Journal of Computational Physics **179,** 268–285 (2002)
doi:10.1006/jcph.2002.7056

A Comparison of Time Discretization Schemes for Two-Timescale Problems in Geophysical Fluid Dynamics

P. Bartello[1]

Department of Atmospheric and Oceanic Science, Department of Mathematics and Statistics, McGill University, 805, Sherbrooke ouest, Montreal, Quebec H3A 2K6, Canada
E-mail: bartello@math.mcgill.ca

Received August 15, 2001; revised February 22, 2002

In many geophysical fluid modeling applications there exist two very different time scales, essentially fast waves and slow vortices. At the largest planetary scales inertial-gravity waves propagate through the fluid at phase speeds much faster than particle velocities, while at small scales fast acoustic modes coexist with much slower vortex motions. If numerical models are to be affordable in this context, schemes which do not explicitly resolve the fast motion must be devised. Although this has traditionally been done for strictly numerical reasons, the result is physically equivalent to a temporal sub-grid-scale model. Examples of such schemes examined here are the semi-implicit, split explicit, and linear-propagator methods. All of these methods fail to resolve the fast motion accurately, but in very different ways. Recent progress in the field of rotating stratified turbulence is employed on a simplified two-time-scale version of the Boussinesq equations in order to determine how well these schemes represent the evolution of the slow variables. It is found that the split explicit and linear-propagator methods exhibit spurious interactions between the slow modes and their numerically retarded fast modes in certain conditions when $\omega\Delta t$ is large. The semi-implicit method is able to maintain a frequency separation, but it mimics the true evolution best when time filtering is applied.

Key Words: implicit; split explicit; linear propagator.

1. INTRODUCTION

Since the passage from quasigeostrophic to primitive-equation numerical weather prediction (NWP) models in the 1970s, there has been a need for numerical approximations that can model the slowly varying meteorological signal using time steps appropriate to its time

[1] Fax: (514) 398-6115.

0021-9991/02 $35.00

scale of variability, even in the presence of fast waves. The addition of inertial-gravity waves introduced a much shorter time scale into the models and traditional explicit time-stepping methods were subject to a new, much more restrictive wave Courant–Friedrichs–Lewy (CFL) criterion (see [1] for the basics),

$$C_W \equiv \omega \Delta t < O(1), \tag{1.1}$$

where C_W is the wave Courant number and the precise value of the upper limit depends on the scheme. The maximum inertial-gravity wave frequency is the Brunt–Vaissalla frequency in GFD models. To circumvent this restriction the semi-implicit method was used in [2] and in a large number of models since. By averaging the fast linear terms in time, it effectively slows down all of the waves with frequencies higher than $O(\Delta t^{-1})$, where Δt is the time step. Since that time horizontal resolutions have increased substantially, implying that today, respecting the advective CFL,

$$C_A \equiv \frac{U_{max} \Delta t}{\Delta x} < O(1), \tag{1.2}$$

where C_A is the advective Courant number, is a more severe restriction (due to the smallness of Δx) and special methods to treat inertial-gravity waves may no longer be needed for high-resolution NWP. For global climate models, however, they will be in place for some time to come. Note that numerical schemes may be able to circumvent the advective CFL constraint as far as stability is concerned (e.g. [3, 4]) but not as far as accuracy is concerned for realistic turbulent flows involving topography ([5, Sect. 4]). For this reason the advective CFL criterion is assumed to determine the maximum useful time step in the following.

On the other end of the range of length scales, fully compressible models are often employed in studies of the mesoscale and below. A popular method of dealing with the resulting fast acoustic modes was proposed in [6] and by [7] in a study of convective storms. Known as the split explicit method, it is a form of partial operator splitting employing two very different time steps. The idea is to identify the terms responsible for the fast oscillations and to treat them with a small time step while holding the other terms constant. The result of this integration is then used to advance the other terms using a much larger time step. Efficiency is gained from the fact that the slow terms are usually more expensive to evaluate [8]. Errors associated with the splitting procedure, the stability of the scheme and its overall efficiency have been debated in [9–11]. Operator splitting has also been used in oceanographic models [12]. In fact, a precursor of this study is that of Higdon and Bennett [13], who demonstrated that splitting errors can lead to numerical instability if the equations intended to model the slow motion actually admit some fast motion. An interesting variant of this method proposed in [14, 15] integrates the fast problem using a low-order scheme and short time steps, and then averages the result over a single long time step for use in a high-order-accurate integration of the slow variables. Although its feasibility has been demonstrated, its efficiency remains an open question.

In numerical models with simpler equations and geometry more progress can be made analytically. For example [16] employed the linear-propagator method on a numerically stiff two-time-scale problem in interfacial flows. By means of an integrating factor this method makes use of the exact solution to the linear equation between evaluations of the nonlinear terms. Since it is efficient only if the linear variability is much faster than that due to the nonlinear terms, it is a candidate for GFD flows at small Rossby, Froude, Rhines, or Mach

numbers. It has also been employed in simulations of rapidly rotating turbulence in [17] as well as in a study of one-dimensional dispersive wave turbulence in [18]. It can be seen to be related to the split explicit method in that it acomplishes analytically what the split explicit method accomplishes in the small-time-step problem. Of course, the analytical step is only possible if the governing equations and geometry are sufficiently simple.

The details of these schemes as applied to the problem studied here are given below. It is underlined that considerations of their efficiency and stability are not the concern here at all. The focus is rather on how well the different approaches represent the temporal subgrid-scale motion. In other words, this work, aims to make a first step in determining the numerical damage caused by them. Since they all fail to represent the fast motion accurately, it must appear numerically as noise. This is not necessarily a problem if, for example, high-frequency modes with zero amplitude are represented numerically as low-frequency modes with zero amplitude. Is this always the case? Do these schemes behave in the same way in this respect? Do any (or all) suffer from interaction between the spurious numerical representation of the fast modes and the physically significant slow modes, whose accuracy is the goal of the entire modeling effort? These are the questions posed at the outset of this study. As a result, the numerical implementations described below are not necesarily the most efficient and wherever possible steps are taken analytically in a way that would not be possible in more realistic model configurations.

In order to proceed, the various schemes need to be implemented in a sufficiently simple model for comparison purposes. Bartello [19] studied the wave-vortex interactions in decaying Boussinesq flow. By projecting solutions onto the normal modes from a linearization about a state of rest, he used conservation of potential vorticity (PV) and the predominance of resonant interactions to argue that when both the rotation and the stratification are sufficiently strong, there is little interaction between fast intertial-gravity waves and the slow geostrophic modes that contribute to the potential vorticity. The slow, or vortical, modes interact according to quasigeostrophic dynamics and as such cascade energy to larger scales, where there is little or no dissipation [20, 21]. On the other hand, the fast modes were observed numerically to cascade to smaller scales, where they were affected by viscosity and thermal diffusion. Much theoretical work has been done on the topic, both before and after [19], and the interested reader is referred to the review article by Lelong and Riley [22]. The important point here is that a nonhydrostatic Boussinesq model with large $N = f$, where f is the Coriolis frequency, is employed in this study. Although $N = f$ is not typical of GFD, this choice yields a single frequency for all fast modes, thereby facilitating the interpretation of the various numerical implementations. It is presented as a simple test case of a two-time-scale problem and as such is intended as the first step in a long process.

In the next section details of the model, its nonlinear "truth," and the numerical schemes are presented. The numerical results are displayed in Section 3. It is found that the semi-implicit method maintains accuracy in the slow modes, but the dissipation of the fast modes (accomplished by the downscale cascade in reality) is insufficient. This can be mimicked by the application of time filtering, which does little to damage the slow modes. On the other hand, the split explicit and linear-propagator methods yield a numerical representation of the fast modes whose frequency is determined by the aliasing inherent in sampling $e^{i\omega t}$ with the large time step. As is shown below for the schemes considered, if $\omega \Delta t/\pi$ is near an integer, the fast motion appears numerically to be slow and, not surprisingly, interaction with the slow modes is significant. In this case the slow motion is not treated accurately. These conclusions are summarized and recommendations for future work are made in Section 4.

2. BOUSSINESQ MODEL AND THE NUMERICAL SCHEMES

2.1. Model Background

The Boussinesq equations can be written as

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u} &= -\nabla\phi - f\hat{\mathbf{z}}\times\mathbf{u} + b\hat{\mathbf{z}} + \mathcal{D}_u(\mathbf{u}), \\ \frac{\partial b}{\partial t} + \mathbf{u}\cdot\nabla b &= -N^2 w + \mathcal{D}_b(b), \\ \nabla\cdot\mathbf{u} &= 0, \end{aligned} \tag{2.1}$$

where $\phi = p/\rho_o$, p is pressure, ρ_o is average density, $\mathcal{D}_q$ represents the linear dissipation of quantity q, buoyancy $b = g\theta/\theta_o$, θ is the potential temperature fluctuation, $\theta_T = \theta_o + \frac{d\bar{\theta}}{dz}z + \theta$, $\mathbf{u} = u\hat{\mathbf{x}} + v\hat{\mathbf{y}} + w\hat{\mathbf{z}}$, and

$$N = \left(\frac{g}{\theta_o}\frac{d\bar{\theta}}{dz}\right)^{\frac{1}{2}}$$

is the Brunt–Väisälä frequency. The quantity $\bar{\theta}$ represents an undisturbed potential temperature profile that is linear in z and is near the reference value θ_o. Following [23] the dissipation operators were set to $\mathcal{D}_q = \nu\Delta^4$. Their influence is confined to the smallest model scales.

The normal modes linearized about a state of rest can be found in [19]. Very briefly, if periodic geometry is used on a domain $(2\pi)^3$, then the Fourier representation is exact, wavevector components are integers, and for each wavevector $\mathbf{k} = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} + k_z\hat{\mathbf{z}}$, there are three modes: a geostrophically balanced mode $G_\mathbf{k}$ with zero linear frequency, and two wave modes denoted $W_\mathbf{k}^\pm$. With the appropriate nondimensionalization, the linear normal modes can be inserted into the fully nonlinear equations, yielding

$$\frac{\partial G_\mathbf{k}}{\partial t} = \epsilon\mathcal{N}_\mathbf{k}^0 - \nu k^8 G_\mathbf{k}, \quad \frac{\partial W_\mathbf{k}^\pm}{\partial t} \mp i\omega_\mathbf{k} W_\mathbf{k}^\pm = \epsilon\mathcal{N}_\mathbf{k}^\pm - \nu k^8 W_\mathbf{k}^\pm, \tag{2.2}$$

where $\epsilon = min[Ro, Fr]$ and $Ro = U/fL$ is the Rossby number and $Fr = U/NH$ is the Froude number, defined using the characteristic scales U, L, H employed in the nondimensionalization. The dispersion relation is

$$\omega_\mathbf{k} = \frac{\left(f^2k_z^2 + N^2k_h^2\right)^{1/2}}{k}, \tag{2.3}$$

where $k_h = (k_x^2 + k_y^2)^{1/2}$ and $k = |\mathbf{k}|$. The nonlinear terms take the general form for a quadratic nonlinearity

$$\mathcal{N}_\mathbf{k}^s = \sum_{\mathbf{k}+\mathbf{p}+\mathbf{q}=0}\ \sum_{s_\mathbf{p}, s_\mathbf{q}} N_1^s G_\mathbf{p}^* G_\mathbf{q}^* + N_2^s G_\mathbf{p}^* W_\mathbf{q}^{s_\mathbf{q}*} + N_3^s W_\mathbf{p}^{s_\mathbf{p}*} W_\mathbf{q}^{s_\mathbf{q}*}, \tag{2.4}$$

where * denotes the complex conjugate, s takes on the labels 0, +, or −, and $s_\mathbf{q}$ and $s_\mathbf{p}$ are + or −. The nonlinearity is seen to occur via interacting triads $\mathbf{k} + \mathbf{p} + \mathbf{q} = 0$. The interaction coefficients, N_i^s, are functions of the triadic wavevectors.

In the limit $\epsilon \to 0$ the amplitude of the fast wave mode varies slowly, e.g., $W_{\mathbf{k}}^{\pm} = \tilde{W}_{\mathbf{k}}^{\pm}(\epsilon t)\mathrm{e}^{\pm i\omega_{\mathbf{k}}t}$. The contribution to the energy from a single wavevector is $|G_{\mathbf{k}}|^2 + |\tilde{W}_{\mathbf{k}}^{+}|^2 + |\tilde{W}_{\mathbf{k}}^{-}|^2$ and its evolution is governed by

$$\frac{\partial |G_{\mathbf{k}}|^2}{\partial(\epsilon t)} = \sum_{\mathbf{k}+\mathbf{p}+\mathbf{q}=0} \sum_{s_{\mathbf{p}},s_{\mathbf{q}}} N_1^0 G_{\mathbf{k}}^* G_{\mathbf{p}}^* G_{\mathbf{q}}^* + N_2^0 G_{\mathbf{k}}^* G_{\mathbf{p}}^* \tilde{W}_{\mathbf{q}}^{s_{\mathbf{q}}^*} \mathrm{e}^{-is_{\mathbf{q}}\omega_{\mathbf{q}}t} + N_3^0 G_{\mathbf{k}}^* \tilde{W}_{\mathbf{p}}^{s_{\mathbf{p}}^*} \tilde{W}_{\mathbf{q}}^{s_{\mathbf{q}}^*} \mathrm{e}^{-i(s_{\mathbf{p}}\omega_{\mathbf{p}}+s_{\mathbf{q}}\omega_{\mathbf{q}})t} + \text{c.c.} - 2\nu k^8 |G_{\mathbf{k}}|^2 \tag{2.5}$$

and

$$\frac{\partial |\tilde{W}_{\mathbf{k}}^{\pm}|^2}{\partial(\epsilon t)} = \sum_{\mathbf{k}+\mathbf{p}+\mathbf{q}=0} \sum_{s_{\mathbf{p}},s_{\mathbf{q}}} N_1^{\pm} \tilde{W}_{\mathbf{k}}^{\pm^*} G_{\mathbf{p}}^* G_{\mathbf{q}}^* \mathrm{e}^{\mp i\omega_{\mathbf{k}}t} + N_2^{\pm} \tilde{W}_{\mathbf{k}}^{\pm^*} G_{\mathbf{p}}^* \tilde{W}_{\mathbf{q}}^{s_{\mathbf{q}}^*} \mathrm{e}^{-i(s_{\mathbf{q}}\omega_{\mathbf{q}}\pm\omega_{\mathbf{k}})t} + N_3^{\pm} \tilde{W}_{\mathbf{k}}^{\pm^*} \tilde{W}_{\mathbf{p}}^{s_{\mathbf{p}}^*} \tilde{W}_{\mathbf{q}}^{s_{\mathbf{q}}^*} \mathrm{e}^{-i(s_{\mathbf{p}}\omega_{\mathbf{p}}+s_{\mathbf{q}}\omega_{\mathbf{q}}\pm\omega_{\mathbf{k}})t} + \text{c.c.} - 2\nu k^8 |\tilde{W}_{\mathbf{k}}^{\pm}|^2, \tag{2.6}$$

where c.c. denotes the complex conjugate. Exchanges between fast W modes and slow G modes can occur through WWG triads and GGW triads. Bartello [19] showed that to leading order as $\epsilon \to 0$ the potential vorticity can be written as a wavevector sum involving only $G_{\mathbf{k}}$ modes. In other words, in this limit fast wave modes do not contibute to the PV. Invoking PV conservation in the above modal energy equations means that $N_3^0 = 0$ since WWG triads cannot change $G_{\mathbf{k}}$ without violating the PV invariance. The other route to wave-vortical exchanges, GGW triads, is off-resonant if all frequencies are large. Therefore, in the limit $\epsilon \to 0$, where N/f is fixed, wave-vortical interactions are weak.

Returning to the time-stepping methods that are the subject here, this schematic summary of the arguments presented in [19] is meant to motivate the choice of the Boussinesq equations with large $N = f$ for the present study. Note that this implies that $\omega_{\mathbf{k}} = \omega =$ constant for all $\mathbf{k}$. There are two favorable aspects to this fact. The first is that three-wave resonances are impossible, since they would require $\pm\omega \pm \omega \pm \omega = 0$, implying that the term multiplied by $N_3^{\pm}$ in (2.6) is off-resonant and therefore weak. The second is that if any of the time-stepping schemes have difficulty with certain linear analytical frequencies (as shown below), it can be demonstrated easily. Of course, in more typical models there will be a range of wave frequencies and problematic frequencies will occur in conjuction with less problematic ones, thereby clouding the issue.

As a result of this parameter choice the leading-order modal energy equations reduce to

$$\frac{\partial |G_{\mathbf{k}}|^2}{\partial(\epsilon t)} = \sum_{\mathbf{k}+\mathbf{p}+\mathbf{q}=0} N_1^0 G_{\mathbf{k}}^* G_{\mathbf{p}}^* G_{\mathbf{q}}^* + \text{c.c.} - 2\nu k^8 |G_{\mathbf{k}}|^2 \tag{2.7}$$

and

$$\frac{\partial |\tilde{W}_{\mathbf{k}}^{\pm}|^2}{\partial(\epsilon t)} = \sum_{\mathbf{k}+\mathbf{p}+\mathbf{q}=0} N_2^{\pm} \tilde{W}_{\mathbf{k}}^{\pm^*} G_{\mathbf{p}}^* \tilde{W}_{\mathbf{q}}^{\mp^*} + \text{c.c.} - 2\nu k^8 \left|\tilde{W}_{\mathbf{k}}^{\pm}\right|^2. \tag{2.8}$$

The first equation describes quasigeostrophic energy transfers as in [20, 21], while the second involves a catalytic transfer between two wave modes, involving a vortical mode, but not altering it. These simple energy transfers give us a nonlinear "truth" to which we can compare the numerical time-stepping schemes. Simulations of the viscous decay

from a random initial field were employed. These arguments then predict a net transfer of vortical energy to larger scales and a corresponding lack of vortical energy dissipation. In addition, simulations and statistical mechanical arguments presented in [19] would lead us to expect a downscale transfer of wave energy, implying a strong dissipation of these modes. The energetics of such a simulation do not tell us everything. The ability of a particular numerical scheme to reproduce this behavior would therefore be a necessary, but not sufficient, condition for numerical accuracy.

2.2. Model Parameters and Time-Stepping Schemes

The numerical model used here was developed from that described in [19]. It employs a de-aliased pseudospectral method with isotropic truncation on a domain of unit aspect ratio. The spatial grid consisted of 90^3 collocation points, and the viscosity coefficient, $\nu = 3.8 \times 10^{-11}$, was set equal to the thermal diffusion. The energy spectrum is defined such that $E_T = \int E(k)\, dk$, where E_T is the total energy. More details on these statistics can be found in [19]. The initial energy spectrum was set to

$$E(k) = a \begin{cases} k^{10}, & \text{if } k \le k_i, \\ k_i^{20}/k^{10}, & \text{if } k > k_i, \end{cases} \tag{2.9}$$

with $k_i = 5$, random Fourier phases, and a chosen such that $E_T = 0.72$. The initial kinetic energy was twice the initial potential energy. For reasons described above, $N = f$ was arbitrarily set to 1000 in model units. More physically, the resulting initial Froude number, based on the root mean square horizontal vorticity, was 1.8×10^{-3} and the initial Rossby number based on vertical vorticity was 3×10^{-3}. All simulations described below were identical in these respects.

There was another common element to the schemes considered here in that they all used second-order centered differences on the time derivative term. The leapfrog scheme is not necessarily the best one, but its use has been and remains very common in GFD. Second-order Adams–Bashforth and Runge–Kutta schemes have also been examined and, as far as the main results of this study are concerned, show the same qualitative behavior. Since the focus is the treatment of the fast linear terms, and for clarity of presentation, only the leapfrog results are presented in detail.

A major difference between the model described here and that described in [19] is the fact that the present model performs the time stepping on the normal-mode amplitudes $\{G_{\mathbf{k}}\}$ and $\{W_{\mathbf{k}}^{\pm}\}$ rather than on the more traditional variables $\mathbf{u}$ and b. The original code actually integrated the three-dimensional vorticity equation along with the b equation of (2.1). In the present study the nonlinear term is evaluated by transforming the normal-mode amplitudes resulting from the time stepping to vorticity and b, using FFTs in the spectral transform method to evaluate nonlinear terms such as $\nabla \cdot (\mathbf{u}b)$, and then projecting the product back onto the normal modes to obtain $\mathcal{N}_{\mathbf{k}}^{s}$. The resulting model equations are therefore given by (2.2). Although going back and forth between normal-mode amplitudes and vorticity and b involves more computation, the projections are only order N operations, where N is the number of model degrees of freedom. The approach has the advantage of making it relatively easy to substitute different time-stepping schemes, as shown below.

The first of these is the explicit small-time-step control run. It simply uses centered-difference time stepping with explicit treatment of the fast linear term and Crank–Nicolson

dissipation. Schematically rewriting the fast-mode equation as

$$\frac{dW}{dt} - i\omega W = N - \nu k^8 W, \tag{2.10}$$

the scheme integrates

$$\frac{W^{n+1} - W^{n-1}}{2\Delta t} - i\omega W^n = N^n - \nu k^8 \frac{1}{2}(W^{n+1} + W^{n-1}),$$

or

$$W^{n+1} = W^{n-1}\frac{\lambda^-}{\lambda^+} + \frac{2\Delta t}{\lambda^+}(N^n + i\omega W^n), \tag{2.11}$$

where n is the time step and $\lambda^\pm = 1 \pm \nu k^8 \Delta t$. The initial step is taken with a second-order Runge–Kutta scheme (see [1]) with the linear term treated analytically. Numerically the time step was set to $\Delta t = 3 \times 10^{-4}$, yielding an initial advective Courant number, $C_A = 8 \times 10^{-3}$, and a wave Courant number, $C_W = 0.3$, implying there were approximately 21 discrete samples per wave oscillation. In all of the simulations the slow modes were integrated as

$$G^{n+1} = G^{n-1}\frac{\lambda^-}{\lambda^+} + \frac{2\Delta t}{\lambda^+}N^n,$$

where the nonlinear term, N, is clearly not the same as in (2.10).

With these parameters the control run was computationally demanding and cheaper alternatives would be a necessity in an operational setting. The oldest of these is the semi-implicit method in which the fast linear term is averaged in time. The advantage here is that when time stepping the normal-mode amplitudes, the linear term is a diagonal matrix and the method reduces to a Crank–Nicolson scheme,

$$\frac{W^{n+1} - W^{n-1}}{2\Delta t} - i\frac{\omega}{2}(W^{n+1} + W^{n-1}) = N^n - \nu k^8 \frac{1}{2}(W^{n+1} + W^{n-1}),$$

or

$$W^{n+1} = W^{n-1}\frac{\phi^-}{\phi^+} + \frac{2\Delta t}{\phi^+}N^n, \tag{2.12}$$

where $\phi^\pm = 1 \pm (\nu k^8 - i\omega)\Delta t$. Consider the behavior of this scheme with $N = \nu = 0$. It is a traditional exercise to show that the factor ϕ^-/ϕ^+ is a second-order approximation to the analytical $e^{2i\omega\Delta t}$ as $\omega\Delta t \to 0$. For the present study we are more interested in the limit $\omega\Delta t \to \infty$. In this case $\phi^-/\phi^+ \to -1$, implying that $W^{n+1} = -W^{n-1}$ and yielding a wave period of $4\Delta t$. It is noted that although this result is only for the leapfrog scheme, other schemes would yield periods of $O(1)\Delta t$. Semi-implicit schemes represent infinitely fast waves as waves with frequencies of $O(\Delta t^{-1})$.

For simplicity of presentation we next consider the linear propagator method. The governing equation is simplified using an integrating factor. Setting $W(t) = \tilde{W}(t)e^{-(\nu k^8 - i\omega)t}$, the resulting ODE is

$$\frac{d\tilde{W}}{dt} = N e^{(\nu k^8 - i\omega)t}. \tag{2.13}$$

In the present small-ϵ context the theory implies that $\tilde{W}$ is a slow variable. However, the nonlinear term is now oscillating rapidly. We discretize this with the leapfrog scheme and an explicit treatment of the nonlinear term. The result, expressed in terms of the original variable and multiplied by $e^{-(\nu k^8 - i\omega)(n+1)\Delta t}$, is

$$W^{n+1} = W^{n-1} e^{-2(\nu k^8 - i\omega)\Delta t} + 2\Delta t N^n e^{-(\nu k^8 - i\omega)\Delta t}. \tag{2.14}$$

In the absence of dissipation and nonlinearity the result is the analytical solution $W^{n+1} = W^{n-1} e^{2i\omega\Delta t}$. At this point one could ask why it should not be. After all, in this simple geometry the fast linear progation of waves does not pose an analytical challenge. Why should it be necessary to slow them down as in the semi-implicit method? The linear-propagator method rather slows down the fast nonlinear interaction terms. This can be seen as follows. In the absence of dissipation the nonlinear term in (2.13) is proportional to $N^n e^{-i\omega t}$. Substituting (2.4) yields

$$\begin{aligned}\frac{d\tilde{W}_{\mathbf{k}}}{dt} \propto \sum_{\mathbf{k}+\mathbf{p}+\mathbf{q}=0} \sum_{s_{\mathbf{p}}, s_{\mathbf{q}}} & N_1^s G_{\mathbf{p}}^* G_{\mathbf{q}}^* e^{-i\omega_{\mathbf{k}} t} + N_2^s G_{\mathbf{p}}^* \tilde{W}_{\mathbf{q}}^{s_{\mathbf{q}}*} e^{-i(\omega_{\mathbf{k}} + s_{\mathbf{q}}\omega_{\mathbf{q}})t} \\ & + N_3^s \tilde{W}_{\mathbf{p}}^{s_{\mathbf{p}}*} \tilde{W}_{\mathbf{q}}^{s_{\mathbf{q}}*} e^{-i(\omega_{\mathbf{k}} + s_{\mathbf{p}}\omega_{\mathbf{p}} + s_{\mathbf{q}}\omega_{\mathbf{q}})t},\end{aligned} \tag{2.15}$$

where we have returned some of the wavevector and modal detail that was dropped in the above schematic equations. In the linear-progagator method we are essentially integrating (2.15) with a large time step for efficiency. If certain interaction terms in the sum are fast, in other words if the frequency sums in the exponential factors are large, then the time step is unable to resolve the rapid driving resulting from these interactions. If, on the other hand, the interactions are close to resonant, i.e., if the frequency sums in the various terms of (2.15) are small, then the contribution from those interactions is well resolved. Since it is these resonant interactions that are important as $\epsilon \to 0$, it is hoped that the result will be able to mimic the explicit time-step truth. It is also hoped that the far-from-resonant interactions, although insufficiently sampled in time, will behave much like a random-noise generator without much of a systematic effect.

Finally, the split explicit scheme identifies the terms responsible for the rapid oscillation, the term multiplied by ω here, and integrates them forward in time while holding the other terms constant. The latter are updated at larger intervals appropriate for describing slower motion of meteorological interest. In the implementation described in this study, we can make the link between the split exlicit method and the linear-propagator method by intergrating the split explicit small-time-step problem analytically. Treating the nonlinear terms, N, as constant over the large-time-step interval in (2.10) yields $W = Ce^{-(\nu k^8 - i\omega)t} + N/(\nu k^8 - i\omega)$, where C is an integration constant, such that

$$W^{n+1} = W^{n-1} e^{-(\nu k^8 - i\omega)\Delta t} + \frac{N^n}{\nu k^8 - i\omega}\left(1 - e^{-2(\nu k^8 - i\omega)\Delta t}\right), \tag{2.16}$$

where n indexes the large time step. Clearly, the analytical integration cannot be performed unless the model is sufficiently simple. In the present study we do so to examine the behavior of this scheme in the limit where absolutely no error is made in the small-time-step problem. Note the similarity between the linear-propagator method (2.14) and the split explicit method (2.16). However, (2.16) does not include the appropriate exponential factor multiplying N to yield accurate resonant interactions when $\omega\Delta t$ is large.

Comparing the linear-propagator and split explicit methods to the semi-implicit method, it is concluded that whereas the semi-implicit method slows down the fast waves, the linear-propagator and split explicit methods do not. The latter two suffer from the aliasing inherent in coarse sampling of $e^{i\omega t}$. In addition, the linear-propagator method slows down the fast nonlinear interactions and treats the resonant interactions accurately. The split explicit method also slows down fast interactions but does not successfully isolate analytically resonant frequencies. At first sight the linear-propagator method would appear superior. The theory of these interactions takes the fast ones as off-resonant and therefore of secondary importance. On the other hand, changing the wave frequencies changes everything with respect to the presence or absence of resonance. One specific effect of this is that the semi-implicit method cannot be expected to (and indeed does not, as is shown below) yield accurate fast mode amplitudes, whereas the linear propagator can. In the next section limiting behavior as $\omega\Delta t \to \infty$ is discussed in parallel with the presentation of numerical results.

3. NUMERICAL RESULTS

We begin with a description of the small-time-step control run based on the explicit scheme (2.11), along with its nonlinear behavior, which we take here to be the truth. It is first noted that all of the schemes described yield results insignificantly different from the control run when integrated with the same small time step ($C_W = 0.3$). Figure 1 shows the time series of the total energy, E, along with its contributions from geostrophic and wave modes, i.e., $E = E_G + E_W$, where $E_G = \sum_{\mathbf{k}} |G_{\mathbf{k}}|^2$ and $E_W = \sum_{\mathbf{k}} |W_{\mathbf{k}}^+|^2 + |W_{\mathbf{k}}^-|^2$. As described above, the geostrophic energy is not dissipated as much as the wave energy.

Figure 2 displays time series of the Rossby and Froude number as defined in [19] in terms of vertical and horizontal root mean square vorticities, respectively. Both Ro and Fr remain well below 10^{-2} and ϵ can safely be considered to be small.

The geostrophic and wave energy spectra, defined such that $E_{G,W} = \int E_{G,W}(k)\,dk$, are shown at $t = 0$ corresponding to (2.9) and at $t = 5$ in Fig. 3. Over this time there has been considerable nonlinear evolution which is different for the two types of modes. Whereas the peak of $E_G(k)$ has moved toward smaller k, consistent with the slow dynamics [20, 21], the

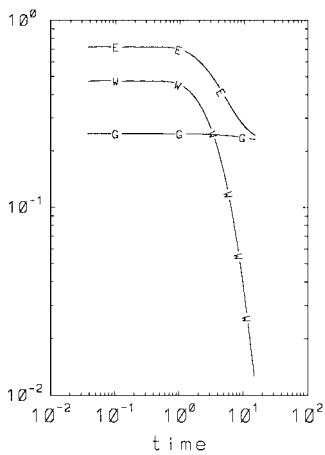

FIG. 1. Time series of the geostrophic, G, wave, W, and total, E, energies in the small-time-step control run.

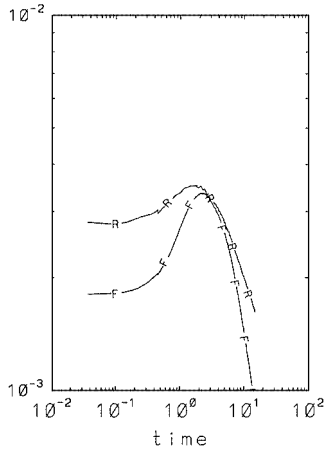

FIG. 2. Time series of the Rossby, R, and Froude, F, numbers in the small-time-step control run.

$E_W(k)$ spectrum has simply flattened out. At intermediate k the geostrophic spectrum decays much more rapidly than the wave spectrum. It is here repeated that these few statistics by no means tell us everything about the flow and agreement is a necessary but not sufficient condition.

Now we turn to the alternative schemes permitting the use of much larger time steps. The first of these is the semi-implicit method (2.12). Figure 4 shows the energy time series as in Fig. 1. The left panel looks insignificantly different from what was obtained by a straightforward implementation of (2.12) with a time step such that $C_W = 2\pi$ (not shown). (The left panel is from a simulation with an extremely weak time filter, as discussed below.) Although the number of floating-point operations in the various schemes considered is not the same, the time step here was increased from the control run by over a factor of 20, implying a considerable reduction in the cost. As can be seen in the figure the model blew up at $t \approx 2.76$. Before it did so it exhibited large amplitude oscillations in energy. The

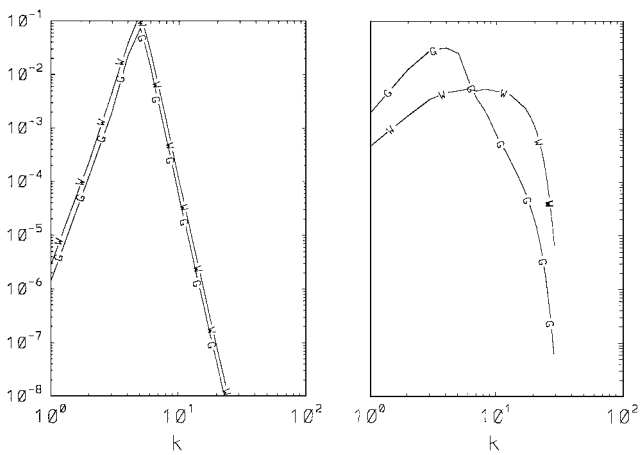

FIG. 3. Geostrophic and wave energy spectra, $E_{G,W}(k)$, in the small-time-step control run. (Left) the initial spectra (2.9) with $E_W(k) > E_G(k)$; (right) $t = 5$, with $E_W(k) > E_G(k)$ only at large k.

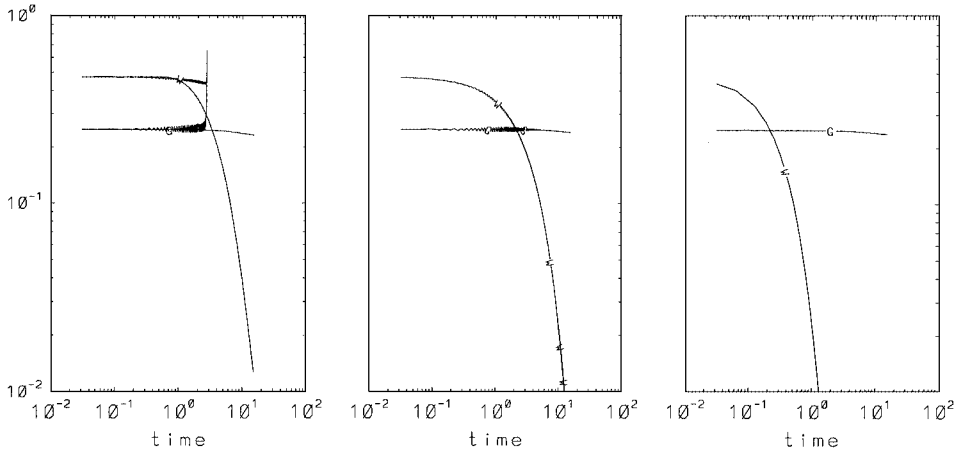

FIG. 4. Time series of geostrophic and ageostrophic energies (as in Fig. 1) using the semi-implicit method with $C_W = 2\pi$ and at various Robert filter parameters, γ. (Left) $\gamma = 10^{-4}$ is the oscillatory curve, while the other curve is the small-time-step control run; (middle) $\gamma = 10^{-3}$; (right) $\gamma = 10^{-2}$.

interpretation is straightforward. It was pointed out above that in this scheme the effective numerical period of an infinitely fast wave was $4\Delta t$ whereas the leapfrog scheme produced a computational mode of period $2\Delta t$. Therefore the numerical fast waves forced the model at very close to the resonant frequency of the computational mode. It is not surprising that this mode grew in amplitude until disaster struck. It is also interesting to note that before the model blew up, wave energy had decayed much more slowly than in the control run (both are displayed in the left panel of Fig. 4). It is tempting to conclude that schemes without computational modes would not blow up, but would also not display the preferential decay of wave energy seen in the control. A semi-implicit version of the second-order Adams–Bashforth scheme, which does have a computational mode but which is more heavily damped, behaved exactly in this manner. Both geostrophic and wave energy decayed very slowly compared to the control. Although this scheme is unstable, the growth rate is very small, such that the simulation energy did not grow significantly.

Most geophysical fluid modelers have dealt with problems relating to the computational mode in centered-difference schemes by applying time filtering. The most popular is due to Robert [24] and is applied as

$$\overline{W^n} = W^n + \gamma(W^{n+1} - 2W^n + \overline{W^{n-1}})$$

after completing (2.12). Here γ is an adjustable constant regulating the strength of the filter, which damps the high frequencies in a model (see [1] for more detail). A succesion of Robert filters [24] were applied to the present problem and they form the sequence of panels in Fig. 4. It can be seen that if γ is sufficiently large, the model does not blow up, the wave energy decays (since it appears with period near $4\Delta t$), and the slow geostrophic dynamics is not much affected. The extent of wave energy decay is determined by the filter, as is the amplitude of the energy oscillations seen in both types of mode.

Figure 5 shows the $t = 5$ energy spectra for the two values of γ that survived to this time. It shows that the wave energy's decay is entirely due to the Robert filter and that the geostrophic energy is not affected. Compare Fig. 5 to the true geostrophic spectrum in

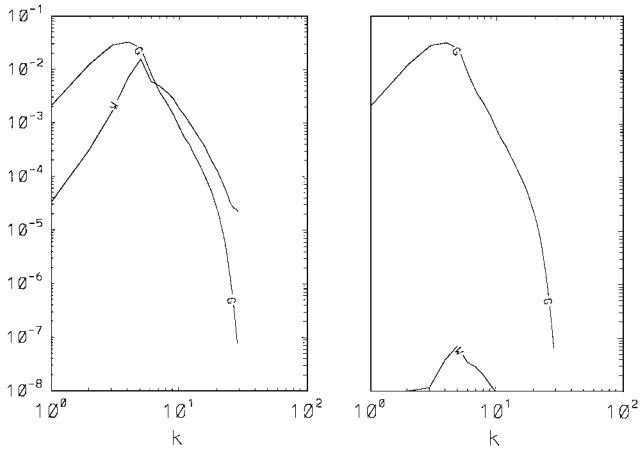

FIG. 5. Geostrophic and ageostrophic energy spectra at $t = 5$ using the semi-implicit method and $C_W = 2\pi$. (Left) $\gamma = 10^{-3}$; (right) $\gamma = 10^{-2}$.

Fig. 3. It also shows that the wave energy spectrum bears no resemblence to the true one, as expected.

We turn now to the linear-propagator method (2.14). The energy time series in the same format are found in Fig. 6. The three panels show results using three different time steps bracketing $C_W = 2\pi$. The surprising result is that the behavior of the scheme is not monotonic. Values both lower and higher than 2π yield curves which are virtually indistinguishable from the control run, while the $C_W = 2\pi$ run shows both geostrophic and wave energy decaying at approximately the same rates, suggesting that they are interacting in a significant way. Although at first sight this may seem surprising, it is clear that if $C_W = 2\pi$, then (2.14) reduces to

$$W^{n+1} = W^{n-1} e^{-2\nu k^8 \Delta t} + 2\Delta t N^n e^{-\nu k^8 \Delta t},$$

which is precisely the scheme one would employ if there were no waves! To see this one

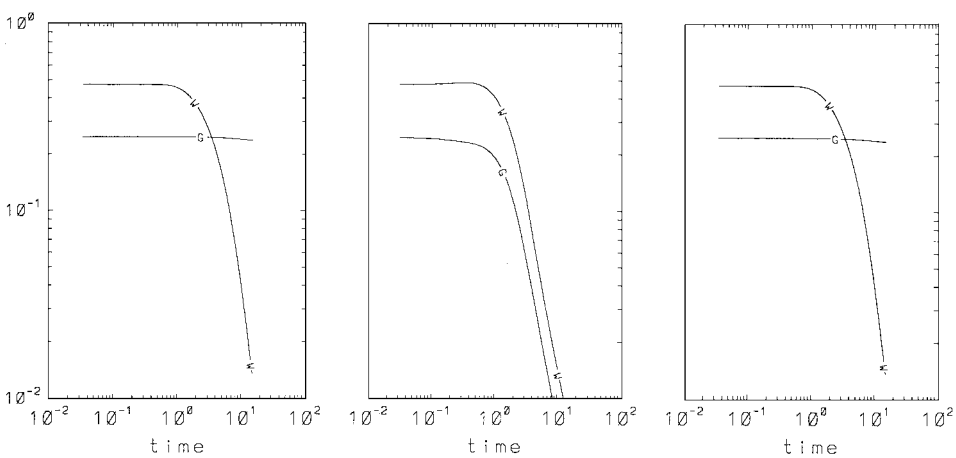

FIG. 6. Time series of geostrophic and ageostrophic energies (as in Fig. 1) using the linear propagator method. (Left) $C_W = 0.9(2\pi)$; (middle) $C_W = 2\pi$; (right) $C_W = 1.1(2\pi)$.

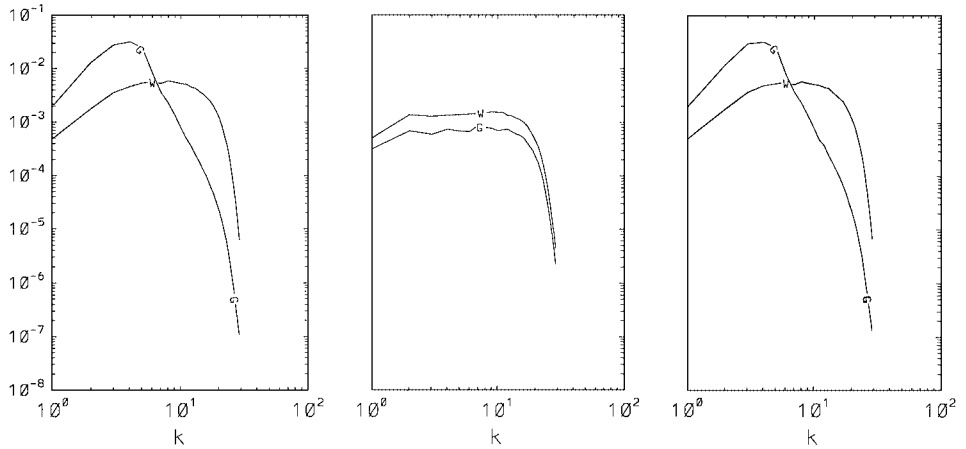

FIG. 7. Geostrophic and ageostrophic energy spectra at $t = 5$ using the linear propagator method. (Left) $C_W = 0.9(2\pi)$; (middle) $C_W = 2\pi$; (right) $C_W = 1.1(2\pi)$.

has only to set $\omega = 0$ in the explicit control scheme (2.11), the semi-implicit scheme (2.12), the linear-propagator scheme (2.14) or take the limit $\omega \to 0$ in the split explicit scheme (2.16). There are, of course, differences in the treatments of the dissipative term, but it should be kept in mind that $\nu k^8 \Delta t \ll 1$ in the large model scales, where most of the energy is. In this case the Boussinesq equations reduce to isotropic flow advecting a passive scalar. In general, if $C_W = \omega \Delta t$ is an integral multiple of π, infinitely fast waves appear as a constant. In fact, whereas in the semi-implicit method infinitely fast waves appear as high-frequency waves in the model, in the linear-propagator method they appear as waves of all frequencies (including zero), with the result being determined by the aliasing inherent in sampling the analytical frequency with a coarse time step. Instead of the set of analytical resonant interactions described above, this scheme at an inappropriate C_W can yield spurious nonlinear resonances.

The energy spectra from the linear-propagator method are displayed in Fig. 7. Again, it can be seen that the $C_W = 2\pi$ run is problematical. However, at a slightly different time step, the results are virtually indistinguishable from the control. Unlike in the semi-implicit method, the last statement also applies to the high-frequency wave spectrum.

Since the linear-propagator method yields a sensitive dependence on the time step, it is an interesting excercise to relax the condition $N = f$ described above. Figure 8 displays the energy time series obtained with $f \Delta t = 0.9(2\pi)$ and $N \Delta t = 2\pi$. Here, the wave Courant number ranges from a value where spurious resonances occur and the scheme fails, to a value where it works remarkably well in the $\omega = \text{const}$ case. Figure 8 shows a poor agreement with the control run and suggests that as long as some frequencies are problematical (C_w is a multiple of π), the model performs poorly. Of course, if N/f is much larger, as in typical GFD flows, there may not be any discrete wavevector modes where spurious resonances occur. An estimate of the "width" of the problem zone of $\omega \Delta t$ around integer multiples of π is presented in the Conclusion.

A quick look at (2.16) will suffice to show that the split explicit method also has problems with $\omega \Delta t$ at integral multiples of π. In fact if $\nu = 0$, (2.16) implies that there is no variability in time for this case. In the numerical implementation of this scheme only the fast modes are integrated using (2.16). The slow modes are calculated using (2.11) with $\omega = 0$, as would

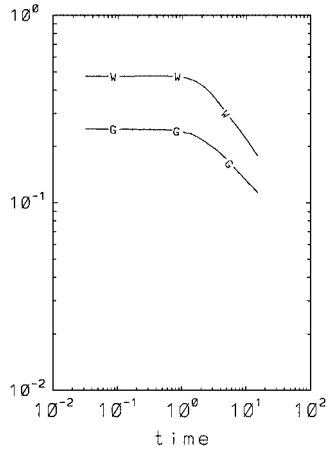

FIG. 8. Time series of geostrophic and ageostrophic energies (as in Fig. 1) using the linear-propagator method with $f\Delta t = 0.9(2\pi)$ and $N\Delta t = 2\pi$.

be the case in a realistic GFD model. Figure 9 shows the energy time series for the case $\omega\Delta t = 2\pi$, as well as a pair of other time steps. As expected, the wave energy shows only a very slight viscous decay at the problem frequency. The slow geostrophic energy, however, increases until the model blows up. To understand this behavior it is necessary to return to (2.2) and (2.4). In other words, before it was assumed that waves were so fast that only resonant interactions were important. In the present case the assumption is rather that the $W_{\mathbf{k}}^{\pm}$ are constant. The $W_{\mathbf{k}}^{\pm}$ appearing in the nonlinear terms of the $G_{\mathbf{k}}$ equation are then like instability terms and conservation of energy is no longer possible.

When $\omega\Delta t$ is not an integer multiple of π the behavior of this scheme shows a remarkable insensitivity to the time step. The other two panels of Fig. 9 show similar results at two very different time steps. The geostrophic energy differs insignificantly from the truth, whereas the wave energy does not decay fast enough. Since (2.16) does not contain the appropriate

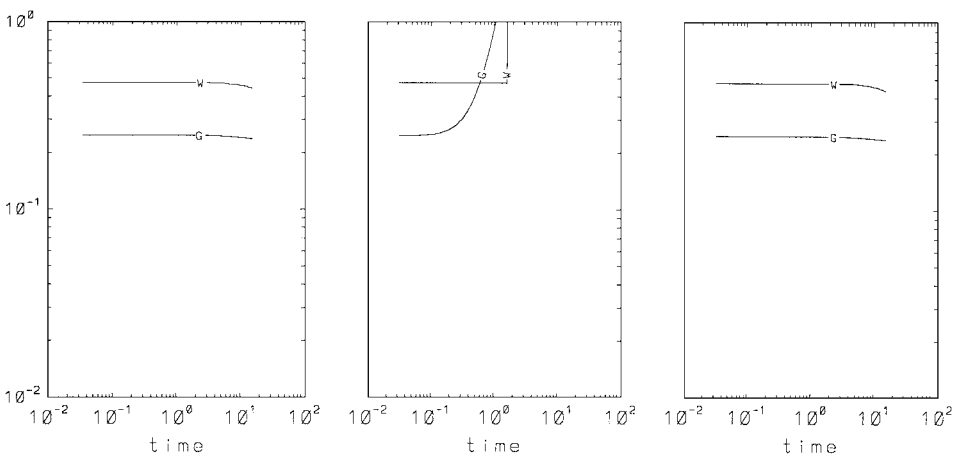

FIG. 9. Time series of geostrophic and ageostrophic energies (as in Fig. 1) using the split implicit method. (Left) $C_W = 0.9(2\pi)$; (middle) $C_W = 2\pi$; (right) $C_W = 7\pi/2$.

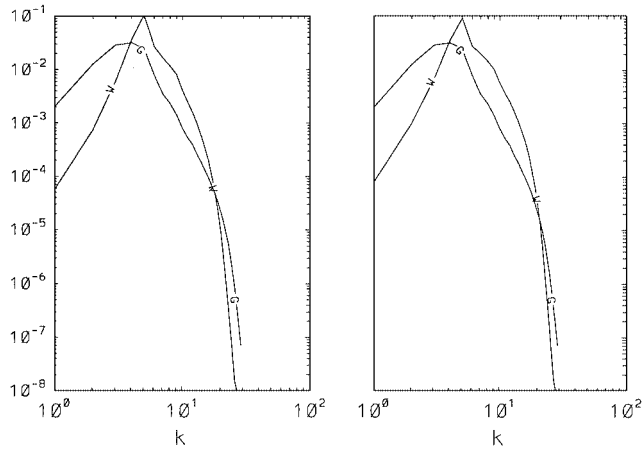

FIG. 10. Geostrophic and ageostrophic energy spectra at $t = 5$ using the split-explicit method. (Left) $C_W = 0.9(2\pi)$; (right) $C_W = 7\pi/2$.

exponential factor multiplying N to isolate the analytical resonant frequencies in the wave-mode equations, we cannot expect accuracy in these modes. However, the spectra displayed in Fig. 10 for the two runs with nonintegral C_W/π show that the slow geostrophic modes are relatively unaffected. In typical applications of this scheme a time filter such as that described above is applied. It serves to mimic the dissipation of wave energy, which is due to its downscale cascade in reality.

4. CONCLUSIONS

As a result of the normal-mode decomposition and numerical investigation of [19], it was a relatively simple matter to program a Boussinesq model to integrate the normal-mode amplitudes. In this context replacing time-stepping schemes is rather easy. In this study particular emphasis was placed on currently popular schemes to integrate two-time-scale problems without explicitly resolving the fast time scale.

The semi-implicit scheme represents high-frequency waves as waves with as high a frequency as permitted by the time step. A reduced frequency separation between these and the model's slow modes can be maintained provided Δt is small enough. As would be expected, the wave dynamics are not accurate in the model, but this did not affect the slow mode dynamics at the level of the statistics examined here. In realistic GFD models with irreversible parameterization schemes representing such things as rainfall, the rapid motion of the spurious fast modes may cause problems. They can be damped by using a time filter such as that due to Robert. If such models are based on other schemes not producing computational modes, time filtering may still be desirable.

The linear-propagator scheme showed the best agreement with the truth at large $\omega\Delta t$ for both the fast mode amplitudes and the slow dynamics, provided $\omega\Delta t/\pi$ is not an integer. In problems with wave dispersion relations such that there is a large range of frequencies, this may be hard to avoid. A dimensional estimate can be made of the width of problematical time steps around integer C_W/π. If we set $\omega\Delta t = \pi m + \delta$, where m is any integer and δ is this width, then the numerical period in time steps is $2\pi/\delta$. Interaction between fast waves and slow vortices was seen to occur when the numerical representation of the fast modes

varied on something approaching the slow time scale. The slow vortical motions are usually characterized by the advective time scale, $\tau = L/U$, where L is the typical eddy size and U is its velocity. If we invert (1.2) to obtain the time step, then the minimum numerical period of the vortices, measured in time steps, is

$$\frac{\tau_{min}}{\Delta t} = \frac{1}{C_A}\frac{L}{\Delta x},$$

where the advective Courant number, C_A, is taken to be order unity or less. Significant interaction with the numerical representation of the fast modes can be expected if the two periods are similar, i.e., if $2\pi/\delta \sim \tau_{min}/\Delta t$, or if

$$\delta < 2\pi C_A \frac{\Delta x}{L}. \tag{4.1}$$

The width of the problem zone of $\omega\Delta t$ near integer multiples of π is linearly proportional to the model scale ratio, $\Delta x/L$, which is the inverse of the number of collocation points per eddy scale, i.e., the best measure of model resolution. In current climate models this is barely 10, whereas in ocean models it is of order unity, considering that one should not include the range of model scales devoted to dissipation. For these applications, and assuming $2\pi C_A$ to be $O(1)$, it is concluded that there is a wide range of time steps that will cause problems of the type described above. For high-resolution NWP and mesoscale meteorological models, $\Delta x/L \sim O(10^{-2})$, and spurious resonances would be less likely. On the other hand, given the wide range of frequencies in real geophysical problems, model testing at several time steps seems warranted.

The split explicit method is subject to the same considerations as the linear-propagator method, with the difference being that the split explicit method does not get the waves right. If these two schemes are effectively limited to $\omega\Delta t < \pi$, their utility must be called into question, especially as a means of dealing with acoustic modes. There are many orders of magnitude difference between sound wave frequencies and vortical frequencies. This speculation should be tested in a future study involving the compressible Navier–Stokes equations.

A reviewer has questioned the applicability of these results to GFD models due to the choice $N = f$, stating that the implication is that the group velocity is zero. Unfortunately, collapsing all frequencies down to a single value provides the means of clarifying the effect of these schemes, but it falsifies the dynamics somewhat. As has been shown, with $N = f$, if certain frequencies are problematic, the entire simulation fails. This can be easily demonstrated using a few global statistics. In a more realistic model, this behavior carries over only for certain wave modes in a frequency range whose width was estimated above. Whether these modes have sufficient energy to influence the slow part of the solution depends on the dynamics. However, the nonlinear interactions in the form of spurious resonances is as described in the simpler $N = f$ case. There will simply be fewer of them. Clearly, further investigation of the case $N \neq f$ is warranted.

Finally, it is necessary to point out that this study examined only the centered-difference leapfrog scheme for the time derivative terms. Many alternatives exist and have proven to have better performance. The restriction to leapfrog here was meant only to simplify the presentation and to employ the techniques that have traditionally been used in GFD. Some exploration has been made with the second-order Adams–Bashforth and Runge–Kutta

schemes. The results have been similar to those presented, with the obvious differences in order-one constants. For example, problems that occur in three-time-level schemes when $\omega \Delta t$ is an integer multiple of π will occur at 2π in a two-time-level scheme, and so forth.

ACKNOWLEDGMENTS

I thank Andrew Majda for provoking this line of research and Steve Thomas and David Straub for many useful suggestions. Comments by Gerald Browning, David Dietrich, Brian Sanderson, and the three Reviewers are also appreciated. Financial support was received from the Natural Sciences and Engineering Research Council, while the Meteorological Service of Canada generously contributed the supercomputer resources.

REFERENCES

1. D. R. Durran, *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics* (Springer-Verlag, New York, 1999).
2. M. Kwizak and A. J. Robert, A semi-implicit scheme for grid point atmospheric models of the primitive equations, *Mon. Weather Rev.* **99**, 32 (1971).
3. A. J. Robert, A stable numerical integration scheme for the primitive meteorological equations, *Atmos. Ocean* **19**, 35 (1981).
4. B. P. Leonard, Note on the Von-Neumann stability of explicit one-dimensional advection schemes, *Comput. Methods Appl. Mech. Eng.* **118**, 29 (1994).
5. P. Bartello and S. J. Thomas, The cost-effectiveness of semi-Lagrangian advection, *Mon. Weather Rev.* **124**, 2883 (1996).
6. G. I. Marchuk, *Numerical Methods in Weather Prediction* (Academic Press, New York, 1974).
7. J. B. Klemp and R. Wilhelmson, The simulation of three-dimensional convective storm dynamics, *J. Atmos. Sci.* **35**, 1070 (1978).
8. R. J. LeVeque and J. Oliger, Numerical methods based on additive splittings for hyperbolic partial differential equations, *Math. Comput.* **40**, 469 (1983).
9. W. C. Skamarock and J. B. Klemp, The stability of time-split numerical methods for the hydrostatic and nonhydrostatic elastic equations, *Mon. Weather Rev.* **120**, 2109 (1992).
10. W. C. Skamarock and J. B. Klemp, Efficiency and accuracy and the Klemp-Wilhelmson time-splitting technique, *Mon. Weather Rev.* **122**, 2623 (1994).
11. G. Browning and H.-O. Kreiss, Splitting methods for problems with different timescales, *Mon. Weather Rev.* **122**, 2614 (1994).
12. P. D. Killworth, D. Stainforth, D. J. Webb, and S. M. Paterson, The development of a free-surface Byan-Cox-Semtner Ocean model, *J. Phys. Oceanogr.* **21**, 1333 (1991).
13. R. L. Higdon and A. F. Bennett, Stability analysis of operator splitting for large-scale ocean modeling, *J. Comput. Phys.* **123**, 311 (1996).
14. R. V. Madala, Efficient time integration schemes for atmosphere and ocean models, in *Finite Difference Techniques for Vectorized Fluid Calculations*, edited by D. Book (Springer-Verlag, New York, 1981), p. 56.
15. B. T. Nadiga, M. W. Hecht, L. G. Margolin, and P. K. Smolarkiewicz, On simulating flows with multiple time scale using a method of averages, *Theoret. Comput. Fluid Dyn.* **9**, 281 (1997).
16. T. Y. Hou, J. S. Lowengrub, and M. Shelley, Removing the stiffness from interfacial flows with surface tension, *J. Comput. Phys.* **114**, 312 (1994).
17. R. S. Rogallo, *Numerical Experiments in Homogeneous Turbulence*, NASA TM-81315 (1981).
18. A. J. Majda, D. W. McLaughlin, and E. G. Tabak, A one dimensional model for dispersive wave turbulence, *J. Nonlinear Sci.* **6**, 9 (1997).
19. P. Bartello, Geostrophic adjustment and inverse cascades in rotating stratified turbulence, *J. Atmos. Sci.* **52**, 4410 (1995).

20. R. Fjørtoft, On the changes in the spectral distribution of kinetic energy for two-dimensional, non-divergent flow, *Tellus* **5**, 225 (1953).
21. J. G. Charney, Geostrophic turbulence, *J. Atmos. Sci.* **28**, 1087 (1971).
22. M.-P. Lelong and J. J. Riley, Fluid motions in the presence of strong stable stratification, *Annu. Rev. Fluid Mech.* **32**, 613 (2000).
23. P. Bartello, O. Métais, and M. Lesieur, Geostrophic versus wave eddy viscosities in atmospheric models, *J. Atmos. Sci.* **53**, 564 (1996).
24. R. Asselin, Frequency filter for time integrations, *Mon. Weather Rev.* **100**, 487 (1972).