -g' \frac{\partial \eta}{\partial a} + f_{\text{mid}} u_\theta,$$

where u_θ is the tangential velocity at a distance a from the center of the monopole, and f_{mid} is the Coriolis parameter at midbasin. The evolution of the monopole at days 80, 160, and 200 are shown in Figure 1. An algorithmic description of the numerical scheme used in this explicit computation that resolves all relevant time scales to second-order accuracy is presented at the end of this section.

A synopsis of the phenomenology as described by Milliff and McWilliams (1994) follows: Since the initial condition is balanced on the f -plane, there is an initial readjustment due to the β effect. After this, the monopole evolves in a nonlinear fashion, propagating mostly westward at a rate controlled by the dispersion of the Rossby wave wake. At day 80, the strong interaction of the monopole with the western boundary results in a positive boundary-trapped Kelvin wave occupying the entire basin perimeter, whose expanding eastern boundary structure can be interpreted as a Rossby wave front propagating at approximately the

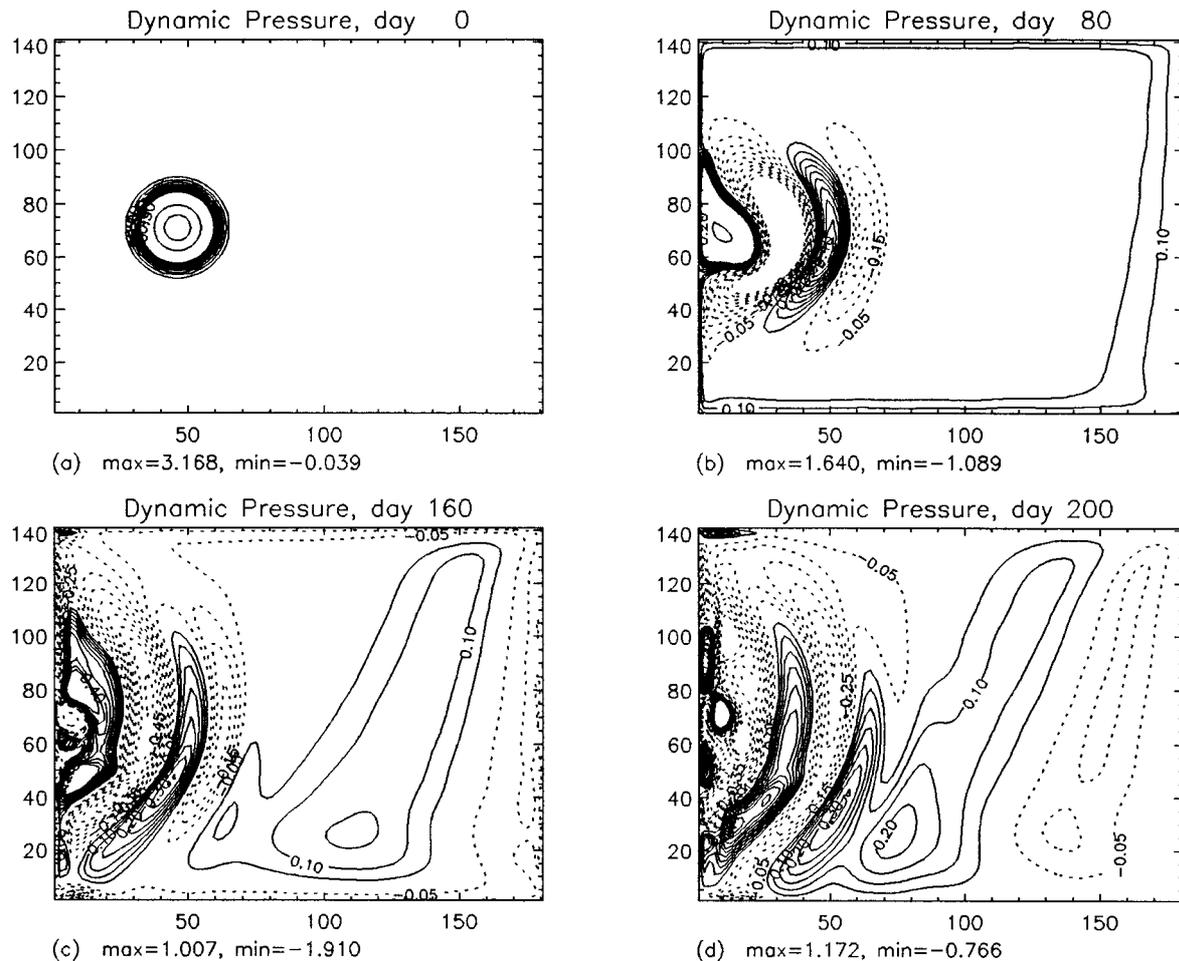


Figure 1. Dynamic pressure anomaly for the reference case, based on explicit time stepping (as in (3b)) of step size δt with MPDATA. Contour intervals are at $-1.5, -0.5, -0.45, \dots, -0.05, 0.05, 0.10, \dots, 0.5, 1.5, 2.5 \text{ m}^2 \text{ s}^{-2}$. Negative contours are dashed. Every second point (at $i = 1, 3, 5, \dots$) is shown in order to avoid obscuring the results by including a $2\delta x$ mode, which is a null mode of the pressure gradient operator.

local, long-wave phase speed $c_1(y) \approx \beta r_1^2(y)$ (faster in the south than in the north). At day 120, the large-scale offshore structure has transitioned into a steeper, smaller scale, boundary-trapped structure, and the first negative lobe of the Rossby wave now strongly interacts with the western boundary, resulting in the initiation of a negative structure at the eastern boundary similar in morphology to the positive structure preceding it. At day 200, the second basin-scale structure is detaching from the eastern boundary and the southern front of the first structure has now propagated far enough west to catch up with the trailing edge of the Rossby wave wake formed from the initial condition.

2.2. The Explicit Second-Order Scheme

The governing equations ((1a) and (1b)) may be written symbolically as

$$\frac{\partial \Psi}{\partial t} + \nabla \cdot \Psi \mathbf{u} = \mathbf{R}, \quad (3a)$$

where Ψ is the vector of dependent variables. Following the development in Smolarkiewicz and Margolin (1993), a second-order time discretization of the above equations may be written as

$$\Psi^{n+1} = \mathcal{A}_2 \left(\Psi^n + \frac{\delta t}{2} \mathbf{R}^{n+1/2} \right) + \frac{\delta t}{2} \mathbf{R}^{n+1}, \quad (3b)$$

where $\alpha = (u\delta t/\Delta x, v\delta t/\Delta y)$, is the vector of local Courant numbers, and where \mathcal{A}_2 is a second-order forward-in-time (FT) scheme for the flux-form, homogenous advection equation ($\mathbf{R} = 0$ in (3a)). The analogous operator that is first-order accurate in space and time is referred to as \mathcal{A}_1 .

We use an A grid (all variables reside at the same location) with the boundaries passing through the grid points and MPDATA (Smolarkiewicz and Grabowski (1990) and references therein) for \mathcal{A}_2 . In addition to the boundary conditions (2), $\nabla^4(\mathbf{u} \cdot \hat{\mathbf{s}})$ at a boundary point is obtained as an extrapolation from within the domain. The actual update of (1) proceeds in three steps as follows

1. The calculation of the advective velocities at time $n + \frac{1}{2}$:

$$\mathbf{u}^{n+1/2} = \left(\tilde{\mathbf{u}} \equiv \mathbf{u}^n + \frac{\delta t}{2} \mathbf{R}_u^n \right) - \frac{\delta t}{2} \mathbf{u}^n \cdot \nabla \tilde{\mathbf{u}}, \quad (4a)$$

where \mathbf{R}_u is the momentum forcing divided by the depth, and where first-order accurate upwind differencing is used in the convective term.

2. The update of the height field:

$$\eta^{n+1} = \mathcal{A}_2(\eta^n, \alpha^{n+1/2}). \quad (4b)$$

3. The update of the momentum field, and the force computation

$$\begin{aligned} \tilde{\mathbf{q}} &= \mathcal{A}_2 \left(\mathbf{q}^n + \frac{\delta t}{2} \mathbf{R}^n, \alpha^{n+1/2} \right), \\ \mathbf{q}^{n+1} &= \tilde{\mathbf{q}} - \frac{\delta t}{2} (\mathbf{R}^{n+1} \equiv -g' \eta^{n+1} \nabla \eta^{n+1} - \nu_4 \nabla^4 \tilde{\mathbf{q}} - f \hat{\mathbf{z}} \times \mathbf{q}^{n+1}). \end{aligned} \quad (4c)$$

3. The Method of Averaging

The algorithm (3b) is a general recipe for second-order accurate integrations of (3a). Its particular realization depends on the selection of the advection scheme \mathcal{A}_2 as well as the specific approximations used for $\alpha^{n+1/2}$ and \mathbf{R}^{n+1} . Those summarized in (4) provide simple second-order accurate solutions to (1a) and (1b). However, for computational stability, they require a time step limited by the propagation speeds of the fastest modes (see Smolarkiewicz and Margolin, 1993). In order to circumvent this restrictive condition on the time step and improve computational efficiency, here we consider alternate realizations of $\alpha^{n+1/2}$ and \mathbf{R}^{n+1} .

Since the main concern is the handling of the fast modes in (1), we first consider the highly simplified case of one-dimensional gravity wave propagation in a fluid layer of depth η_0 and which is initially at rest. With the choice of reference quantities

$$x_r = \Delta x, \quad \eta_r = \eta_0, \quad t_r = \delta t, \quad u_r = \sqrt{g'\eta_0},$$

the nondimensionalized equations for this simplified case are

$$\frac{\partial \eta}{\partial t} = -\beta \frac{\partial u}{\partial x}, \quad (5a)$$

$$\frac{\partial u}{\partial t} = -\beta \frac{\partial \eta}{\partial x}, \quad (5b)$$

where $\beta = \sqrt{g'\eta_0}\delta t/\Delta x$ is the Courant number based on the long gravity wave (long with respect to the undisturbed depth η_0) speed and the time step δt . Without loss of generality, assume $u = u_s + u_f$, where u_s varies only on the slow time scale and u_f varies on the fast time scale. We further choose initial and end conditions for u_f such that $u_f(t + \Delta t) = u_f(t)$ where Δt is a long time step equal to $\mathcal{M}\delta t$ with $\mathcal{M} > 1$. We decompose η similarly into its fast and slow components. Integration of (5) over the long time step Δt then results in

$$\frac{\eta_s(t + \Delta t) - \eta_s(t)}{\Delta t} \left(\approx \frac{\partial \eta_s}{\partial t} \right) = -\beta \frac{\partial \bar{u}}{\partial x}, \quad (6a)$$

$$\frac{u_s(t + \Delta t) - u_s(t)}{\Delta t} \left(\approx \frac{\partial u_s}{\partial t} \right) = -\beta \frac{\partial \bar{\eta}}{\partial x}, \quad (6b)$$

where the bar indicates averaging over the $\mathcal{M} + 1$ values (see (7)). Thus the scheme for evolving the slow variables (η_s, u_s) is first to evolve (5) over \mathcal{M} successive gravity-wave resolving time steps, each of length δt , to obtain the sequence of solutions (η^{n+m}, u^{n+m}), $m = 0, \dots, \mathcal{M}$. Then we form the weighted averages

$$\bar{u} = \sum_{m=0}^{\mathcal{M}} w_m u^{n+m}, \quad \bar{\eta} = \sum_{m=0}^{\mathcal{M}} w_m \eta^{n+m}, \quad (7)$$

with weights w_m , and evolve (6) over the large time step Δt .

3.1. Linear Stability Analysis of (6)

The linear stability of (6) proceeds in two steps: the linear stability analysis of (5) and the consequent analysis of (6). The details of the linear stability for the scheme (4) in the simplified situation of (5) are summarized below (for details see Appendix B of Smolarkiewicz and Margolin (1993)). Let $\alpha = \beta \sin \theta$, where $\theta = (2\pi/\lambda)\Delta x \in [0, \pi]$ is a dimensionless wave number corresponding to the wavelength λ resolved on the grid and β is the gravity wave Courant number as defined earlier. The complex amplification factors $r_{1,2}$ (one each for the left- and right-going waves) over the time step δt are given by the two roots of the quadratic equation

$$r^2 - 2r \left(1 - \frac{\alpha^2}{2} \right) + 1 = 0.$$

They are

$$r_{1,2} = 1 - \frac{\alpha^2}{2} \pm i\alpha \sqrt{1 - \frac{\alpha^2}{4}}.$$

For gravity wave Courant numbers β such that $\beta \leq 2$, the two amplification factors are complex conjugates, each with unit magnitude and arguments $\pm \delta$ such that

$$\cos \delta = 1 - \frac{\alpha^2}{2}, \quad \sin \delta = \alpha \sqrt{1 - \frac{\alpha^2}{4}}, \quad (8)$$

with the square root assumed to yield a positive value. Unless otherwise mentioned, we assume $\beta \leq 2$ in what follows.

Forming the time averages $\bar{\eta}$ and \bar{u} now reduces to forming the time averages of the above complex roots:

$$[\bar{\eta}, \bar{u}] = [\eta_s^n, u_s^n] \bar{r}_{1,2}(\alpha).$$

Considering the simplest case of extended trapezoidal rule for the time averaging,

$$w_0 = w_{\mathcal{M}} = \frac{0.5}{\mathcal{M}}, \quad w_m = \frac{1}{\mathcal{M}}, \quad 1 \leq m < \mathcal{M}, \quad (9)$$

$$\bar{r}_{1,2} = \sum_{m=0}^{\mathcal{M}} w_m r_{1,2}^m = \frac{r_{1,2}^{\mathcal{M}} - 1}{2\mathcal{M}} \frac{r_{1,2} + 1}{r_{1,2} - 1}, \quad (10)$$

where the formula for the summation of a geometric series has been used. Then from (8), the expressions for $\bar{r}_{1,2}$ can be simplified into

$$\bar{r}_{1,2} \equiv \bar{r}_r \pm i\bar{r}_i = \frac{1}{\mathcal{M}} \cot \frac{\delta}{2} \sin \frac{\delta\mathcal{M}}{2} \exp\left(\pm \frac{\delta\mathcal{M}}{2}\right). \quad (11)$$

A simple discretization of (6) over the long time step Δt that is second-order accurate in space and time is

$$\eta_{si}^{n+1} = \eta_{si}^n - \frac{\mathcal{M}\beta}{2} (\bar{u}_{i+1} - \bar{u}_{i-1}), \quad (12a)$$

$$u_{si}^{n+1} = u_{si}^n - \frac{\mathcal{M}\beta}{2} (\bar{\eta}_{i+1} - \bar{\eta}_{i-1}), \quad (12b)$$

where the second-order time accuracy comes from identifying the averaged values of u and η on the right-hand side of (12) as representative first-order estimates of u_s and η_s at time $n + \frac{1}{2}$. Following the standard procedure of von Neumann linear analysis, the necessary condition for the existence of a solution leads to the following complex amplification factors $R_{1,2}$ over the time step Δt :

$$R_{1,2} = 1 \pm i\mathcal{M}\alpha\bar{r}_{1,2}(\alpha), \quad (13)$$

where the plus sign goes with \bar{r}_1 and the minus sign goes with \bar{r}_2 to consistently account for the right-going and left-going waves. From the above it also follows that R_1 and R_2 are complex conjugates, each with magnitude $|R|$.

It now remains to show that $|R| \leq 1$ to demonstrate the stability of the overall method. Toward this, from (11) we have

$$|R|^2 = 1 + \frac{\alpha^2}{4} |\bar{r}|^2 - \alpha\bar{r}_i.$$

Thus for stability we need to show that

$$0 \leq \frac{\alpha|\bar{r}|^2}{4\bar{r}_i} \leq 1. \quad (14)$$

Manipulation of (11), however, shows that

$$\frac{\alpha|\bar{r}|^2}{4\bar{r}_i} = \frac{\alpha}{2} \cot \frac{\delta}{2},$$

and further simplification using (8) leads to

$$\frac{\alpha|\bar{r}|^2}{4\bar{r}_i} = \sqrt{1 - \frac{\alpha^2}{4}}.$$

Considering that $\alpha = \beta \sin \theta$, $0 \leq \beta \leq 2$ and $0 \leq \theta \leq \pi$, the inequality (14) is satisfied and hence the method is stable for any \mathcal{M} . The amplification factors themselves are, however, dependent on \mathcal{M} directly (see (13)) and through the dependence of $\bar{r}_{1,2}$ on \mathcal{M} .

The above analysis ignores completely any slow terms in (1), which leads to the conclusion of unconditional stability of the outer loop. A more realistic example would include additional terms such as advection and hyperviscosity, which vary on the slow scale. A more complete stability analysis of the equations with advection added shows that the stability of the overall scheme requires the limitation of the time step of the outer loop by a complicated function of the Courant numbers associated with gravity waves and material advection. However the ratio of the outer time step to the inner time step, \mathcal{M} is of the order of the ratio of the time scales.

3.2. Extension of the Method to (1)

In analogy with (6), the full nonlinear equations for the evolution of the slow variables, i.e., the slow version of (1), may be written in the flux form as

$$\frac{\eta_s(t + \Delta t) - \eta_s(t)}{\Delta t} \left(\approx \frac{\partial \eta_s}{\partial t} \right) = -\nabla \cdot \eta_s \bar{\mathbf{u}}, \quad (15a)$$

$$\frac{\mathbf{q}_s(t + \Delta t) - \mathbf{q}_s(t)}{\Delta t} \left(\approx \frac{\partial \mathbf{q}_s}{\partial t} \right) + \nabla \cdot \mathbf{q}_s \bar{\mathbf{u}} = -\overline{g' \eta \nabla \eta} - \overline{f \eta \hat{\mathbf{z}} \times \mathbf{u}} - \nu_4 \eta \nabla^4 \mathbf{u}_s \equiv \bar{\mathbf{R}} + \mathbf{R}_s. \quad (15b)$$

As before, the subscript s indicates that that variable varies only on the slow time scale. On the right-hand side of (1b), though, it is clear that only the pressure forces vary on the fast time scale, the Coriolis terms are also considered in the initial first-order integration and the ensuing averaging procedure in order to maintain the (primary) geostrophic balance of the system.

The basis scheme for the evolution of the full equations (1) is as follows:

1. Evolve (1a) and (1b') over \mathcal{M} time steps using the first-order equivalent of the explicit scheme (4). The time step δt used in this evolution is gravity wave resolving.
2. Form the time averages $\bar{\mathbf{R}}$ and $\bar{\mathbf{u}}$ as in (7).
3. Update the slow depth field η_s using (15a):

$$\eta_s^{n+1} = \mathcal{A}_2(\eta_s^n, \bar{\alpha}), \quad (16a)$$

where $\alpha = (\bar{u} \delta t / \Delta x, \bar{v} \delta t / \Delta y)$, is the vector of local Courant numbers and \mathcal{A}_2 is the same operator as in (3b).

4. Update the slow momentum field \mathbf{q}_s using (15b) in a similar fashion to above correctly accounting for the accuracy of the forces (see Smolarkiewicz and Margolin, 1993):

$$\mathbf{q}_s^{n+1} = \mathcal{A}_2(\mathbf{q}_s^n, \bar{\alpha}) + \mathcal{A}_1(\bar{\mathbf{R}} + \mathbf{R}_s^{n+1/2}, \frac{1}{2} \bar{\alpha}). \quad (16b)$$

These four steps are referred to as the method of averaging (MOA) in the rest of the paper. Step 1 is referred to as the subcycle and steps 3 and 4 are termed the supercycle. The results and the efficiency gain resulting from the above scheme is discussed in the next section. We, however, wish to point out that in step 1 of the above description, the advection and viscous forces were considered on the fast time scale for reasons of simplicity of description. Since these two processes clearly occur on the slow time scale, we have computationally verified that they need only be evaluated on the slow time scale. That is, it is sufficient to include extrapolations of these terms based on their slow time scale values in the fast time equations.

4. Simulation and Comparisons

A hierarchy of speeds exists within the monopole problem, with the gravity waves traveling at the fastest speed of about 9 m s^{-1} , the flow velocity reaching maximum speed of about 0.4 m s^{-1} , and with the long Rossby wave phase speed $c_1(y) \approx \beta r_1^2(y)$ reaching about 0.2 m s^{-1} at the southernmost latitude. Within the MOA we would expect the supercycle, which performs advection over the long time step Δt , to be time step limited mainly by the flow velocity, rather than by the much more restrictive gravity wave speed. In the previous section we demonstrated the stability of MOA in handling gravity waves for any value of \mathcal{M}

in the absence of advection, viscosity, and other slow terms and indicated that an inclusion of these slow terms would lead to a maximum value of \mathcal{M} for which the gravity waves would be stable. For purpose of argument, if we assume the former (that the gravity waves are stable for any \mathcal{M}), then the only restriction on the supercycle time step is due to \mathcal{A}_2 . Using a subcycle gravity wave Courant number β of 0.95 in the subcycle, the theoretical maximum value of \mathcal{M} for this two-dimensional problem is estimated to be 16.

Although the boundary Kelvin waves propagate rapidly in the monopole problem, they are of very long wavelengths—comparable with the basin perimeter—and therefore their period is long relative to the time Δt over which temporal averages are computed. For this reason we expect that the essence of the Kelvin wave dynamics, including their coupling to the Rossby waves that are emitted from the eastern boundary, will be preserved. We do expect some differences between the solutions produced with the explicit code and those produced with MOA to exist. The temporal smoothing of the advecting velocities and of the forcing that results from the averaging done in the subcycle will result in differences in transport, though the solution will still be of second-order accuracy in space and time if $\bar{\mathbf{u}}$ is at least a first-order accurate estimate of $\mathbf{u}_s^{n+1/2}$. As FT methods are less dissipative at higher Courant numbers, a longer time step will in this respect lead to an improvement in the accuracy of the solution.

We evaluate the effectiveness of MOA by judging the degree to which we capture the quality of the solution demonstrated by the explicit, second-order accurate model. Consideration of a first-order accurate, donor cell-based solution also highlights the degree to which the solution to the monopole problem depends on the method that is used.

The MOA produces stable results at our estimated maximum value of $\mathcal{M} = 16$ when uniform weights are used to form the average. The advective Courant number reaches 0.49 at this value of \mathcal{M} , indicating that this Courant number is sufficient to limit the time step Δt . We prefer to show results at $\mathcal{M} = 8$, well within the stable region of parameter space, although results are not very different than at $\mathcal{M} = 16$. The MOA-based model was run with a coefficient of biharmonic viscosity in the supercycle equal to that of the explicit model, no explicit viscosity in the subcycle (where we rely on the high implicit viscosity of the first-order accurate differencing), and with weighting coefficients for the averaging as in (9).

Figure 2 shows the results for the case $\mathcal{M} = 8$ at the same three times as in Figure 1. The overall comparison between the two figures is excellent. The Kelvin wave signal that encompasses the boundary is well reproduced, and so are the Rossby waves that are emitted from the eastern boundary. The principal difference between the two cases is in the details associated with the intensified western boundary regions. For example, a close inspection of Figures 2(c) and 1(d) reveals a slightly faster disappearance of the extremum of negative sign near the midpoint of the western boundary in MOA.

A clearer indication of the accuracy with which the Kelvin waves are reproduced is seen in the record of the dynamic height at a point on the boundary (the midpoint of the southern boundary), in Figure 3. The MOA reproduces the lower-frequency behaviour of the reference case well. The amplitude of the high-frequency component of the signal is reduced in MOA. This high-frequency component disappears when higher viscosities are used.

We use the kinetic energy, summed over the domain, as another diagnostic. The kinetic energy for the reference case rises as the monopole impacts the western boundary, and then settles at a slightly lower level through the time of interaction of the first, negative anomaly associated with the Rossby wave wake. The kinetic energy time series of MOA tracks that of the reference case well at most times, as seen in Figure 4. The difference in kinetic energies between the two cases is very small until shortly before day 200, which was taken to be the end of the simulation in MM94. The first-order accurate case, also plotted in Figure 4, is much more dissipative than either of the higher-order cases.

A cost analysis was performed through code profiling on a Sun Sparc 10 workstation. Relative to the cost of the MPDATA-based reference code, the costs of running at $\mathcal{M} = 4, 8, \text{ and } 16$ are 0.36, 0.27, and 0.23, respectively. The asymptotic cost limit, corresponding formally to $\mathcal{M} = \infty$ (which is only the limit of MOA based on its present implementation), is 0.21. The first-order accurate explicit code has a cost factor of 0.35; it is more expensive than the limit given above because the conservative differencing is more expensive than the nonconservative advective-form subcycling scheme, and because we are not applying the biharmonic operator in the subcycle. Comparison of Figures 1, 2, and 5 clearly demonstrates that the MOA-produced result resembles that of the second-order accurate code much more than that of the first-order accurate code. Consideration of the dynamic heights recorded in Figure 3 and of the kinetic energy recorded in Figure 4 reinforces this conclusion.

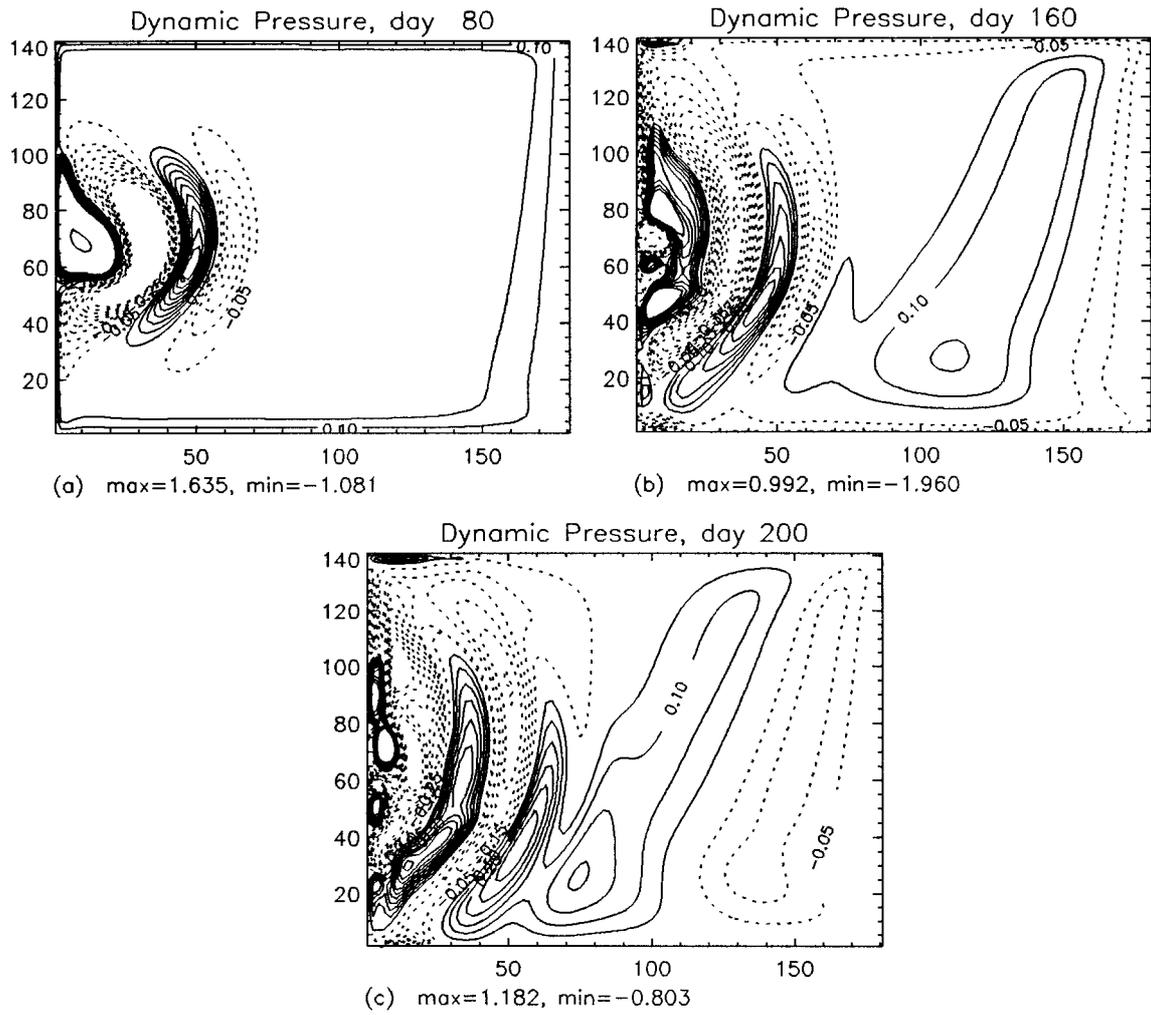


Figure 2. Dynamic pressure anomaly produced through application of MOA. Contour intervals are as in Figure 1.

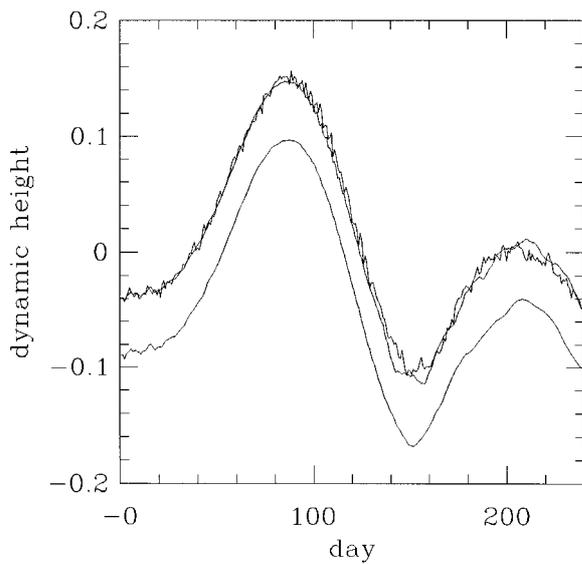


Figure 3. Dynamic pressure anomaly time series at a point located at the midpoint of the southern boundary, sampled every 72,000 s. The reference and MOA cases are overlaid, with MOA being the smoother of the two curves. The time series corresponding to a second MOA case, run with increased biharmonic viscosity of $3 \times 10^{10} \text{ m}^4 \text{ s}^{-1}$ and plotted with an offset of $-0.05 \text{ m}^2 \text{ s}^{-2}$, is smoother yet.

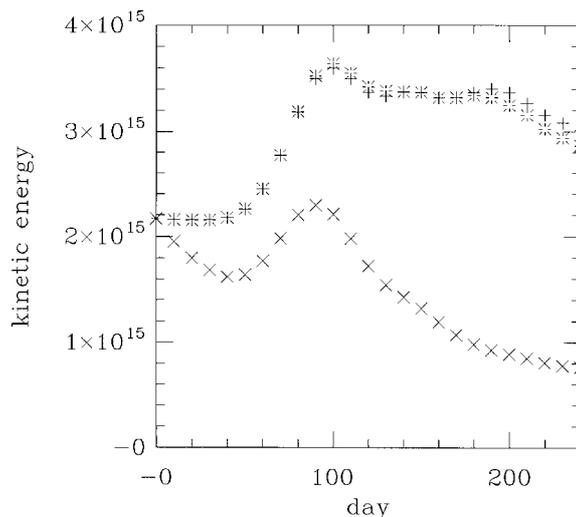


Figure 4. Kinetic energy, summed over the domain. Plus symbols indicate the reference, second-order accurate case, bursts the MOA case, and crosses the donor cell case.

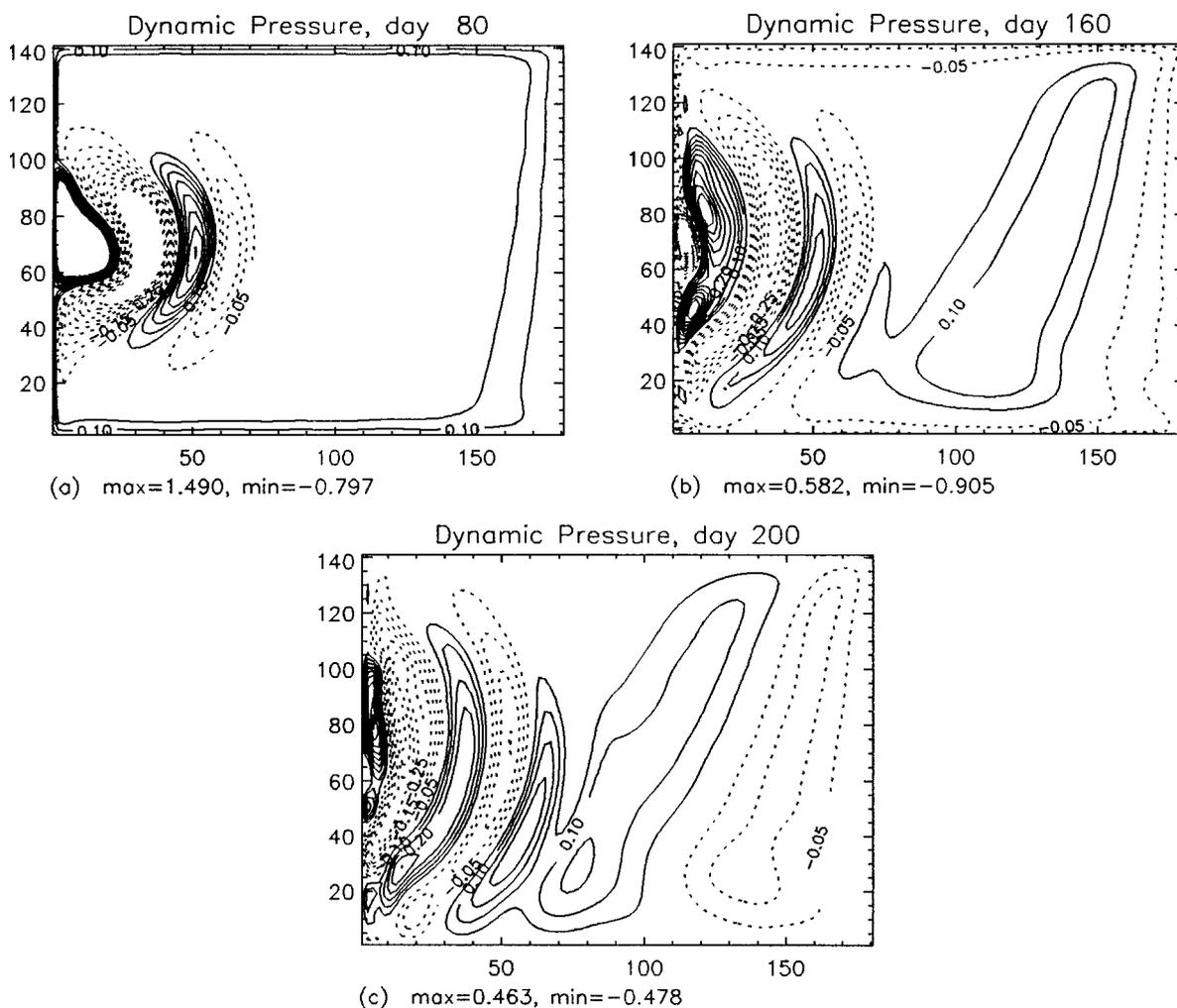


Figure 5. Dynamic pressure anomaly produced through simple time stepping of step size δt with first-order accurate donor cell differencing. Contour intervals are as in Figure 1.

5. Conclusions

The presence of very fast modes in certain systems with multiple time scales requires the time step of an explicit treatment to be much smaller than that required from accuracy considerations and this leads to computational inefficiency. This is, for example, the case with the primitive equations used in the simulation of the circulations of the world oceans. In this paper we have described a method of averaging for accelerating computational simulations of such systems. We have analytically demonstrated the linear stability of the method in a simplified context, and verified the stability of the full system through simulation.

The results obtained using MOA are comparable with those from the reference, explicit second-order accurate code, and are much better than those from the first-order accurate code. The improved performance is obtained without additional cost, relative to the first-order accurate model (and for about a fourth of the cost of the second-order accurate explicit code). We are able to run our model close to the theoretical time step limit.

In conventional primitive equation ocean models based on barotropic/baroclinic splitting, the first baroclinic mode generally limits the model time step to one significantly smaller than that required by an advective Courant limit. The essential dynamics that arise from this mode are usually associated with features that change on a relatively slow period, and hence these features will be captured by MOA, as demonstrated in the test problem. Thus it is possible that MOA will form the basis for developing a competitive computational scheme for modeling oceans. Such an approach would also avoid the potentially destabilizing effect of approximate modal decompositions.

References

- R. Bleck and L.T. Smith (1990): A wind-driven isopycnic coordinate model of the North and equatorial Atlantic. *J. Geophys. Res.*, **95**, 3273–3285.
- G. Browning and H.-O. Kreiss (1987): Reduced systems for the shallow water equations. *J. Atmos. Sci.*, **44**, 2813–2822.
- G. Browning and H.-O. Kreiss (1994): Splitting methods for problems with different timescales. *Monthly Weather Rev.*, **122**, 2614–2622.
- K. Bryan (1969): A numerical method for the study of the circulation of the world ocean. *J. Comput. Phys.*, **4**, 347–376.
- J.K. Dukowicz, R.D. Smith, and R.C. Malone (1993): A reformulation and implementation of the Bryan–Cox–Semtner ocean model on the connection machine. *J. Atmos. Ocean Technol.*, **10**, 195–208.
- R.L. Higdon and A.F. Bennett (1996): Stability analysis of operator splitting for large-scale ocean modeling. *J. Comput. Phys.*, **123**, 311–329.
- J.B. Klemp and R. Wilhelmson (1978): The simulation of three-dimensional convective storm dynamics. *J. Atmos. Sci.*, **35**, 1070–1096.
- R.J. LeVeque and J. Olinger (1983): Numerical methods based on additive splittings for hyperbolic partial differential equations. *Math. Comp.*, **40**, 469–497.
- R.V. Madala (1981): Efficient time integration schemes for atmosphere and ocean models. In *Finite-Difference Techniques for Vectorized Fluid Calculations*, D. Book, ed. Springer-Verlag, New York, pp. 56–74.
- G.I. Marchuk (1974): *Numerical Methods in Weather Prediction*. Academic Press, New York. (Russian edn., 1967).
- F. Mesinger, Z.I. Janjić, S. Ničković, D. Gavrilov, and D.G. Deaven (1988): The step-mountain coordinate: model description and performance for cases of Alpine lee cyclogenesis and for a case of Appalachian redevelopment. *Tellus*, **25**, 444–458.
- R.F. Milliff and J.C. McWilliams (1994): The evolution of boundary pressure in ocean basins. *J. Phys. Ocean.*, **24**, 1317–1338.
- W.C. Skamarock and J.B. Klemp (1992): The stability of time-split numerical methods for the hydrostatic and nonhydrostatic elastic equations. *Monthly Weather Rev.*, **120**, 2109–2127.
- W.C. Skamarock and J.B. Klemp (1994): Efficiency and accuracy and the Klemp–Wilhelmson time-splitting technique. *Monthly Weather Rev.*, **122**, 2623–2630.
- P.K. Smolarkiewicz and W.W. Grabowski (1990): The multidimensional positive definite advection transport algorithm: nonoscillatory option. *J. Comput. Phys.*, **86**, 355–375.
- P.K. Smolarkiewicz and L.G. Margolin (1993): On forward-in-time differencing for fluids: extension to a curvilinear framework. *Monthly Weather Rev.*, **121**, 1847–1859.
- G. Strang (1968): On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.*, **5**, 516–517.