Journal de Mécanique théorique et appliquée,
Journal of theoretical and applied mechanics
Special issue, supplement n° 2 to vol. 7-1988

The numerical modelling of Saint-Venant equations

M. FARGE and J.-F. LACARRA

Laboratoire de Météorologie Dynamique, 24, rue Lhomond, 75231 Paris Cedex 05

Abstract

We discuss the problems encountered when numerically integrating Saint-Venant (Shallow Water) equations in the inviscid case. Firstly, we show that the truncated system does conserve energy but not potential enstrophy, as it is the case for the continuous system. These invariants being both non quadratic, desaliasing becomes discussible because we find that, in this case, we would also loose the energy conservation. Secondly, we study the numerical stability and show that the Courant-Friedrichs-Lewy criterium becomes insufficient when inertio-gravity waves are excited, due to the effect of numerical dispersion errors. Thirdly, we discuss the questions of initialisation and resynchronisation of the leapfrog scheme, showing that the classical techniques do not work here, and we then define a procedure better suited to the wave behaviour of Saint-Venant equations.

0750-7240/1988/63 23/$ 4.30/

1. Introduction

In order to study the interactions between geostrophic eddies and inertio-gravity waves in atmospheric and oceanographic flows, we have developed a numerical model of the Saint-Venant (Shallow Water) equations. This type of flow can be termed 'compressible two-dimensional', in the sense that the variable free-surface height acts as a variable two-dimensional density and produces wave motions: consequently their behaviour is very different from the dynamics of incompressible two-dimensional flows. We want to test the conservative properties of this model and we therefore consider the case of an inviscid fluid. A measure problem is encountered here, due to the fact that the invariants are no more quadratic and, contrary to the incompressible case, the truncated system no longer conserves the invariants of the non-truncated equations. In this paper, we should discuss as separate questions the problem of numerical conservation, treated in paragraphs 4. and 5., and the problem of numerical stability, treated in paragraph 6. Concerning numerical stability, the compressible case behaves very differently from the incompressible case, as soon as gravity waves are not negligible, and, in particular, we will see that the Courant-Friedrichs-Lewy stability criteria is then no longer sufficient. We will also present new techniques to initialise and resynchronise the leapfrog time scheme, in order to guarantee numerical stability and control numerical dispersion when inertio-gravity waves are present.

2. Equations, invariants and eigenmodes

2.1 Equations

We model the dynamics of a free-surface fluid layer in uniform rotation, on a doubly periodic domain D^2, under the following assumptions:

vertical velocity << horizontal velocity,
Mach number[1] < 1,
Rossby number < 1,
barotropic fluid,

using shallow water (Saint-Venant's) equations:

$$\begin{cases} \dfrac{\partial \mathbf{U}}{\partial t} + (f + \xi)\mathbf{n} \times \mathbf{U} + \nabla\left(\Phi + \dfrac{V^2}{2}\right) = \mathbf{0} \\ \dfrac{\partial \Phi}{\partial t} + \nabla . (\Phi \mathbf{U}) = 0 \end{cases} \tag{1}$$

with the notations:

$\mathbf{U}$: horizontal velocity
Φ : free surface potential or geopotential
f : Coriolis parameter
$\mathbf{n}$: vertical unitary vector
ξ : relative vorticity in the rotating frame

They can be rewritten in terms of vorticity $\xi = \nabla \times \mathbf{U}$, divergence $\delta = \nabla . \mathbf{U}$ and geopotential Φ:

$$\begin{cases} \dfrac{\partial \xi}{\partial t} + \nabla . (\xi + f)\mathbf{U} = 0 \\ \dfrac{\partial \delta}{\partial t} - \nabla \times (\xi + f)\mathbf{U} + \nabla^2\left(\Phi + \dfrac{V^2}{2}\right) = 0 \\ \dfrac{\partial \Phi}{\partial t} + \nabla . (\Phi \mathbf{U}) = 0 \end{cases} \tag{2}$$

[1] This two-dimensional incompressible flow with free surface behaves as a barotropic three-dimensional compressible flow under the condition $\gamma = C_p/C_v = 2$, the free-surface height acting as a variable density.

2.2 Invariants

The equations have three global invariants:

total mass $$M = \int_D \Phi \qquad (3)$$

total energy (potential plus kinetic) $$E = \int_D (\Phi^2 + \Phi V^2) \qquad (4)$$

total potential enstrophy $$S = \int_D \frac{(\xi+f)^2}{\Phi} \qquad (5)$$

and one Lagrangian invariant:

potential vorticity

$$q = \frac{\xi+f}{\Phi} \qquad (6)$$

All these invariants are nonquadratic, contrary to the incompressible case.

2.3 Eigenmodes

If we linearize equations (2) around a minimal energy state, defined by $\Phi = \bar{\Phi}$, mean geopotential, and $V = 0$, we obtain:

$$\begin{cases} \dfrac{\partial \Phi}{\partial t} + \bar{\Phi}\delta = 0 \\ \dfrac{\partial \xi}{\partial t} + f\delta = 0 \\ \dfrac{\partial \delta}{\partial t} - f\xi + \nabla^2\Phi = 0 \end{cases} \qquad (7)$$

These equations have three eigenvalues $\lambda = 0$, $\lambda = +i\omega$ and $\lambda = -i\omega$ with $\omega^2 = k^2\Phi + f^2$, and therefore three eigenmodes per wavenumber $|\mathbf{k}|$:

. $\lambda = 0$, corresponds to one stationary non-divergent and geostrophic mode, whose components in spectral space are:

$$\mathcal{X}^R \begin{cases} \Phi^R = \dfrac{k^2 + k_0^2\hat{\varphi}}{k^2+k_d^2} \\ \hat{\chi}^R = 0 \\ \hat{\varphi}^R = \Phi^R \end{cases} \tag{8}$$

with
$\Psi = \nabla^{-2}\xi$, stream function,
$\chi = \nabla^{-2}\delta$, potential velocity,
$\varphi = \Phi/f$, geostrophic stream function,
k , modulus of the wavevector $\mathbf{k}$,
$k_d = 2\pi f/\sqrt{\Phi}$, Rossby deformation wavenumber,

. $\lambda = \pm i\omega$, we obtain two ageostroshic modes, having zero linearized potential vorticity ($\nabla^2\Psi - \lambda^2\Phi = 0$), which corresponds to dispersive inertia-gravity waves, whose components in spectral space are:

$$\mathcal{X}^G : \begin{cases} \Phi^G = \dfrac{k_d^2(\Phi-\hat{\varphi})}{k^2 + k_d^2} \\ \hat{\chi}^G = \hat{\chi} \\ \hat{\varphi}^G = \dfrac{k^2(\hat{\varphi}-\Phi)}{k^2 + k_d^2} \end{cases} \tag{9}$$

We can then rewrite the shallow water equations in eigenmodes, such as:

$$\frac{\partial \mathcal{X}}{\partial t} = i\lambda\mathcal{X} + N(\mathcal{X}) \tag{10}$$

$$\text{with } \mathcal{X} \begin{cases} \mathcal{X}^R \\ \mathcal{X}^{G+} \\ \mathcal{X}^{G-} \end{cases}$$

and $N(\mathcal{X})$ representing the nonlinear terms, i.e. the transport properties of the equations and the energy exchanges between modes.

4. Conservation laws and their discretisation

41. Discretisation

Let us now consider the behaviour of the continuous equations' invariants under the spatial discretisation.

While by discretising the equations in such a way that aliasing errors are eliminated, the time derivative of the PDE system is modified only for scales smaller than k_m. We can then deduce that, if an invariant of the continuous equations has a zero differential in the subgrid scale directions at every truncated state, its time derivative will not be modified by the spatial discretisation. It will still be an invariant of the discretised system. In such a case, the interaction between the represented and the subgrid scales only affects the second derivative of the invariant, which is canceled by filtering the time derivative of the state in phase space.

An important example, for which the previous argument applies, is incompressible turbulence. In this case the energy or the enstrophy dependence on subgrid scale is quadratic, which implies that these quantities have an extrema for the filtered states.

As we have mentioned in a previous section, the shallow water equations have three invariants. The total mass is of the type described above : since it depends only on resolved scales, it is therefore an invariant of the discretised pseudo-spectral model also. For the two other invariants the preceding condition is not satisfied. We will compute the time derivative of the energy, which is quite easy to obtain, and briefly mention the problems which arise for potential enstrophy.

42. Conservation of Energy

Let us first deduce from (1) the equation for the evolution of the energy (E) in the continuous case:

$$E = \int_D (\Phi^2 + \Phi V^2)$$

$$\frac{dE}{dt} = \int_D ((2\Phi + V^2) \frac{\partial \Phi}{\partial t} + 2(\Phi \mathbf{U}) . \frac{\partial \mathbf{U}}{\partial t})$$

. P, which denotes the non orthogonal projection that applies a given function to the one which coincides on the grid and which contains only scales larger than T. From a practical point of view, P acts as the FFT by applying subgrid scales on large scales, which are not necessarily zero.

Having defined these operators, the pseudo-spectral equations can be rewritten as follows:

$$\begin{cases} \dfrac{\partial \mathbf{U}}{\partial t} + FP((f + \xi)\mathbf{n} \times \mathbf{U}) + \nabla(FP(\Phi + \dfrac{V^2}{2})) = \mathbf{0} \\ \dfrac{\partial \Phi}{\partial t} + \nabla.(FP(\Phi \mathbf{U})) = 0 \end{cases} \tag{11}$$

where the variables are assumed to be filtered (i.e. $F\mathbf{U}=\mathbf{U}$ and $F\Phi = \Phi$).

The only difference with the spectral model lies in the presence of the projector P. The condition that must be satisfied, if we want to avoid aliasing errors, is then:

$$FP\ (F(f).F(g)) = F(F(f).F(g)) \tag{12}$$

for all functions f and g, which means that the represented scales, namely the image of F, must be correctly computed in the product of two truncated functions.

As F cancels Fourier components outside the wavenumber range $[-k_m, k_m]$, the product of two filtered functions has components in $[-2k_m, 2k_m]$. The only components which can be confounded with the ones filtered by P correspond to wavenumbers belonging to $[-k_m, -2lT]$, l being any positive integer. In order to avoid this alias problem, the two sets must be disjoint, which means that $2k_m < -k_m + 2T$, or equivalently that $k_m < 2T/3$.

More generaly, if the nonlinearities are of degree d, the condition is: $k_m < 2T/(1+d)$.

f-plane. But the Fourier components of these derivatives can differ from zero for wavenumbers greater than T. In order to obtain a closed system, these components must be cancelled at each time step, the others being left unchanged. In other words, an orthogonal projection operator should be applied to the time derivative.

32. Aliasing

We have just presented what is done in spectral models. But our model is in fact pseudo-spectral, which brings aliasing problems as we shall see now. In both, spectral and pseudo-spectral models, the spatial derivatives are computed in Fourier space. But the quadratic terms, needed for the computation of the time derivatives, are calculated in the Fourier space for the spectral method, and on the physical grid before going back to Fourier space for the pseudo-spectral method.

As we mentioned before, the products of truncated functions can give rise to scales smaller than T while calculating the time derivatives. In physical space, these scales cannot be distinguished from larger ones by the grid G. Thus, when going back to Fourier space, the coefficients of the scales we want to solve can be altered by the subgrid scales. The projection operator is not orthogonal to the subgrid scales, therefore they are not cancelled by it.

This is the aliasing phenomena. It can be avoided by a further truncation in Fourier space without changing the grid resolution, so that the represented scales of the product of two truncated functions would still be recognized by the grid in physical space, at the expenses of modifying the grid values.

We will now examine more closely the link between the original PDE and the discretised ODE, in order to be able to discuss what happens with the invariants of the original equations. It will be useful to introduce two operators:

. F, which denotes the orthogonal projection that truncates the functions above a certain wavenumber k_m ($k_m \leq T$),

3. Spatial discretisation and aliasing

31. Spatial discretisation

The simulation of fluid dynamics problems needs the numerical resolution of partial differential equation (PDE) systems governing their evolution. These systems possess an infinite number of degrees of freedom since the instantaneous state of the fluid is defined by the value of the variables at each point of the physical domain. In other words, the set of all possible states (phase space) is the infinite dimensional space of functions which are smooth enough on the given domain.

Since we can only simulate on a computer systems with finite number of degrees of freedom, we have to reduce the dimensionality of phase space. The spatial discretisation consists to choose a finite space which we will call the discretised phase space, and an associated ordinary differential equations (ODE) system describing the time evolution in this space.

By doing this, the original PDE system is approximated by an ODE system. The underlying assumption is that, due to viscous effects, the real evolution of the fluid takes place in a small neighborhood of a finite dimensional subspace of the whole phase space. One of the important problems in the choice of the spatial discretisation lies in the fact that, since this space is not invariant under the original PDE system, a somehow arbitrary closure equation must be applied in order to obtain the ODE system.

An element of the discretised phase space is a continuous function on the domain. It can equivalently be defined by its restriction to a finite grid, or in our case, the phase space may be a set of functions which Fourier components, for wavenumbers greater than a certain truncation T, are zero. A regular grid G with $N = 2T$ points being chosen, the isomorphism between our phase space and the grid functions is given by the Fast Fourier Transform (FFT) algorithm.

The instantaneous state having a finite number of Fourier components and the nonlinearity of the equations being polynomial, the time derivatives of the PDE can be computed exactly in the case of the

$$\frac{dE}{dt} = -\int_D ((2\Phi + V^2)\nabla.(\Phi\mathbf{U}) + (\Phi\mathbf{U}).\nabla(2\Phi + V^2) + (\Phi\mathbf{U}).((f + \xi)\mathbf{n} \times \mathbf{U})) \quad (13)$$

There are three terms. The first two are cancelled in the integration by parts, while the third one is zero, due to the orthogonality between $\mathbf{U}$ and $\mathbf{n} \times \mathbf{U}$. In the pseudo spectral model the equations are modified by the presence of P, such as:

$$\frac{dE}{dt} = -\int_D ((2\Phi + V^2)\nabla.(FP(\Phi\mathbf{U})) + (\Phi\mathbf{U}).\nabla(FP(2\Phi + V^2)) + (\Phi\mathbf{U}).(FP((f + \xi)\mathbf{n} \times \mathbf{U}))) \quad (14)$$

F commutes with the gradient operator and is symmetric with respect to L^2 scalar product. It then does not cause any trouble for the cancellation of the first two terms if there is no aliasing error (the operator P having no effect in that case). The third therm in turn can be non zero. There is no reason for $\Phi\mathbf{U}$ to be orthogonal to $FP(\xi\mathbf{n}\times\mathbf{U})$ outside the grid. The problem comes from the degrees of freedom due to the presence of Φ, because the orthogonality of the non filtered functions is not conserved for the filtered part (see appendix 1). In the incompressible case, only the integral of $\Phi\mathbf{U}.FP(\xi\mathbf{n}\times\mathbf{U})$ acts in the time derivative of the energy, which means that only the Fourier component associated with $k = 0$ must vanish for the energy to be an invariant. In the compressible case this expression is weighted by the density Φ, which implies that the conservation holds only when every Fourier components of the same expression vanishes.

Since conservation of orthogonality relation plays a part in the time equation for E, as well as the integration by part relation, we can try to find a discretised energy function E_d related to the grid G. It can be otained as follows:

$$E_d = \sum_\Theta (\Phi^2 + \Phi V^2) = \int_D P(\Phi^2 + \Phi V^2) \quad (15)$$

The advantage of this choice is that the relation between discrete sum and integral ensures the validity of the integration by parts relation:

$$\sum_{\Theta} f.\nabla(Pg) = \int_{D} (Pf).\nabla(Pg) = -\int_{D} (Pg).\nabla(Pf) = -\sum_{\Theta} g.\nabla(Pf) \quad (16)$$

Computing the time derivative of E_d, we will obtain the equivalent of the previous term:

$$\frac{dE_d}{dt} = \int_{D} P(\Phi U).FP((\xi)\mathbf{n} \times \mathbf{U}) \quad (17)$$

The non-vanishing of the energy time derivative indicates the presence of an energy exchange with subgrid scales, even when aliasing errors are avoided. This seems to imply that the non conservation of energy is not related to an incorrect space discretisation scheme, but rather to a physical property of the simulated equations which cannot maintain energy confined to the large scales.

43. Conservation of potential enstrophy

We briefly mention now what happens to potential enstrophy S. Its conservation in the continuous case is due to the Lagrangian invariance of the potential vorticity $q = (\xi + f)/\Phi$. Due to its fractional expression q can have an infinite number of non-zero Fourier components and the time derivative in the truncated model is thus more difficult to control. In the continuous system we have:

$$\frac{dS}{dt} = 2\int_{D} ((FP(q\Phi U) - qFP(\Phi U)).\nabla(\eta)) \quad (18)$$

which becomes in the truncated system:

$$\frac{dS_d}{dt} = \int_{D} ((2P(P(q)F\nabla(P(q))) - F\nabla(Pq^2))).P(\Phi U)) \quad (19)$$

5. Conserving schemes, advantages and disadvantages

We have seen that, when suppressing aliasing errors, the ODE system loses the invariants present in the original PDE system. On the other hand, we can see from (4) that, if F is omitted, the discretised energy has a zero derivative. This is due to the expression of E_d in terms of grid values and to the fact that the orthogonal relation, needed to ensure the energy conservation, is respected on the grid G. We have here two possibilities: either the spatial discretisation is made without aliasing error, and the equations do not conserve the invariants E and S, or energy is conserved at the expense of aliasing errors, when the filter F is not applied. The question of knowing which of the two solutions is better is not trivial, as we shall discuss now.

We can first argue that there is no physical reason for the conservation of energy and potential enstrophy by the truncated equations, since in the real system there can be energy and enstrophy exchanges with subgrid scales. This leads us to prefer a non-aliased scheme, which correctly represents transfers when the subgrid scales are zero. This choice is also sustained by the fact that, once the model variables and truncation are chosen, all the discretisation schemes avoiding aliasing errors are equivalent. Therefore, in our case, any model of the shallow-water equations, written with the same set of variables, would not conserve E or S without aliasing errors.

The previous argument must be tempered because the notion of aliasing error is not so universal as it can seem at first. It is in fact related to the choice of the model variables. If we make a change of the non-linear variables, then the scale analysis associated to the Fourier decomposition has a different meaning in the two systems of coordinates. These new variables define a new orthogonal structure and a new projector in the transformation of PDE into ODE. This can be illustrated by the following choice of variables: Φ, $\nu=\Phi U$, for which the nonlinear terms of the equations are no more polynomial. Then, as soon as the spatial discretisation scheme avoids aliasing errors, the energy, which is now quadratic, is therefore conserved by the model.

So we see that the question of what to do with aliasing errors, as well as the problem of the conservation of invariants, does not appear clearly. It is of no use to refer to the physical properties of the simulated fluid, since the comparison with the model cannot be disjoined from the choice of the subgrid scale parametrisation. The meaning of energy exchanges, between represented and subgrid scales, cannot be correctly interpreted if the subgrid scales are maintained at a non-physical zero level.

In practice, people prefer conserving schemes, real problems being left to the subgrid scale parameterisation. This attitude can be explained by several reasons. The principal one is that there is almost no control on the accuracy of the time discretisation scheme, and in particular on the influence of the time step chosen. The conserved quantity, if there is any, plays then a very important part as a diagnostic variable in order to validate these choice, even when the subgrid scales are parameterized by known schemes; for instance, instability problems mentionned by Phillips, 1959 were cancelled by the use of a conserving scheme proposed by Arakawa, 1966. From an experimental point of view, when time discretisation schemes are not adequate, there is a numerical blow up which can be first detected by an increase of the theoretical invariants. Another reason is that aliasing errors are not easy to detect and eliminate, in particular for grid-point models. The correctness of the spatial discretisation being difficult to assert, a conserving scheme seems at least able to limit numerical errors. In the shallow water pseudo-spectral case we have just shown that two possibilities exist. The differences between the two models do not seem very important. In the non conserving scheme, it is found experimentally that the energy tends to decay on very long time scales, although a closer examination of equation (4) shows that the sign of the energy derivative does not need to be negative (details of the different tests are given in Appendix 2).

6. Time integration, stability and numerical dispersion

61. Numerical stability

For the time integration we have chosen the leapfrog scheme, which is second order and explicit, because it presents the advantage of being neutral as soon as it is stable, and therefore well adapted to follow the wave behaviour of shallow-water equations without numerical damping. We first consider only the linearized shallow-water equations written in eigenmodes:

$$\frac{dX}{dt} = i\omega(k)X, \tag{20}$$

integrated with a leapfrog time scheme:

$$X^{n+1} = X^{n-1} + 2\Delta t \frac{dX^n}{dt}, \tag{21}$$

Δt being the time step. We then get the amplification relation:

$$\begin{pmatrix} X^{n+1} \\ X^n \end{pmatrix} = \begin{pmatrix} 2i\omega(k)\Delta t & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} X^n \\ X^{n-1} \end{pmatrix} \tag{22}$$

and find that the amplification matrix has two eigenvalues:

$$\lambda^{\pm} \sqrt{1-\omega^2\Delta t^2} + \sqrt{-1}\omega\Delta t = e^{\sqrt{-1}\omega_{LF}\Delta t} \quad \text{with } \omega_{LF} = \frac{\text{arc sin}\omega\Delta t}{\Delta t} \tag{23}$$

ω_{LF} acting for the discrete equations as ω for the continuous equations,

. the firsts corresponds to the physical mode

$$\Lambda_+ = 1 + i\omega(k)\Delta t - \frac{\omega^2}{2}\Delta t^2 + O(\Delta t^3), \tag{24}$$

. the second corresponds to the numerical mode

$$\Lambda_- = -1 + i\omega(k)\Delta t + \frac{\omega^2}{2}\Delta t^2 + O(\Delta t^3) \tag{25}$$

whose modulus is one if we respect the Courant-Friedrichs-Lewy stability criteria:

$$\Delta t \leqslant (\omega(k_{max}))^{-1} \tag{26}$$

with k_{max} smallest wavenumber represented by the grid.
This may be rewritten as:

$$\Delta t \leqslant \frac{\Delta x}{\pi V_p(k_{max})} \tag{27}$$

with

$\Delta x = \dfrac{\pi}{k_{max}}$, grid size,

$V_p(k) = c\left(1 + \dfrac{k_d^2}{k^2}\right)^{1/2}$, phase velocity of inertia - gravity waves,

$c = (\bar{\Phi})^{1/2}$, phase and group velocity of gravity waves in absence of rotation, which is equivalent to the sound velocity in this system,

or:

$$\Delta t \leqslant \frac{\Delta x}{\pi c^2} V_g(k_{max}) \tag{28}$$

with $V_g(k) = \dfrac{c^2}{V_p(k)} = c(1+\dfrac{k_d^2}{k^2})^{1/2}$, $V_p(k)$ group velocity of inertia - gravity waves.

The previous analysis, developed in order to choose the maximum time step allowed for stability, was purely linear. Therefore, if we want to take into account the nonlinear properties of shallow-water equations, resulting from the transport term, we should consider a

modified CFL stability criteria:

$$\Delta t \leqslant \frac{\Delta x}{\pi(V_p(k_{max}) + V_{max})} \tag{29}$$

with V_{max}, maximum of the velocity modulus,

which can be rewritten as an equality:

$$\Delta t = C \frac{\Delta x}{\pi(V_p(k_{max}) + V_{max})} \tag{30}$$

with $C \leqslant 1$, CFL number.

62. Numerical dispersion

When $C < 1$, although the leapfrog time scheme is stable and neutral, it presents the disadvantage of being numerically dispersive. If we rewrite the eigenvalues of the amplification matrix in the following way:

$$\Lambda^{\pm} = a(\cos\theta_{LF} + i\sin\theta_{LF}) \tag{31}$$

$$\text{With} \left| \begin{array}{l} a = 1 \\ \cos\theta_{LF} = \pm\sqrt{1 - \omega^2\Delta t^2} \\ \sin\theta_{LF} = \omega\Delta t \end{array} \right.$$

we find that, after one time step of the leapfrog scheme, the phase angle is:

$$\theta_{LF} = \arcsin(\omega\Delta t) = \omega_{LF}\Delta t, \tag{32}$$

while it should only be $\theta = \omega\Delta t$ for a wave in the physical system. Therefore, the leapfrog time scheme tends to accelerate the phase velocity of inertia-gravity waves and this proportionally to their frequency. So, if we consider the evolution of a wave packet, the leapfrog time scheme will tend to spread it out, in accelerating the high frequency waves . For instance, if $C = 0.75$, $\theta_{LF} - \theta = 0.09$ rd, i.e. the information concerning the phase of high frequency waves will be

lost after 33 time steps, while, if C = 0.075, $\theta_{LF}-\theta=7.10^{-15}$rd, in which case it will be lost only after 45000 time steps.

63. Tests

In order to test the stability of the leapfrog time scheme, we have performed several integrations, considering different CFL numbers, and studied the evolution of both total energy E and potential enstrophy S. We found that the conservation of these invariants presents a typical nonlinear behaviour with fast augmentations followed by long stable phases during which they are conserved (cf. Figure 1). We therefore have conducted those integrations on long time scales, i.e. during 55000 time steps for C = 0.75 and 110000 time steps for C=0.075. Our first conclusion is that the CFL criteria (corresponding to C=1) is not sufficient to control those nonlinear instabilities when inertia-gravity wave are not negligible (cf. Figure 1). For instance, if we consider a case where the inertia-gravity modes are 100 times more energetic than the geostrophic modes, we found for C=0.75 that, after 200 eddy turn-over times, i.e. respectively after 11000 and 110000 time steps, total energy is increased by 5% and potential enstrophy by 0.1%, while for C=0.075 their augmentation is respectively 0.008% and 0.02%, only. This complex nonlinear behaviour is probably partially caused by the modification of the dispersive properties of shallow water equations due to numerical dispersion errors brought by the leapfrog time scheme. Those dispersive properties play indeed an important role to control the apparition of singularities because they disperse away any strong perturbation as soon as it develops in the flow. In practice, for the inviscid case, when inertio-gravity waves are dominant, we propose to use a very small CFL number, for instance C=0.1.

Our second series of tests concerns the initialization and resynchronization of the leapfrog scheme. Being second order in time, it is necessary to know the solution at two distinct instants, in order to start the computation. But, as we have seen in section 61., the leapfrog scheme presents a numerical mode, which tends to separate

the even and odd solutions. After having tested different initialization schemes, among them Euler, forward-backward and Heun's schemes (see Farge, 1985), we found that the best initialization procedure consists to exactly integrate the linearized equations written in eigenmodes (21), neglecting the nonlinear effects, which evolve on a much longer time scale. We then recalculate the solution obtained for the first time step in such a way that we cancel the numerical leapfrog mode, and we obtain:

$$X^1 = X^\circ \, e^{\sqrt{-1}\omega_{LF}\Delta t} \qquad (34)$$

with X° solution at $t=0$,
X^1 solution at $t=\Delta t$.

On Figure 2, we compare three different initializations and we see that the integration in eigenmodes is the only one which does not present a leapfrog separation.

Concerning the risk of separation between the odd and even solutions, i.e. the development of the leapfrog instability after a certain time due to nonlinear effects, we have tested different resynchronisation methods, among them the 1/4 - 1/2 1/4 averaging, Asselin's filter and Heun's scheme (see Farge, 1985). We found that, in presence of inertio-gravity waves, they all were, or too dissipative or too exciting (see Figure 3). We therefore decide, as soon as the separation between the even and odd solutions will become larger than 1% for anyone of the invariants, to restart the integration from one of the two solutions, using the same eigenmode integration as we have already applied for the initialization.

Acknowledgements

The computing has been done on the Cray-1 of C_2VR, Palaiseau, using as front-end the IBM3090 of CIRCE. The manuscript has been typed by M.C. Cally.

References

A. Arakawa, 1966
Computational design for long-term numerical integration of the equations of fluid motion : two-dimensional incompressible flow.
J. Comp. Phys., **1**, 119-143

M. Farge, 1985
Stabilité numérique des équations de Saint-Venant
Rapport interne L.M.D., n°131

N.A. Phillips, 1959
The atmosphere and sea in motion
Rockfeller Institute Press and the Oxford University Press, New-York, pp. 501-504

Appendix 1

As example, we will exhibit a truncated state for which the energy time derivative is non zero, even when aliasing errors are avoided.

Let $\mathbf{k}$ be a wave vector parallel to the x axis, such that $2\mathbf{k}$ is filtered by F. We define then the state of the model as follow:

$$\mathbf{U} = \cos(kx).\mathbf{U}_1 + \sin(kx).\mathbf{U}_2 \text{ and } \Phi = \cos(kx)$$

$\mathbf{U}_1$ and $\mathbf{U}_2$ being two fixed vectors. Let us compute explicitly the energy time derivative from equation (11):

$$\xi = k(-v_1 \sin(kx) + v_2 \cos(kx))$$

$$\xi \mathbf{n}\times\mathbf{U} = \frac{k}{2}\,((-v_1 \sin(2kx) + v_2(1+\cos(2kx)))\mathbf{n}\times\mathbf{U}_1 + (-v_1(1-\cos(2kx)+v_2\sin(2kx))\mathbf{n}\times\mathbf{U}_2)$$

$$F(\xi \mathbf{n}\times\mathbf{U}) = \frac{k}{2}\,\mathbf{n}\times(v_2\mathbf{U}_1 - v_1\mathbf{U}_2)$$

$$(\Phi\mathbf{U}).F(\xi \mathbf{n}\times\mathbf{U}) = \frac{k}{4}\,(v_1(1 + \cos(2kx)) - v_2\sin(2kx))\mathbf{n} \times \mathbf{U}_2$$

$$\frac{dE_d}{dt} = \frac{k}{4}\, v_1 \mathbf{n}.\mathbf{U}_1 \times \mathbf{U}_2$$

v_1 and v_2 being the second coordinates of $\mathbf{U}_1$ and $\mathbf{U}_2$, respectively. By a careful choice of $\mathbf{U}_1$ and $\mathbf{U}_2$, the E_d derivative does not vanish.

Appendix 2

List of the different cases we have tested

The results are shown on the table below.

1. truncated compressible model, k_{max} = 16, Δt = 20s, C=0.1
2. desaliased compressible model, k_{max} = 10, Δt = 20s, C=0.1
3. desaliased compressible model, k_{max} = 10 with suppression of the term $\xi n \times U$ in the equations, Δt = 20s, C = 0.1
4. non truncated incompressible model, k_{max} = 16, Δt=60s, C=0.3
5. truncated incompressible model, k_{max} = 15, Δt = 60s, C=0.3
6. desaliased incompressible model, k_{max} = 10, Δt = 120s, C=0.6
7. desaliased incompressible model, k_{max} = 10, Δt = 60s, C=0.3
8. desaliased incompressible model, k_{max} = 10, Δt = 20s, C=0.1
9. desaliased incompressible model, k_{max} = 10, Δt = 10s, C=0.05
10. desaliased incompressible model, k_{max} = 10, Δt = 5s, C=0.025
11. desaliased incompressible model, k_{max} = 10, Δt = 2s, C=0.01

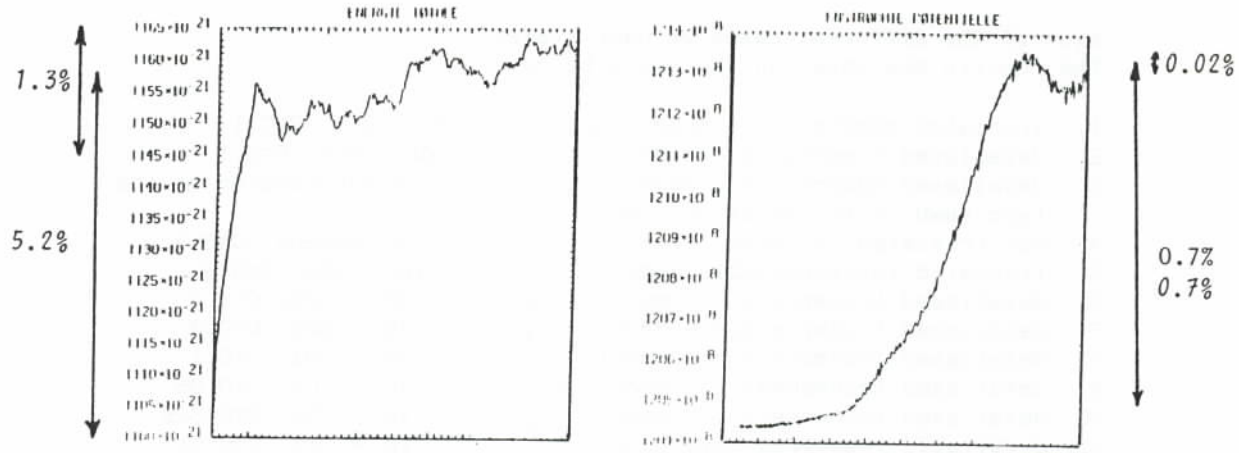
The mean value of E and S had been subtracted to study their fluctuations, even very small. For each invariant, 3 important quantities must be observed:

- the time derivative due to the spatial discretisation: $(\frac{dE}{dt})_d$,
- the time derivative due to the whole model: $(\frac{dE}{dt})_m$ based on the variations of E in the integration,
- the characteristic time after which the errors on E are of the same order of magnitude than E: $T_E = E/[(dE/dt)_m]$.

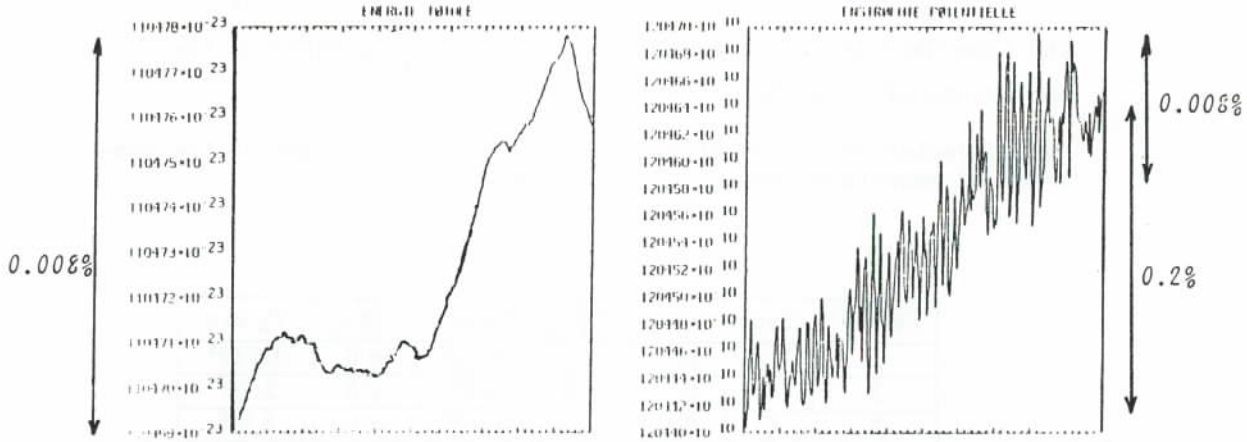
Test N°	Type	*c*	$(\frac{dE}{dt})_d$	T_E *in s*	$(\frac{dZ}{dt})_d$	T_Z *in s*
1	C	0.1	0	$5.\,10^6$	$\neq 0$	$1.5\,10^5$
2	C	0.1	$\neq 0$	10^8	$\neq 0$	$2.\,10^7$
3	C	0.1	0	$3.\,10^9$	$\neq 0$	$5\,10^5$
4	I	0.3	0	10^7	$\neq 0$	$2\,10^5$
5	I	0.3	0	10^9	$\neq 0$	$1.2\,10^6$
6	I	0.6	0	10^{10}	0	$1.5\,10^9$
7	I	0.3	0	$3.\,10^{10}$	0	10^{10}
8	I	0.1	0	10^{11}	0	$2.\,10^{11}$
9	I	0.05	0	$5.\,10^{11}$	0	$1.5\,10^{12}$
10	I	0.025	0	$2.\,10^{12}$	0	6à40 10^{12}
11	I	0.01	0	$1.2\,10^{13}$	0	4à10 10^{13}

where *C* *means compressible*
I *means incompressible*

Tableau 1



Foreward-backward initialization and
then leapfrog scheme
for C= 0.75
integrated during 55 000 time steps



Foreward-backward initialization and
then leapfrog scheme
for C= 0.075
integrated from timestep 55 000 to
110 000

Figure 1

Courant-Friedrichs-Lewy number tests