INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2008; **57**:389–416
Published online 14 January 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fld.1617

Fourier analysis of a class of upwind schemes in shallow water systems for gravity and Rossby waves

A. M. Mohammadian*,† and D. Y. Le Roux

Mathématiques et Statistique, Université Laval, Quebec City, Quebec, Canada

SUMMARY

A Fourier analysis has been performed for a class of upwind finite volume schemes, including the study of phase speed, group velocity, damping and dispersion. In the first part, pure gravity waves are investigated. As expected, most upwind schemes lead to a significant damping, but they exhibit a better phase behavior than most centered schemes. In the second part, the Coriolis parameter is considered and the Rossby modes are studied. In this case, all selected upwind schemes lead to a severe damping. The numerical results are also compared with those obtained by using a slope limiter approach. It is concluded that most upwind schemes with or without slope limiters present poor results for an accurate calculation of the Rossby modes. Copyright © 2008 John Wiley & Sons, Ltd.

Received 22 August 2006; Revised 16 August 2007; Accepted 16 August 2007

KEY WORDS: finite volume method; upwind schemes, shallow water equations; gravity waves; Rossby waves

1. INTRODUCTION

Finite volume schemes are well known due to their inherent conservation properties. In addition, upwind finite volume schemes have become popular for hyperbolic systems during the past two decades due to their ability to capture discontinuities with a low level of numerical diffusion and oscillations. For these methods, the critical stage is the calculation of the numerical flux and various schemes have been developed based on different approaches for estimation of those fluxes (see, e.g. [1, 2]).

Upwind finite volume schemes use the characteristic information of the hyperbolic system to calculate numerical fluxes. In the case of the scalar advection equation, this simply leads to a biased discretization of the equation in the flow direction. However, this is not the case for coupled systems, such as the shallow water equations. This is because the flow direction is not the only

*Correspondence to: A. M. Mohammadian, Mathématiques et Statistique, Université Laval, Quebec City, Quebec, Canada.
†E-mail: majidm@mat.ulaval.ca

factor to be considered in the calculation of the numerical fluxes. In fact, for coupled systems, the flux vector must be decomposed in the basis of eigenvectors and each component is then calculated in an upwind manner based on the sign of the corresponding eigenvalue.

Extensive research has been dedicated to the study of upwind schemes for shallow water systems, especially in the convection-dominated cases, i.e. for hyperbolic formulations (see e.g. [2]). Upwind finite volume schemes have been successfully employed for the simulation of some challenging problems for shallow flows such as dam break flows and supercritical flows over spillways (e.g. [3–5]).

Fourier analysis is a useful tool to study the effect of discrete schemes on some quantities such as wave amplitude, phase speed and group velocity. The analytical form of those quantities should be preserved by an ideal numerical method. Further, the Fourier analysis may give guidelines for a proper selection of numerical parameters such as the Courant–Friedrich–Lewy (CFL) number. Note that the Fourier analysis has been applied to the discretized form of the shallow water equations using finite difference (e.g. [6–11]), and finite element methods [12–17]. Note that Sankaranarayanan and Spaulding [11] considered the boundary fitted coordinates and the Fourier analysis was performed in transformed plane. The dispersion relation for the least-squares-mixed formulation of the shallow water equations has also been analyzed by Le Roux and Carey [18] and the results are compared with those of the Galerkin scheme. They concluded that the method should be used with care particularly for long-term simulations due to its inherent numerical damping.

The discretization of the shallow water system using upwind schemes has been rarely studied by the Fourier analysis. This is because slope/flux limiters are inherent to most upwind formulations and cannot be included in a Fourier-type analysis. Szpilka and Kolar [19] have numerically studied the phase and damping errors of some slope limiters. However, such a study can only be performed for a few sets of parameters due to the excessive amount of required numerical computations.

Gossard and Kolar [20] performed Fourier and dispersion analyses of the first-order upwind finite volume scheme for the shallow water equations, where slope limiters are not involved. They have concluded that such a scheme exhibits a good phase behavior. Owing to the inherent conservation of finite volume schemes, the work of Gossard and Kolar [20] motivates the study of higher order upwind finite volume schemes. Moreover, such a study seems necessary because, as shown in the following, the use of slope limiters is not necessarily recommended for all types of waves. Indeed, shallow water equations permit various waves, including fast gravity and slow Rossby modes. The later are due to the Earth rotation and usually need to be computed quite accurately [21]. This is because most energy transfer in the ocean scale is due to these waves. As shown later, the amplitude of the Rossby modes is considerably damped by the use of slope limiters in upwind finite volume schemes, which has not been reported earlier in the literature, up to our knowledge. Note that the study of Gossard and Kolar [20] is concerned only with the propagation of gravity waves. Such waves are important for small scale flows, e.g. estuaries, but not for large-scale flows such as global ocean circulation. Indeed, unless an accurate representation of the fast modes is important, gravity waves may be regarded as small amplitude noise superimposed on the slow solution and they can be justifiably retarded or damped, which is usually the case in ocean modeling.

The present paper is dedicated to dispersion and Fourier analyses of high-order upwind finite volume schemes for gravity and Rossby waves and it is organized as follows. The model equations for the pure gravity waves are introduced in Section 2. Various finite volume schemes are presented in Section 3 for these waves, and their Fourier analysis in a semi-discretization framework is performed in Section 4. The temporal discretization method is explained in Section 5 and the behavior of the resulting fully discrete schemes is studied in Section 6. The model equations for

DOI: 10.1002/fld

Rossby waves are then introduced in Section 7 and a Fourier analysis is also performed. Some numerical test cases are presented in Section 8 and concluding remarks complete the study.

2. MODEL PROBLEM FOR GRAVITY WAVES

The one-dimensional, inviscid, linearized form of the shallow water equations may be expressed in Cartesian coordinates [21] as

$$\eta_t + Hu_x = 0 \tag{1}$$

$$u_t + g\eta_x = 0 \tag{2}$$

where u is the velocity variable, η is the surface elevation with respect to the reference level $z=0$, g is the gravitational acceleration and the mean depth H is assumed constant.

In the present analysis, (1)–(2) are solved on an infinite channel (i.e. with periodic boundary conditions) subject to initial conditions. System (1)–(2) could be written in the following conservative vector form:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = 0 \tag{3}$$

where

$$\mathbf{U} = \begin{bmatrix} \eta \\ u \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} Hu \\ g\eta \end{bmatrix} \tag{4}$$

System (1)–(2) can also be written in the following non-conservative form

$$\frac{\partial \mathbf{U}}{\partial t} + A\frac{\partial \mathbf{U}}{\partial x} = 0 \tag{5}$$

where

$$A = \begin{bmatrix} 0 & H \\ g & 0 \end{bmatrix} \tag{6}$$

Matrix A has two real eigenvalues

$$\lambda_1 = \sqrt{gH}, \quad \lambda_2 = -\sqrt{gH} \tag{7}$$

with the following corresponding eigenvectors:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ +\sqrt{g/H} \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 1 \\ -\sqrt{g/H} \end{bmatrix} \tag{8}$$

Matrix A is decomposed as

$$A = PDP^{-1} \tag{9}$$

DOI: 10.1002/fld

where D is the diagonal matrix including the eigenvalues of A as

$$D=\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}=\begin{bmatrix}\sqrt{gH} & 0\\ 0 & -\sqrt{gH}\end{bmatrix} \tag{10}$$

and P is the matrix including the eigenvectors of A as

$$P=[\mathbf{e}_1,\mathbf{e}_2]=\begin{bmatrix}1 & 1\\ \sqrt{g/H} & -\sqrt{g/H}\end{bmatrix} \tag{11}$$

and the inverse matrix P^{-1} is given by

$$P^{-1}=\frac{1}{2}\begin{bmatrix}1 & \sqrt{H/g}\\ 1 & -\sqrt{H/g}\end{bmatrix} \tag{12}$$

Define

$$D^+=\begin{bmatrix}\max(\lambda_1,0) & 0\\ 0 & \max(\lambda_2,0)\end{bmatrix} \tag{13}$$

$$D^-=D-D^+ \tag{14}$$

or

$$D^+=\begin{bmatrix}\sqrt{gH} & 0\\ 0 & 0\end{bmatrix},\quad D^-=\begin{bmatrix}0 & 0\\ 0 & -\sqrt{gH}\end{bmatrix} \tag{15}$$

and

$$|D|=\begin{bmatrix}|\lambda_2| & 0\\ 0 & |\lambda_2|\end{bmatrix}=\begin{bmatrix}\sqrt{gH} & 0\\ 0 & \sqrt{gH}\end{bmatrix} \tag{16}$$

Therefore, matrix A is decomposed as

$$A=A^+ + A^- \tag{17}$$

where

$$A^+=PD^+P^{-1}=\begin{bmatrix}\sqrt{gH}/2 & H/2\\ g/2 & \sqrt{gH}/2\end{bmatrix} \tag{18}$$

and

$$A^-=PD^-P^{-1}=\begin{bmatrix}-\sqrt{gH}/2 & H/2\\ g/2 & -\sqrt{gH}/2\end{bmatrix} \tag{19}$$

DOI: 10.1002/fld

3. NUMERICAL ALGORITHM

In the finite volume method, the equations are integrated in each computational cell. This leads to

$$\int_{\Omega_c}\left(\frac{\partial \mathbf{U}}{\partial t}+\frac{\partial \mathbf{F}}{\partial x}\right)\mathrm{d}x=0 \tag{20}$$

where Ω_c represents the area of a 1D time–space control volume. The divergence theorem is then employed to replace the volume integral by a surface one

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_c}\mathbf{U}\,\mathrm{d}x+\int_{\Gamma_c}\mathbf{F}\cdot\mathbf{n}\,\mathrm{d}\Gamma_c=0 \tag{21}$$

where Γ_c represents the boundary of a control volume and $\mathbf{n}$ is its unit outward normal vector. For the one dimensional case, (21) is rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{U}_j(t)=-\frac{\mathbf{F}_{j+1/2}(t)-\mathbf{F}_{j-1/2}(t)}{\Delta x} \tag{22}$$

where $\mathbf{U}_j(t)$ represents the cell-averaged value of conserved variables and $\mathbf{F}_{j+1/2}(t)$ is the numerical flux. For the various numerical schemes employed in this study, the corresponding numerical fluxes (see, e.g. [2]) are given in the following.

For the centered scheme we have

$$\mathbf{F}_{j+1/2}=0.5(\mathbf{F}_{j+1}+\mathbf{F}_j) \tag{23}$$

The first-order upwind scheme (1st) gives

$$\mathbf{F}_{j+1/2}=A^-\mathbf{U}_{j+1}+A^+\mathbf{U}_j \tag{24}$$

or equivalently

$$\mathbf{F}_{j+1/2}=0.5(\mathbf{F}_{j+1}+\mathbf{F}_j)-0.5|A|(\mathbf{U}_{j+1}-\mathbf{U}_j) \tag{25}$$

where

$$|A|=P|D|P^{-1} \tag{26}$$

In (25), the first-order upwind flux is equal to the centered flux in (23) plus an artificial diffusive flux, which stabilizes the numerical scheme. A class of higher order schemes may be constructed by calculating the interface values more accurately, e.g. as in the κ scheme. In this method, which includes a family of schemes, the numerical flux is calculated as

$$\mathbf{F}_{j+1/2}=0.5(\mathbf{F}^{\mathrm{L}}_{j+1/2}+\mathbf{F}^{\mathrm{R}}_{j+1/2})-0.5|A|(\mathbf{U}^{\mathrm{L}}_{j+1/2}-\mathbf{U}^{\mathrm{R}}_{j+1/2}) \tag{27}$$

where the superscripts R an L represent the evaluation of the right and left sides of the interface, respectively, with

$$\mathbf{F}^{\mathrm{L}}_{j+1/2}=\mathbf{F}(\mathbf{U}^{\mathrm{L}}_{j+1/2}) \tag{28}$$

$$\mathbf{F}^{\mathrm{R}}_{j+1/2}=F(\mathbf{U}^{\mathrm{R}}_{j+1/2}) \tag{29}$$

DOI: 10.1002/fld

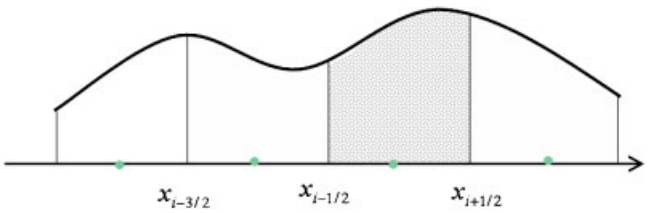

Figure 1. Upwind polynomial fitting for the 3rdCell scheme. A cubic curve is fitted such that the mass is conserved in four consecutive cells.

The interface values $\mathbf{U}^{\mathrm{L}}_{j+1/2}$ and $\mathbf{U}^{\mathrm{R}}_{j+1/2}$ are calculated at the interface $j+1/2$ as

$$\mathbf{U}^{\mathrm{L}}_{j+1/2}=\mathbf{U}_j+\tfrac{1}{4}((1-\kappa)(\mathbf{U}_j-\mathbf{U}_{j-1})+(1+\kappa)(\mathbf{U}_{j+1}-\mathbf{U}_j)) \tag{30}$$

$$\mathbf{U}^{\mathrm{R}}_{j+1/2}=\mathbf{U}_{j+1}-\tfrac{1}{4}((1-\kappa)(\mathbf{U}_{j+2}-\mathbf{U}_{j+1})+(1+\kappa)(\mathbf{U}_{j+1}-\mathbf{U}_j)) \tag{31}$$

where the choice of κ leads to the following schemes:

$$\kappa=\begin{cases}-1 & \text{second-order upwind}\\ 0 & \text{simplified Fromm scheme}\\ 1/6 & \text{cell-based third-order upwind}\\ 1/3 & \text{third-order upwind}\\ 1/2 & \text{Quick scheme}\\ 1 & \text{centered scheme}\end{cases} \tag{32}$$

Among the schemes given above, the centered and second-order upwind (2nd) schemes are well documented in the literature (see, e.g. [1]), the Quick scheme is a simplified version of the method presented in Leonard [22] where the third-order upwind scheme (3rd) is also introduced, and the Fromm scheme is a limiting case of the method proposed in Fromm [23].

The cell-based (formally) third-order accurate discretization (3rdCell) for a 1D problem, which is introduced here for the first time to our knowledge, is obtained by fitting a cubic curve over four consecutive computational cells, such that the total mass of each cell is correctly represented by the corresponding cubic curve for each conserved variable, i.e. (Figure 1)

$$\int_{x_{i-1/2}}^{x_{i+1/2}}\mathbf{U}^{\mathrm{up}}_j(x)\,\mathrm{d}x=\mathbf{U}_i\Delta x,\quad \text{for } i=j-2,j-1,j,j+1 \tag{33}$$

where $\mathbf{U}^{\mathrm{up}}_j(x)$ denotes the constructed upwind polynomial for the cell j. Note that the upwind polynomial fitting is performed by using two cells at the upstream side of the cell j and one cell in the downstream side. This leads to the following cubic polynomial:

$$\mathbf{U}^{\mathrm{up}}_j(x)=\mathbf{a}_3x^3+\mathbf{a}_2x^2+\mathbf{a}_1x^2+\mathbf{a}_0,\quad -\Delta x/2\leqslant x\leqslant\Delta x/2 \tag{34}$$

where

$$\mathbf{a}_3=\frac{1}{24\Delta x^3}(4\mathbf{U}_{j+1}-12\mathbf{U}_j+12\mathbf{U}_{j-1}-4\mathbf{U}_{j-2}) \tag{35}$$

DOI: 10.1002/fld

$$\mathbf{a}_2 = \frac{1}{24\Delta x^2}(12\mathbf{U}_{j+1} - 24\mathbf{U}_j + 12\mathbf{U}_{j-1}) \tag{36}$$

$$\mathbf{a}_1 = \frac{1}{24\Delta x}(7\mathbf{U}_{j+1} + 15\mathbf{U}_j - 27\mathbf{U}_{j-1} + 5\mathbf{U}_{j-2}) \tag{37}$$

$$\mathbf{a}_0 = \tfrac{1}{24}(-\mathbf{U}_{j+1} + 26\mathbf{U}_j - \mathbf{U}_{j-1}) \tag{38}$$

The interface values is then adjusted such that the condition

$$\left.\frac{\partial \mathbf{U}_j^{\mathrm{up}}(x)}{\partial x}\right|_{x=0} = \frac{\mathbf{U}_{j+1/2}^{\mathrm{L}} - \mathbf{U}_{j-1/2}^{\mathrm{L}}}{\Delta x} \tag{39}$$

is satisfied. That is, the left-hand side in (39) is known using the (conservative) cubic polynomial in (34) as

$$\left.\frac{\partial \mathbf{U}_j^{\mathrm{up}}(x)}{\partial x}\right|_{x=0} = \frac{7\mathbf{U}_{j+1} + 15\mathbf{U}_j - 27\mathbf{U}_{j-1} + 5\mathbf{U}_{j-2}}{24\Delta x} \tag{40}$$

and the left interface values $\mathbf{U}_{j+1/2}^{\mathrm{L}}$ and $\mathbf{U}_{j-1/2}^{\mathrm{L}}$ in the right-hand side of (39) are sought such that condition (39) is satisfied. This leads to the following (left) interface value

$$\mathbf{U}_{j+1/2}^{\mathrm{L}} = \tfrac{1}{24}(7\mathbf{U}_{j+1} + 22\mathbf{U}_j - 5\mathbf{U}_{j-1}) \tag{41}$$

In a similar way, the downwind polynomial $\mathbf{U}_j^{\mathrm{dn}}(x)$ is constructed for the cell j as

$$\int_{x_{i-1/2}}^{x_{i+1/2}} \mathbf{U}_j^{\mathrm{dn}}(x)\,\mathrm{d}x = \mathbf{U}_i \Delta x \quad \text{for } i = j-1, j, j+1, j+2 \tag{42}$$

and the (right) interface value is obtained as

$$\mathbf{U}_{j+1/2}^{\mathrm{R}} = \tfrac{1}{24}(-5\mathbf{U}_{j+2} + 22\mathbf{U}_{j+1} + 7\mathbf{U}_j) \tag{43}$$

Note that the cell-based third-order accurate method may be written in the κ scheme form, with $\kappa = \frac{1}{6}$, as mentioned in (32). We remind the reader that the 3rdCell scheme is conservative because it is written in the flux form (22) and the above-mentioned polynomial fitting is only performed to compute the interface values. Indeed, in the general flux formula in (27), the left and right interface values $\mathbf{U}_{j+1/2}^{\mathrm{L}}$ and $\mathbf{U}_{j+1/2}^{\mathrm{R}}$ are used to compute the flux and the resulting scheme is conservative because the flux across the cell interfaces is continuous. That is, the mass exiting from a cell enters into the adjacent cell; hence, the total mass is conserved.

Finally, we remark here that special treatments are needed for the implementation of the numerical schemes at boundaries. For the upwind schemes, the boundary conditions are usually specified based on the theory of the characteristics which is beyond the scope of present paper and in the Fourier analyses performed here, the boundaries are assumed to be periodic. That is, the general behavior of the upwind schemes with no effects from boundaries is studied in this paper.

DOI: 10.1002/fld

4. FOURIER ANALYSIS FOR THE SPATIAL DISCRETIZATION SCHEMES

Substituting periodic solutions of the form $u=\mathrm{Re}[\widetilde{u}e^{i(kx+\omega t)}]$ and $\eta=\mathrm{Re}[\widetilde{\eta}e^{i(kx+\omega t)}]$ into (1) and (2), where $\widetilde{u}$ and $\widetilde{\eta}$ are amplitudes, k is the wave number in the x-direction and ω is the angular frequency, we obtain

$$\begin{pmatrix} \omega & gk \\ Hk & \omega \end{pmatrix} \begin{pmatrix} \widetilde{\eta} \\ \widetilde{u} \end{pmatrix} = 0 \tag{44}$$

Noting that the determinant of the matrix in the left-hand side of (44) must be zero for a nontrivial solution, an equation for the frequency or the so-called dispersion relation is obtained:

$$\omega^2 - gHk^2 = 0 \tag{45}$$

The two (analytic) solutions $\omega_{\mathrm{AN}}=\pm\sqrt{gH}k$ correspond to the free-surface gravitational modes. The gravity waves can propagate along the $-x$ axis in both directions at a speed $c=\omega_{\mathrm{AN}}/k=\sqrt{gH}$, independent of the wave number k; hence, there is no dispersion of the waves.

Phase (C) and group (G) velocities are calculated using

$$C = \frac{\omega_{\mathrm{r}}}{k} \tag{46}$$

$$G = \frac{\partial \omega_{\mathrm{r}}}{\partial k} \tag{47}$$

where ω_{r} represents the real part of ω. For the analytical case, we obviously have

$$\omega_{\mathrm{r}} = \omega_{\mathrm{AN}} \tag{48}$$

As for the continuum case, the dispersion relation for the discrete scheme is found through a Fourier expansion. An ideal numerical scheme should present similar dispersion relation to the continuous mode in (45). A monotonic curve for $\omega(k)$ indicates that the numerical solution is free of spurious $2\Delta x$ oscillations [24]. Such oscillations have been observed in a number of finite difference grids [9, 25] and finite element schemes [13].

Usually, numerical models are set up [26] such that the desired wavelengths are resolved at least by 20 cells (i.e. $k\Delta x/\pi \leqslant 0.1$). Therefore, in the following, the range $k\Delta x/\pi \leqslant 0.1$ will be referred as the region of interest. The remaining region is generally concerned with the shorter waves, which do not transfer much energy. However, phase error for these waves, particularly for the $2\Delta x$ waves, could lead to oscillatory results unless they are effectively damped. Indeed, it is expected that an ideal numerical scheme preserves the amplitude of the waves, i.e. with zero damping error. This is the case when all wave numbers are transferred with the correct speed. However, zero damping error is not always desirable because most schemes have phase speed errors for high frequencies; hence, it is desirable to damp them in this case. This is particularly true for the $2\Delta x$ waves. This damping does not seriously affect the whole numerical simulation because most energy is transferred *via* waves of intermediate and small wave numbers rather than high-frequency waves.

In order to obtain the dispersion relation for the discretized equations, the solutions

$$(\eta_j, u_j) = \mathrm{Re}[(\widetilde{\eta}, \widetilde{u})e^{i(kx_j+\omega t)}] \tag{49}$$

DOI: 10.1002/fld

Table I. d^+ and d^- for the first-order upwind and the κ scheme.

	The first-order upwind scheme	The κ scheme
d^+	$\mathrm{i}\sin(k\Delta x)$	$[\sin(k\Delta x)(6-2\kappa)-\sin(2k\Delta x)(1-\kappa)]/4$
d^-	$1-\cos(k\Delta x)$	$[3-4\cos(k\Delta x)+\cos(2k\Delta x)](1-\kappa)/4$

are sought at nodes $j=1,2,3,\ldots,$ where (η_j, u_j) are the nodal unknowns that appear in the selected discrete equations and $(\widetilde{\eta}, \widetilde{u})$ are amplitudes. The x_j coordinates are expressed in terms of a distance to a reference node. Note that owing to the special discretization limitation, the wave vector space for the discrete schemes presented above is truncated at $k\Delta x=\pi$ and the maximum resolvable wavelength is $2\Delta x$. Substituting (49) in the discrete equations (22) leads to a square matrix system for the Fourier amplitudes as

$$\begin{bmatrix} \mathrm{i}\omega+d^-\sqrt{gH}/\Delta x & Hd^+/\Delta x \\ gd^+/\Delta x & \mathrm{i}\omega+d^-\sqrt{gH}/\Delta x \end{bmatrix} \begin{bmatrix} \widetilde{\eta} \\ \widetilde{u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{50}$$

where d^+ and d^- are given in Table I for the first-order upwind and κ schemes.

The dispersion relation is then obtained by setting the determinant of the matrix system to zero, and this leads to

$$\mathrm{i}\omega+\sqrt{gH}/\Delta x d^- = \pm\sqrt{gH}d^+/\Delta x \tag{51}$$

or

$$\omega=\mathrm{i}\frac{\sqrt{gH}}{\Delta x}(d^-\pm d^+) \tag{52}$$

For the first-order upwind scheme, the resulting equation for the frequency is written as [20]

$$\omega=\frac{\sqrt{gH}}{\Delta x}[\mathrm{i}(1-\cos(k\Delta x))\pm\sin(k\Delta x)] \tag{53}$$

and for the κ scheme we obtain

$$\omega=\frac{\sqrt{gH}}{4\Delta x}[\mathrm{i}(3-4\cos(k\Delta x)+\cos(2k\Delta x))(1-\kappa)\pm(\sin(k\Delta x)(6-2\kappa)-\sin(2k\Delta x)(1-\kappa))] \tag{54}$$

The two roots of ω have equal imaginary parts but their real parts are of different signs, i.e. the two waves move in different directions as in the continuous system. The imaginary part of ω deals with damping, which is equal here for both waves and this means that the symmetry of the continuous system is preserved in the numerical method. A positive imaginary part for ω leads to amplification factor less than one due to the selected form of the solution as $\mathrm{e}^{\mathrm{i}(kx_j+\omega t)}$. The imaginary part of ω for the first-order upwind scheme in (53) is always positive; hence, the results of the first-order upwind scheme remain bounded. The imaginary part of ω for the κ scheme in (54) becomes

$$\omega_{\mathrm{Im}}=\frac{\sqrt{gH}}{4\Delta x}[(3-4\cos(k\Delta x)+\cos(2k\Delta x))(1-\kappa)] \tag{55}$$

DOI: 10.1002/fld

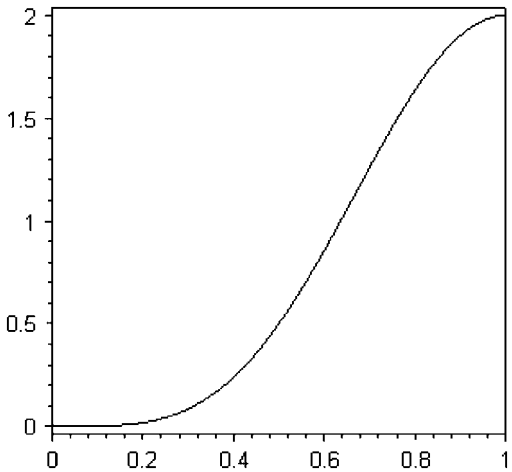

Figure 2. The term $[3-4\cos(k\Delta x)+\cos(2k\Delta x)]/4$ in the real part of the frequency in (54). The horizontal axis represents $k\Delta x/\pi$.

The term $[3-4\cos(k\Delta x)+\cos(2k\Delta x)]$ in (55) is always positive as graphed in Figure 2. It rapidly increases for intermediate waves with $k\Delta x/\pi$ roughly between 0.4 and 0.8 and reaches its highest values for the shortest resolvable waves (with $k\Delta x=\pi$) where both terms $-4\cos(k\Delta x)$ and $\cos(2k\Delta x)$ are maximized, as graphed in Figure 2. Hence, the short waves are largely damped, which is desirable as explained above to prevent numerical osculation due to phase errors at high-frequency waves. Moreover, $\lambda_I \to 0$ for long waves; thus, the long waves that transfer most of energy are damped much less than the short waves. In order to have $\omega_{\mathrm{Im}} \geqslant 0$, the term $(1-\kappa)$ must be also positive and this leads to a stability region given by $\kappa \leqslant 1$. Note that for all selected schemes, we have $\kappa \leqslant 1$, which ensures that the imaginary part of ω is positive and the numerical solution remains bounded. The particular case of $\kappa=1$, corresponding to the standard centered difference scheme, leads to no damping or neutral stability as expected. Moreover, since ω_{Im} is proportional to $(1-\kappa)$, a low level of damping is expected for the schemes with larger κ (within the stability region $\kappa \leqslant 1$). For instance, the second-order scheme with $\kappa=-1$ leads to the highest level of damping among the selected methods and the Quick scheme leads to a lower level of damping than the 3rd scheme. This is clearly confirmed in Figure 3(a) where the damping for all schemes is graphed for various wave numbers.

The results of phase speed and group velocity are presented in Figure 3(b) and (c). The term $\lambda_{\mathrm{R}}=(\sin(k\Delta x)(6-2\kappa)-\sin(2k\Delta x)(1-\kappa))/4\Delta x$ in the real part of the frequency in (54) is responsible for phase error. It could be easily verified that $\lambda_{\mathrm{R}} \to 1$ for long waves (whatever κ is); hence, all schemes perform well at the large wave number limit. On the other hand, all schemes decelerate the short waves. This is because for short waves, the terms $\sin(k\Delta x)$ and $-\sin(2k\Delta x)$ collaborate since they both have the same sign. Recall that the two coefficients $(6-2\kappa)$ and $(1-\kappa)$ are both positive in the stability region.

Of fundamental importance is the observation that waves of length $2\Delta x$ have zero frequency, i.e. $\omega=0$, due to the terms $\sin(k\Delta x)$ and $\sin(2k\Delta x)$ in the real part of the frequency equations in (53) and (54), which are both zero for the shortest resolvable waves ($k\Delta x=\pi$). Hence $\lambda_{\mathrm{R}}=0$, implying zero phase velocity for the shortest resolvable waves, i.e. they do not propagate at all. Generally,

DOI: 10.1002/fld

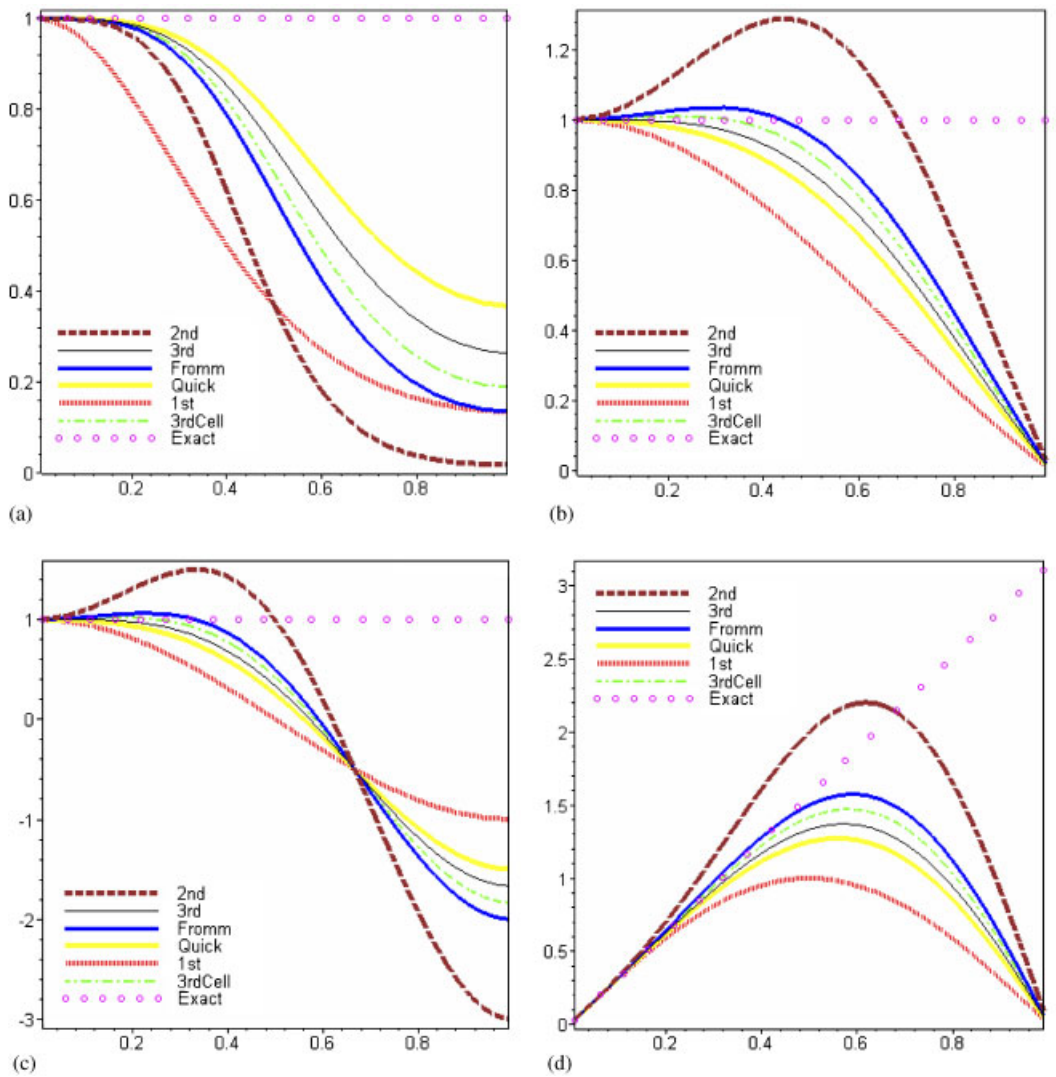

Figure 3. Various numerical properties of different spatial discretization schemes: (a) damping ($\exp(-\mathrm{Im}(w\Delta x)/\sqrt{gH})$); (b) phase speed ($C/\sqrt{gH}$); (c) group velocity ($G/\sqrt{gH}$); and (d) frequency ($\mathrm{Re}(\omega\Delta x)/\sqrt{gH}$). Horizontal coordinate represents $k\Delta x/\pi$.

this may lead to numerical oscillations; however, it is not the case here due to the damping of short waves as discussed above.

The 3rd and quick schemes decelerate all waves, while 3rdCell, Fromm and 2nd accelerate long waves and decelerate short waves. Hence, a switching algorithm may be useful to balance accelerations with deceleration (e.g. alternately using 3rd and Fromm schemes during time steps). The second-order upwind scheme leads to poor group velocity results as it is expected from its phase behavior discussed before. The Quick method leads to significant phase and group velocity

DOI: 10.1002/fld

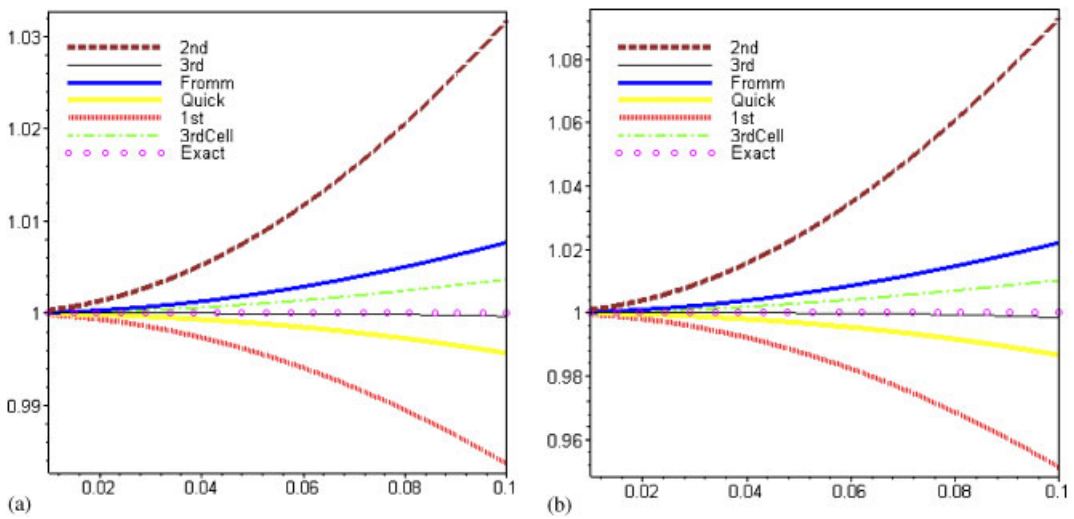

Figure 4. Phase speed $C/\sqrt{gH}$ (left) and group velocity $G/\sqrt{gH}$ (right) for various spatial discretization schemes in the region of interest. Horizontal coordinate is $k\Delta x/\pi$.

error with little damping. Hence, it may lead to a high level of numerical oscillations. Although the Quick scheme presents higher accuracy for the group velocity of short waves, this is not very important because most of energy (which propagates with group velocity) is due to long waves. The 3rdCell and Fromm schemes give very close phase and group velocities. However, the 3rdCell has better phase and group velocities than the Fromm scheme in the region of interest, as shown in Figure 4, where the results of the former method are closer to the analytical ones. Moreover, in Figure 3(a), the 3rdCell appears less diffusive than the Fromm scheme. Hence, by considering phase and group velocities and damping, the 3rdCell may be preferred to the Fromm scheme at this stage of the argument.

Note that in Figure 4, the 3rd scheme gives the best results for the phase speed and group velocity in the region of interest (long waves); it nearly matches the analytical one. Moreover, it has less numerical damping than the 3rdCell. Hence, the 3rd scheme is superior to the 3rdCell for long waves. However, the 3rdCell scheme presents less phase error and more damping for intermediate and short waves; hence, it leads to less numerical oscillations than the 3rd scheme. Moreover, as will be shown in Section 6, when the effect of time-discretization scheme is also considered, such an accurate phase behavior of the 3rd scheme remains the same only for small CFL numbers.

The numerical results for $\mathrm{Re}(\omega)$ are plotted in Figure 3(d). All selected numerical methods lead to a folded dispersion plot. This is due to the presence of the terms $\sin(k\Delta x)$ and $\sin(2k\Delta x)$ in the real part of the frequency equations (53) and (54) which are not monotonic. We remind the reader that a folded dispersion plot implies aliasing problems since two separate wave vectors can contribute to the same frequency [24] and thus, all the above-mentioned schemes may encounter aliasing problems.

For most numerical schemes, the cutoff frequency corresponds to zero group velocity and maximum frequency i.e. the folding point of the frequency plot. After the cutoff frequency, the

DOI: 10.1002/fld

group velocity becomes negative. Since energy propagates with the group velocity, this means that energy moves in a wrong direction after the cutoff frequency, which is not desirable. In Figure 3(c) and (d), the cutoff frequency ranges from $\pi/2$ to $\pi-\cos^{-1}(-\frac{1}{2}+\frac{\sqrt{3}}{2})$ or approximately 0.6193 π and those values correspond to the first- and second-order upwind schemes, respectively. The cutoff frequency of the cell-based third-order upwind scheme is larger than the classical (point-wise) third-order upwind method. The former gives here intermediate results between the 3rd and Fromm schemes. The cutoff frequency is larger for the 2nd scheme than the other schemes. However, the 2nd scheme largely overestimates the frequency of long waves; hence, the Fromm scheme may be preferred in terms of dispersion relation in Figure 3(d). However, it should be stressed that the temporal integration scheme may have a significant impact on the dispersion relation (see, e.g. [27]), this will be examined in the next section.

We now consider the influence of the temporal discretization.

5. TEMPORAL DISCRETIZATION

At a given time step $\Delta t=t^{n+1}-t^n$ we use a general time discretization of (22) of the form

$$u_j^{n+1}+\frac{\Delta t}{\Delta x}\sigma(F_{j+1/2}^{n+1}-F_{j-1/2}^{n+1})=u_j^n+\frac{\Delta t}{\Delta x}(1-\sigma)(F_{j+1/2}^n-F_{j-1/2}^n) \tag{56}$$

where σ is a real parameter such that $0\leqslant\sigma\leqslant 1$. Observe that the standard choices $\sigma=0,\frac{1}{2},1$ yield the respective forward Euler, trapezoidal Crank–Nicolson and backward Euler-type schemes. Other choices such as Runge–Kutta schemes are possible, but they would make the present stability/dispersion analysis much less tractable.

6. FOURIER ANALYSIS FOR THE FULLY DISCRETIZED SCHEMES

A Fourier analysis is now conducted for both spatial and temporal schemes. By substituting periodic solutions of the form $(\eta_j^n,u_j^n)=\mathrm{Re}[(\widetilde{\eta},\widetilde{u})\mathrm{e}^{\mathrm{i}(kx_j+\omega t^n)}]$ into (56), we obtain

$$\begin{bmatrix} E-1+ad^-(\sigma E+1-\sigma) & Hbd^+(\sigma E+1-\sigma) \\ gbd^+(\sigma E+1-\sigma) & E-1+ad^-(\sigma E+1-\sigma) \end{bmatrix}\begin{bmatrix}\widetilde{\eta}\\ \widetilde{u}\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix} \tag{57}$$

with $a=\sqrt{gH}\Delta t/\Delta x$ (the CFL number) and $b=\Delta t/\Delta x$.

An equation (also called dispersion relation) for the propagation factor, defined by $E=\mathrm{e}^{\mathrm{i}\omega\Delta t}$, is then obtained by setting the determinant of the matrix in the left-hand side of (57) equal to zero for a nontrivial solution. For both the first-order and the κ schemes, the dispersion relation leads to

$$E=\frac{1+\mathrm{i}\omega\Delta t(1-\sigma)}{1-\mathrm{i}\omega\Delta t\sigma} \tag{58}$$

where ω is the frequency corresponding to the semi-discrete first-order and κ schemes in (53) and (54), respectively. It is straightforward to show that the first-order upwind scheme remains stable for $\sqrt{gH}\Delta t/\Delta x\leqslant 1$, even with a fully explicit time-discretization scheme ($\sigma=0$). However,

DOI: 10.1002/fld

all other schemes may lead to unstable results ($|E|>1$) for $\sigma=0$. Indeed, by setting $\sigma=0$ in (58), we obtain $E=1+\mathrm{i}\omega\Delta t$ or

$$E=1+\frac{a}{4}[-(3-4\cos(k\Delta x)+\cos(2k\Delta x))(1-\kappa)\pm\mathrm{i}(\sin(k\Delta x)(6-2\kappa)-\sin(2k\Delta x)(1-\kappa))] \tag{59}$$

The amplification factor $|E|$ is given in Figure 5, which shows unbounded growth even for small CFL numbers. This could be also verified by expanding (59) around $k\Delta x=0$ as

$$E=1+\mathrm{i}\frac{\sqrt{gH}\Delta t}{\Delta x}\left[-k\Delta x+\left(-\frac{1}{12}+\frac{\kappa}{4}\right)(k\Delta x)^3\right]+O((k\Delta x)^4) \tag{60}$$

which leads to $|E|>1$ for all CFL numbers at long waves limit, as confirmed in Figure 5(b) for a small CFL number. On the other hand, for σ ranging from $\frac{1}{2}$ to 1, we observe a significant damping for the κ scheme (results not shown). Therefore, we use $\sigma=\frac{1}{2}$ for the κ scheme. However, for the first-order upwind scheme, a fully explicit method ($\sigma=0$) is employed in the following because the use of a Crank–Nicolson scheme would bring an excessive damping. The amplification factor ($|E|$), the group velocity, the frequency (real part) and the phase speed ratio for various schemes are plotted in Figures 6–8 for CFL$=0.1, 0.5$ and 0.9, respectively, and will be discussed in the following. Note that the behaviors in those plots could be also justified by examining the corresponding formulas as performed for the semi-discrete case or by expansions as in (60). However, such a justification is not necessary for the purpose of this paper and only the behavior of various schemes is considered here.

Note that the results for small CFL number (CFL$=0.1$) are close to those of the semi-discretized method. This is expected because the semi-discretized scheme could be viewed as the limit of the fully discretized method for small time steps.

On the basis of Fourier analysis results, it is concluded that the numerical performance of upwind schemes highly depends on the CFL number. The first-order upwind scheme is somehow better than higher order schemes for CFL numbers close to one. This is expected because for CFL$=1$, the first-order upwind scheme leads to point-by-point advection of the Riemann invariants, which is the exact solution. However, in actual simulations the CFL is not constant over entire system (due to the topographic changes, etc.); hence, the overall performance of schemes should be considered.

For the κ scheme, the damping is smaller for short time steps (small CFL numbers) than long time steps. However, it should be noted that, for a given total time of simulation, more time steps are needed when CFL is small. Hence, the total damping after a given time of simulations may be even larger when small CFL numbers are employed.

For CFL<0.5, e.g. in Figure 6(d) with CFL$=0.1$, all the curves are folded and waves of wavelength $2\Delta x$ (i.e. $k\Delta x=\pi$) do not propagate. For CFL$\geqslant 0.5$, as shown in Figures 7(d) and 8(d), the curves corresponding to the first- and second-order methods are no longer folded and they are close to the monotonic continuous solution. In particular, for CFL$=0.5$, the first-order upwind scheme exactly coincides with the analytical frequency and phase speed. This could be also easily verified from the corresponding discrete dispersion relation. However, both 1st and 2nd schemes have other serious drawbacks. Indeed, in the region of interest ($k\Delta x<\pi/10$), the first-order upwind scheme leads to an excessive damping and the second-order upwind scheme shows significant phase and group velocity errors; the phase speed and group velocities of the 2nd scheme at intermediate and high CFL numbers are non-monotonic with large deviations, and this could

DOI: 10.1002/fld

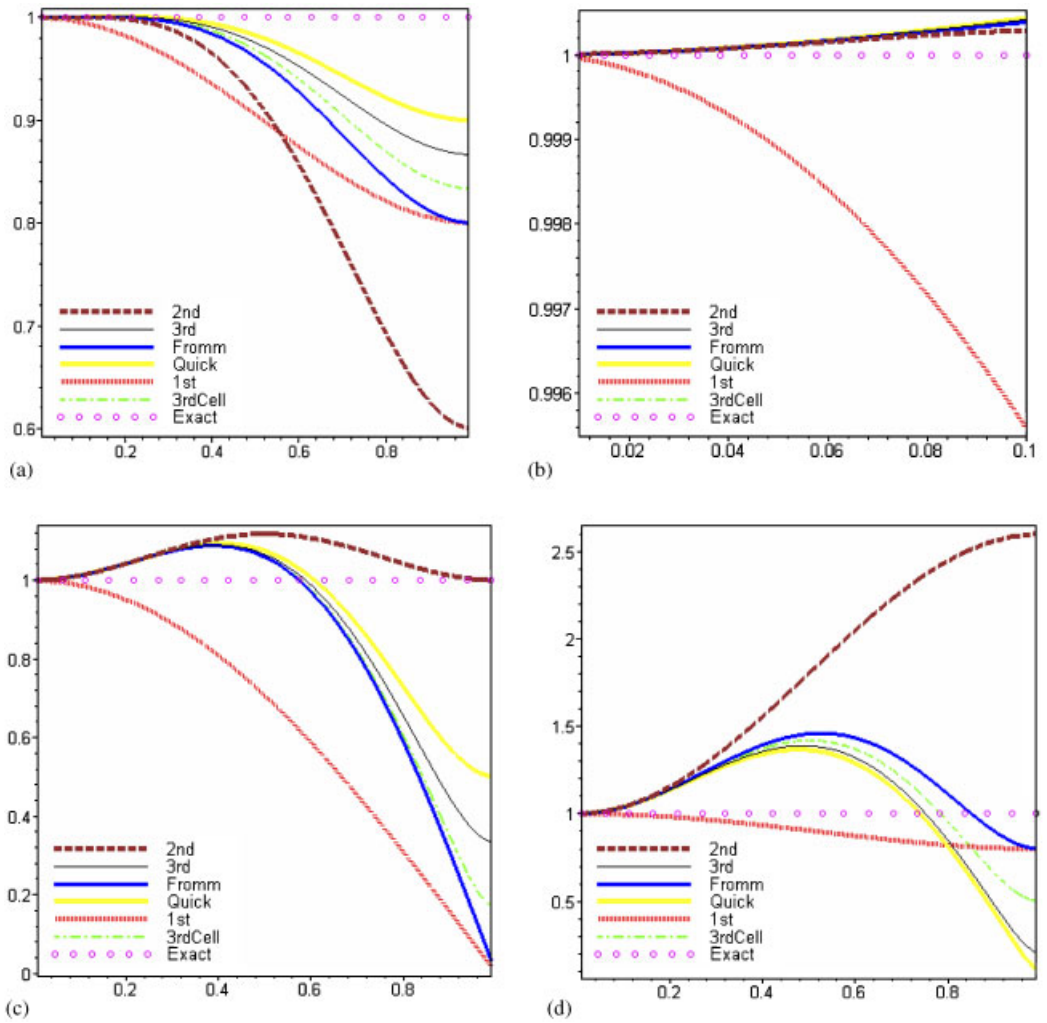

Figure 5. Amplification factor $|E|$ for $\sigma = 0$ and various CFL numbers. Horizontal coordinate represents $k\Delta x/\pi$: (a) $|E|$ for CFL $= 0.1$; (b) $|E|$ for CFL $= 0.1$, with $k\Delta x/\pi \leqslant 0.1$; (c) $|E|$ for CFL $= 0.5$; and (d) $|E|$ for CFL $= 0.9$.

lead to oscillatory results due to the mixed acceleration and deceleration of wave components. Therefore, the first-order and the second-order upwind schemes will no longer be considered in the following.

The remaining schemes (i.e. 3rdCell, 3rd, Fromm and Quick) are now studied in more details, as the HOA (high-order accurate) schemes. The cutoff frequency is progressively increasing as the CFL number increases as shown in Figures 6(d), 7(d) and 8(d). In this respect, the Fromm scheme exhibits the best behavior and the Quick one gives the worst results. Similar conclusions have already been drawn for the semi-discrete case.

DOI: 10.1002/fld

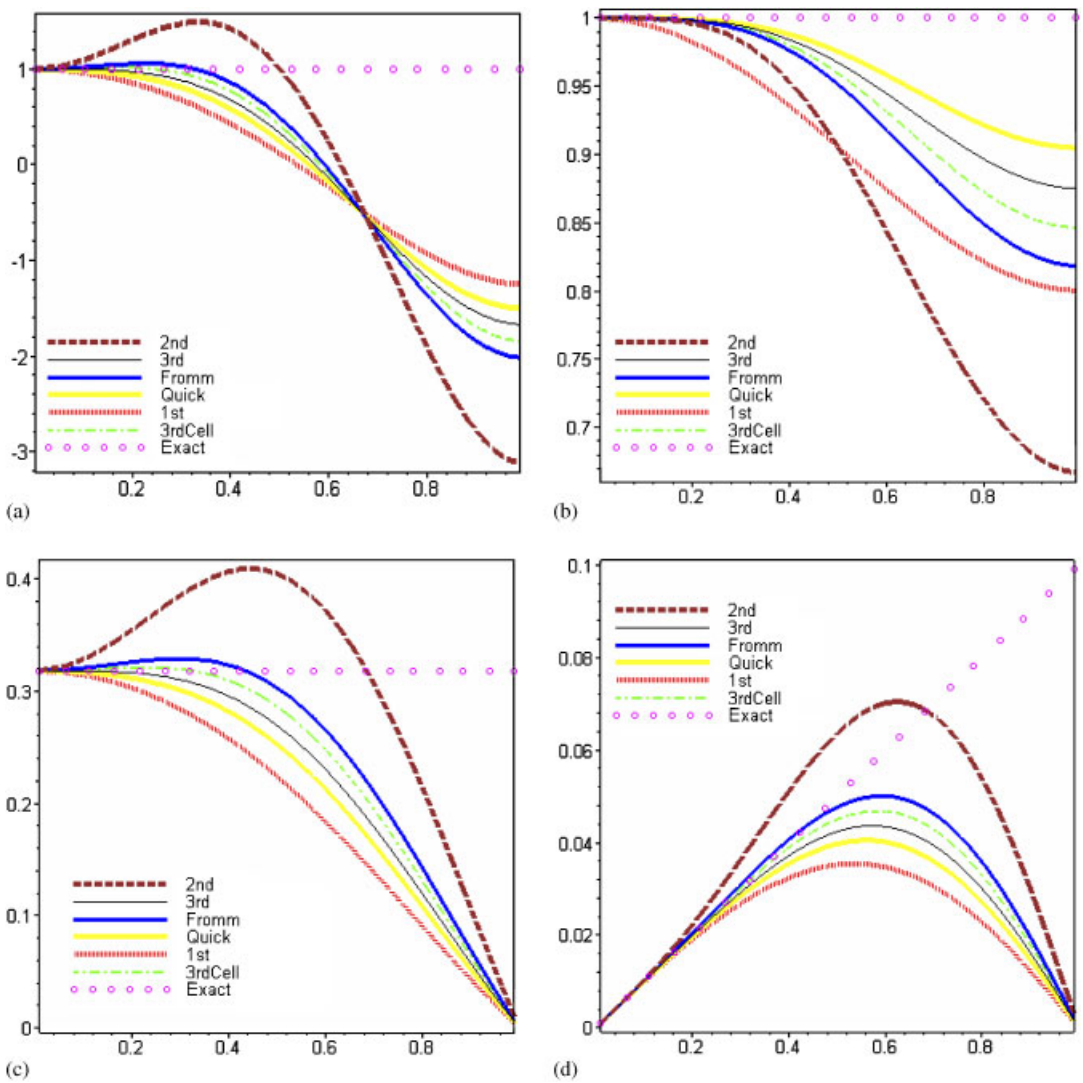

Figure 6. Various numerical properties of the fully discretization schemes for CFL = 0.1: (a) group velocity ($G/\sqrt{gH}$ for CFL = 0.1); (b) damping ($|E|$ for CFL = 0.1); (c) phase speed ($C/(\pi\sqrt{gH})$ for CFL = 0.1); and (d) frequency for CFL = 0.1 ($\omega\Delta t/\pi$ for CFL = 0.1). Horizontal coordinate represents $k\Delta x/\pi$.

As shown in Figures 6–8, the performance of the HOA methods is improved regarding the phase speed and the frequency for all wave numbers as the CFL number increases. As an example, if we expand the phase velocity of the Fromm scheme around $k\Delta x = 0$, for CFL = 0.5 we obtain

$$\frac{C}{\pi\sqrt{gH}} = \frac{1}{\pi} + \frac{1}{16\pi}(k\Delta x)^2 + O((k\Delta x)^4) \tag{61}$$

DOI: 10.1002/fld

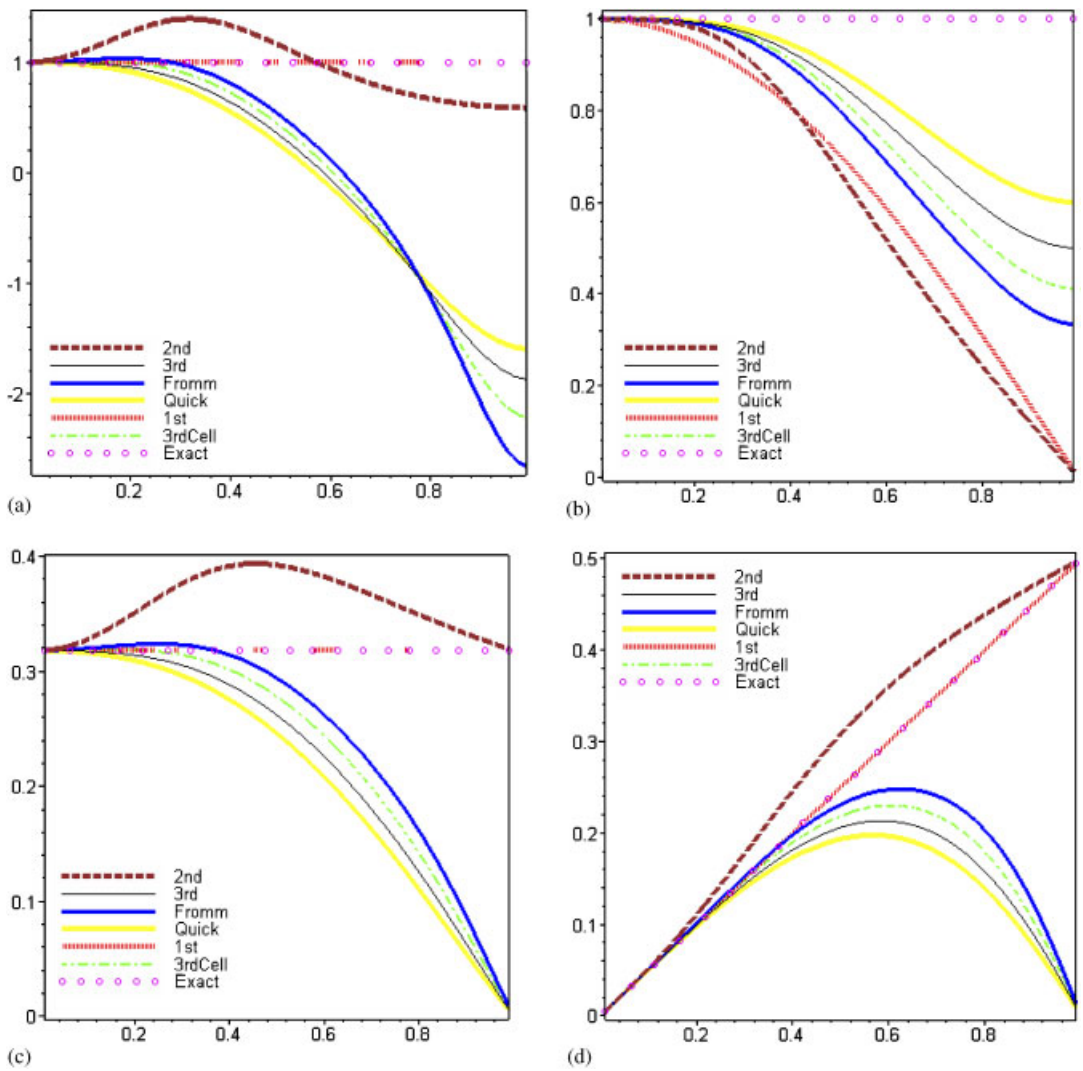

Figure 7. Various numerical properties of the fully discretization schemes for CFL = 0.5: (a) group velocity ($G/\sqrt{gH}$ for CFL = 0.5); (b) damping ($|E|$ for CFL = 0.5); (c) phase speed ($C/(\pi\sqrt{gH})$ for CFL = 0.5); and (d) frequency for CFL = 0.5 ($\omega\Delta t/\pi$ for CFL = 0.5). Horizontal coordinate represents $k\Delta x/\pi$.

while the phase speed for CFL = 0.9 leads to

$$\frac{C}{\pi\sqrt{gH}} = \frac{1}{\pi} + \frac{19}{1200\pi}(k\Delta x)^2 + O((k\Delta x)^4) \tag{62}$$

Obviously, the second-order term $(1/16\pi)(k\Delta x)^2$ in (61), which is positive, is responsible for acceleration of long waves at CFL = 0.5, while it becomes smaller $(19/1200\pi)(k\Delta x)^2$ at CFL = 0.9 in (62).

DOI: 10.1002/fld

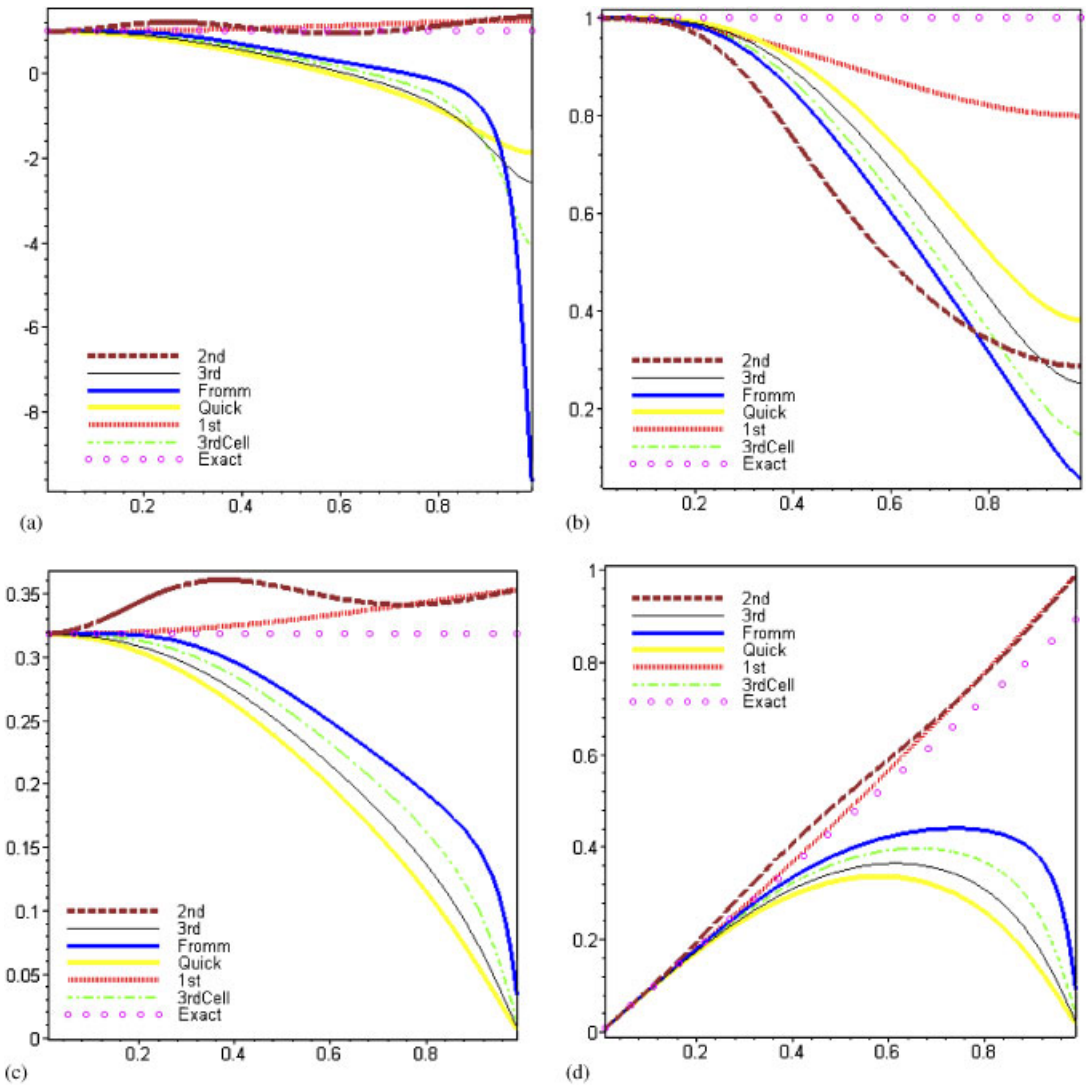

Figure 8. Various numerical properties of the fully discretization schemes for CFL = 0.9: (a) group velocity ($G/\sqrt{gH}$ for CFL = 0.9); (b) damping ($|E|$ for CFL = 0.9); (c) phase speed ($C/(\pi\sqrt{gH})$ for CFL = 0.9); and (d) frequency for CFL = 0.9 ($\omega\Delta t/\pi$ for CFL = 0.9). Horizontal coordinate represents $k\Delta x/\pi$.

The phase error of the Quick scheme is larger than the other HOA schemes. Moreover, it exhibits less damping of short waves than the others. Hence, the Quick scheme is expected to present the most oscillatory results compared with the other HOA schemes. Note that the numerical oscillations are mostly due to short waves. Similarly, the 3rd scheme is more oscillatory than 3rdCell and Fromm schemes. The Fromm and the 3rdCell schemes have the most accurate phase speed for high CFL numbers, as shown in Figure 8(c). Although they lead to zero phase speed for $2\Delta x$ waves, they also induce a significant damping for those waves that largely prevents numerical

DOI: 10.1002/fld

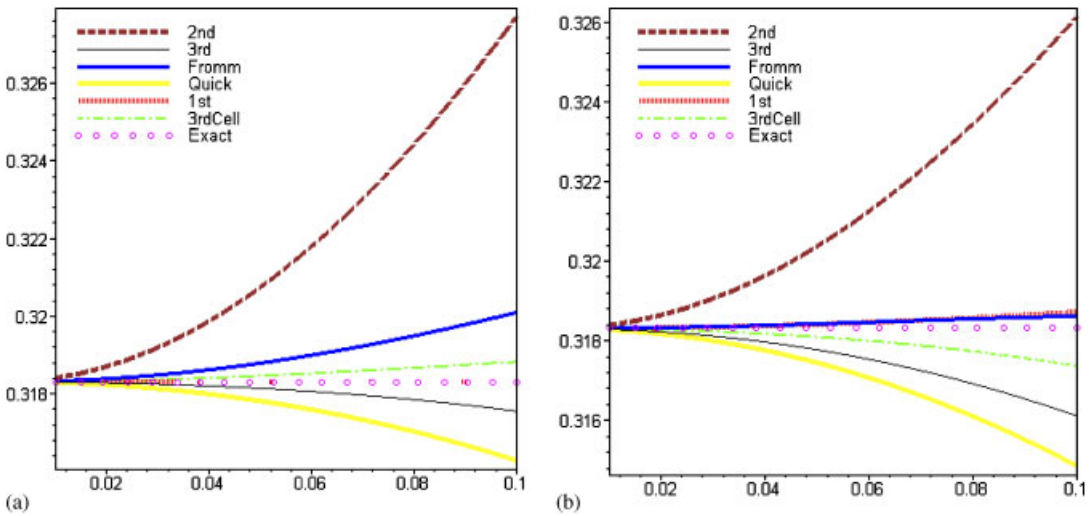

Figure 9. Phase speed $C/(\pi\sqrt{gH})$ at CFL$=$0.5 (left) and at CFL$=$0.9 (right) for long waves. Horizontal coordinate represents $k\Delta x/\pi$.

oscillations. In particular, the 3rdCell scheme gives better phase speed results than the Fromm method in the region of interest, whatever the CFL is.

Recall that the 3rd scheme gives the most accurate phase speed in the region of interest for the semi-discrete case. However, for the fully discrete case, as the CFL number increases, the phase speed of the 3rdCell scheme becomes better than the 3rd scheme at the region of interest as shown in Figure 9. In particular, for large CFL number in Figures 8(c) and 9, the 3rdCell scheme gives better results than the 3rd scheme both for the short and long waves. Hence, when the overall performance for phase speed is considered, the 3rdcell scheme may be preferred to the 3rd and Fromm schemes at large CFL numbers.

When considering the group velocity and the damping errors, the Quick scheme exhibits the best results while the Fromm scheme has the largest errors, as shown in Figure 8(a) and (b). In particular, the Fromm scheme shows a significant group velocity error for the $2\Delta x$ waves at high CFL numbers. This is because the phase speed curve presents a sharp gradient for the $2\Delta x$ waves. However, the Fromm scheme also shows the largest damping for those waves. Thus, short waves do not lead to serious problems since energy transfer (with the group velocity) is mostly performed by long waves rather than short ones. Again, for all HOA schemes, the group velocity accuracy is improved as the CFL increases in the region of interest. The 3rdCell and 3rd schemes give intermediate results between those of the Quick and Fromm schemes as shown in Figures 6–8. This was largely expected due to the ordering of κ values in (32). In particular, the 3rdCell method gives better group velocity results than the Fromm scheme.

It should be emphasized that none of the selected schemes could be considered as the 'best' one since this choice largely depends on which quantity (frequency, damping, phase or group velocity) is more important regarding the physical problem at hand. Again, a switching algorithm (e.g. alternately using 3rd and 3rdCell schemes during time steps) is expected to be useful.

DOI: 10.1002/fld

After having discussed the propagation of pure gravity waves, we now consider the Rossby modes by including the Coriolis term in the model equations. In the following, we are particularly interested in the dissipative effects for Rossby waves in ocean and atmosphere modeling applications.

7. MODEL PROBLEM AND MODAL BEHAVIOR FOR ROSSBY MODES

In this section we consider stationary Rossby modes, i.e. we assume that the Coriolis parameter is constant. In Section 8.3, we will consider the more general case where the model equations are modified to include a β term to permit the propagation of the slow Rossby waves.

When the Coriolis term is taken into account, the 1D, inviscid, linearized form of the shallow water equations may be expressed in Cartesian coordinates [24] as

$$\eta_t + Hu_x = 0 \tag{63}$$

$$u_t - fv + g\eta_x = 0 \tag{64}$$

$$v_t + fu = 0 \tag{65}$$

where the Coriolis parameter f is assumed constant. Periodic solutions of the form $u = \widetilde{u}e^{i(kx+\omega t)}$, $v = \widetilde{v}e^{i(kx+\omega t)}$ and $\eta = \widetilde{\eta}e^{i(kx+\omega t)}$ are sought again and we obtain

$$\begin{pmatrix} i\omega & iHk & 0 \\ igk & i\omega & -f \\ 0 & f & i\omega \end{pmatrix} \begin{pmatrix} \widetilde{\eta} \\ \widetilde{u} \\ \widetilde{v} \end{pmatrix} = 0 \tag{66}$$

Letting the determinant of the matrix in the left-hand side of (66) equal to zero to obtain non-trivial solution leads to

$$\omega(\omega^2 - f^2 - gHk^2) = 0 \tag{67}$$

The two first roots $\omega = \pm\sqrt{f^2 + gHk^2}$ correspond to inertia–gravity modes and the third one, $\omega = 0$, is the geostrophic mode (or stationary Rossby mode) and it would correspond to the slow Rossby mode on a β-plane.

The finite volume method and the temporal scheme (presented in Sections 4 and 5) are again employed to obtain the discrete form of (63)–(65). The discretization of (63) exactly coincides with the pure gravity wave case. Equation (65) only contains a source term (fu) and it is discretized as

$$v_j^{n+1} = v_j^n - \gamma \Delta t f u_j^{n+1} - (1-\gamma)\Delta t f u_j^n \tag{68}$$

with $0 \leqslant \gamma \leqslant 1$. In this study, on the basis of numerical experiments, a fully implicit case for the Coriolis term ($\gamma = 1$) has been chosen. The case $\gamma = \frac{1}{2}$ leads to similar damping while producing more oscillations. Finally, in (64) all fluxes are basically calculated as for pure gravity wave case, but now with the Coriolis term $(-fv)$ treated as the source term (fu) in (68). Again, the first-order upwind and the κ schemes are used in the following.

DOI: 10.1002/fld

A Fourier analysis is conducted at the discrete level by substituting periodic solutions of the form $(\eta_j^n, u_j^n, v_j^n) = \mathrm{Re}[(\tilde{\eta}, \tilde{u}, \tilde{v})\mathrm{e}^{\mathrm{i}(kx_j+\omega t^n)}]$ into discretized form of (63)–(65). This leads to a square 3×3 matrix system for the wave amplitudes as

$$\begin{bmatrix} E-1+ad^-(\sigma E+1-\sigma) & Hbd^+(\sigma E+1-\sigma) & 0 \\ gbd^+(\sigma E+1-\sigma) & E-1+ad^-(\sigma E+1-\sigma) & -f\Delta t(\gamma E+1-\gamma) \\ 0 & f\Delta t(\gamma E+1-\gamma) & E-1 \end{bmatrix} \begin{bmatrix} \tilde{\eta} \\ \tilde{u} \\ \tilde{v} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{69}$$

By setting the determinant of the matrix in the right-hand side of (69), a cubic equation for the propagation factor E is obtained as follows:

$$c_3E^3+c_2E^2+c_1E+c_0=0 \tag{70}$$

where the coefficients c_0, c_1, c_2 and c_3 for $\sigma=\frac{1}{2}$ are given by

$$c_0=-1+ad^- -\frac{\Delta t^2 f^2}{4}-\frac{a^2d^{-2}}{4}+\frac{ad^-\Delta t^2 f^2}{8}+\frac{d^{+2}a^2}{4} \tag{71}$$

$$c_1=3-ad^- -\frac{\Delta t^2 f^2}{4}-\frac{a^2d^{-2}}{4}+\frac{3ad^-\Delta t^2 f^2}{8}+\frac{d^{+2}a^2}{4} \tag{72}$$

$$c_2=-3-ad^- +\frac{\Delta t^2 f^2}{4}+\frac{a^2d^{-2}}{4}+\frac{3ad^-\Delta t^2 f^2}{8}-\frac{d^{+2}a^2}{4} \tag{73}$$

$$c_3=1+ad^- +\frac{\Delta t^2 f^2}{4}+\frac{a^2d^{-2}}{4}+\frac{ad^-\Delta t^2 f^2}{8}-\frac{d^{+2}a^2}{4} \tag{74}$$

Two roots of (70) are complex conjugate and correspond to inertia–gravity modes. For these modes, $|E|$ is graphed in Figure 10(a) with CFL$=0.9$ and a typical case $f\Delta t=0.1$. The results are

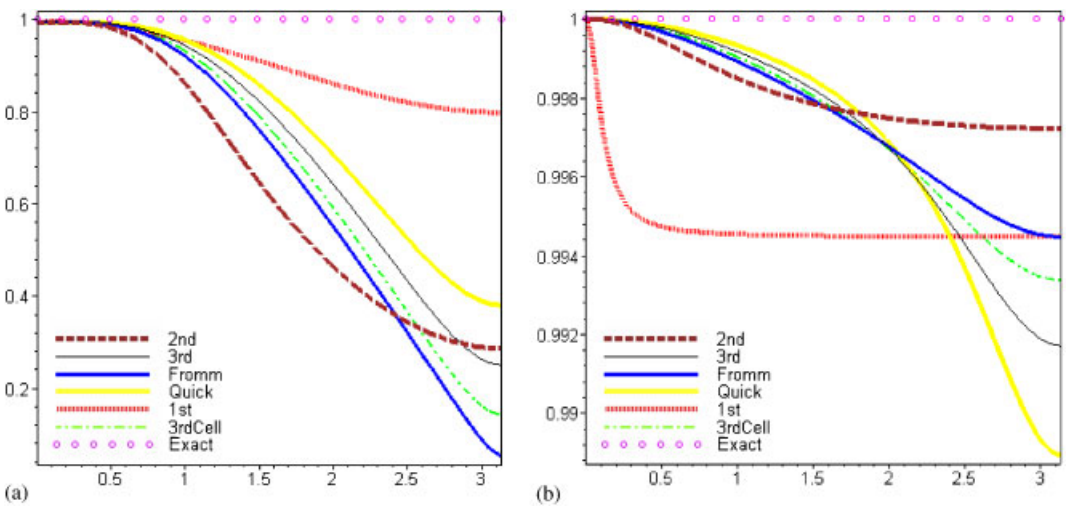

Figure 10. Damping error $|E|$ for various schemes for inertial gravity waves (a) and the Rossby waves (b). Horizontal coordinate represents $k\Delta x/\pi$.

DOI: 10.1002/fld

basically similar to those of the pure gravity wave case (Figure 8(b)), with a shift depending on $f\Delta t$. The third root of (70) is real and corresponds to the stationary Rossby wave. $|E|$ is graphed for this mode with CFL$=0.9$ and $f\Delta t=0.1$ in Figure 10(b) for the first-order and κ schemes. It should be mentioned that, for Rossby waves, the κ scheme is less sensitive to the CFL number than the first-order upwind scheme. Therefore, only the results for CFL$=0.9$ are shown in Figure 10(b). The Quick scheme shows less damping than the other methods in the region of interest, while the first- and second-order upwind schemes lead to the highest level of damping. However, all schemes examined in Figure 10(b) exhibit a high level of damping for the Rossby mode. The observed level of damping should occur at each time step in a numerical simulation and would lead to an excessive damping for long-term simulations. Indeed, a high-order spatial accuracy is usually desirable for the treatment of the slow Rossby modes and has proven practical and beneficial for atmospheric and oceanic prediction models.

8. NUMERICAL RESULTS

The Fourier analysis results obtained in the previous section for the Rossby modes are now computationally verified by considering three numerical tests. The first one is concerned with a single wave solution, while the second test corresponds to a Gaussian initial condition. In the third test, which generalizes the first one, we include a β term in the model equations; hence, we are able to observe the propagation of slow Rossby modes. In all cases periodic boundary conditions are used and (63)–(65) are numerically solved to examine the level of damping induced by the first-order and κ schemes.

8.1. Stationary wave solution

The wave solution of (63)–(65) corresponding to the stationary Rossby wave is

$$\eta=\widetilde{\eta}\cos(kx) \tag{75}$$

$$u=0 \tag{76}$$

$$v=-\frac{gk}{f}\widetilde{\eta}\sin(kx) \tag{77}$$

Equations (75)–(77) are used here as the initial condition with $\widetilde{\eta}=0.05\,\mathrm{m}$, CFL$=0.9$, $f\Delta t=0.1$ and $k\Delta x/\pi=0.1$. The latter parameter corresponds to the region of interest.

The level of damping (per time step) that occurs in the computation of the Rossby waves is examined in the case of the Fourier analysis and the numerical test performed here. As expected, a very good matching between the damping for the velocity and surface elevation was found. In Table II, the results corresponding to the damping of both fields are presented and a close agreement between the numerical and analytical results is observed. The (explicit) first-order upwind scheme leads to unstable results, although the amplification factors for both gravity and Rossby modes are less than one. Indeed, as also observed in Foreman [26, 28], instability may occur even if $|E|_{\max}<1$ (where $|E|_{\max}$ is the maximum amplification factor for all modes).

DOI: 10.1002/fld

Table II. Damping ($|E|$) per time step of the Rossby waves for different schemes obtained from Fourier analysis and numerical test corresponding to CFL$=0.9$ and $f\Delta t=0.1$.

Scheme	Fourier analysis	Numerical test
1st	0.99509	Unstable
2nd	0.99977	0.99975
3rd	0.99992	0.99991
Fromm	0.99988	0.99987
Quick	0.99994	0.99993
3rd Cell	0.99990	0.99989

8.2. A Gaussian water surface distribution

The following test is performed to give a physical idea about the level of damping of the stationary Rossby waves in a more realistic case. A Gaussian distribution is prescribed at initial time and the initial velocity field is taken to be in geostrophic equilibrium, i.e. the Coriolis and the pressure gradient terms are balanced and we have

$$\eta(x,0)=\mu e^{-\nu x^2} \tag{78}$$

$$u(x,0)=0 \tag{79}$$

$$v(x,0)=gf\eta_x(x,0) \tag{80}$$

The mesh length parameter $\Delta x=30\,\text{km}$ is taken and the domain is chosen sufficiently long to prevent the Gaussian from approaching the boundaries. The Coriolis parameter is evaluated at 25°N ($f=6.163465\text{e}-5\,\text{s}^{-1}$) with $H=1.63\,\text{m}$ and the radius of deformation is thus $\lambda=\sqrt{gH}/f\approx 65\,\text{km}$. The parameters μ and ν are chosen such that the e-folding radius of the initial Gaussian is resolved by $3\Delta x$ and the initial maximum azimuthal velocity is $1\,\text{m/s}$. A simple scaling can provide more realistic atmospheric parameters without substantially modifying the results.

Numerical results corresponding to the damping of the surface elevation are shown for various schemes in Table III(a) and (b) with $\Delta t=15\,\text{min}$ ($f\Delta t\approx 0.055$) and $\Delta t=30\,\text{min}$ ($f\Delta t\approx 0.111$), respectively, up to 30 days of simulation. It is observed that the Rossby modes are highly damped and the relative level of damping for the various schemes is in accordance with the results of the Fourier analysis shown in Figure 10(b). The level of damping in Table III(a) and (b) is almost identical for the finite volume schemes, while the case $\Delta t=30\,\text{min}$ is marginally more dissipative. Although the Quick scheme has the smallest level of damping, the results are still too damped for most practical application. For the two-dimensional case, a higher level of damping is expected.

Finally, in finite volume methods, slope limiters are widely used to reduce the oscillations resulting from phase errors. In order to examine the consequence of the slope limiter on the accuracy of the Rossby mode approximation, the third-order upwind scheme using a typical slope limiter [29] is employed here. This scheme performs very well in modeling shock waves [29]. The interface values (30) are now rewritten on the form

$$\mathbf{U}^{\text{L}}_{j+1/2}=\mathbf{U}_j+\tfrac{1}{4}((1-\kappa s)(\mathbf{U}_j-\mathbf{U}_{j-1})+(1+\kappa s)(\mathbf{U}_{j+1}-\mathbf{U}_j)) \tag{81}$$

DOI: 10.1002/fld

Table III. Damping of the Rossby waves for different schemes for Gaussian initial conditions ($\Delta t = 15$ min) (a) and $\Delta t = 30$ min (b).

Scheme	10 days	20 days	30 days
(a)			
2nd	0.80855	0.72197	0.66763
Fromm	0.87819	0.80672	0.75708
3rd Cell	0.89370	0.82711	0.77956
3rd	0.91097	0.85061	0.80605
Quick	0.93041	0.87824	0.83806
1st	0.28529	0.19982	0.16258
3rd Monotone	0.69892	0.57115	0.50429
Least-squares	0.99297	0.98625	0.97974
Galerkin	1.00000	1.00000	1.00000
(b)			
2nd	0.80854	0.72197	0.66763
Fromm	0.87817	0.80672	0.75708
3rd Cell	0.89368	0.82711	0.77956
3rd	0.91095	0.85061	0.80605
Quick	0.93039	0.87823	0.83806
1st	0.28524	0.19979	0.16256
3rd Monotone	0.69892	0.57116	0.50427
Least-squares	0.98623	0.97343	0.96142
Galerkin	1.00000	1.00000	1.00000

and the slope limiter is

$$s = \frac{2\Delta_- \Delta_+ + \varepsilon}{\Delta_-^2 + \Delta_+^2 + \varepsilon} \tag{82}$$

where ε is a very small number to avoid division by zero in the regions of zero slope and the quantities Δ_- and Δ_+ are computed for each component of $\mathbf{U}$, e.g. η, as

$$\Delta_- = \eta_j - \eta_{j-1} \tag{83}$$

$$\Delta_+ = \eta_{j+1} - \eta_j \tag{84}$$

The interface value $\mathbf{U}^{\mathrm{R}}_{j+1/2}$ is computed in a similar manner.

The level of damping of the third-order scheme, using the above slope limiter approach, for the stationary Rossby modes, is shown in Table III(a) and (b) (3rdMonotone). As observed, the level of damping of the 3rdMonotone scheme is significantly higher than the 3rd one (with $\sigma = \frac{1}{2}$). This is also expected for most other upwind finite volume schemes as a consequence of slope limiting. In Table III(a) and (b), the previous results are compared with those of the least-squares and the Galerkin methods using linear approximations, obtained with $\sigma = \gamma = \frac{1}{2}$ [30]. By taking into account the dissipative nature of the least-squares scheme, the upwind finite volume schemes appear to give over damped results in Table III(a) and (b). For real applications, Rossby modes are computed through nonlinear models. However, nonlinear finite volume methods usually require the use of slope limiters. Therefore, for real applications with nonlinear models, upwind finite volume schemes should excessively damp the Rossby waves.

DOI: 10.1002/fld

8.3. A 1D Rossby wave

Here, we generalize the test of Section 8.1 by considering a modification to the model equations (63)–(65), which has already been used by Vreugdenhil [21]. This enables us to examine the propagation of one dimensional analogous of slow moving Rossby waves. Following Vreugdenhil [21] the model equations are rewritten as

$$\eta_t + Hu_x = \frac{\beta H}{f} v \tag{85}$$

$$u_t - fv + g\eta_x = 0 \tag{86}$$

$$v_t + fu = 0 \tag{87}$$

where $\beta = \mathrm{d}f/\mathrm{d}y$. Assuming traveling wave solutions of the form $u = \mathrm{Re}[\widetilde{u}\mathrm{e}^{\mathrm{i}(kx+\omega t)}]$, $v = \mathrm{Re}[\widetilde{v}\mathrm{e}^{\mathrm{i}(kx+\omega t)}]$ and $\eta = Re[\widetilde{\eta}\mathrm{e}^{\mathrm{i}(kx+\omega t)}]$ we obtain

$$\begin{pmatrix} \mathrm{i}\omega & \mathrm{i}Hk & -\dfrac{\beta H}{f} \\ \mathrm{i}gk & \mathrm{i}\omega & -f \\ 0 & f & \mathrm{i}\omega \end{pmatrix} \begin{pmatrix} \widetilde{\eta} \\ \widetilde{u} \\ \widetilde{v} \end{pmatrix} = 0 \tag{88}$$

which leads to the following dispersion relation:

$$\omega^3 - f^2\omega - gHk^2\omega - \beta gHk = 0 \tag{89}$$

For slow motions (corresponding to small values of ω), we can neglect the first term in (89) and we obtain

$$\omega = -\frac{\beta gHk}{f^2 + gHk^2} \tag{90}$$

or

$$\omega = -\frac{\beta k}{\dfrac{1}{\lambda^2} + k^2} \tag{91}$$

where $\lambda = \sqrt{gH}/f$ is the Rossby radius of deformation. Equation (89) is indeed the 1D equivalent of the Rossby wave frequency. The amplitudes are then approximately

$$\widetilde{\eta} = \frac{\omega^2 - f^2}{k\omega g} \tag{92}$$

$$\widetilde{u} = 1 \tag{93}$$

$$\widetilde{v} = -\mathrm{i}\frac{f}{\omega} \tag{94}$$

DOI: 10.1002/fld

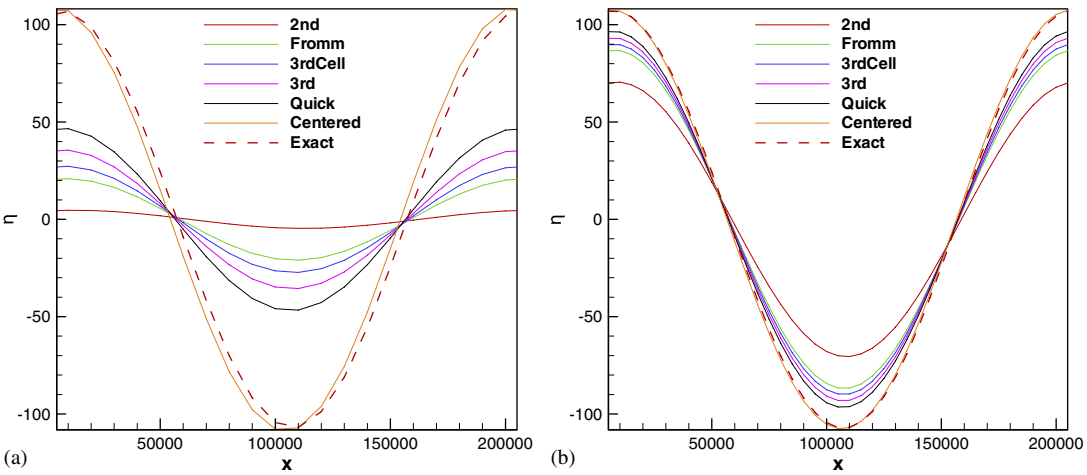

Figure 11. Numerical and exact solutions for Rossby wave after one waver period for coarse (left) and fine (right) grids.

We now consider a particular case with the following data suggested by Vreugdenhil [21]: length $L=200\,\text{km}$, depth $H=20\,\text{m}$, Coriolis parameter $f=10^{-4}\,\text{s}^{-1}$, $\beta=10^{-11}$. The initial conditions are specified using (92)–(94). The numerical results obtained after one waver period using various schemes with a coarse grid ($\Delta x=10\,\text{km}, \Delta t=315\,\text{s}$) and a fine grid ($\Delta x=5\,\text{km}, \Delta t=167\,\text{s}$) are shown in Figure 11. A fully centered scheme is also employed for comparison. As it is observed in Figure 11, while the centered scheme gives very accurate results, the upwind schemes again lead to a high level of damping even for the fine grid, as expected. The results of the 3rdMonotone scheme (not presented here) are even more damped than those of the 2nd scheme and this confirms the results of the previous sections.

9. CONCLUSION

A class of upwind finite volume schemes including first-, second- and third-order, as well as Quick, Fromm and a cell-based third-order methods have been studied here using a Fourier analysis approach. Various numerical aspects including group velocity, damping, phase speed and frequency have been considered and discussed for these schemes first for gravity waves. For most methods, folded frequency curves with stationary $2\Delta x$ waves have been observed in both the semi and fully discrete cases. However, the $2\Delta x$ waves are effectively damped although they encounter high phase errors.

Secondly, when the effect of the Coriolis term is taken into account, all selected upwind schemes lead to a very high (and unacceptable) damping error for the Rossby waves. An academic test case was designed to accurately verify the Fourier analysis results. By performing three numerical tests, it has been numerically shown that the damping of Rossby waves is even more significant when slope limiters are used. On the basis of these results, it is concluded that most upwind finite volume schemes, with or without slope limiters, should be used with care (if at all) for Rossby waves and restricted to short simulations.

DOI: 10.1002/fld

REFERENCES

1. LeVeque RJ. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press: Cambridge, U.K., 2002.
2. Toro EF. *Shock Capturing Methods for Free Surface Shallow Flows*. Wiley: New York, 2000.
3. Mohammadian A, Le Roux D, Tajrishi M, Mazaheri K. A mass conservative scheme for simulating shallow flows over variable topography using unstructured grid. *Advances in Water Resources* 2005; **28**:523–537.
4. Mohammadian A, Le Roux D. Simulation of shallow flows over variable topographies using unstructured grids. *International Journal for Numerical Methods in Fluids* 2006; **52**(5):473–498.
5. Lai J-S, Lin G-F, Guo W-D. An upstream flux-splitting finite-volume scheme for 2D shallow water equations. *International Journal for Numerical Methods in Fluids* 2005; **48**(10):1149–1174.
6. Batteen ML, Han YJ. On the computational noise of finite-difference schemes used in ocean models. *Tellus* 1981; **33**:387–396.
7. Wajsowicz RC. Free planetary waves in finite-difference numerical models. *Journal of Physical Oceanography* 1986; **16**:773–789.
8. Neta B, Williams RT. Rossby wave frequencies and group velocities for finite element and finite difference approximations to the vorticity–divergence and the primitive forms of the shallow water equations. *Monthly Weather Review* 1989; **117**:1439–1457.
9. Randall DA. Geostrophic adjustment and the finite-difference shallow-water equations. *Monthly Weather Review* 1994; **122**:1371–1377.
10. Adcroft AJ, Hill CN, Marshall JC. A new treatment of the coriolis terms in C-grid models at both high and low resolutions. *Monthly Weather Review* 1999; **127**(8):1928–1936.
11. Sankaranarayanan S, Spaulding ML. Dispersion and stability analyses of the linearized two-dimensional shallow water equations in boundary-fitted co-ordinates. *International Journal for Numerical Methods in Fluids* 2003; **42**(7):741–763.
12. Foreman MGG. A two-dimensional dispersion analysis of selected methods for solving the linearized shallow water equations. *Journal of Computational Physics* 1984; **56**(2):287–323.
13. Walters R, Carey GF. Analysis of spurious oscillation modes for the shallow water and Navier–Stokes equations. *Computers and Fluids* 1983; **11**:51–68.
14. Atkinson JH, Westerink JJ, Luettich RA. Two-dimensional dispersion analyses of finite element approximations to the shallow water equations. *International Journal for Numerical Methods in Fluids* 2004; **45**(7):715–749.
15. Le Roux DY, Rostand V, Pouliot B. Analysis of numerically-induced oscillations in 2D finite-element shallow water models; Part I: inertia–gravity waves. *SIAM Journal on Scientific Computing* 2007; **29**:331–360.
16. Le Roux DY, Pouliot B. Analysis of numerically-induced oscillations in 2D finite-element shallow water models; Part II: free planetary waves. *SIAM Journal on Scientific Computing* 2006, submitted.
17. Le Roux DY. Dispersion relation analysis of the $P_1^{NC}-P_1$ finite-element pair in shallow-water models. *SIAM Journal on Scientific Computing* 2005; **27**:394–414.
18. Le Roux DY, Carey GF. Stability/dispersion analysis of the discontinuous Galerkin linearized shallow-water system. *International Journal for Numerical Methods in Fluids* 2005; **48**(3):325–347.
19. Szpilka CM, Kolar RL. Numerical analogs of Fourier and dispersion analysis: development, verification, and application to the shallow water equations. *Advances in Water Resources* 2003; **26**:649–662.
20. Gossard CM, Kolar RL. Phase behavior of a finite volume shallow water algorithm. *Proceedings of CMWR XIII Volume 2*: *Computational Methods*, *Surface Water Systems and Hydrology*, Calgary, Canada, Bentley LR, Sykes JF (eds). 2000; 921–928.
21. Vreugdenhil CB. *Numerical Methods for Shallow-Water Flow*. Kluwer Academic Publishers: Dordrecht, 1994.
22. Leonard BP. A stable and accurate convective modeling procedure based on quadratic upstream interpolation. *Computer Methods in Applied Mechanics and Engineering* 1979; **19**:59–98.
23. Fromm JE. A method for reducing dispersion in convective difference schemes. *Journal of Computational Physics* 1968; **3**:176–189.
24. Platzman GW. Some response characteristics of finite-element tidal models. *Journal of Computational Physics* 1981; **40**:36–63.
25. Arakawa A, Lamb VR. Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods in Computational Physics* 1977; **17**:173–265.
26. Foreman MGG. An analysis of two-step time discretizations in the solution of the linearized shallow water equations. *Journal of Computational Physics* 1983; **51**(3):454–483.

DOI: 10.1002/fld

27. Beckers JM, Deleersnijder E. Stability of a FBTCS scheme applied to the propagation of shallow water inertia–gravity waves on various space grids. *Journal of Computational Physics* 1993; **108**(1):95–104.
28. Foreman MGG. An analysis of the 'wave equation' model for finite element tidal computations. *Journal of Computational Physics* 1983; **52**:290–312.
29. Batina J. Implicit flux-split Euler schemes for unsteady aerodynamics analysis involving unstructured dynamic meshes. *AIAA* 1990; **29**(11):1836.
30. Le Roux DY, Carey GF. Dispersion analysis of the least-squares finite-element shallow-water system. *International Journal for Numerical Methods in Fluids* 2003; **42**(6):607–622.

DOI: 10.1002/fld