JOURNAL OF COMPUTATIONAL PHYSICS 37, 1–18 (1980)

A Linear ADI Method for the Shallow-Water Equations

G. FAIRWEATHER

Department of Mathematics, University of Kentucky, Lexington, Kentucky 40506

AND

I. M. NAVON

National Research Institute for Mathematical Sciences, CSIR, P.O. Box 395, Pretoria 0001, South Africa

Received August 30, 1977; revised June 6, 1979

A new ADI method is proposed for the approximate solution of the shallow-water equations. Based on a perturbation of a linearized Crank–Nicolson-type discretization, this method is algebraically linear and also locally second-order correct in time. Its performance on a test problem given in [1] demonstrates that it requires less computer time per time step than the fastest quasi-Newton method proposed in [1], and for given mesh parameters it appears to be marginally more accurate. The results of several long runs are also reported, and phenomena similar to those described in [14] are exhibited. In particular, it is shown that neither the new method nor the best of the methods of [1] conserves potential enstrophy and, as a result, after a finite time, these methods always "blow up."

1. INTRODUCTION

To avoid the Courant–Friedrichs–Levy (CFL) stability condition, restricting the time step in explicit finite-difference approximations to quasi-linear hyperbolic partial differential equations, implicit methods must be used. It is known [2, 3] that in hyperbolic problems any dynamical degree of freedom that is stabilized by an implicit scheme in time, is treated inaccurately, viz., its phase speed is slowed down so that the explicit stability criteria are satisfied. On the other hand, in meteorology, the high-speed oscillations which impose the severe upper limit on the time interval are unimportant, so that the deceleration of their phase speed poses no problem.

The numerical solution of nonlinear hyperbolic initial/boundary value problems in two space dimensions using an implicit method constitutes a formidable computational task. However, when the spatial region is rectangular this task can be simplified by using an alternating-direction implicit (ADI) method. Such methods reduce multidimensional problems to systems of one-dimensional problems [5, 8, 9]. In this paper a new ADI method is proposed for the approximate solution of the

0021-9991/80/100001–18$02.00/0

shallow-water equations, the primitive equations for an incompressible, inviscid fluid with a free surface, using the β-plane approximation on a rectangular domain. Based on a perturbation of a linearized Crank–Nicolson-type discretization, this method is algebraically linear, that is, at each time step it requires the solution of systems of linear algebraic equations whereas the method of [1] involves systems of nonlinear equations. Like the method of [1], the new procedure is locally second-order correct in time.

An outline of the remainder of the paper is as follows. The differential equations and boundary conditions are given in Section 2. The salient features of Gustafsson's method [1] are described in Section 3, and in Section 4 the new ADI method is derived. In Section 5 the new method is tested on the same problem as that considered in [1] and using the same computer system (CDC 6600). It is shown that the new method requires less computer time per time step than the best of Gustafsson's methods [1], QNEX1, and is marginally more accurate. The results of several long runs are of particular interest since they exhibit phenomena similar to those described in [14]. In particular it is shown that neither the new method nor Gustafsson's method, QNEX1, conserve potential enstrophy, and, as a result, after a certain time, which can be delayed by increasing the resolution, these methods "blow up."

2. The Differential Equations and Boundary Conditions

Throughout this paper we shall adopt the notation used in [1]. We denote by w the vector function

$$w = (u, v, \Phi)^T, \tag{1}$$

where $(w)^T$ denotes the transpose of w; u, v are the velocity components in the x and y directions, respectively; and

$$\Phi = 2(gh)^{1/2},$$

where h is the depth of the fluid and g is the acceleration due to gravity. The shallow-water equations can then be written in the form (see [4])

$$\frac{\partial w}{\partial t} = A(w)\frac{\partial w}{\partial x} + B(w)\frac{\partial w}{\partial y} + C(y)w, \tag{2}$$
$$0 \leqslant x \leqslant L, \qquad 0 \leqslant y \leqslant D, \qquad t \geqslant 0.$$

In (2) A, B, and C are matrices given by

$$A = -\begin{bmatrix} u & 0 & \Phi/2 \\ 0 & u & 0 \\ \Phi/2 & 0 & u \end{bmatrix} \qquad B = -\begin{bmatrix} v & 0 & 0 \\ 0 & v & \Phi/2 \\ 0 & \Phi/2 & v \end{bmatrix} \qquad C = \begin{bmatrix} 0 & f & 0 \\ -f & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{3}$$

where f is the Coriolis term given by

$$f = \hat{f} + \beta(y - D/2) \tag{4}$$

with $\hat{f}$ and β constants. Periodic boundary conditions are assumed in the x-direction.

$$w(x, y, t) = w(x + L, y, t), \tag{5}$$

while in the y-direction

$$v(x, 0, t) = v(x, D, t) = 0. \tag{6}$$

Initially

$$w(x, y, 0) = \psi(x, y). \tag{7}$$

It is easy to see (cf. [3, 4]) that

$$E = \tfrac{1}{2} \int_0^L \int_0^D (u^2 + v^2 + \Phi^2/4)\, \Phi^2/(4g)\, dx\, dy \tag{8}$$

is independent of time, i.e., the total energy is conserved. Also, the average value of the height of the free surface, $\bar{h}$, is independent of time, i.e.,

$$\bar{h} = \int_0^L \int_0^D h\, dx\, dy \Big/ \bar{A} \tag{9}$$

is conserved, $\bar{A}$ being the area of the spatial domain.

3. Gustafsson's ADI Method

In order to describe Gustafsson's method and the new ADI method we require the following notation.

Let N_x and N_y be positive integers and set

$$\Delta x = L/N_x, \qquad \Delta y = D/N_y. \tag{10}$$

We shall denote by w_{jk}^n an approximation to $w(j\Delta x, k\Delta y, n\Delta t)$ and by D_{0x}, D_{+x}, D_{-x} the basic difference operators

$$\begin{aligned} D_{0x} w_{jk}^n &= (w_{j+1,k}^n - w_{j-1,k}^n)/(2\Delta x), \\ D_{+x} w_{jk}^n &= (w_{j+1,k}^n - w_{jk}^n)/\Delta x, \\ D_{-x} w_{jk}^n &= (w_{jk}^n - w_{j-1,k}^n)/\Delta x, \end{aligned} \tag{11}$$

respectively, with similar definitions for D_{oy}, D_{+y}, and D_{-y}. Also define the operators P_{jk}^n and Q_{jk}^n by

$$P_{jk}^n = \frac{\Delta t}{2}\left[A(w_{jk}^n)\, D_{0x} + C_k^{(1)}\right], \tag{12}$$

$$Q_{jk}^n = \frac{\Delta t}{2}\left[B(w_{jk}^n)\, D_k + C_k^{(2)}\right], \tag{13}$$

where

$$C_k^{(1)} = \begin{bmatrix} 0 & 0 & 0 \\ -f_k & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad C_k^{(2)} = \begin{bmatrix} 0 & f_k & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{14}$$

and, because of the boundary conditions in the y-direction,

$$D_k = \begin{cases} D_{0y} & \text{for} \quad k = 1, 2, \ldots, N_y - 1; \\ D_{+y} & \text{for} \quad k = 0; \\ D_{-y} & \text{for} \quad k = N_y. \end{cases} \tag{15}$$

For the approximate solution of the initial/boundary value problem consisting of (2) and (5)–(7), Gustafsson [1] proposed an ADI difference scheme, which takes the form

$$(1 - P_{jk}^{n+1/2})\, w_{jk}^{n+1/2} = (1 + Q_{jk}^n)\, w_{jk}^n, \tag{16a}$$

$$(1 - Q_{jk}^{n+1})\, w_{jk}^{n+1} = (1 + P_{jk}^{n+1/2})\, w_{jk}^{n+1/2} \tag{16b}$$

where $j = 1, \ldots, N_x$, $k = 0, \ldots, N_y$, and $n \geqslant 0$. (Because of boundary condition (6), Eqs. (16) do not apply to the v-component when $k = 0$, N_y). The first step of this procedure (16a) requires the solution of a sequence of one-dimensional problems in the x-direction, i.e., along the grid lines $k = \text{constant}$, while the second (16b) gives rise to a sequence of one-dimensional problems in the y-direction, i.e., along the grid lines $j = \text{constant}$. However, at each time step of this scheme, several systems of nonlinear algebraic equations must be solved. Two methods of solution are proposed in [1]. In the first, the systems are written in the form

$$w = r(w),$$

and solved by a simple iteration technique

$$w^{(m+1)} = r(w^{(m)}), \qquad m = 0, 1, 2, \ldots, p, \tag{17}$$

where the superscript denotes iteration index. The number of iterations in each half time step has to be chosen from the sequence 3, 4, 7, 8, 11, 12,..., in order to avoid

instability for the linear case, (cf. [15]). In [1], the iterative procedure (17) is called the GIp method.

The second method is a quasi-Newton method. If a system of nonlinear equations is written in the form

$$g(\alpha) = 0,$$

then Newton's method takes the form

$$\alpha^{(m+1)} = \alpha^{(m)} - \mathscr{J}^{-1}(\alpha^{(m)})\, g(\alpha^{(m)}),$$

where the superscript again denotes iteration index, and $\mathscr{J}(\alpha)$ is the Jacobian

$$\mathscr{J}(\alpha) = \partial(g, \alpha).$$

In order to determine $\mathscr{J}^{-1}g$, an LU decomposition of $\mathscr{J}$ is performed and $\mathscr{J}^{-1}g$ is then computed by back substitution (see [1]). The quasi-Newton method considered in [1] consists in performing the LU decomposition only every Mth time step, where M is a fixed integer. The quasi-Newton iterative procedure is then

$$\alpha^{(m+1)} = \alpha^{(m)} - \hat{\mathscr{J}}^{-1}(\alpha^{(m)})\, g(\alpha^{m}), \tag{18}$$

where

$$\hat{\mathscr{J}} = \mathscr{J}(\alpha^{(0)}) + O(\Delta t).$$

One of the most satisfactory of the quasi-Newton procedures examined in [1] is the method QNEX1, which consists of (18) with only one iteration. In this procedure $\alpha^{(0)}$ is obtained by linear extrapolation in time using the solutions at the two latest time levels [1].

4. The New ADI Method

The new ADI method is derived by first constructing a scheme which is locally second-order correct in both space and time (at least at all interior nodes). The technique of Douglas and Gunn [5] is then used to generate a perturbation of this scheme which can be factored into a sequence of one-dimensional problems.

For convenience we introduce the operators $P_{jk}^{n^*}$ and $Q_{jk}^{n^*}$ defined by

$$P_{jk}^{n^*} = \frac{\Delta t}{2}\,[A(\hat{w}_{jk}^{n})\, D_{0x} + C_k^{(1)}],$$

$$Q_{jk}^{n^*} = \frac{\Delta t}{2}\,[B(\hat{w}_{jk}^{n})\, D_k + C_k^{(2)}],$$

respectively, where

$$\hat{w}_{jk}^{n} = \tfrac{1}{2}(3w_{jk}^{n} - w_{jk}^{n-1}), \qquad n \geqslant 1,$$
$$\hat{w}_{jk}^{0} = w_{jk}^{0} + (P_{jk}^{0} + Q_{jk}^{0})\, w_{jk}^{0}.$$

Consider the method defined by

$$w_{jk}^{n+1} - w_{jk}^{n} = (P_{jk}^{n^*} + Q_{jk}^{n^*})(w_{jk}^{n+1} + w_{jk}^{n}), \qquad (19a)$$
$$j = 1,..., N_x, \qquad k = 0,..., N_y, \qquad n \geqslant 0,$$

with

$$\begin{aligned} w_{0k}^{n} &= w_{N_x k}^{n}, && k = 0,..., N_y, \quad n \geqslant 0, \\ v_{j0}^{n} &= v_{jN_y}^{n} = 0, && j = 0,..., N_x, \quad n \geqslant 0, \\ w_{jk}^{0} &= \psi(x_j, y_k), && j = 1,..., N_x, \quad k = 0,..., N_y. \end{aligned} \qquad (19b)$$

(Note that because of boundary condition (6), (19a) does not apply to the v-component when $k = 0, N_y$.) It is easy to see that the finite-difference method (19a), which resembles the method of Lees [6] for nonlinear parabolic problems, is locally second-order correct in time for the approximate solution of (2) and, moreover, is algebraically linear. However, it does have one major drawback. Although the totality of difference equations at each time step gives rise to a linear algebraic system with a sparse coefficient matrix, the system is difficult to solve. But if the difference equations are perturbed by a term which gives rise to a $O((\Delta t)^2)$ error locally, the resulting procedure has the same local accuracy as (19a) and it can be written as a system of one-dimensional problems. To determine an appropriate perturbation which will yield such a procedure we use the techniques proposed in [5]. Following (2.7) of [5] we obtain the following ADI method:

$$\begin{aligned} (1 - P_{jk}^{n^*})\, w_{jk}^{(n+1)^*} &= (1 + P_{jk}^{n^*} + 2Q_{jk}^{n^*})\, w_{jk}^{n}, \\ (1 - Q_{jk}^{n^*})\, w_{jk}^{n+1} &= w_{jk}^{(n+1)^*} - Q_{jk}^{n^*} w_{jk}^{n}, \end{aligned} \qquad (20)$$

where $w_{jk}^{(n+1)^*}$ is an auxiliary solution (cf. $w_{jk}^{n+1/2}$ of (16), which is considered as an approximation to $w(j\Delta x, k\Delta y, (n+\frac{1}{2})\Delta t)$). On eliminating this quantity from (20) we obtain

$$(1 - P_{jk}^{n^*})(1 - Q_{jk}^{n^*})\, w_{jk}^{n+1} = (1 + P_{jk}^{n^*})(1 + Q_{jk}^{n^*})\, w_{jk}^{n} \qquad (21)$$

or

$$w_{jk}^{n+1} - w_{jk}^{n} = (P_{jk}^{n^*} + Q_{jk}^{n^*})(w_{jk}^{n+1} + w_{jk}^{n}) - P_{jk}^{n^*} Q_{jk}^{n^*}(w_{jk}^{n+1} - w_{jk}^{n}), \qquad (22)$$

which is clearly a perturbation of (19a).

A procedure which is more convenient than (20), or the usual Peaceman–Rachford-type splitting of (21) for the computation of w_{jk}^{n+1}, is

$$(1 - P_{jk}^{n^*})\, w_{jk}^{(n+1)^*} = V_{jk}^{n}, \qquad (23a)$$
$$(1 - Q_{jk}^{n^*})\, w_{jk}^{n+1} = 2w_{jk}^{(n+1)^*} - V_{jk}^{n}, \qquad (23b)$$

where

$$V_{jk}^{n} = (1 + Q_{jk}^{n^*})\, w_{jk}^{n}. \qquad (24)$$

Upon eliminating $w_{jk}^{(n+1)^*}$ between (23a) and (23b) we obtain Eq. (21), after some manipulation. Note that because of boundary condition (6) the Eqs. (23a) and (23b) do not apply to the v-component when $k = 0, N_y$. In the following this new ADI method is called ADIF.

When system (2) is linear with time-independent coefficients, the new ADI method (21) and Gustafsson's method (16) (with $w_{jk}^{n+1/2}$ eliminated) are identical. Consequently the linear stability analysis given in [1] for (16) also applies to (21).

Consider now the algebraic problem associated with ADIF, (23). Because of the form of the matrices A and B, no more than two variables are coupled to each other on the left-hand sides of the equations in (23). When the unknowns are ordered along horizontal rows for $w^{(n+1)^*}$, we first solve for the coupled variables $(u_{jk}^{(n+1)^*}, \Phi_{jk}^{(n+1)^*})$, $j = 1,..., N_x$, $k = 0,..., N_y$, and only then for the variables $v_{jk}^{(n+1)^*}$, $j = 1,..., N_x$, $k = 1,..., N_y - 1$. The elements of the vector $w^{(n+1)^*}$ are determined by solving for $(u_{jk}^{(n+1)^*}, \Phi_{jk}^{(n+1)^*})$ a sequence of linear systems whose coefficient matrices are cyclic block-tridiagonal matrices, each block being (2×2), and then solving for $v_{jk}^{(n+1)^*}$ a sequence of linear systems whose coefficient matrices are scalar cyclic tridiagonal. The cyclic form of the tridiagonal matrices arises from the assumption of periodic boundary conditions in the x-direction. The scalar cyclic tridiagonal systems are solved using a routine from [13] which is based on the algorithm described in [7] while a generalization of the algorithm of [7] is used for the block cyclic tridiagonal systems.

When solving for w_{jk}^{n+1} from (23b), the unknowns are ordered along vertical lines. The coupled variables $(v_{jk}^{n+1}, \Phi_{jk}^{n+1})$, $k = 0,..., N_y, j = 1,..., N_x$ are determined first, by solving block-tridiagonal systems using the subroutine BT of [10]. The quantities u_{jk}^{n+1}, $k = 0,..., N_y$, $j = 1,..., N_x$, are then obtained by solving scalar tridiagonal systems using the standard routine; see, for example, [12].

5. Numerical Results

5.1. *The Test Problem*

To compare the computational efficiency and accuracy of the new ADI method (23) with those of the method called QNEX1 in [1], the test problem of [1] was used, viz., the initial height field condition No. 1 used by Grammeltvedt [11]

$$h(x, y) = H_0 + H_1 \tanh\left(\frac{9(D/2 - y)}{2D}\right) + H_2 \operatorname{sech}^2\left(\frac{9(D/2 - y)}{D}\right) \sin\left(\frac{2\pi x}{L}\right). \tag{25}$$

The initial velocity fields were derived from the initial height field via the geostrophic relationship

$$u = \left(\frac{-g}{f}\right)\frac{\partial h}{\partial y}, \qquad v = \left(\frac{g}{f}\right)\frac{\partial h}{\partial x}. \tag{26}$$

The constants used were

$$L = 4400 \text{ km}, \qquad g = 10 \text{ msec}^{-2},$$
$$D = 6000 \text{ km}, \qquad H_0 = 2000 \text{ m},$$
$$\hat{f} = 10^{-4} \text{ sec}^{-1}, \qquad H_1 = 220 \text{ m},$$
$$\beta = 1.5 \times 10^{-11} \text{ sec}^{-1} m^{-1}, \qquad H_2 = 133 \text{ m}.$$

The time and space increments used in the short runs (2 days) were

(a) $\Delta x = \Delta y = 200$ km, $\Delta t = 1800$ sec,
(b) $\Delta x = \Delta y = 200$ km, $\Delta t = 3600$ sec.

For the long runs the space and time increments

$$\Delta x = \Delta y = 500 \text{ km}, \qquad \Delta t = 3600 \text{ sec},$$

were also used.

All of the calculations were carried out on a CDC 6600 computer at the South African Meteorological Office. Gustafsson's experiments [1] were performed on a similar computer.

5.2. *Computational Efficiency*

The methods QNEX1 and ADIF were first compared by finding the run times in seconds per full time step. The results are shown in Table I, where OPT = 2 is an optimized CDC version of FORTRAN, and OPT = 0 is a fast two-pass compilation with little optimization of object code. It is evident that ADIF is almost twice as fast as any of the versions of QNEX1 of [1], a result which was not unexpected because, while the methods have similar operation counts ($[115 + 152/\text{M}]\, N_x N_y$ operations per full time step for QNEX1, $143 N_x N_y$ for ADIF) QNEX1 is more complicated to program than ADIF. Only the method GI3 of [1] is faster than ADIF. However, the GI3 method has a convergence criterion imposing a limit on $\lambda_x = \Delta t/\Delta x$, which is only four times as large as the CFL criterion for explicit schemes, and its convergence is very slow if λ_x, λ_y are near the convergence limit.

TABLE I

Method	Run time per full time step (sec)
QNEX1, M = 12	0.43
QNEX1, M = 6	0.49
ADIF (with OPT = 0)	0.25
ADIF (with OPT = 2)	0.23
GI3	0.16

TABLE II

$\| E_{AP} \| / \| w_{EX} \|$ ($t = 2$ days)

Method	$\Delta x = \Delta y = 500$ km		$\Delta x = \Delta y = 200$ km	
	$\Delta t = 3600$ sec	$\Delta t = 1800$ sec	$\Delta t = 3600$ sec	$\Delta t = 1800$ sec
QNEX1 ($M = 6$)	$6..7 \times 10^{-4}$	3.1×10^{-4}	1.0×10^{-4}	4.4×10^{-5}
ADIF	5.4×10^{-4}	2.3×10^{-4}	8.7×10^{-5}	3.9×10^{-5}

5.3. *Accuracy*

In order to provide a basis for comparison between the new method (ADIF) and Gustafsson's QNEX1 [1], we assume that the exact solution of the initial/boundary value problem (2), (5)–(7), w_{EX}, is the solution of (23) computed with a fine discretization, viz., $\Delta x = \Delta y = 50$ km and $\Delta t = 450$ sec. As in [1], the relative error in an approximate solution, w_{AP}, is measured in the norm $\| \cdot \|$ defined by the inner product

$$(\alpha, \beta) = \Delta x \, \Delta y \sum_{j=1}^{N_x} \left\{ \sum_{k=1}^{N_y - 1} \alpha_{jk}^T \beta_{jk} + \tfrac{1}{2} (\alpha_{j0}^T \beta_{j0} + \alpha_{jN_y}^T \beta_{jN_y}) \right\},$$

where α and β are grid functions satisfying the boundary conditions given in (19b). The relative errors in the approximations determined by QNEX1 ($M = 6$) and ADIF are shown in Table II, where $E_{AP} = w_{AP} - w_{EX}$.

The height field was also computed with each method using $\Delta x = \Delta y = 200$ km. Figure 1 shows the initial height field, while Figs. 2 and 3 show the height field after 2 days determined from ADIF with $\Delta t = 1800$ sec and $\Delta t = 3600$ sec, respectively.

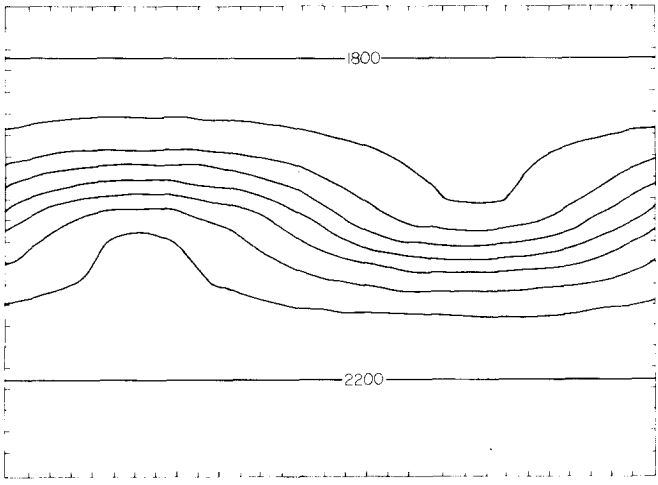

FIG. 1. The initial height field. $\Delta x = \Delta y = 200$ km.

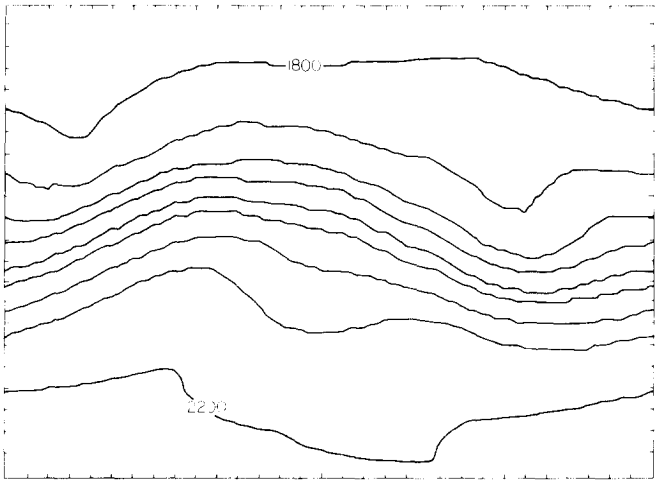

FIG. 2. The height field after 2 days for ADIF. $\Delta t = 1800$ sec, $\Delta x = \Delta y = 200$ km.

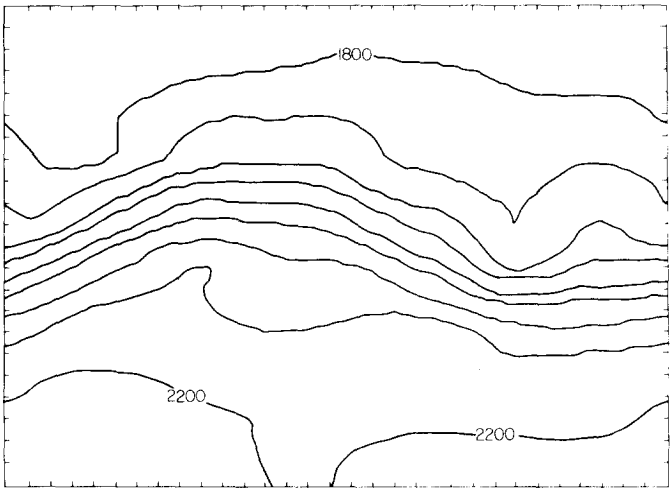

FIG. 3. The height field after 2 days for ADIF. $\Delta t = 3600$ sec, $\Delta x = \Delta y = 200$ km.

Figures 4 and 5 show the corresponding height fields computed with QNEX1. It is interesting to note that Fig. 3 is almost identical to Fig. 9 of [17], which shows the height field computed with a semi-implicit finite-difference method using the same mesh parameters. The ADIF method is more efficient than the semi-implicit method of [17] by a factor of 2.5.

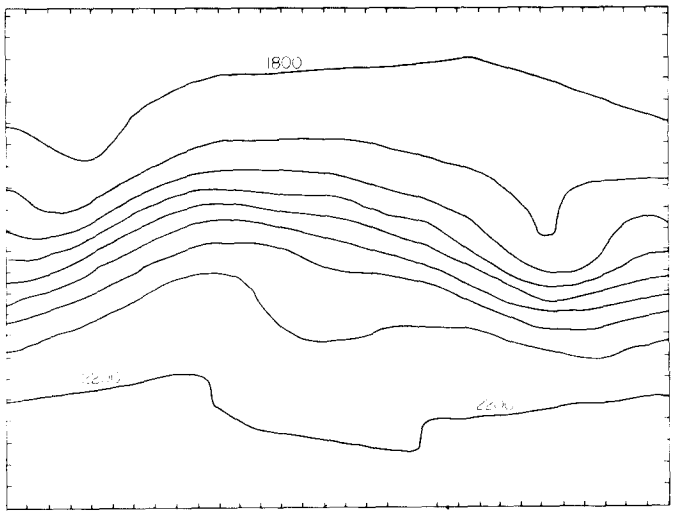

FIG. 4. The height field after 2 days for QNEX1 ($M = 6$). $\Delta t = 1800$ sec, $\Delta x = \Delta y = 200$ km.

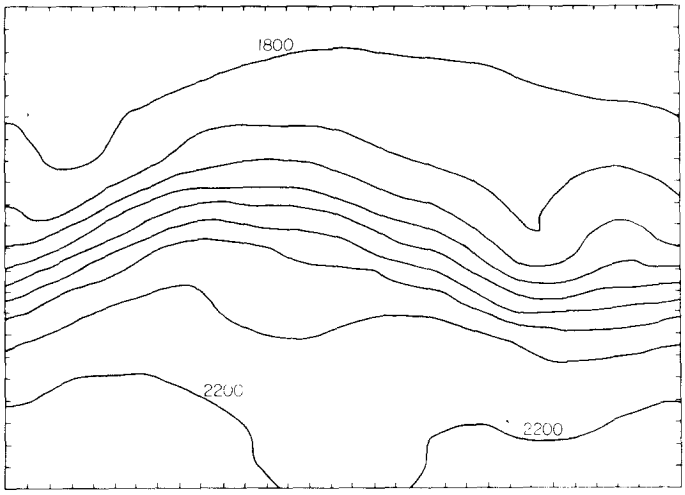

FIG. 5. The height field after 2 days for QNEX1 ($M = 6$). $\Delta t = 3600$ sec, $\Delta x = \Delta y = 200$ km.

5.4. *Long-Time Integrations*

Some long-term runs were made using both ADIF and QNEX1 ($M = 6$). The approximate solutions always "blew up"; with the coarse mesh this occurred after 12–14 days. It was observed that the "blow-up" was preceded by a sudden strong dissipation of energy (which was more pronounced in the case of ADIF than for

QNEX1). This phenomenon appears to be a characteristic of energy-conserving models for the shallow-water equations. Sadourny [14] observed that for an energy-conserving model when this dissipation of energy occurs the potential enstrophy (the mean square potential vorticity) jumps by an order of magnitude. The potential enstrophy, Z, is a third invariant of the shallow-water equations, and is defined in the following way. If we denote by

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial v}{\partial y},$$

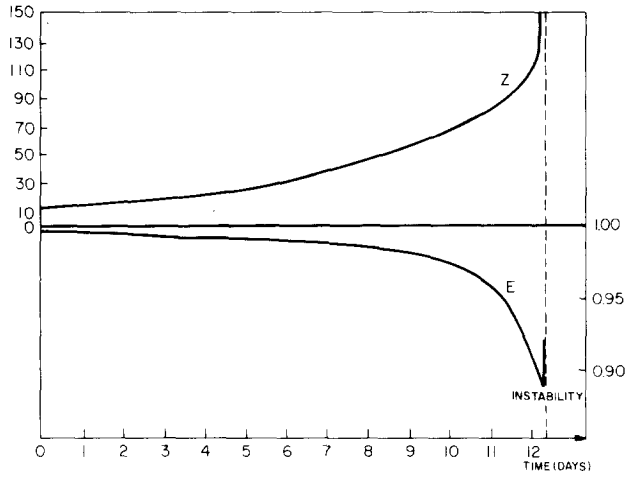

FIG. 6. The time evolution of potential enstrophy (Z) and total energy (E) for ADIF. $\Delta x = \Delta y = 500$ km, $\Delta t = 3600$ sec.

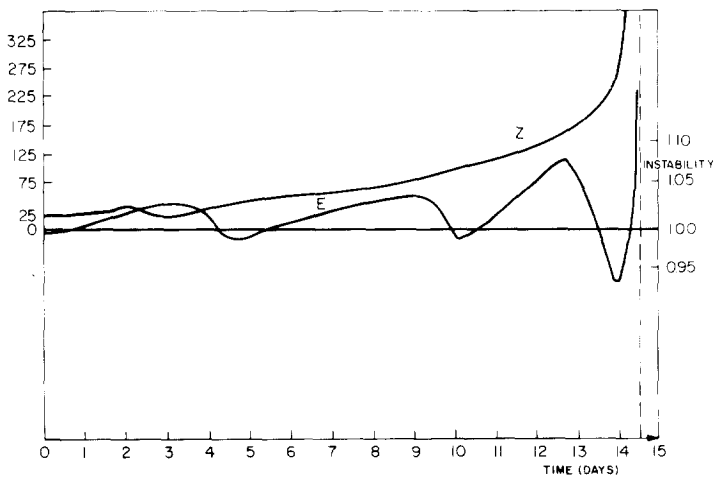

FIG. 7. The time evolution of potential enstrophy (Z) and total energy (E) for QNEX1 ($M = 6$). $\Delta x = \Delta y = 500$ km, $\Delta t = 3600$ sec.

the component of the relative vorticity along the local vertical, the absolute vorticity is defined as

$$Q = \zeta + f.$$

Then

$$Z = \tfrac{1}{2}\int_0^L \int_0^L \frac{Q^2}{h}\, dx\, dy.$$

Several numerical experiments were conducted to examine the time evolution of both the energy and the enstrophy invariants, E and Z, respectively, for the methods ADIF and QNEX1. Two spatial grids were used, namely,

$$\Delta x = \Delta y = 500 \text{ km}, \tag{27}$$

and

$$\Delta x = \Delta y = 200 \text{ km}, \tag{28}$$

with $\Delta t = 3600$ sec in each case. The results obtained from ADIF with (27) are plotted in Fig. 6. In this and subsequent figures the broken line denotes the "blow-up" time, T_c. In this experiment the enstrophy increases linearly during the first 6 days, but in the vicinity of T_c ($\approx$12 days) it jumps by an order of magnitude. The total energy remains reasonably constant for about 7 days after which it slowly decreases before decreasing sharply prior to "blow-up." The discontinuity occurring at $t = T_c$ has been referred to as an "energy catastrophe" [14].

The results for QNEX1 are given in Fig. 7. In this case T_c is around 14 days and hence QNEX1 is marginally more stable than ADIF. The behavior of E is quite different from that displayed in Fig. 6 but, again, a sharp decrease of E accompanied by a rapid increase in Z is observed prior to "blow-up."

The same experiment was repeated with the finer mesh (28), and the time evolution of E and Z in the case of ADIF is plotted in Fig. 8. The behavior of E and Z is similar to that displayed in Fig. 6, with the exception that the "blow-up" time T_c is delayed from 12 to 22 days, approximately. Sadourny [14] also observed that decreasing the mesh sizes delays T_c. The corresponding results for QNEX1 are given in Fig. 9.

In a second series of experiments a dissipation term of the form

$$\varepsilon(\Delta t)^3 [D_{+x}D_{-x} + D_{+y}D_{-y}]\, w_{jk}^n$$

was added to the right-hand side of Eq. (23a) of the ADIF scheme. In [1] dissipation is introduced into QNEX1 by adding the terms $\varepsilon(\Delta t)^3 D_{+y}D_{-y} w_{jk}^n$ and $\varepsilon(\Delta t)^3 D_{+x}D_{-x} w_{jk}^{n+1/2}$ to the right sides of (16a) and (16b), respectively. The time evolution of E and Z for these versions of ADIF and QNEX1 was studied for ε in the vicinity of the critical dissipativity, ε_c (0.003 for ADIF and 0.002 for QNEX1). The critical dissipativity, ε_c, is the value of ε above which the numerical solution is stabilized far beyond T_c, and below which it "blows up" at approximately T_c (cf.

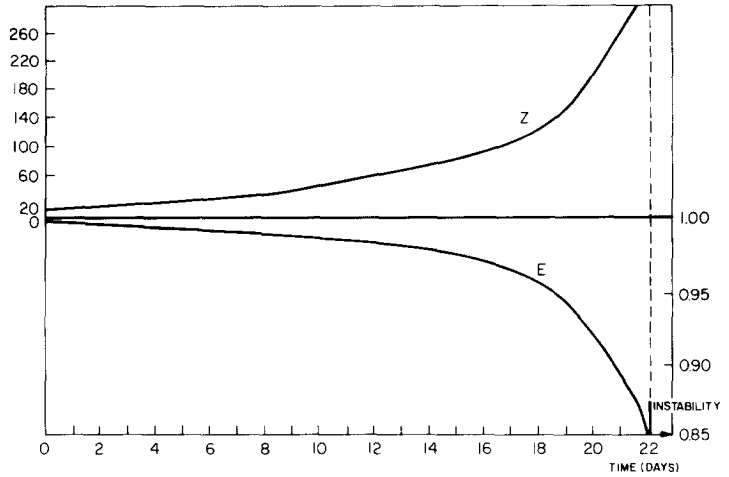

FIG. 8. The time evolution of potential enstrophy (Z) and total energy (E) for ADIF. $\Delta x = \Delta y = 200$ km, $\Delta t = 3600$ sec.

[14]). The results for ADIF with $\varepsilon_c = 0.003$ and QNEX1 with $\varepsilon_c = 0.002$ and the grid (27) are plotted in Figs. 10 and 11, respectively. In each case it is seen that the enstrophy increases in a manner similar to that with $\varepsilon = 0$ until time T_c after which it oscillates. A similar observation was made by Sadourny [14] for his energy-conserving model.

Runs with $\varepsilon > \varepsilon_c$ were also made, and, as observed by Sadourny [14], the time evolution of E was found to be rather insensitive to the value of ε as long as it stays

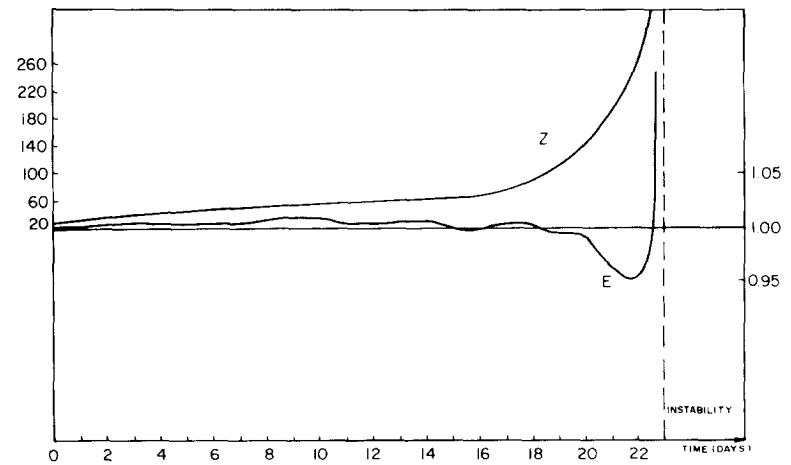

FIG. 9. The time evolution of potential enstrophy (Z) and total energy (E) for QNEX1 ($M = 6$). $\Delta x = \Delta y = 200$ km, $\Delta t = 3600$ sec.

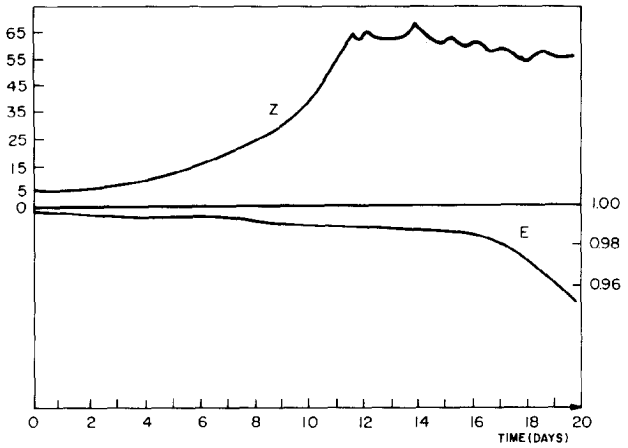

FIG. 10. The time evolution of potential enstrophy (Z) and total energy (E) for ADIF with $\varepsilon = \varepsilon_c = 0.003$. $\Delta x = \Delta y = 500$ km, $\Delta t = 3600$ sec.

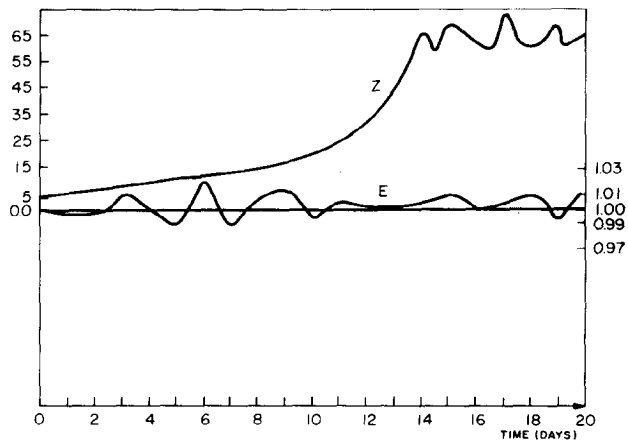

FIG. 11. The time evolution of potential enstrophy (Z) and total energy (E) for QNEX1 ($M = 6$) with $\varepsilon = \varepsilon_c = 0.002$. $\Delta x = \Delta y = 500$ km, $\Delta t = 3600$ sec.

within an order of magnitude of its critical value. In Figs. 12 and 13 the time evolutions of E and Z for ADIF and QNEX1 are plotted, respectively, with the spatial grid (27) and $\varepsilon = 0.015$.

As we have indicated there is a striking similarity between the results of our experiments and those of Sadourny [14] for his energy-conserving model, and it is possible to draw the same analogy between our method and the Navier–Stokes equations in three dimensions that Sadourny drew between his model and these

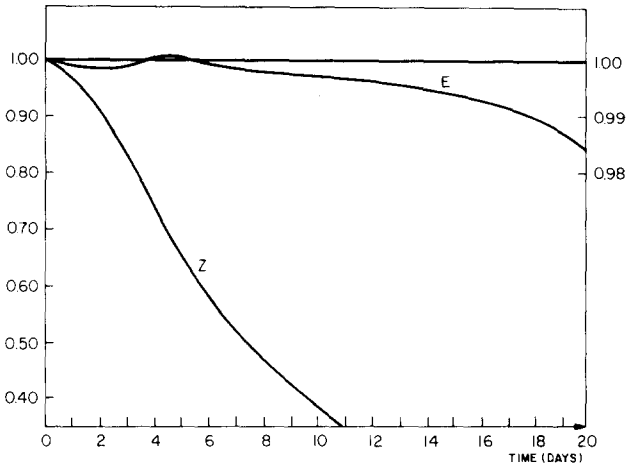

FIG. 12. The time evolution of potential enstrophy (Z) and total energy (E) for ADIF with $\varepsilon = 0.015$. $\Delta x = \Delta y = 500$ km, $\Delta t = 3600$ sec.

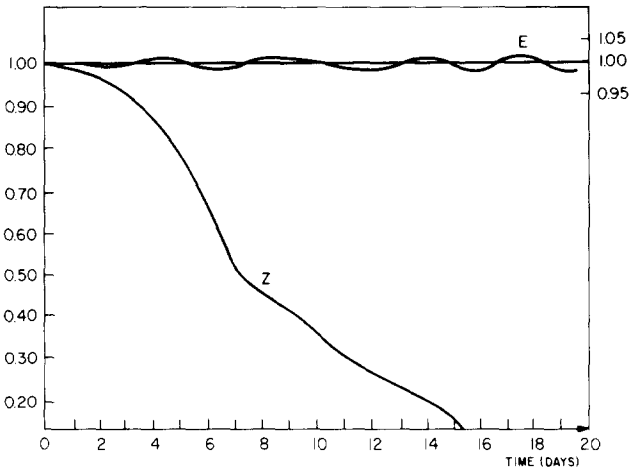

FIG. 13. The time evolution of potential enstrophy (Z) and total energy (E) for QNEX1 ($M = 6$) with $\varepsilon = 0.015$. $\Delta x = \Delta y = 500$ km, $\Delta t = 3600$ sec.

equations (see [14, p. 684]). In particular, our results verify that the critical time T_c is inversely proportional to the square root of the initial enstrophy Z_0. (Compare [14, 16, 19]. In our example $Z_0^{-1/2} = 13$.) In addition our study shows that

$$T_c = C(\Delta)\, Z_0^{-1/2},$$

where $C(\Delta)$ depends on the mesh and increases as the mesh is refined.

It should be emphasized that for the three-dimensional Navier–Stokes equations there exists an intrinsic critical time T_c, independent of resolution. Since potential enstrophy is an invariant of the shallow-water equations, there is no intrinsic critical time for these equations. However, as observed by Sadourny [18], catastrophic behavior may occur at a finite time in numerical models if they do not conserve potential enstrophy. In this case the source of catastrophy lies in the truncation error and becomes less and less effective as resolution increases, and the critical time increases to infinity with resolution.

Finally, we emphasize a point made by Sadourny. There is little to be gained by adding dissipation to ADIF and QNEX1. While dissipative terms stabilize the methods beyond T_c, the catastrophe itself is unavoidable. Since the potential energy is an invariant of the shallow-water equations one should think of T_c as the time beyond which the methods cease to be dynamically consistent with the original equations, and consider calculations performed beyond that time as purely formal.

ACKNOWLEDGMENTS

Thanks are due to Mss J. Hewitt of NRIMS, CSIR, who did the programming.

The work of the first author was partially supported by the National Science Foundation under Grant MCS 75-08331.

REFERENCES

1. B. GUSTAFSSON, *J. Comput. Physics* **7** (1971), 239–254.
2. S. A. ORSZAG AND M. ISRAELI, *Ann. Rev. Fluid Mech.* (1974), 281–318.
3. M. KWIZAK, "Semi-implicit Integration of a Grid-Point Model of the Primitive Equations," Department of Meteorology, McGill University Publ. in Meteorology No. 98, 1970.
4. D. HOUGHTON, A. KASAHARA, AND W. WASHINGTON, *Mon. Weather Rev.* **94** (1966), 141–150.
5. J. DOUGLAS, JR., AND J. E. GUNN, *Numer. Math.* **6** (1964), 428–453.
6. M. LEES, *in* "Nonlinear Partial Differential Equations" (W. F. Ames, Ed.), pp. 193–201, Academic Press, New York, 1967.
7. H. H. AHLBERG, E. N. NILSON, AND J. L. WALSH, "The Theory of Splines and Their Applications," Academic Press, New York, 1967.
8. N. N. YANENKO, "The Method of Fractional Steps," Springer-Verlag, Berlin, 1971.
9. G. I. MARCHUK, "Numerical Methods in Weather Prediction," Academic Press, New York, 1974.
10. A. C. HINDMARSH, "Solution of Block-Tridiagonal Systems of Linear Algebraic Equations," Lawrence Livermore Lab., UCID 30150, 1977.
11. A. GRAMMELTVEDT, *Mon. Weather Rev.* **97**, No. 5 (1969), 384–404.
12. E. ISAACSON AND H. B. KELLER, "Analysis of Numerical Methods," Wiley, New York, 1966.
13. I. M. NAVON, "Algorithms for the Solution of Scalar and Block Cyclic Tridiagonal Systems," CSIR Special Report WISK 265, Pretoria, South Africa, 1977.
14. R. SADOURNY, *J. Atmos. Sci.* **32** (1975), 680–689.
15. J. GARY, *Math. Comput.* **18** (1964), 1–18.
16. A. BRISSAUD, U. FRISCH, J. LEORAT, M. LESIEUR, A. MAZURE, A. POUQUET, R. SADOURNY, AND P. L. SULEM, *Ann. Geophys.* **29** (1973), 539–546.

17. I. M. NAVON, *Beit. Phys. Atmos.* **51** (1978), 281–305.
18. R. SADOURNY, private communication.
19. C. BARDOS, *in* "Nonlinear Partial Differential Equations," pp. 1–46, Lecture Notes in Mathematics No. 648, Springer-Verlag, New York, 1978.