Energy Conserving and Potential-Enstrophy Dissipating Schemes for the Shallow Water Equations

AKIO ARAKAWA AND YUEH-JULIAN G. HSU*

Department of Atmospheric Sciences, University of California, Los Angeles, California

(Manuscript received 5 February 1990, in final form 5 February 1990)

ABSTRACT

To incorporate potential enstrophy dissipation into discrete shallow water equations with no or arbitrarily small energy dissipation, a family of finite-difference schemes have been derived with which potential enstrophy is guaranteed to decrease while energy is conserved (when the mass flux is nondivergent and time is continuous). Among this family of schemes, there is a member that minimizes the spurious impact of infinite potential vorticities associated with infinitesimal fluid depth. The scheme is, therefore, useful for problems in which the free surface may intersect with the lower boundary.

1. Introduction

As Arakawa (1966) showed, we can construct finite-difference Jacobians that maintain important integral constraints on the continuous Jacobian. When applied to the vorticity equation governing two-dimensional incompressible inviscid flow, maintaining these constraints guarantees conservation of energy and enstrophy in the discrete system. Arakawa (1970) and Arakawa and Lamb (1977) further pointed out that in long-term integrations with energy conserving (but not enstrophy conserving) schemes, enstrophy increases considerably due to spurious energy cascade into smaller scales. On the other hand, enstrophy conserving (but not energy conserving) schemes approximately conserve energy. This is reasonable because the amount of energy in smaller scales, where discretization errors are large, is effectively constrained by enstrophy conservation.

Similar situations exist in finite-difference schemes for the shallow water equations, as Sadourny (1975) pointed out. Based on the vector-invariant form of the shallow water equations without Coriolis force, he presented two schemes, one conserves energy and the other conserves potential enstrophy. He then showed that the energy conserving scheme produces an unrealistic increase of potential enstrophy, while the potential-enstrophy conserving scheme prevents any build up of energy in smaller scales.

Compared to enstrophy conservation, therefore, energy conservation appears to be a less effective computational constraint on the nonlinear (potential) vorticity advection term. Energy conservation, however, can be a very important constraint on some other terms, such as the Coriolis force term. A scheme for the shallow water equation (with the Coriolis force) that conserves both energy and (absolute) potential enstrophy when the mass flux is nondivergent was derived by Sadourny and tested by the ECMWF (Burr ridge and Haseler 1977; see also Hollingsworth et al. 1983). Arakawa and Lamb (1981) derived a scheme that conserves both potential enstrophy and energy for the general case of divergent mass flux. Arakawa and Lamb also derived a family of schemes that satisfy the same conservation requirements when the mass flux is nondivergent. Takano and Wurtele (1982) extended the Arakawa-Lamb scheme to a partially fourth-order potential-enstrophy and energy conserving scheme.

In a real fluid, however, both energy and potential enstrophy can cascade into smaller scales and eventually dissipate. Dynamical processes responsible for energy cascade and those for potential enstrophy cascade, however, can be quite different. In quasi-geostrophic turbulence, for example, it is primarily the potential enstrophy not the energy that cascades into smaller scales and eventually dissipates. (For excellent reviews of quasi-geostrophic turbulence, see Sadourny 1984, 1985).

In this paper, we first derive a family of finite-difference schemes for the shallow water equations with which potential enstrophy is guaranteed to be either conserved or decreased in time while energy is conserved (when the mass flux is nondivergent and time is continuous). With these schemes, potential enstrophy dissipation can be incorporated with no or arbi-
trarily small energy dissipation. The amount of potential enstrophy decrease is controlled by a time scale, which can be specified through physical reasoning. The result is a generalization of Arakawa (1966) and Arakawa and Lamb (1981) to a (potential) enstrophy dissipative system. It can also be interpreted as an application of the simplest version of the "anticipated potential vorticity method" (Sadourny and Basdevant 1985) to a discrete system.

When the free surface can intersect with the bottom surface, the discrete continuity equation must be positive definite. Also, special care must be taken in choosing a scheme from the family of schemes we have derived. In these schemes, products of potential vorticity and mass flux at different grid points appear in the momentum equation. These products do not necessarily remain finite as the fluid depth at the neighboring grid points approaches zero. To avoid this situation, we choose a unique member from the family, with which the spurious impact of infinite potential vorticity associated with infinitesimal fluid depth is minimized.

In section 2, the continuous shallow water equations and some of their consequences are reviewed. Section 3 (along with appendix A) presents an energy conserving and potential-kestrony dissipating scheme for the general case of divergent mass flux, while section 4 (along with appendix B) presents a family of energy conserving and potential-kestrony dissipating schemes for the case of nondivergent mass flux. Section 5 discusses requirements for minimizing the spurious impact of infinite potential vorticities and shows that a unique member of the family can satisfy these requirements. Finally, section 6 gives a summary of this paper and further comments.

2. Continuous equations

The governing equations for quasi-static motions in a homogeneous incompressible fluid with a free surface (shallow water equations) can be written as

\[
\frac{\partial v}{\partial t} + \nabla \times \mathbf{v}^* + \nabla (K + \Phi) = 0, \quad (2.1)
\]

\[
\frac{\partial h}{\partial t} + \nabla \cdot \mathbf{v}^* = 0. \quad (2.2)
\]

Here \(\mathbf{v}\) is the horizontal velocity, \(t\) the time, \(q = (f + \xi)/h\) the potential vorticity, \(f\) the Coriolis parameter, \(\xi = k \cdot \nabla \times \mathbf{v}\) the relative vorticity, \(h\) the depth of a fluid column above the bottom surface, \(k\) the vertical unit vector, \(\mathbf{v}^* = \mathbf{h} \cdot \mathbf{v}\) the horizontal mass flux, \(\nabla\) the horizontal del operator, \(K = \mathbf{v}^* \cdot \nabla\) the horizontal kinetic energy per unit mass, \(\Phi = g(h + h_i)\) the geopotential at the free surface, \(g\) the gravitational acceleration and \(h_i\) the height of the bottom surface.

Multiplying (2.1) by \(\mathbf{v}^*\) and using (2.2), we obtain the equation for the time change of kinetic energy,

\[
\frac{\partial}{\partial t} (hK) + \nabla \cdot (\mathbf{v}^* K) + \mathbf{v}^* \cdot \nabla \Phi = 0. \quad (2.3)
\]

Multiplying (2.2) by \(\Phi\), on the other hand, we obtain the equation for the time change of potential energy,

\[
\frac{\partial}{\partial t} \left[ \frac{1}{2} gh^2 + gh_i \right] + \nabla \cdot (\mathbf{v}^* \Phi) - \mathbf{v}^* \cdot \nabla \Phi = 0. \quad (2.4)
\]

The sum of (2.3) and (2.4) yields conservation of total energy,

\[
\frac{\partial}{\partial t} \left[ h \left( K + \frac{1}{2} gh + gh_i \right) \right] = 0, \quad (2.5)
\]

where the overbar denotes the area mean over a periodic domain or a domain with no inflow or outflow through the boundaries.

Operating \(\mathbf{k} \cdot \nabla\times\) on (2.1), we obtain the vorticity equation that is equivalent to the flux form of the potential vorticity equation given by

\[
\frac{\partial}{\partial t} (hq) + \nabla \cdot (\mathbf{v}^* q) = 0. \quad (2.6)
\]

Subtracting (2.2) times \(q\) from (2.6), we obtain

\[
\frac{\partial q}{\partial t} + \mathbf{v} \cdot \nabla q = 0, \quad (2.7)
\]

that is the advective form of the potential vorticity equation. In the case of nondivergent mass flux, for which \(\nabla \cdot \mathbf{v}^* = 0\) (and \(\partial h/\partial t = 0\)), we can define a streamfunction \(\psi^*\) for the mass flux by

\[
\mathbf{v}^* = \mathbf{k} \times \nabla \psi^*. \quad (2.8)
\]

Then, using Cartesian coordinates \(x\) and \(y\), we can express (2.6) as

\[
\frac{\partial}{\partial t} (hq) + J(\psi^*, q) = 0, \quad (2.9)
\]

where \(J\) is the Jacobian defined by

\[
J(a, b) = \frac{\partial a}{\partial x} \frac{\partial b}{\partial y} - \frac{\partial a}{\partial y} \frac{\partial b}{\partial x}. \quad (2.10)
\]

Multiplying (2.7) by \(hq\) and using (2.2), we obtain the potential enstrophy equation

\[
\frac{\partial}{\partial t} \left( h \frac{1}{2} q^2 \right) + \nabla \cdot \left( \mathbf{v}^* \frac{1}{2} q^2 \right) = 0 \quad (2.11)
\]

that leads to conservation of potential enstrophy

\[
\frac{\partial}{\partial t} \left( h \frac{1}{2} q^2 \right) = 0. \quad (2.12)
\]

In this paper, we are interested in including a dissipative effect on potential enstrophy while conserving energy. To do so, let us replace \(q\) in (2.1) by \(q + q'\). Obviously this does not influence (2.5), therefore, en-
ergy is still conserved. Equation (2.11), however, is modified to
\[
\frac{\partial}{\partial t} \left( h \frac{1}{2} q^{2} \right) + \nabla \cdot \left( v^* \frac{1}{2} q^{2} \right) = -\nabla \cdot (q v^* q^*) + q v^* \cdot \nabla q.
\] (2.13)

If we choose
\[
q' = -(\tau/h)(v^* \cdot \nabla q),
\] (2.14)
where \( \tau \) is a time scale to be specified, (2.13) gives
\[
\frac{\partial}{\partial t} \left( h \frac{1}{2} q^{2} \right) = -\frac{\tau}{h} (v^* \cdot \nabla q)^2.
\] (2.15)
Thus, the potential enstrophy is conserved with \( \tau = 0 \) and decreases in time with \( \tau > 0 \). The choice of \( q' \) given by (2.14) is a simple example of the anticipated potential vorticity method proposed by Sadourny and Basdevant (1985).

3. An energy conserving and potential-enstrophy dissipating scheme for the general case of divergent mass flux

Using the C grid (Arakawa and Lamb 1977) shown by Fig. 1, a space finite-difference scheme for the continuity equation (2.2) can be written as
\[
\frac{\partial}{\partial t} h^{i+1/2,j+1/2} + (\nabla \cdot v^*)_{i+1/2,j+1/2} = 0,
\] (3.1)
where
\[
(\nabla \cdot v^*)_{i+1/2,j+1/2} = \frac{1}{d} \left[ u^*_{i+1/2,j+1/2} - u^*_{i,j+1/2} + v^*_{i+1/2,j+1} - v^*_{i+1/2,j+1/2} \right],
\] (3.2)
\[ u^*_{i,j+1/2} = [h(u)u]_{i,j+1/2}, \] (3.3)
\[ v^*_{i+1/2,j} = [h(v)v]_{i+1/2,j}, \] (3.4)
\[ d \] is the grid size, \( u \) and \( v \) are the \( x \) and \( y \) components of \( v \), and \( h(u) \) and \( h(v) \) are \( h \) defined at \( u \) and \( v \) points, respectively.

Following Arakawa and Lamb (1981), we write an energy-conserving space finite-difference scheme for the components of the momentum equation (2.1) as
\[
\frac{\partial}{\partial t} u_{i,j+1/2} - \alpha_{i,j+1/2} v^*_{i+1/2,j+1} = -\beta_{i,j+1/2} v^*_{i+1/2,j+1} - \gamma_{i,j+1/2} v^*_{i+1/2,j+1/2} - \delta_{i,j+1/2} u^*_{i+1/2,j+1/2} + \epsilon_{i+1/2,j+1/2} u^*_{i+1/2,j+1/2} - \epsilon_{i+1/2,j+1} u^*_{i+1/2,j+1} + \frac{1}{d} [(K + \Phi)_{i+1/2,j+1/2} - (K + \Phi)_{i+1/2,j+1/2}] = 0,
\] (3.5)
\[
\frac{\partial}{\partial t} v_{i+1/2,j} + \gamma_{i+1/2,j+1/2} u^*_{i+1/2,j+1} + \delta_{i+1/2,j} u^*_{i+1/2,j+1} + \alpha_{i+1/2,j} u^*_{i+1/2,j+1} - \beta_{i+1/2,j} u^*_{i+1/2,j+1} + \frac{1}{d} [(K + \Phi)_{i+1/2,j+1/2} - (K + \Phi)_{i+1/2,j+1/2}] = 0.
\] (3.6)
Here, the coefficients \( \alpha, \beta, \gamma, \delta, \epsilon \) and \( \phi \) are linear combinations of \( q \) at neighboring grid points. We can see that the terms involving these coefficients have no net contribution to \( \sum_{ij} [u^*_{i,j+1/2} \times (3.5) + v^*_{i+1/2,j} \times (3.6)] \) for a periodic domain.

Arakawa and Lamb (1981) used
\[
h^{(u)}_{i,j+1/2} = (\bar{h})_{i,j+1/2} \] (3.7)
\[ h^{(v)}_{i+1/2,j} = (\bar{h})_{i+1/2,j}, \] (3.8)
where overbars \(-i\) and \(-j\) denote the arithmetic mean over two neighboring points in \( x \) and \( y \) directions, respectively. Then the scheme given by (3.5) and (3.6) conserves total energy for the general case of divergent mass flux if \( K \) is specified as
\[
K_{i+1/2,j+1/2} = \frac{1}{2} [\bar{u}^{2i} + \bar{v}^{2j}]_{i+1/2,j+1/2}.
\] (3.9)
For more details, see Arakawa and Lamb (1981). In this paper, we include other possibilities in specifying \( h^{(u)} \) and \( h^{(v)} \) (see section 5).

From (3.5) and (3.6), we can obtain a discrete version of (2.6) as
\[
\frac{\partial}{\partial t} (h^{(v)}q)_{i,j} = \frac{1}{d} \left[ -v^*_{i+1/2,j+1} (-\alpha_{i,j+1/2} + \phi_{i+1/2,j+1/2}) \right. \]
\[ - v^*_{i-1/2,j+1/2} \left. (\beta_{i,j+1/2} - \phi_{i-1/2,j+1/3}) + v^*_{i+1/2,j} \right]
\]
\[ q_{i,j} = -\frac{1}{6a} \left( \frac{\tau}{\gamma} \right)_i \left[ 2U_{i+1/2,j}^*(q_{i+1,j} - q_{i,j}) + 2U_{i-1/2,j}^*(q_{i-1,j} - q_{i,j}) + 2V_{i,j+1/2}^*(q_{i,j+1} - q_{i,j}) + 2V_{i,j-1/2}^*(q_{i,j-1} - q_{i,j}) + U_{i+1/2,j+1/2}^*(q_{i+1,j+1} - q_{i,j+1}) + U_{i+1/2,j-1/2}^*(q_{i+1,j-1} - q_{i,j}) + V_{i,j+1/2}^*(q_{i,j+1} - q_{i-1,j+1}) + V_{i,j-1/2}^*(q_{i,j-1} - q_{i-1,j-1}) \right], \] (3.20)

where

\[ U_{i+1/2,j}^* = \frac{1}{4} \left( u_{i+1,j+1/2}^* + u_{i,j+1/2}^* \right) \]
\[ + u_{i+1,j-1/2}^* + u_{i,j-1/2}^*, \] (3.21)

\[ V_{i,j+1/2}^* = \frac{1}{4} \left( v_{i+1/2,j+1}^* + v_{i+1/2,j}^* \right) \]
\[ + v_{i,j+1/2}^* + v_{i-1/2,j}^*, \] (3.22)

\[ U_{i+1/2,j+1/2}^* = \frac{1}{4} \left( u_{i+1,j+1/2}^* + u_{i,j+1/2}^* \right) \]
\[ + v_{i+1/2,j+1}^* + v_{i+1/2,j}^*, \] (3.23)

\[ V_{i,j+1/2}^* = \frac{1}{4} \left( v_{i+1/2,j+1}^* + v_{i+1/2,j}^* \right) \]
\[ + v_{i-1/2,j+1}^* - u_{i,j+1/2}^* - u_{i,j-1/2}^*. \] (3.24)

As in (2.14), \( \tau \) in (3.20) is a time scale to be specified.

The expression of \( q' \) given by (3.20) is rather complicated. In order to derive a simpler scheme, we need more flexibility. This can be obtained for the case of nondivergent mass flux as shown in the next section.

4. Energy conserving and potential-entropy dissipating schemes for the case of nondivergent mass flux

In this section, we derive a family of schemes with which energy is conserved and potential entropy is guaranteed to be either conserved or decreased in time when the mass flux is nondivergent. Following (2.8), we define \( \psi^* \) by

\[ u_{i+1/2,j+1/2}^* = \frac{1}{d} (\psi_{i+1,j}^* - \psi_{i-1,j+1}^*), \] (4.1)

\[ v_{i+1/2,j-1/2}^* = \frac{1}{d} (\psi_{i+1,j+1}^* - \psi_{i+1,j}^*). \] (4.2)

Then, as in the continuous case, an energy conserving and potential-entropy dissipating scheme can be obtained by replacing \( q \) in (3.14)–(3.19) by \( q + q' \) and by properly choosing \( q' \). Following the procedure shown in appendix A, we can show that the potential entropy in this system is guaranteed to decrease in time with the following choice of \( q' \):

The coefficients \( \alpha \) through \( \phi \) are yet to be specified. Arakawa and Lamb (1981) showed that the following choice of these coefficients satisfies potential entropy conservation for the general case of divergent mass flux:

\[ \alpha_{i,j+1/2} = \frac{1}{24} (2q_{i+1,j+1} + q_{i-1,j+1} + 2q_{i+1,j}), \] (3.14)

\[ \beta_{i,j+1/2} = \frac{1}{24} (q_{i+1,j+1} + 2q_{i-1,j+1} + q_{i-1,j} + 2q_{i,j}), \] (3.15)

\[ \gamma_{i,j+1/2} = \frac{1}{24} (2q_{i+1,j+1} + q_{i-1,j+1} + 2q_{i-1,j} + q_{i,j}), \] (3.16)

\[ \delta_{i,j+1/2} = \frac{1}{24} (q_{i+1,j+1} + 2q_{i,j+1} + q_{i,j} + 2q_{i,j-1}), \] (3.17)

\[ \epsilon_{i+1/2,j+1/2} = \frac{1}{24} (q_{i+1,j+1} + q_{i,j+1} - q_{i,j} - q_{i,j-1}), \] (3.18)

\[ \phi_{i+1/2,j+1/2} = \frac{1}{24} (-q_{i+1,j+1} + q_{i+1,j} + q_{i,j} - q_{i,j+1}). \] (3.19)

It is well known that, for a horizontal nondivergent flow, the use of the Arakawa-Jacobian (Arakawa 1966) for the vorticity advection \( J(\psi, \zeta) \) maintains conservation of enstrophy and kinetic energy. This Jacobian can be applied directly to the Jacobian in (2.9) to obtain a finite-difference scheme for the potential vorticity equation that conserves potential enstrophy and energy for a nondivergent mass flux. However, since we are
now interested in deriving a family of schemes with which potential enstrophy can decrease in time, while energy is conserved, we generalize the Arakawa-Jacobian $J_d$ so that

$$\frac{\partial}{\partial t} (h^{(a)} q)_{i,j} = - \frac{1}{6 d^2} \{ (\psi^{*}_{i-1,j+1} + \psi^{*}_{i+1,j-1} - \psi^{*}_{i,j+1} - \psi^{*}_{i,j-1}) \tilde{q}_{i-1/2,j} \\
- (\psi^{*}_{i,j+1} - \psi^{*}_{i,j+1,j+1}) \tilde{q}_{i+1/2,j} \\
- (\psi^{*}_{i-1,j+1} - \psi^{*}_{i-1,j+1}) \tilde{q}_{i+1/2,j} \\
+ (\psi^{*}_{i+1,j} + \psi^{*}_{i+1,j+1} - \psi^{*}_{i-1,j} - \psi^{*}_{i+1,j+1}) \tilde{q}_{i+1/2,j+1/2} \\
- (\psi^{*}_{i,j} - \psi^{*}_{i,j+1} - \psi^{*}_{i,j+1}) \tilde{q}_{i-1/2,j+1/2} \\
+ (\psi^{*}_{i+1,j} - \psi^{*}_{i+1,j+1}) \tilde{q}_{i+1/2,j+1/2} - (\psi^{*}_{i,j} - \psi^{*}_{i,j}) \tilde{q}_{i-1/2,j+1/2} \\
\times \tilde{q}_{i-1/2,j+1/2} + (\psi^{*}_{i+1,j} - \psi^{*}_{i+1,j+1}) \tilde{q}_{i+1/2,j+1/2} - (\psi^{*}_{i,j} - \psi^{*}_{i,j}) \tilde{q}_{i-1/2,j+1/2} \}.$$

(4.3)

Here $\tilde{q}$ and $\tilde{q}$ are linear combinations of $q$ at neighboring grid points. If we choose

$$\tilde{q}_{i+1/2,j} = \frac{1}{2} (q_{i+1,j} + q_{i,j}),$$

(4.4)

$$\tilde{q}_{i,j+1/2} = \frac{1}{2} (q_{i,j+1} + q_{i,j}),$$

(4.5)

the rhs of (4.3) becomes $[J_d(q, \psi^*)]_{i,j}$. For simplicity, we let

$$\epsilon_{i+1/2,j+1/2} = \phi_{i+1/2,j+1/2} = 0.$$  

(4.8)

Rewriting the rhs of (3.10) using (4.1), (4.2), and (4.8), we obtain

$$\begin{align*}
\alpha_{i,j+1/2} - \gamma_{i+1,j+1/2} &= \frac{1}{6} (-\tilde{q}_{i+1/2,j} + \tilde{q}_{i,j+1/2}), \\
-\alpha_{i-1,j-1/2} + \gamma_{i,j-1/2} &= \frac{1}{6} (-\tilde{q}_{i-1/2,j} + \tilde{q}_{i,j-1/2}), \\
-\beta_{i,j+1/2} + \delta_{i+1,j+1/2} &= \frac{1}{6} (\tilde{q}_{i-1/2,j} - \tilde{q}_{i,j+1/2}), \\
\beta_{i+1,j-1/2} - \delta_{i,j-1/2} &= \frac{1}{6} (\tilde{q}_{i+1/2,j} - \tilde{q}_{i,j-1/2});
\end{align*}$$

(4.10)

$$\begin{align*}
-\alpha_{i,j+1/2} + \beta_{i,j+1/2} + \gamma_{i,j+1/2} - \delta_{i,j+1/2} &= \frac{1}{6} (-\tilde{q}_{i+1/2,j} + \tilde{q}_{i-1/2,j} - \tilde{q}_{i+1/2,j+1/2} + \tilde{q}_{i-1/2,j+1/2}), \\
\alpha_{i,j-1/2} - \beta_{i,j-1/2} + \gamma_{i,j-1/2} + \delta_{i,j-1/2} &= \frac{1}{6} (\tilde{q}_{i+1/2,j} - \tilde{q}_{i-1/2,j} - \tilde{q}_{i+1/2,j-1/2} + \tilde{q}_{i-1/2,j-1/2}); \\
-\alpha_{i,j-1/2} + \beta_{i,j-1/2} + \gamma_{i,j+1/2} + \delta_{i,j+1/2} &= \frac{1}{6} (\tilde{q}_{i,i+1/2} - \tilde{q}_{i,j-1/2} + \tilde{q}_{i+1/2,j} - \tilde{q}_{i+1/2,j-1/2}); \\
\alpha_{i-1,j+1/2} + \beta_{i-1,j+1/2} - \gamma_{i,j+1/2} - \delta_{i,j+1/2} &= \frac{1}{6} (-\tilde{q}_{i+1/2,j} + \tilde{q}_{i,j-1/2} + \tilde{q}_{i-1/2,j} - \tilde{q}_{i-1/2,j+1/2}).
\end{align*}$$

(4.14)

(4.15)

(4.16)

(4.17)

Appendix B shows that these pairs of equations are satisfied by

$$\tilde{q}_{i+1/2,j} = \frac{1}{2} (D_{i+1,j} + D_{i,j}),$$

(4.18)

$$\tilde{q}_{i,j+1/2} = \frac{1}{2} (D_{i+1,j} + D_{i,j}),$$

(4.19)

$$\tilde{q}_{i+1/2,j+1/2} = \frac{1}{4} (E_{i+1/2,j+1/2} + D_{i+1,j+1} + D_{i,j} - D_{i,j+1} - D_{i+1,j}).$$

(4.20)
and
\[ \tilde{q}_{i+1/2,j+1/2} = \frac{1}{4} \left( E_{i+1/2,j+1/2} - D_{i+1,j+1} \right) - D_{i,j} + D_{i,j+1} + D_{i+1,j+1}. \] (4.21)

In summary, the use of (4.18)–(4.21) in the rhs of (4.3) gives a generalization of \([J_A(q, \psi^*)]_{i,j}\), while energy conservation for a nondivergent mass flux is maintained. From (4.18–4.21), \(D \to q\) and \(E \to 4q\) as \(d \to 0\) are required for consistency with the potential vorticity equation (2.6). These conditions, however, should be relaxed if the effect of potential enstrophy dissipation is added to the original equation. Two examples of specifying \(D\) and \(E\), with which potential enstrophy is guaranteed to be decreased in time, are given below.

a. D-scheme

If we choose \(E_{i+1/2,j+1/2} = D_{i+1,j+1} + D_{i,j} + D_{i,j+1} + D_{i+1,j+1}\), \(\tilde{q}\) and \(\tilde{q}\) given by (4.20) and (4.21) become

\[ \frac{\partial}{\partial t} \left( h(q) q_{i,j} \right) = -\frac{1}{12d^2} \left[ (\psi_{i,j-1}^* + \psi_{i,j-1}^* - \psi_{i,j+1}^* - \psi_{i,j+1}^*)(q_{i+1,j} + q_{i,j}) \right] \\
- (q_{i+1,j} + q_{i,j}) \left( \psi_{i,j}^* - \psi_{i,j}^* - \psi_{i,j}^* - \psi_{i,j}^* \right) \left( q_{i,j} \right) + (\psi_{i,j}^* + \psi_{i,j+1}^* - \psi_{i,j}^* + \psi_{i,j}^*) \left( q_{i+1,j} + q_{i,j} \right) \left( \psi_{i,j}^* + \psi_{i,j}^* \right) \\
\times \left( q_{i+1,j} + q_{i,j} + \frac{1}{2} E_{i+1/2,j+1/2} \right) - (\psi_{i,j}^* - \psi_{i,j}^*) \left( q_{i+1,j} + q_{i,j} - \frac{1}{2} E_{i+1/2,j+1/2} \right). \] (4.25)

The rhs of (4.25) becomes \([J_A(q, \psi^*)]_{i,j}\) when \(E' = 0\). Multiplying (4.25) by \(q_{i,j}\), taking the area mean and using (A.1), we obtain

\[ \frac{\partial}{\partial t} \left( h(q) \frac{1}{2} \frac{q^2}{q_{i,j}} \right) = \frac{1}{24d^2} \left[ (\psi_{i,j+1}^* - \psi_{i,j}^*) (q_{i+1,j+1} - q_{i,j}) + (\psi_{i,j+1}^* - \psi_{i,j}^*) (q_{i,j+1} - q_{i+1,j}) \right]. \] (4.26)

From (4.26) it is clear that potential enstrophy is decreased in time if we choose

\[ E_{i+1/2,j+1/2} = -\frac{\tau}{h_{i+1/2,j+1/2}} \]

\[ \times \frac{6}{d^2} \left[ (\psi_{i,j+1}^* - \psi_{i,j}^*) (q_{i,j+1} - q_{i,j}) \right] \\
+ (\psi_{i,j+1}^* - \psi_{i,j}^*) (q_{i,j+1} - q_{i+1,j}). \] (4.27)

Here again, \(\tau\) is a time scale to be specified. The coefficient of the rhs of (4.27) is chosen so that (4.26) is consistent with (2.15).

What remains now is to find expressions for the coefficients \(\alpha, \beta, \gamma\) and \(\delta\). Following a procedure parallel to that used in Arakawa and Lamb (1981, section 4), we can derive the following expressions:

\[ \alpha_{i,j+1/2} = \frac{1}{48} \left[ 5q_{i,j+1} + 3(q_{i+1,j} + q_{i+1,j+1} + q_{i+1,j+1} + E_{i+1/2,j+1/2}) \right], \] (4.28)

\[ \beta_{i,j+1/2} = -\frac{1}{48} \left[ 5q_{i,j+1} + 3(q_{i+1,j} + q_{i+1,j+1} + q_{i+1,j+1} + E_{i+1/2,j+1/2}) \right], \] (4.29)

\[ \gamma_{i,j+1/2} = \frac{1}{48} \left[ 5q_{i,j} + 3(q_{i,j} + q_{i+1,j}) + q_{i,j} + q_{i+1,j} + E_{i+1/2,j+1/2} \right], \] (4.30)

\[ \delta_{i,j+1/2} = -\frac{1}{48} \left[ 5q_{i,j} + 3(q_{i,j} + q_{i+1,j}) + q_{i,j} + q_{i+1,j} + E_{i+1/2,j+1/2} \right]. \] (4.31)
Here, $C$ is a linear combination of $q$ at neighboring grid points.

5. Minimizing the impact of infinite potential vorticities associated with infinitesimal fluid depth

Since $C$ in (4.28)–(4.31) is yet to be specified, the scheme still has to satisfy additional requirements. For example, by choosing a proper expression for $C$ (and proper expressions for $\epsilon$ and $\phi$ instead of (4.8)), we can reconstruct the scheme derived in section 3.

In this section, we present another unique member of the family of schemes derived in section 4. When the depth of fluid, $h$, becomes infinitesimally small, $q = (f + \xi)/h$ generally becomes infinite. (It can be finite if $f + \xi$ also becomes infinitesimally small, as in the case when a finite initial $q$ is conserved with time. With the effect of surface stress, however, $|\xi|$ should become small as $h$ decreases so that $f + \xi$ should remain finite.) In the continuous momentum equation (2.1), the term $gk \times v^*$ is still finite since $q$ is multiplied by $v^* = hv$. In the discrete case, however, the corresponding terms are not necessarily finite because they involve multiplications of $q$ by $u^*$ or $v^*$ at different grid points, as in (3.5) and (3.6). Thus, $\partial u/\partial t$ and $\partial v/\partial t$ are not necessarily finite.

To minimize the spurious impact of infinite potential vorticities at neighboring grid points on $\partial u/\partial t$ and $\partial v/\partial t$ at finite-mass grid points, we impose the following conditions:

\[
\begin{align*}
\frac{h^{(u)}}{\partial t} & \text{ is finite,} \\
\frac{h^{(v)}}{\partial t} & \text{ is finite.} 
\end{align*}
\] (5.1)

Then, for example, $\partial u/\partial t$ cannot be infinite unless $h$ at that grid point is infinitesimally small (i.e., there is no mass). Moreover, since $u$ and $v$ are finite, (5.1) with (3.3) and (3.4) indicates

\[
\begin{align*}
\frac{u^*}{\partial t} & \text{ is finite,} \\
\frac{v^*}{\partial t} & \text{ is finite.} 
\end{align*}
\] (5.2)

This guarantees that the kinetic energy conversion between $u$ and $v$ components is finite.

From (3.11) and (3.13), $q_{i,j}$ can become infinite only when $h_{i,1/2,j,1/2}$, $h_{i-1/2,j,1/2}$, and $h_{i,1/2,j-1/2}$ all approach zero. Then, when (3.5) and (3.6) are used, $u_{i,j+1/2}$, $u_{i,j-1/2}$, $v_{i+1/2,j}$, and $v_{i-1/2,j}$ all approach zero. Thus $q_{i,j}$ can be used when and only when it is multiplied by these mass fluxes. More generally, we find the following conditions:

(1) For $\alpha_{i,j+1/2}u_{i,j+1/2}v_{i+1/2,j+1}$ to be always finite, only $q_{i,j+1}$, $q_{i,j}$, $q_{i+1,j+1}$ can be used in $\alpha_{i,j+1/2}$; (5.3)

(2) For $\beta_{i,j+1/2}u_{i,j+1/2}v_{i-1/2,j+1}$ to be always finite, only $q_{i,j+1}$, $q_{i,j}$, $q_{i-1,j+1}$ can be used in $\beta_{i,j+1/2}$; (5.4)

(3) For $\gamma_{i,j+1/2}u_{i,j+1/2}v_{i-1/2,j}$ to be always finite, only $q_{i,j}$, $q_{i,j+1}$, $q_{i,j-1}$ can be used in $\gamma_{i,j+1/2}$; (5.5)

(4) For $\delta_{i,j+1/2}u_{i,j+1/2}v_{i+1/2,j}$ to be always finite, only $q_{i,j}$, $q_{i,j+1}$, $q_{i-j+1}$ can be used in $\delta_{i,j+1/2}$. (5.6)

We have developed an advection scheme which is positive definite when $h$ is continuous (Hsu 1988; Hsu and Arakawa 1990). When this scheme is applied to the continuity equation (3.1), mass fluxes approach zero as $h$ at the upstream side approaches zero; in other words,

\[
\begin{align*}
\frac{u_{i,j+1/2}}{\partial t} & \rightarrow 0 \\
\text{as } h_{i+1/2,j+1/2} & \rightarrow 0 \text{ for } u_{i,j+1/2} < 0, \\
\text{as } h_{i-1/2,j+1/2} & \rightarrow 0 \text{ for } u_{i,j+1/2} > 0;
\end{align*}
\] (5.7)

\[
\begin{align*}
\frac{v_{i+1/2,j}}{\partial t} & \rightarrow 0 \\
\text{as } h_{i+1/2,j+1/2} & \rightarrow 0 \text{ for } v_{i+1/2,j} < 0, \\
\text{as } h_{i+1/2,j-1/2} & \rightarrow 0 \text{ for } v_{i+1/2,j} > 0.
\end{align*}
\] (5.8)

It is easy to see that the conditions (5.3)–(5.6) are valid also for this scheme.

We can show that the $E$-scheme can satisfy conditions (5.3)–(5.6) if $E$ and $C$ are properly chosen. Using (4.1) and (4.2), we can express (4.27) as

\[
E_{i+1/2,j+1/2} = -\left(\frac{\tau}{h}\right)_{i+1/2,j+1/2} 
\]

\[
\times \frac{3}{d} [(u_{i,j+1/2}^{*} + u_{i+1,j+1/2}^{*})(q_{i+1,j} + q_{i+1,j+1}) 
\]

\[
- q_{i,j} - q_{i,j+1}) + (v_{i+1/2,j}^{*} + v_{i+1/2,j+1}^{*}) 
\]

\[
\times (q_{i,j+1} + q_{i+1,j+1} - q_{i,j} - q_{i,j+1})].
\] (5.9)

Let us choose

\[
C_{i+1/2,j+1/2} = \frac{1}{48} (q_{i,j} + q_{i+1,j+1} - q_{i,j+1} 
\]

\[
- q_{i+1,j} + C_{i+1/2,j+1/2}),
\] (5.10)

where

\[
C'_{i+1/2,j+1/2} = \left(\frac{\tau}{h}\right)_{i+1/2,j+1/2} 
\]

\[
\times \frac{3}{d} [(u_{i,j+1/2}^{*} + u_{i+1,j+1/2}^{*})(q_{i,j} + q_{i+1,j}) 
\]

\[
- q_{i,j+1} - q_{i+1,j+1}) + (v_{i+1/2,j}^{*} + v_{i+1/2,j+1}^{*}) 
\]

\[
\times (q_{i,j} + q_{i+1,j} - q_{i,j+1} - q_{i+1,j+1})].
\] (5.11)

Substituting (5.9) and (5.10) with (5.11) into (4.28)–(4.31), we obtain
\[
\alpha_{i,j+1/2} = \frac{1}{12} \left( q_{i,j+1} + q_{i,j} + q_{i+1,j+1} - \frac{\tau}{h} \right)_{i+1/2,j+1/2} \\
\times \frac{1}{2d} U^w_{i+1/2,j+1/2} (q_{i+1,j+1} - q_{i,j}), \quad (5.12)
\]

\[
\beta_{i,j+1/2} = \frac{1}{12} \left( q_{i,j+1} + q_{i,j} + q_{i-1,j+1} - \frac{\tau}{h} \right)_{i-1/2,j+1/2} \\
\times \frac{1}{2d} V^w_{i-1/2,j+1/2} (q_{i-1,j+1} - q_{i,j}), \quad (5.13)
\]

\[
\gamma_{i,j+1} = \frac{1}{12} \left( q_{i,j} + q_{i+1,j} + q_{i-1,j} - \frac{\tau}{h} \right)_{i-1/2,j+1/2} \\
\times \frac{1}{2d} U^w_{i-1/2,j+1/2} (q_{i+1,j} - q_{i-1,j}), \quad (5.14)
\]

\[
\delta_{i,j+1/2} = \frac{1}{12} \left( q_{i,j} + q_{i+1,j} + q_{i+1,j} - \frac{\tau}{h} \right)_{i+1/2,j+1/2} \\
\times \frac{1}{2d} V^w_{i+1/2,j+1/2} (q_{i+1,j} - q_{i,j}), \quad (5.15)
\]

where \( U^w \) and \( V^w \) are defined by (3.23) and (3.24). It is clear that these \( \alpha, \beta, \gamma, \) and \( \delta \) satisfy the conditions (5.3)–(5.6). When \( \tau = 0 \), these choices correspond to the scheme derived by Sadourny and subsequently tested by the European Centre for Medium Range Weather Forecasts (Burrage and Haseler 1977).

6. Summary and further comments

In this paper we have derived a family of finite-difference schemes for the shallow water equations, with which potential enstrophy is guaranteed to be either conserved or decreased in time while energy is conserved (when the mass flux is nondivergent and time is continuous). The amount of potential enstrophy decrease is controlled by a time scale that can be arbitrarily specified.

By adding an energy-dissipating term if necessary, a scheme in this family can be used as a basis for discrete shallow water equations in which the amount of energy dissipation and that of potential-enstrophy dissipation can be independently controlled based on physical reasoning.

A member of the family that uses (5.12)–(5.15) in (3.5) and (3.6) is unique in that the spurious impact of infinite potential vorticity associated with infinitesimal fluid depth is minimized. The scheme is, therefore, especially useful when the free surface can intersect the bottom boundary. In their numerical model of the atmosphere with an isotropic vertical coordinate, Hsu and Arakawa (1990) used this scheme combined with a discrete continuity equation that satisfies (5.7) and (5.8) (see section 6 and appendix B of their paper).

Acknowledgments. This paper is based on a part of the Ph.D. dissertation by the second author (Hsu 1988), whose graduate study at UCLA was supported by the National Science Council of the Republic of China. The research was also supported by NSF under Grants ATM-8515013, ATM-8617146, and ATM-8910564, and by NASA under Grant NAG-5-789. The authors would like to thank Ms. Mary Grace Bigornia for typing the manuscript.

APPENDIX A

A Potential-Enstrophy Dissipating Scheme

For simplicity, let us assume that the domain is periodic so that, for an arbitrary variable \( A \),

\[
\overline{A_{i,j}} = \overline{A_{i+1,j}}, \quad (A.1)
\]

where an overbar denotes the average over all grid points.

Using the coefficients (3.14)–(3.19) and notations (3.21)–(3.24) in the rhs of (3.10) and replacing \( q \) by \( D \), we obtain

\[
\frac{\partial}{\partial t} (\rho q)_{i,j} = -\frac{1}{6d} [2U^*_{i+1/2,j}(D_{i+1,j} + D_{i,j}) \\
- 2U^*_{i-1/2,j}(D_{i,j} + D_{i+1,j}) + 2V^*_{i,j+1/2}(D_{i,j+1} + D_{i,j}) \\
- 2V^*_{i,j-1/2}(D_{i,j} + D_{i,j-1}) + U^*_{i+1/2,j+1/2} \\
\times (D_{i+1,j+1} + D_{i,j}) - U^*_{i-1/2,j-1/2}(D_{i,j} + D_{i+1,j-1}) \\
+ V^*_{i+1/2,j+1/2}(D_{i+1,j+1} + D_{i,j}) \\
- V^*_{i-1/2,j-1/2}(D_{i+1,j} + D_{i+1,j-1})]. \quad (A.2)
\]

Using (3.1), (3.2), (3.3) and (3.21–3.24), on the other hand, we may write the continuity equation at \( q \)-points as

\[
\frac{\partial}{\partial t} k_{i,j}^{(q)} = -\frac{1}{d} [U^*_{i+1/2,j} - U^*_{i-1/2,j} \\
+ V^*_{i,j+1/2} - V^*_{i,j-1/2}] \quad (A.3)
\]

or

\[
\frac{\partial}{\partial t} k_{i,j}^{(q)} = -\frac{1}{d} [U^w_{i+1/2,j+1/2} - U^w_{i-1/2,j-1/2} \\
+ V^w_{i,j+1/2} - V^w_{i,j-1/2}] . \quad (A.4)
\]

Let

\[
D_{i,j} = q_{i,j} + q'_{i,j}. \quad (A.5)
\]

Then, using (A.1) in

\[
\overline{q_{i,j} \times (A.2) - (1/6)q_{i,j}^2(2 \times (A.3) + (A.4))},
\]

we obtain
\[
\frac{\partial}{\partial t} \left( h^{(2)} \left( \frac{1}{2} q^2 \right) \right)_{ij} = \frac{1}{6d} \left[ 2U^*_{i+1/2,j}(q'_{i+1,j} + q'_j)(q_{i+1,j} - q_{i,j}) + 2U^*_{i-1/2,j}(q_{i,j} - q_{i-1,j}) 
+ 2V^*_{i,j+1/2}(q'_{i,j+1} + q'_i)(q_{i,j+1} - q_{i,j}) 
+ U^*_{i+1/2,j+1/2}(q'_{i+1,j+1} + q'_i)(q_{i+1,j+1} - q_{i,j}) 
+ V^*_{i-1/2,j+1/2}(q'_{i-1,j+1} + q'_i)(q_{i-1,j+1} - q_{i,j}) \right]
\]

(A.6)

Using (A.1) again in (A.6),

\[
\frac{\partial}{\partial t} \left( h^{(2)} \left( \frac{1}{2} q^2 \right) \right)_{ij} = \frac{1}{6d} \left[ q'_{i,j}[2U^*_{i+1/2,j}(q_{i+1,j} - q_{i,j}) + 2U^*_{i-1/2,j}(q_{i,j} - q_{i-1,j}) 
+ 2V^*_{i,j+1/2}(q_{i,j+1} - q_{i,j}) + 2V^*_{i,j-1/2}(q_{i,j} - q_{i,j-1}) 
+ U^*_{i+1/2,j+1/2}(q_{i+1,j+1} - q_{i,j}) + U^*_{i-1/2,j-1/2}(q_{i,j} - q_{i-1,j}) 
+ V^*_{i-1/2,j+1/2}(q_{i-1,j+1} - q_{i,j}) + V^*_{i+1/2,j-1/2}(q_{i,j} - q_{i+1,j}) \right]
\]

(A.7)

From (A.7), it is clear that the rhs of (A.7) is negative with \( q'_{i,j} \) given by (3.20).

**APPENDIX B**

**Derivation of (4.18)-(4.21)**

For (4.10), (4.12), (4.14) and (4.16) to be consistent with (4.11), (4.13), (4.15) and (4.17), respectively, the following conditions are required:

\[
\begin{align*}
\hat{q}_{i+1/2,j} - \hat{q}_{i,j+1/2} + \hat{a}_{i+1/2,j+1} - \hat{a}_{i+1,j+1/2} &= 0, \quad (B.1) \\
\hat{q}_{i-1/2,j} - \hat{a}_{i-1/2,j+1/2} + \hat{a}_{i-1,j+1/2} &= 0, \quad (B.2) \\
-\hat{q}_{i+1/2,j} + \hat{q}_{i-1/2,j} + \hat{a}_{i+1/2,j+1/2} - \hat{a}_{i-1/2,j+1/2} &= 0, \quad (B.3) \\
\hat{a}_{i+1,j} - \hat{a}_{i-1,j} + \hat{a}_{i+1/2,j+1} - \hat{a}_{i-1/2,j+1} &= 0, \quad (B.4)
\end{align*}
\]

Note that (B.1) and (B.2) are identical and satisfied by

\[
\hat{a}_{i+1/2,j+1/2} = \hat{a}_{i+1/2,j} + \hat{a}_{i+1/2,j+1} = \hat{a}_{i,j+1/2} + \hat{a}_{i+1,j+1/2}.
\]

From (B.5) and (B.6), we obtain

\[
\hat{q}_{i+1/2,j} = \frac{1}{2} (c - a)_{i+1/2,j+1/2}, \quad (B.8)
\]

and

\[
\hat{q}_{i+1/2,j+1} = \frac{1}{2} (c + a)_{i+1/2,j+1/2}.
\]

For consistency of (B.8) and (B.8'),

\[
(c - a)_{i+1/2,j+1/2} = (c + a)_{i+1/2,j+1/2}. \quad (B.9)
\]

From (B.5) and (B.7), on the other hand, we obtain

\[
\hat{a}_{i,j+1/2} = \frac{1}{2} (c - b)_{i+1/2,j+1/2}, \quad (B.10)
\]

\[
\hat{a}_{i+1,j+1/2} = \frac{1}{2} (c + b)_{i+1/2,j+1/2}. \quad (B.10')
\]

For consistency of (B.10) and (B.10'),

\[
(c - b)_{i+1/2,j+1/2} = (c + b)_{i+1/2,j+1/2}. \quad (B.11)
\]

To find expressions for \( a \) and \( b \), we substitute (B.6) into (B.3) and (B.7) into (B.4). The results are

\[
\begin{align*}
a_{i+1/2,j+1/2} - a_{i-1/2,j+1/2} - (\hat{q} - \hat{a})_{i+1/2,j+1/2} \\
- (\hat{q} - \hat{a})_{i-1/2,j+1/2} &= 0, \quad (B.12) \\
b_{i+1/2,j+1/2} + b_{i+1,j+1/2} + (\hat{q} - \hat{a})_{i+1/2,j+1/2} \\
+ (\hat{q} - \hat{a})_{i-1/2,j+1/2} &= 0. \quad (B.13)
\end{align*}
\]

Since \( \Sigma_i (B.12) \) and \( \Sigma_j (B.13) \) with (A.1) give

\[
-2 \sum_i (\hat{q} - \hat{a})_{i+1/2,j+1/2} = 0
\]

and

\[
2 \sum_j (\hat{q} - \hat{a})_{i+1/2,j+1/2} = 0,
\]

\( \hat{q} - \hat{a} \) must have a form of differences in both \( x \) and \( y \) directions. Thus we may write

\[
(\hat{q} - \hat{a})_{i+1/2,j+1/2} = \frac{1}{2} (D_{i+1,j+1} + D_{i,j} - D_{i,j+1} - D_{i+1,j}), \quad (B.14)
\]

where \( D \) is a linear combination of \( q \) at neighboring grid points. Using (B.14) in (B.12),
\[ 2a_{i+1/2,j+1/2} - D_{i+1,j+1} + D_{i+1,j} - D_{i,j+1} + D_{i,j} = 2a_{i-1/2,j+1/2} - D_{i,j+1} + D_{i,j} - D_{i-1,j+1} + D_{i-1,j}. \] (B.15)

Thus the expression on each side of (B.15) is constant as \( i \) is changed. Taking the constant as zero,

\[ a_{i+1/2,j+1/2} = \frac{1}{2} (D_{i+1,j+1} - D_{i+1,j} + D_{i,j+1} - D_{i,j}). \] (B.16)

Following a similar argument using (B.14) in (B.13),

\[ b_{i+1/2,j+1/2} = \frac{1}{2} (D_{i+1,j+1} - D_{i,j+1} + D_{i+1,j} - D_{i,j}). \] (B.17)

With (B.16), (B.9) is satisfied by

\[ c_{i+1/2,j+1/2} = \frac{1}{2} (D_{i+1,j+1} + D_{i,j+1} + D_{i+1,j} + D_{i,j}). \] (B.18)

Another expression for \( c \) may be obtained from (B.17) and (B.11). The result is, however, identical to (B.18).

The expressions for \( \tilde{q} \) and \( \tilde{\tilde{q}} \) in terms of \( D \) can now be obtained. Using (B.16) and (B.18) in (B.8), we obtain (4.18). Using (B.17) and (B.18) in (B.10), on the other hand, we obtain (4.19). To express \( \tilde{q} \) and \( \tilde{\tilde{q}} \), let

\[ (\tilde{q} + \tilde{\tilde{q}})_{i+1/2,j+1/2} = \frac{1}{2} E_{i+1/2,j+1/2}. \] (B.19)

Then, from (B.14) and (B.19), we obtain (4.20) and (4.21).

REFERENCES


