OPEN BOUNDARY CONDITIONS: FACT AND FICTION

D.R. DURRAN
University of Washington
Department of Atmospheric Sciences
Box 351640
Seattle, WA 98195, USA

1. Introduction

Many atmospheric and oceanic phenomena occur in a localized region. When numerically simulating such phenomena it is not practical to include all of the surrounding fluid in the numerical domain. As a case in point, one would not simulate an isolated thunderstorm with a global atmospheric model just to avoid possible problems at the lateral boundaries of a limited domain. Moreover, in a fluid such as the atmosphere there is no distinct upper boundary, and most numerical representations of the atmosphere's vertical structure terminate at some arbitrary level.

When the computational domain is terminated at an arbitrary location within a larger body of fluid, the conditions imposed at the edge of the domain are intended to mimic the presence of the surrounding fluid. These boundary conditions should allow outward-traveling disturbances to pass through the boundary without generating spurious reflections that propagate back toward the interior. Boundary conditions designed to minimize spurious backward reflection are known as nonreflecting, open, wave-permeable, or radiation boundary conditions. The terminology "radiation boundary condition" is due to Sommerfeld, who defined it as the condition that "the sources must be sources, not sinks, of energy. The energy which is radiated from the sources must scatter to infinity; no energy may be radiated from infinity into... the field."

As formulated by Sommerfeld, the radiation condition applies at infinity; however, in all practical computations a boundary condition must be imposed at some finite distance from the energy source, and this creates two problems. The first problem is that the radiation condition itself may not properly describe the physical behavior occurring at an arbitrarily designated location within the fluid when that location is only a finite distance from the energy source. The second problem is that it is often very difficult to express the radiation condition mathematically at a boundary that is only a finite distance from the energy source. It is so difficult to translate Sommerfeld's physical description into a mathematical formula, that in most practical problems it is not known how to express the radiation condition as a solvable algebraic or differential equation involving the prognostic variables in the neighborhood of the boundary. Instead the exact radiation condition must be replaced by an approximation, and this approximation introduces an error in the mathematical model of the physical system that is distinct from the errors subsequently incurred when constructing a discrete approximation to the mathematical model.

There are several factors that prevent numerical models of the atmosphere and ocean from actually converging toward the exact solution as the spatial and temporal resolution is repeatedly refined. The foremost impediment to convergence is surely the enormous range of scales that must be included in geophysical models of high-Reynolds-number flow (i.e., in a hypothetical model without any parameterization of subgrid-scale eddies). The numerical solution cannot be expected to begin to converge until the spatial resolution becomes sufficiently fine that the true Reynolds number is an order of magnitude smaller than the "numerical Reynolds" number computed using the effective numerical diffusivity in place of true molecular viscosity. Nevertheless, it should be emphasized that errors in the specification of open boundary conditions also pose a limitation on the ability to replicate the true solution in an otherwise highly refined model. Due to the approximations that are generally required to mathematically formulate open boundary conditions for the nondiscretized continuous equations, the numerical solutions to those equations will not converge to the correct solution to the underlying physical problem as the spatial and temporal mesh is refined, but rather to the correct solution to the approximate mathematical problem determined by the approximate open boundary conditions.

2. Is a radiation boundary condition appropriate?

The unmodified radiation condition is obviously not appropriate in situations where information must be transmitted inward through the boundary, as might be the case when the evolution of the large-scale structure of the solution on the boundary is determined by observations or by data from a coarse-resolution simulation conducted on a larger domain. Nevertheless, even if one wishes to force the boundary tendencies toward values determined by an external dataset, it is still necessary to allow waves originating within a limited-area domain to pass outward through the boundary without spurious reflection. In the remainder of this paper we will therefore focus on the basic problem of developing wave-permeable boundary conditions without making specific provision for the inclusion of externally specified boundary data.

Consider a problem in which the surrounding fluid is initially quiescent and all the initial disturbances are contained within the computational domain. One might suppose that the radiation condition, applied at the edge of the computational domain, would constitute a correct open boundary condition for such a problem—but in general it does not. Two nonlinear waves that propagate past an artificial boundary within a fluid may interact outside the artificial boundary and generate an inward-propagating disturbance that should reenter the domain. This point was emphasized by Hedstrom (1979), who provided a simple example from compressible gas dynamics demonstrating that a shock overtaking a contact discontinuity must generate an echo that propagates back toward the wave generator.

Numerical simulations of nonlinear vertically-propagating mountain waves provide another example where the documented sensitivity of the numerical solution to the loca-

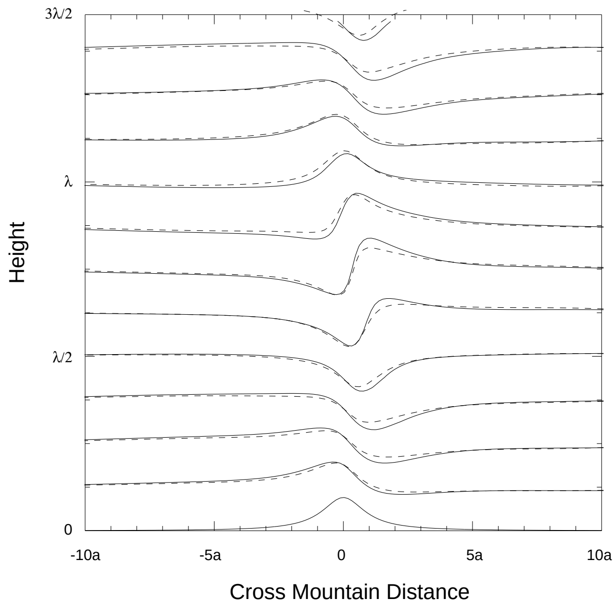

Figure 1. Streamlines in flow over a Witch of Agnesi mountain at nondimensional time $\tilde{t} = 44$ with the upper boundary at $z = 1.5\lambda$ (solid) or 2λ (dashed).

tion of the upper boundary (Durran and Klemp 1983, Klemp and Durran 1983, Scinocca and Peltier 1989, Epifanio and Durran 2000) may be less a consequence of inadequately approximating the radiation condition at the upper boundary than a failure to adequately capture important feedbacks on the low-level waves produced by downward propagating signals generated through wave-wave interactions occurring outside the numerical domain. A representative illustration of this sensitivity is shown in Fig. 1, which plots the streamlines in Boussinesq flow over a two-dimensional Witch of Agnesi mountain, $h(x) = h_0 a^2/(x^2 + a^2)$, from two numerical solutions computed using the model described in Epifanio and Durran (2000). The upstream stability and wind speed were constant with $N = 0.01$ s^{-1}, $U = 10$ ms^{-1}; the terrain parameters were $a = 10$ km, $h_0 = 600$ m. Thus $Nh_0/U = 0.6$, so the flow is moderately nonlinear, and $Na/U = 10$, so the flow is almost hydrostatic. Coriolis effects are ignored. In these simulations $\Delta x = 0.15a$ and $\Delta z = 0.4\lambda_z$, where $\lambda_z = 2\pi U/N$ is the vertical wavelength of a steady hydrostatic mountain wave. The lateral boundaries are placed extremely far upstream and downstream, at $x = -500a$ and $x = 700a$. The waves are initialized by gradually accelerating the flow from rest over $4\tilde{t}$, where $\tilde{t} = Ut/a$.

The solutions shown in Fig. 1 differ solely with respect to the location of the upper boundary. The solution shown by the solid lines was computed with an approximate radiation condition at $z = 1.5\lambda$; that shown by the dashed lines was computed with the upper boundary condition imposed at $z = 2\lambda$. One might suppose that this sensitivity to the location of the upper boundary is simply due to an inadequate approximation to the radiation condition, which was implemented using the local-in-time-and-space technique

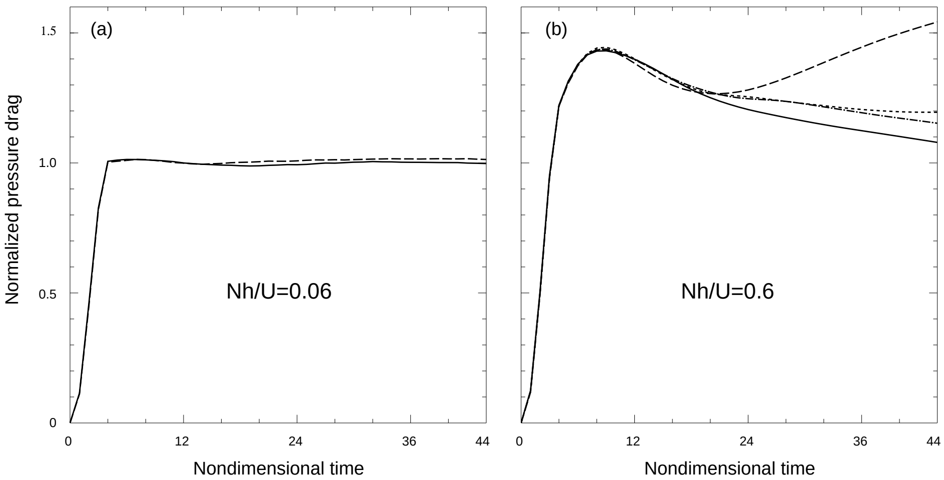

Figure 2. Cross mountain pressure drag, normalized by the drag associated with the corresponding linear solution for (a) $Nh_0/U = 0.06$ and (b) $Nh_0/U = 0.6$ when the upper boundary is at $z = 1.5\lambda$ (long dashed) or 2λ (solid). Also shown in (b) is the pressure-drag for cases with the upper boundary at 3λ (dot-dashed) and 3.5λ (short-dashed).

given in Durran (1999, p. 428), with $k_1 = 2\pi/a$, $k_2 = k_1/4$ and $k_3 = k_1/8$. However, a similar sensitivity is also obtained if the radiation condition is imposed by adding a very thick smoothly-varying sponge layer to the top of the model (Durran and Klemp 1983, Fig. 6). Moreover, the sensitivity to the location of the upper boundary is drastically reduced if the waves are made more linear by reducing h_0 by a factor of ten, as may be seen in the plots of the cross-mountain pressure drag versus time shown in Fig. 2 for the two cases: $Nh_0/U = 0.06$ and $Nh_0/U = 0.6$. The values plotted in Fig. 2 have been normalized by dividing the actual drag by the drag associated with the linear solution for flow over the same obstacle ($\pi\rho_0 NUh_0^2/4$). When $Nh_0/U = 0.06$, the solution generated by the nonlinear model is almost identical to the linear solution and there is almost no sensitivity to the height at which the radiation upper boundary condition is imposed.

When $Nh_0/U = 0.6$, finite-amplitude effects are no longer negligible, but the waves remain laminar and the subgrid-scale turbulence parameterization in the numerical model remains inactive. A semi-analytic steady-state solution for the nonlinear problem can be computed from Long's equation (Long 1953, Lilly and Klemp 1979). Since the waves are well-resolved, one might expect that the numerical solution could be made to converge to Long's solution by repeatedly refining the space-time mesh; however, difficulties with the boundary conditions prevent the numerical model from obtaining the correct result. Nevertheless, as suggested by those curves for the evolution of the drag in deeper domains plotted in Fig. 2b, increasing the depth of the domain decreases the sensitivity of the solution to the location of the upper boundary and delays the time at which the various solutions begin to diverge. At $\tilde{t} = 44$, the drags for the solutions for the two deepest domains (3λ and 3.5λ) bracket the correct value of 1.17 obtained from semi-analytic solutions to Long's equation.

Since similar sensitivities to the location of the upper boundary occur using very different numerical approximations to the radiation boundary condition (a local-in-space-and-time formulation and a deep "gentle" sponge layer), and since this sensitivity disappears in the small-amplitude limit, one can reasonably hypothesize that the basic problem arises from the use of the radiation condition itself. Nonsteady, nonlinear mountain waves appear to be generating downward propagating waves through wave-wave interactions. The truncation of the numerical domain at any finite height inevitably eliminates or distorts some of the internally generated downward propagating transients that should be present in the correct, nonsteady solution.

Errors resulting from a failure to incorporate inward-propagating signals generated by real physical processes occurring outside the boundaries of the computational domain cannot be avoided without enlarging the domain. Nevertheless, nonreflecting boundary conditions often become reasonable approximations to the true physical boundary condition as the size of the computational domain increases, provided that the local energy density of a disturbance arriving at the boundary is reduced as the result of wave dispersion or absorption in the large domain. When the disturbances arriving at the boundary are sufficiently weak, the governing equations in the region near the boundary can be approximated by their linearized equivalents, and it can be relatively easy to ensure that the radiation condition correctly describes the boundary conditions for the linearized system.

In summary, if the surrounding fluid is initially quiescent, the radiation boundary condition is usually the best statement of a physical condition that one might attempt to impose at an artificial boundary, but it is often only an approximation to the actual conditions that would hold at that boundary. As a consequence, even if a perfect implementation of the radiation boundary condition is present in some numerical model, one should not expect the model-generated solution to an arbitrary nonlinear problem to be approximately independent of the domain size unless the domain is sufficiently large.

3. Are One-Dimensional Test Problems Misleading?

One must learn to walk before learning to run, and in that spirit a great deal of effort has been devoted to the study of radiation boundary conditions for problems with two independent variables: time t and a single spatial coordinate x. While it is certainly true that open-boundary formulations that fail to work in x–t problems cannot be expected to perform well in higher-dimensional applications, it is not true that the most successful techniques for x–t problems can be trivially generalized to higher dimensions. Correct radiation boundary conditions for linear hyperbolic partial differential equations at each spatial boundary in an x–t domain can be obtained by setting the incoming characteristic variables to zero. Numerical details must still be considered to ensure that the equations for the out-going characteristic variables are treated in a stable, accurate manner, but prior to discretizing the governing equations, the radiation boundary conditions can be expressed mathematically without approximation.

On the other hand, in most time-dependent problems involving two or more spatial dimensions concepts related to characteristic variables are considerably less useful for the formulation of open boundary conditions. Two misconceptions that can easily develop after casual study of simple hyperbolic partial differential equations in an x–t problem are (1) that waves typically propagate along characteristics and (2) that setting the amplitude of an incoming wave to zero along the boundary is equivalent to specifying zero amplitude in the incoming characteristic variable in higher dimensional problems.

3.1. RAY PATHS AND CHARACTERISTIC CURVES

The ray paths along which geophysically significant waves such as gravity or Rossby waves propagate have essentially no relation to characteristic curves or their higher-dimensional generalizations (such as the Monge cones). Even in simple x–t problems, waves need not propagate along characteristics. As a first example, consider the linearized shallow-water equations for one-dimensional flow on an f-plane. Assuming the basic state is at rest, the linearized system reduces to

$$\frac{\partial u}{\partial t} + g\frac{\partial h}{\partial x} - f_0 v = 0, \tag{1}$$

$$\frac{\partial v}{\partial t} + f_0 u = 0, \tag{2}$$

$$\frac{\partial h}{\partial t} + H\frac{\partial u}{\partial x} = 0, \tag{3}$$

where u and v are the perturbation horizontal velocity components, h is the free-surface displacement about its mean depth H, and f_0 is the (constant) Coriolis parameter. Diagonalize the coefficient matrix of the spatial derivative by defining a vector of new unknown functions

$$\mathbf{v} = \begin{pmatrix} u - gh/c \\ v \\ u + gh/c \end{pmatrix},$$

with respect to which (1)–(3) transform to

$$\frac{\partial \mathbf{v}}{\partial t} + \begin{pmatrix} -c & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & c \end{pmatrix} \frac{\partial \mathbf{v}}{\partial x} + \begin{pmatrix} 0 & -f_0 & 0 \\ f_0/2 & 0 & f_0/2 \\ 0 & -f_0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0}.$$

The characteristics for this system are the curves satisfying $dx/dt = \pm c$ and $dx/dt = 0$.

Wave solutions to (1)–(3) have the form

$$(u', v', h') = \Re\left\{(u_0, v_0, h_0)e^{i(kx-\omega t)}\right\}, \tag{4}$$

provided that the frequency ω and wave number k satisfy the dispersion relation

$$\omega^2 = c^2k^2 + f_0^2, \tag{5}$$

as may be demonstrated by substituting (4) into (1)–(3). Lines of constant phase, such as the locations of the troughs and crests, propagate at the *phase speed* ω/k, which from (5) is

$$\frac{\omega}{k} = \pm c\left(1 + \frac{f_0^2}{c^2k^2}\right)^{1/2}.$$

A compact group of waves travels at the group velocity $\partial\omega/\partial k$, which can also be computed from (5):

$$\frac{\partial\omega}{\partial k} = \pm c\left(1 + \frac{f_0^2}{c^2k^2}\right)^{-1/2}.$$

In the limit $|k| \gg f_0/c$, the phase speed and group velocity both approach the slope of a characteristic along which $|dx/dt| = c$. Nevertheless, for any finite value of k,

$$\left|\frac{\partial\omega}{\partial k}\right| < c < \left|\frac{\omega}{k}\right|,$$

and neither the lines of constant phase nor the wave groups follow trajectories that coincide with the characteristic curves. The magnitude of the group velocity, which is the rate at which energy propagates in a wave, is nevertheless bounded by c. Thus, if waves are generated by a compact source at some initial time, the disturbance will lie within the portion of the x–t plane bounded by characteristic curves emanating from each edge of the source region, i.e., the disturbance will lie within the conventionally defined "domain of dependence" for the hyperbolic system (Durran 1999, p. 45).

The loose connection between wave propagation and the characteristics in the preceding example can disappear altogether if the Coriolis parameter is a function of the spatial coordinate. Then a second type of wave, the Rossby wave, may appear as an additional solution. If f increases linearly in proportion to y, Rossby-wave solutions may exist with phase speeds in the negative-x direction. Neither the phase speeds nor the group velocities of these waves have any relation to the characteristic curves. It is not surprising that Rossby waves do not propagate along the characteristics, because the terms involving the undifferentiated functions of u and v play no role in the determination of the characteristics of (3), yet those same terms are essential for the maintenance of the Rossby waves.

The Euler equations governing inviscid isentropic motion in a density stratified fluid provide another example of a hyperbolic system that supports two types of waves, one of which propagates along ray paths that are completely unrelated to any characteristic curve. The essential properties of this system can be most simply examined in a two-dimensional context. Let x and z be the horizontal and vertical coordinates, and decompose the thermodynamic fields into a vertically varying basic state and a perturbation such that

$$\begin{aligned}\pi(x,z,t) &= \overline{\pi}(z) + \pi'(x,z,t),\\ \theta(x,z,t) &= \overline{\theta}(z) + \theta'(x,z,t),\\ c_p\overline{\theta}\frac{d\overline{\pi}}{dz} &= -g.\end{aligned} \tag{6}$$

Here π is nondimensional Exner function defined as $(p/p_0)^{R/c_p}$ and $\theta = T/\pi$ is the potential temperature. The velocity components are decomposed as

$$u(x,z,t) = U + u'(x,z,t), \qquad w(x,z,t) = w'(x,z,t). \tag{7}$$

The basic-state vertical velocity is zero to ensure that the basic state is a steady solution to the nonlinear equations. The linearized governing equations become

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right)u' + c_p\overline{\theta}\frac{\partial\pi'}{\partial x} = 0, \tag{8}$$

$$\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial x}\right)w'+c_p\overline{\theta}\frac{\partial\pi'}{\partial z}=g\frac{\theta'}{\overline{\theta}}, \tag{9}$$

$$\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial x}\right)\theta'+\frac{\overline{\theta}}{g}N^2w'=0, \tag{10}$$

$$\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial x}\right)\pi'+w'\frac{\partial\overline{\pi}}{\partial z}+\frac{R\overline{\pi}}{c_v}\left(\frac{\partial u'}{\partial x}+\frac{\partial w'}{\partial z}\right)=0, \tag{11}$$

where

$$N^2=\frac{g}{\overline{\theta}}\frac{d\overline{\theta}}{dz}$$

is the square of the Brunt–Väisälä frequency.

Suppose that the reference state is isothermal. Then N^2 and the speed of sound $c_\mathrm{s}=(c_pRT/c_v)^{1/2}$ are constant, and the preceding system can be simplified by removing the influence of the decrease in the mean density with height via the transformation

$$\tilde{u}=\left(\frac{\overline{\rho}}{\rho_0}\right)^{1/2}u',\qquad \tilde{\pi}=\left(\frac{\overline{\rho}}{\rho_0}\right)^{1/2}\frac{c_p\overline{\theta}}{c_\mathrm{s}}\pi', \tag{12}$$

$$\tilde{w}=\left(\frac{\overline{\rho}}{\rho_0}\right)^{1/2}w',\qquad \tilde{\theta}=\left(\frac{\overline{\rho}}{\rho_0}\right)^{1/2}\frac{g}{N\overline{\theta}}\theta'. \tag{13}$$

Note that $\tilde{\theta}$ represents a scaled buoyancy and $\tilde{\pi}$ a scaled pressure. Let

$$\mathbf{v}=\left(\begin{array}{cccc}\tilde{u} & \tilde{w} & \tilde{\theta} & \tilde{\pi}\end{array}\right)^T;$$

then the transformed equations have the form

$$\frac{\partial\mathbf{v}}{\partial t}+\mathbf{A}_1\frac{\partial\mathbf{v}}{\partial x}+\mathbf{A}_2\frac{\partial\mathbf{v}}{\partial z}+\mathbf{B}\mathbf{v}=\mathbf{0}, \tag{14}$$

in which

$$A_1=\begin{pmatrix}U & 0 & 0 & c_\mathrm{s}\\ 0 & U & 0 & 0\\ 0 & 0 & U & 0\\ c_\mathrm{s} & 0 & 0 & U\end{pmatrix},\qquad A_2=\begin{pmatrix}0 & 0 & 0 & 0\\ 0 & 0 & 0 & c_\mathrm{s}\\ 0 & 0 & 0 & 0\\ 0 & c_\mathrm{s} & 0 & 0\end{pmatrix},$$

$$B=\begin{pmatrix}0 & 0 & 0 & 0\\ 0 & 0 & -N & -S\\ 0 & N & 0 & 0\\ 0 & S & 0 & 0\end{pmatrix},\qquad S=c_\mathrm{s}\left[\frac{1}{2\overline{\rho}}\frac{\partial\overline{\rho}}{\partial z}+\frac{1}{\overline{\theta}}\frac{\partial\overline{\theta}}{\partial z}\right].$$

The eigenvalues of $\mathbf{A}_1$ are $U, U, U+c_\mathrm{s}$, and $U-c_\mathrm{s}$; those of $\mathbf{A}_2$ are $0, 0, c_\mathrm{s}$, and $-c_\mathrm{s}$. The eigenvalues involving c_s give the speed at which sound waves in an unstratified fluid propagate parallel to the x and z coordinate axes. As will be demonstrated below, sound waves in an isothermally stratified atmosphere actually propagate at slightly different speeds due to the influence of the zero-order term in (14). The remaining eigenvalues relate to the speed at which fluid parcels are advected horizontally and vertically by the mean flow. None of these eigenvalues have any relation to the speed of propagation of gravity (or buoyancy) waves, which are the second type of fundamental wave motion supported by (14).

When the basic state is isothermal, S is constant, and wave solutions to (14) exist in the form

$$(\tilde{u}, \tilde{w}, \tilde{\theta}, \tilde{\pi}) = \Re\left\{(u_0, w_0, \theta_0, \pi_0)e^{i(kx+mz-\omega t)}\right\}, \tag{15}$$

provided that ω, k, and m satisfy the dispersion relation

$$(\omega - Uk)^2 = \frac{c_s^2}{2}\left(k^2 + m^2 + \frac{N^2 + S^2}{c_s^2}\right) \pm \frac{c_s^2}{2}\left[\left(k^2 + m^2 + \frac{N^2 + S^2}{c_s^2}\right)^2 - \frac{4N^2k^2}{c_s^2}\right]^{1/2}, \tag{16}$$

which is obtained by substituting (15) into (14). The second term inside the square root is much smaller than the first term in most applications (Durran 1999, p. 351), so (16) can be separated into a pair of approximate dispersion relations for the sound waves and the gravity waves. The dispersion relation for the sound waves,

$$(\omega - Uk)^2 = c_s^2\left(k^2 + m^2\right) + S^2 + N^2, \tag{17}$$

is obtained by taking the positive root in (16). In a manner analogous to the effect of the Coriolis force on gravity waves in the shallow-water system, the terms involving the product of N or S with the zero-order derivatives of the unknown variables introduce a slight discrepancy between the phase speeds and group velocities of the actual sound waves and those that might be suggested by the eigenvalues of $\mathbf{A}_1$ and $\mathbf{A}_2$.

The dispersion relation for the gravity waves is obtained by taking the negative root in (16), which to a good approximation yields

$$(\omega - Uk)^2 = \frac{N^2k^2}{k^2 + m^2 + (S^2 + N^2)/c_s^2}. \tag{18}$$

Neither the phase speeds nor the group velocities of these waves have any relation to the eigenvalues of $\mathbf{A}_1$ and $\mathbf{A}_2$. Unlike sound waves, *gravity waves do not even approximately propagate along the characteristics*. There is no relation between the characteristics and the paths of the gravity-waves because some of the physical processes essential for gravity-wave propagation are mathematically represented by undifferentiated functions of the unknown variables, and as such exert no influence on the shape of the characteristics.

3.2. SPECIFICATION OF THE INCOMING “CHARACTERISTIC” VARIABLE

As suggested previously, exact and practically useful mathematical formulae for the specification of nonreflecting boundary conditions are seldom available in multidimensional problems. The difficulties are readily apparent in simple two-dimensional shallow-water flow. The two-dimensional shallow-water equations, linearized about a basic state at rest with depth H, may be written

$$\begin{pmatrix} u \\ v \\ h \end{pmatrix}_t + \begin{pmatrix} 0 & 0 & g \\ 0 & 0 & 0 \\ H & 0 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ h \end{pmatrix}_x + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & g \\ 0 & H & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ h \end{pmatrix}_y = 0,$$

where the notation follows that used previously and Coriolis forces have been neglected. Suppose that a solution is sought in the limited domain $0 \le x \le L$, $0 \le y \le L$. In contrast to the one-dimensional case, the two-dimensional system cannot be reduced to a set of three scalar equations because no transformation of variables will simultaneously diagonalize both coefficient matrices. Nevertheless, the coefficient matrix multiplying the x-derivative can be diagonalized through the same change of variables used in the one-dimensional problem. Defining $d \equiv u - gh/c$ and $e \equiv u + gh/c$, the two-dimensional system becomes

$$\begin{pmatrix} d \\ v \\ e \end{pmatrix}_t + \begin{pmatrix} -c & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} d \\ v \\ e \end{pmatrix}_x + \begin{pmatrix} 0 & -c & 0 \\ -c/2 & 0 & c/2 \\ 0 & c & 0 \end{pmatrix} \begin{pmatrix} d \\ v \\ e \end{pmatrix}_y = 0. \quad (19)$$

Since only one coefficient matrix is a diagonal matrix, (19) is not as simple as the corresponding relation for one-dimensional flow. Nevertheless, (19) is useful for determining the number of boundary conditions that should be specified at a x-boundary. The signal in v and e is propagating inward through the boundary at $x = 0$, and the signal in d is propagating inward through the boundary at $x = L$. In order to obtain a well-posed problem, one might therefore attempt to specify two conditions at $x = 0$ of the form

$$e(0,t) = \alpha_1 d(0,t) + f_1(t), \qquad v(0,t) = \alpha_2 d(0,t) + f_2(t),$$

and one condition at $x = L$ of the form

$$d(L,t) = \alpha_3 e(L,t) + \alpha_4 v(L,t) + f_3(t). \quad (20)$$

This approach follows the guideline that the number of conditions specified at each boundary should be equal to the number of eigenvalues associated with outward propagation through each boundary. Even so, not all choices of $\alpha_1, \ldots, \alpha_4$ yield a well-posed problem. Oliger and Sundström (1978) and Sundström and Elvius (1979) provide details about various allowable values for $\alpha_1, \ldots, \alpha_4$. One way to obtain a well-posed problem is by choosing $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$, which specifies e and v at inflow and d at outflow.

Unlike the one-dimensional case, it is not clear what values of e, v, and d should be prescribed to prevent waves impinging on the boundary from partially reflecting into an inward-propagating mode. Suppose that the incoming variable d is set to zero at $x = L$, and suppose that a wave of of the form $(u, v, h) = (\hat{u}, \hat{v}, \hat{h}) \exp[i(kx + \ell y - \omega t)]$ is approaching the boundary at $x = L$ as shown in Fig. 3. The amplitudes in u and h are related by the linearized x-momentum equation, which for this case with no mean flow yields

$$-i\omega\hat{u} + ikg\hat{h} = 0,$$

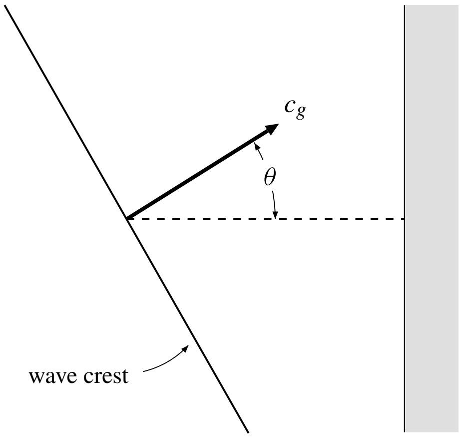

Figure 3. Wave crest approaching the "east" boundary at $x = L$.

or

$$\hat{u} = \frac{gk}{\omega}\hat{h}. \tag{21}$$

Using (21) and substituting for ω from the shallow-water dispersion relation for waves moving to the "east",

$$\omega = c(k^2 + \ell^2)^{1/2}, \tag{22}$$

the value of d at the right boundary becomes

$$d(L, y, t) = \left(\frac{gk}{\omega} - \frac{g}{c}\right)\hat{h}e^{i(kL+\ell y-\omega t)} = \frac{g}{c}\left(\frac{k}{(k^2+\ell^2)^{1/2}} - 1\right)\hat{h}e^{i(kL+\ell y-\omega t)}. \tag{23}$$

The condition $d(L, y, t) = 0$ cannot be satisfied without the simultaneous presence of a second reflected wave unless $\ell = 0$, or equivalently, unless the outward propagating wave is traveling at right angles to the boundary.

An expression for the amplitude of the spurious reflected wave generated by the $d(L, y, t) = 0$ boundary condition can be derived as follows. Let r be the amplitude of this reflected wave, and note that the condition $d(L, y, t) = 0$ cannot be satisfied at every point along $x = L$ unless the incident and reflected waves are linearly dependent functions of y and t. The frequency and the wave number parallel to the y-axis must therefore be the same in the incident and reflected waves, so according to the dispersion relation (22), the wave numbers parallel to the x-axis must have the same magnitude and opposite sign. The sum of the incident and reflected waves may therefore be expressed in the form

$$h(x, y, t) = \hat{h}e^{i(kx+\ell y-\omega t)} + r\hat{h}e^{i(-kx+\ell y-\omega t)}. \tag{24}$$

Here the first term represents the wave approaching the boundary, and the second term represents the reflected wave. Using (23) and (24) the condition $d(L, y, t) = 0$ becomes

$$\left(\frac{gk}{\omega} - \frac{g}{c}\right)\hat{h}e^{i((kL+\ell y-\omega t)} + r\left(\frac{-gk}{\omega} - \frac{g}{c}\right)\hat{h}e^{i((-kL+\ell y-\omega t)} = 0.$$

Solving for r yields

$$r = -\left(\frac{\omega - kc}{\omega + kc}\right) e^{2ikL}. \tag{25}$$

The strength of the reflected wave can be alternatively expressed as a function of the angle of incidence as follows. The lines of constant phase in Fig. 3 are parallel to the wave troughs and crests and satisfy $F(x, y) = 0$, where

$$F(x, y) = kx + \ell y - C;$$

C is a constant; and in the case shown in Fig. 3, k and ℓ are positive. The group velocity of the wave is both perpendicular to the lines of constant phase and parallel to the wave-number vector (k, ℓ), as may be confirmed from the relations

$$\left(\frac{\partial\omega}{\partial k}, \frac{\partial\omega}{\partial\ell}\right) = \frac{c^2}{\omega}(k, \ell) = \frac{c^2}{\omega}\nabla F.$$

As in Fig. 3, let θ be the angle by which the propagation of the wave deviates from the direction normal to the boundary, then $\tan\theta = \ell/k$. Using the dispersion relation (22) it follows that

$$k = \frac{\omega}{c}\cos\theta, \quad \ell = \frac{\omega}{c}\sin\theta. \tag{26}$$

Finally, substituting (26) into (25), one obtains

$$|r| = \left|\frac{\omega - kc}{\omega + kc}\right| = \left|\frac{1 - \cos\theta}{1 + \cos\theta}\right|,$$

again showing that this condition is perfectly nonreflecting only when the waves approach perpendicular to the boundary.

When d is set to zero at $x = L$ in two-dimensional shallow-water flow, at least one obtains an approximation to the correct radiation condition. If one attempts the same procedure with the continuous-stratified x–z problem governed by (14), the resulting boundary conditions will serve as an approximate radiation condition for sound waves but will not do anything to eliminate the spurious reflection of gravity waves impinging on the boundary. To be specific, the coefficient matrix of $\partial\mathbf{v}/\partial x$ in (14) will be diagonalized if $\mathbf{v}$ is transformed to $(\tilde{u} - \tilde{\pi}, \tilde{w}, \tilde{\theta}, \tilde{u} + \tilde{\pi})$. Assuming the flow is subsonic, the variable $\tilde{u} - \tilde{\pi}$ would be transmitted inward through an "east" boundary at $x = L$ at speed $U - c_s$. Will the relation $\tilde{u} - \tilde{\pi} = 0$ at $x = L$ therefore constitute a correct formulation of the radiation boundary condition?

As in the shallow-water case, consider a basic state with no mean flow for simplicity and suppose that a wave of the form $(\tilde{u}, \tilde{w}, \tilde{\theta}, \tilde{\pi}) = (\hat{u}, \hat{w}, \hat{\theta}, \hat{\pi}) \exp[i(kx + mz - \omega t)]$ is approaching the boundary at $x = L$. The x-momentum equation implies that

$$\hat{u} = \frac{c_s k}{\omega}\hat{\pi},$$

which, using the condition that $\tilde{u} - \tilde{\pi} = 0$ at $x = L$, may be expressed as

$$\left(\frac{c_s k}{\omega} - 1\right)\hat{\pi} e^{i(kL + mz - \omega t)} = 0. \tag{27}$$

Nontrivial solutions to (27) require $c_s = \omega/k$, or equivalently that *the trace speed*[1] *along the x-axis must be equal to the speed of sound.* If the outgoing wave is a sound wave propagating perpendicular to the boundary (so that $m = 0$), and if buoyancy effects are neglected ($N = S = 0$), then according to the sound wave dispersion relation (17), the wave will be transmitted without reflection. On the other hand, if the angle of incidence is not perpendicular to the boundary, or if buoyancy effects are not negligible, the condition $\tilde{u} - \tilde{\pi} = 0$ at $x = L$ is only an approximation.

By far the most serious problems associated with the strategy of setting the incoming variable $\tilde{u} - \tilde{\pi}$ to zero at $x = L$ involve the failure of this boundary condition to transmit gravity waves. The only gravity waves that will be transmitted with minimal reflection are those whose horizontal trace speeds are approximately equal to the speed of sound! These are certainly not the gravity waves of primary interest in most atmospheric and oceanic applications.

4. Are Dispersive Waves a Particular Challenge?

Previous researchers have often assumed that the Sommerfeld radiation condition at a lateral boundary may be expressed mathematically as

$$\frac{\partial \phi}{\partial t} + c^* \frac{\partial \phi}{\partial x} = 0, \tag{28}$$

where ϕ is an arbitrary prognostic variable, c^* is some effective wave speed, and (28) is applied at an x boundary. In the case of one-dimensional subcritical linearized shallow-water flow, (28) is a correct statement of the radiation condition provided that at the "east" boundary, $c^* = U + c$, or at the "left" boundary, $c^* = U - c$.

Equation (28) is not, however, a correct formulation of the Sommerfeld radiation condition in more general physical systems. Suppose a wave of the form $\hat{\phi} \exp[i(kx + \ell y + mz - \omega t)]$ is approaching the boundary at $x = L$. If this wave is to be transmitted without reflection, it must satisfy (28) at $x = L$, in which case

$$(\omega - c^* k)\hat{\phi} e^{i(kL + \ell y + mz - \omega t)} = 0.$$

For a nontrivial solution, $\omega/k = c^*$; that is c^* *must be identical to the x-trace speed of the wave*. In general $\omega/k = c^*$ is a function of the vertical and/or horizontal wavenumbers and, in contrast to one-dimensional shallow-water flow, there is no single value of c^* that can transmit an arbitrary superposition of several waves through the boundary without spurious reflection. Note that *the difficulty in specifying c^* is not due to wave dispersion per se;* the same essential difficulty arises when formulating boundary conditions for nondispersive two-dimensional shallow-water waves.

5. Can c^* be effectively diagnosed from (28)?

One example, in which c^* is a function of wavelength, is gravity wave propagation in stratified flow. Pearson (1974) suggested that (28) may be used to radiate internal gravity

[1] The x trace speed, ω/k, is the apparent phase speed of the wave along the x-axis. Unless the wave is actually propagating parallel to the x-axis, the x trace speed exceeds the true phase speed $\omega(k^2 + m^2)^{-1/2}$.

waves in stratified flow by fixing c^* at the Doppler-shifted phase speed of the dominant vertical mode. As an alternative, Orlanski (1976) suggested calculating c^* at a point just inside the boundary from the relation

$$c^* = -\frac{\partial\phi/\partial t}{\partial\phi/\partial x}. \tag{29}$$

Variants of the Orlanski lateral boundary condition have become very widely used in limited-area geophysical modeling.

The Orlanski scheme appears to have the virtue of avoiding the specification of c^* as an arbitrary parameter, but in fact it seldom yields a meaningful value for c^*. In particular, the computed value of c^* often falls outside the range allowed by the CFL stability criteria for the numerical approximation to (28) and must be artificially reset to a stable value. For example, the value of c^* would need to be limited to the interval $0 \le c^* \le \Delta x/\Delta t$ to preserve the stability of an upstream approximation to (28) at the "east" boundary.

Some indication of the difficulties encountered in the calculation of c^* were documented by Durran, Yang, Slinn and Brown (1993), who investigated the accuracy with which (29) diagnosed the phase speed of numerically simulated shallow-water waves, for which the correct value of c^* is simply $\sqrt{gH}$. Durran et al. (1993) solved a a linearized shallow-water model governed by the following finite-difference equations

$$\frac{u_{i+1/2}^{n+1} - u_{i+1/2}^{n-1}}{2\Delta t} + U\left(\frac{u_{i+3/2}^{n} - u_{i-1/2}^{n}}{2\Delta x}\right) + g\left(\frac{h_{i+1}^{n-1} - h_{i}^{n-1}}{\Delta x}\right) = 0,$$

$$\frac{h_{i}^{n+1} - h_{i}^{n-1}}{2\Delta t} + U\left(\frac{h_{i+1}^{n} - h_{i-1}^{n}}{2\Delta x}\right) + H\left(\frac{u_{i+1/2}^{n+1} - u_{i-1/2}^{n+1}}{\Delta x}\right) = 0.$$

The velocity and height variables are spatially staggered such that the edge of the domain coincides with the outermost velocity points. At the initial time the velocity is zero and

$$\eta(x,0) = g\sin^4(\pi x).$$

The solution was integrated forward 200 time steps on the domain $0 \le x \le 4$ km, with $\Delta x = 80$ m, $\Delta t = 0.48$ s, $c = 40$ ms^{-1}, $U = 10$ ms^{-1}. The boundary condition (28) was applied to each field at $x = 0$ and $x = 4$ km using upstream differencing to step between time levels $t - \Delta t$ and $t + \Delta t$. A reference solution was also computed on a large periodic domain -8 $\le x \le$ 12 km. The large-domain solution may be considered free from boundary-induced errors, since the periodic domain is too large to allow any disturbance exiting the region $0 \le x \le 4$ km to reenter that region before the end of the integration.

The value of c^* at $x = 4$ km was calculated using using Miller and Thorpe's (1981) formula

$$c^* = -\frac{\Delta x}{\Delta t}\left(\frac{\phi_{b-1}^{n} - \phi_{b-1}^{n-1}}{\phi_{b-1}^{n-1} - \phi_{b-2}^{n-1}}\right), \tag{30}$$

where ϕ represents either u or η, b is the index of the gridpoint at the right boundary, and the final value of c^* is constrained to lie in the interval $0 < c^* < 0.95\Delta x/2\Delta t$. A similar

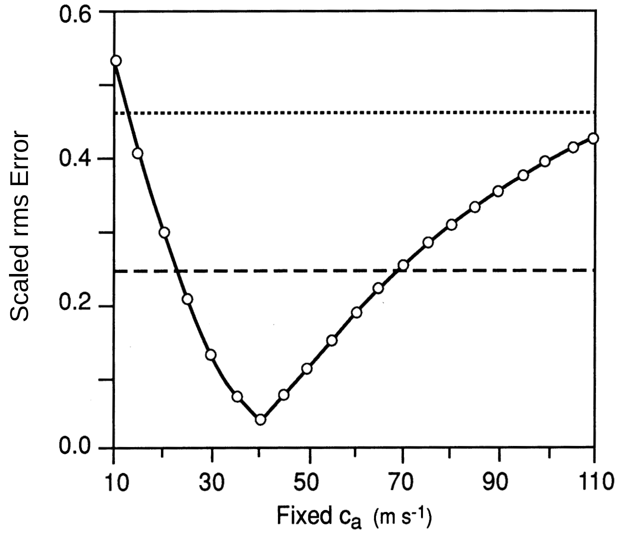

Figure 4. Cumulative rms boundary-induced error generated by the mis-specification of c^* as $U \pm c_a$, plotted as a function of c_a (solid curve). Also shown is the error for two variants of the Orlanski scheme which, being independent of c_a, appear as horizontal lines. The result using Orlanski's original formula for c^* is shown as the short-dashed line, the long-dashed line is obtained by computing c^* from (30). From Durran et al., (1993).

expression is evaluated at $x = 0$. This numerical approximation to (29) was found to give better results than the original formula proposed by Orlanski (1976).

Additional simulations were also conducted in which the gravity-wave phase speed was approximated by an arbitrary constant value c_a, and c^* was set to $U + c_a$ (or $U - c_a$) at the right (or left) boundary. Even when c_a was a relatively poor approximation to $\sqrt{gH}$, the constant-c_a approach produced better results than those obtained by continually estimating c^* from (30). This superiority is illustrated in Fig. 4, in which the scaled cumulative root-mean-square boundary-induced error is plotted as a function of c_a and compared with the error that develops when c^* is calculated using (30). The boundary-induced error is the difference between the limited-area and large-periodic-domain solutions for u and h. Details of the scaled rms error computation are given in Durran et al. (1993).

As evident in Fig. 4, the optimal choice of $c_a = \sqrt{gh} = 40$ ms^{-1} gives substantially less error than that obtained by calculating c^* at the boundary. It is certainly not surprising that fixing c_a at its theoretically correct value gives the best result. What is surprising is that fixed values of c_a are superior to those obtained through the evaluation of c^* via (30) over the wide range $25 \leq c_a \leq 65$ ms^{-1}. To understand why the diagnosis of c^* is inferior over such a wide range of c_a, it is instructive to examine the actual c^* values calculated at the boundaries. The values of c^* computed from (30) at the right-hand-boundary are plotted as a function of the time step in Fig. 5. At the initial time, a wave trough is present at the boundary; every 20 time steps thereafter another trough or crest arrives at the boundary. During the interval between the passage of the initial trough and the first crest (the first 20 steps), the c^* calculation yields a somewhat reasonable approximation to the correct value of $U + \sqrt{gH} = 50$ ms^{-1}. During the interval between the passage of

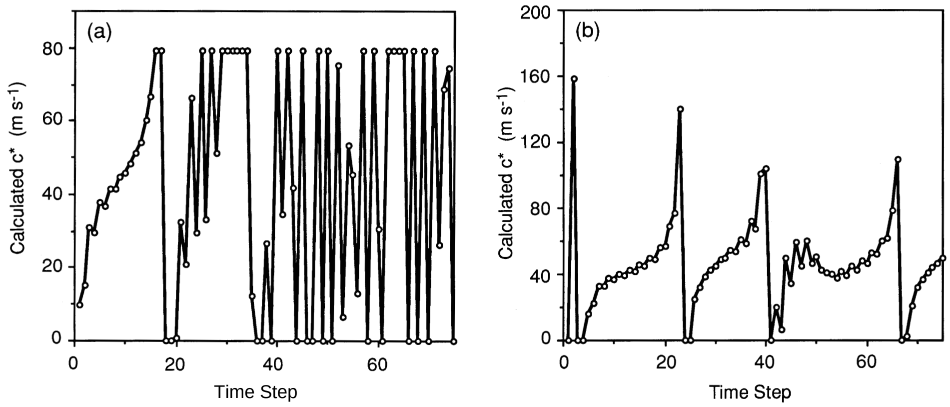

Figure 5. Plot of c^* as a function of time step as calculated from (30) using the solution on: (a) the limited domain, and (b) the wide periodic domain. After Durran et al., (1993)

the first crest and the second trough (the second 20 steps), the computed c^* is quite noisy. After the first 40 steps, c^* is drastically over- or under-predicted and set to its limiting values of 0 or 80 ms^{-1} in an almost $2\Delta t$ oscillation.

The most serious problem with the c^* calculation is not the fidelity with which the finite-difference formula (30) approximates the continuous expression (29), but the susceptibility of the calculation to a positive feedback between the generation of error at the lateral boundary and further degradations in the computed values of c^*. The effect of the cumulative boundary-induced error on the calculation of c^* may be appreciated by comparing Fig. 5a with Fig. 5b. Fig. 5b shows the right-boundary c^* obtained when the large-periodic-domain solution (instead of the limited-domain solution) is used in the evaluation of (30). When c^* is diagnosed from the large-periodic-domain solution, the estimation of c^* beyond the 20th time step is much improved and there is no systematic degradation in accuracy as the number of time steps increases. The largest errors are associated with the passage of troughs or crests in which case the denominator in (29) becomes small and the c^* calculation becomes very sensitive to numerical error.

The positive feedback between errors in c^* and errors in u and h is not surprising. Since the full equations support waves moving at both $U + c$ and $U - c$, (28) does not correctly describe the dynamics of the flow at the point point just inside the right boundary at which c^* is evaluated via (30). Unless only one of the two admissible waves is present, the c^* calculation cannot be expected to yield a meaningful result. Very early in the simulation there are no inward propagating waves adjacent to the lateral boundary, and a reasonable calculation of c^* is possible (compare the first 20 steps in Figs. 5a and 5b; note the difference in the vertical scales). As the simulation proceeds, initially small errors develop at the lateral boundary and induce inward propagating waves. These erroneously generated waves grow until (28) is no longer applicable, and the local calculation of the phase speed via (29) becomes essentially meaningless.

The difficulties encountered using an Orlanski-style open boundary condition in the one-dimensional shallow water system are discouraging since this is one of the few situations where there is a unique wave speed that might be calculable from (29). On the other hand, the one-dimensional shallow-water system is also one of the few systems in which there is no need for the Orlanski approach, because it is easy to specify c^* correctly. In systems that support waves moving at different speeds, it is not possible to specify a single value of c^* that will transmit both waves without reflection and one might suppose that it could be more effective to calculate some (presumably averaged) c^* via (29) than to fix c_a at a constant value. Nevertheless, as shown in Durran et al. (1993), distinctly better results are still obtained using a fixed value of c_a [instead of (29)] in a two-layer shallow-water system in which two waves with different speeds impinge on the boundary.

6. Better Formulations

The preceding sections have emphasized the shortcomings and limitations of some popular methods and their underlying assumptions. What approach might yield an improvement? Perhaps the most promising techniques are those that build of the pioneering work by Engquist and Majda (1977) who suggested replacing the true governing equation at an open boundary by a locally solvable partial differential equation whose dispersion relation is approximately satisfied by the outward directed waves arriving at the boundary. These partial differential equations are "one-way wave equations" in the sense that they are not even approximately satisfied by any inward directed wave, thereby minimizing the development of spurious reflections at the boundary.

The original formulations of Engquist and Majda (1977), which were proposed for two-dimensional shallow-water flow, have been generalized in a series of papers by Higdon. In particular, Higdon (1994) has suggested that an improved approximation to the radiation boundary condition for waves with dispersive x-trace speeds can be obtained by replacing the basic one-way wave equation (28) with the product of a series of one-way operators of the form

$$\left[\prod_{j=1}^{p}\left(\frac{\partial}{\partial t}+c_j\frac{\partial}{\partial x}\right)\right]\phi = 0, \tag{31}$$

where the set of c_j is chosen to span the range of x trace speeds associated with the waves appearing at the boundary. The ratio of the amplitude of the reflected wave generated by this boundary condition to that of an incident wave traveling parallel to the x-axis at trace speed c is

$$|r| = \prod_{j=1}^{p}\left|\frac{c-c_j}{c+c_j}\right|.$$

Although this scheme does not seem to have been used as a boundary condition for limited-area models of stably stratified flow, Higdon (1994) has successfully used to it simulate dispersive shallow-water waves on an f-plane. Experiments using (31) as a lateral boundary condition in simulations of stratified flow conducted by this author have suggested that a substantial improvement over the simplest one-way wave equation (28) can be obtained using only two phase speeds, i.e., by increasing p from one to two. Further in-depth analysis of the formulation of one-way wave equations for the approximation of open boundary conditions in both shallow-water and stratified flow is provided in Durran (1999, Chpt. 8).

Acknowledgments. This research was sponsored by NSF grant ATM-9817728.

References

Durran, D. R.: 1999, *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*, Springer-Verlag, New York. 465 p.

Durran, D. R. and Klemp, J. B.: 1983, A compressible model for the simulation of moist mountain waves, *Mon. Wea. Rev.* **111**, 2341–2361.

Durran, D. R., Yang, M.-J., Slinn, D. N. and Brown, R. G.: 1993, Toward more accurate wave-permeable boundary conditions, *Mon. Wea. Rev.* **121**, 604–620.

Engquist, B. and Majda, A.: 1977, Absorbing boundary conditions for the numerical simulation of waves, *Math. Comp.* **31**, 629–651.

Epifanio, C. C. and Durran, D. R.: 2000, Three-dimensional effects in high-drag-state flows over long ridges, *J. Atmos. Sci.* in press.

Hedstrom, G.: 1979, Nonreflecting boundary conditions for nonlinear hyperbolic systems, *J. Comp. Phys.* **30**, 222–237.

Higdon, R.: 1994, Radiation boundary conditions for dispersive waves, *SIAM J. Numer. Anal.* **31**, 64–100.

Klemp, J. B. and Durran, D.: 1983, An upper boundary condition permitting internal gravity wave radiation in numerical mesoscale models, *Mon. Wea. Rev.* **111**, 430–444.

Lilly, D. K. and Klemp, J. B.: 1979, The effects of terrain shape on nonlinear hydrostatic mountain waves, *J. Fluid Mech.* **95**, 241–261.

Long, R. R.: 1953, Some aspects of the flow of stratified fluids I. A theoretical investigation, *Tellus* **5**, 42–58.

Oliger, J. and Sundström, A.: 1978, Theoretical and practical aspects of some initial boundary value problems in fluid dynamics, *SIAM J. Appl. Math* **35**, 419–446.

Orlanski, I.: 1976, A simple boundary condition for unbounded hyperbolic flows, *J. Comp. Phys.* **21**, 251–269.

Pearson, R.: 1974, Consistent boundary conditions for numerical models of systems that admit dispersive waves, *J. Atmos. Sci.* **31**, 1481–1489.

Scinocca, J. F. and Peltier, W. R.: 1989, Pulsating downslope windstorms, *J. Atmos. Sci.* **46**, 2885–2914.

Sundström, A. and Elvius, T.: 1979, Computational problems related to limited area modelling, *Numerical Methods Used in Atmospheric Models*, Vol. II of *GARP Publication Series No. 17*, World Meteorological Organization, Geneva, pp. 379–416.