Journal of Computational Physics **172,** 592–608 (2001)
doi:10.1006/jcph.2001.6843, available online at http://www.idealibrary.com on IDEAL®

A Nonlinear Galerkin Method for the Shallow-Water Equations on Periodic Domains

Saulo R. M. Barros* and José W. Cárdenas†,[1]

**Departamento de Matemática Aplicada, Universidade de São Paulo, R. do Matão, 1010, 05508-900 São Paulo, Brazil; and †Laboratório Nacional de Computação Científica, Av. Getúlio Vargas, 333, 25651-070, Petrópolis, R. J., Brazil*
E-mail: saulo@ime.usp.br; cardenas@lncc.br

Received September 14, 2000; revised May 23, 2001

A nonlinear Galerkin method for the shallow-water equations is developed, based on spectral transforms. The scheme is compared to a pseudo-spectral Galerkin method. Our numerical results indicate that the nonlinear scheme has the potential advantage of providing similar accuracy at a lower cost than the Galerkin method. The nonlinear method has also less restrictive stability conditions. © 2001 Academic Press

1. INTRODUCTION

Nonlinear Galerkin methods were first introduced by Marion and Temam [14] for Navier–Stokes equations, from a theoretical point of view. The technique relies on the theory of approximate inertial manifolds [9, 12] (see also [2] for the shallow-water equations) and employs a decomposition of the solution into its small- and large-scale components. Applications of the method have been developed, among others, for the Burgers equation [5, 10] and Navier–Stokes equations [4, 7, 11].

In this article we propose a nonlinear Galerkin method for the shallow-water equations on two-dimensional domains with periodic boundary conditions (the equations are formulated as in the model proposed by Lorenz [13], on a f-plane). The shallow-water equations differ from the incompressible Navier–Stokes equations by the inclusion of the Coriolis force and by a different mass-conservation equation. In particular, a shallow-water flow is not divergence-free and the equations can not be projected on the space of nondivergent velocities, as tipically done for the Navier–Stokes equations. In the scheme we propose, the velocity components and the geopotential height are expanded as double Fourier series, truncated at a certain resolution and the solution space is decomposed into low and high modes. Some nonlinear interaction terms are neglected and, through

[1] Partially supported by the Brazilian Research Council (CNPq).

0021-9991/01 $35.00

projection of the equations onto the solution spaces, Galerkin equations for the low and high modes are derived. The implementation of the scheme is made efficient by the use of the spectral transform method [8, 15] to compute the projections. These are computed exactly, with no aliasing, by choosing auxiliary grids of appropriate sizes. The time-discretization is based on a semi-implicit method, leading to CFL stability constraints guided by the flow velocities (and not by the high phase speeds of the gravity-wave modes of the shallow-water equations). The stability conditions are less restrictive than for the Galerkin method. These facts are shown in a linear stability analysis and verified in the numerical experiments.

A nonlinear Galerkin method for the shallow-water equations has been proposed in [6]. There, two families of basis functions for the velocity field are employed, one composed by purely rotational and one by purely divergent fields (these basis functions have to be precomputed). Their approach also differs from ours in the time discretization (they employ predictor corrector schemes) and in the numerical treatment of the equations, since they don't use spectral transforms (which we feel to be fundamental for efficiency).

The appropriate truncation number for a nonlinear Galerkin scheme depends on the spectral distribution of energy [4, 6]. In any case, if the high modes are resolved only up to a certain wave number, a corresponding Galerkin method using the same truncation should provide results at least as good, since it uses more information than the nonlinear scheme (in which some nonlinear interactions are neglected). The potential advantage of the nonlinear Galerkin scheme is to provide essentially the same accuracy at a reduced computational cost. We compared the nonlinear method with a corresponding Galerkin scheme and obtained numerical evidence of this advantage of the nonlinear method. We feel the technique attractive for applications and we are investigating its use on global models on the sphere, aiming at numerical weather prediction.

The paper is structured as follows. It begins with the description of the equations in Section 2, followed by the presentation of the Galerkin method in Section 3. Section 4 is dedicated to the nonlinear Galerkin method, and in Section 5 we develop the linear stability analysis of the Galerkin schemes. Numerical results are presented in Section 6 and the paper is closed with some conclusions.

2. THE SHALLOW-WATER EQUATIONS

We consider the shallow-water equations in nondimensional form as proposed by Lorenz [13], on a two-dimensional rectangular domain (the so called f-plane), with periodic boundary conditions. Diffusion and a mass-forcing term are explicitly included. The equations are given by

$$\begin{aligned} \frac{\partial u}{\partial t} + uu_x + vu_y - v + z_x - \nu_0 \Delta u &= 0 \\ \frac{\partial v}{\partial t} + uv_x + vv_y + u + z_y - \nu_0 \Delta v &= 0 \\ \frac{\partial z}{\partial t} + uz_x + vz_y + (z_0 + z)(u_x + v_y) - \kappa_0 \Delta z &= F, \end{aligned} \tag{1}$$

where the unknows are the two velocity components u and v and the geopotential height z; the domain is $\Omega = [0, 2\pi] \times [0, 2\pi]$. The terms $-v$ and u in the first two equations

correspond to the nondimensional form of the Coriolis force, ν_0 and κ_0 are the diffusion coefficients and F is the (time independent) mass-forcing term.

The shallow-water equations are distinguished from the $2D$-incompressible Navier–Stokes equations by including effects of the Earth's rotation in the Coriolis terms and by having a different mass continuity equation. In particular, the flow is not divergence-free and a Galerkin method for these equations cannot employ a projection on the space of nondivergent velocities. The shallow-water equations are often employed in models for ground-water, oceanic, and atmospheric flows.

3. A PSEUDO-SPECTRAL GALERKIN METHOD FOR THE EQUATIONS

The shallow-water equations admit solutions evolving in different time scales. They present slower modes, the Rossby waves, and the much faster evolving gravity waves. While one is normally interested in following the large-scale motion of the Rossby waves, the presence of the gravity waves, in spite of the fact that they usually carry little energy, poses severe stability restrictions for explicit schemes, because of their high-phase speed. Therefore, it is important to adopt some degree of implicitness in the numerical schemes for these equations, in order to attenuate the CFL stability constraints.

We propose here a semi-implicit pseudo-spectral Galerkin method for system (1). The prognostic fields u, v, and z will be expanded as double Fourier series,

$$\begin{bmatrix} u_N(x,y,t) \\ v_N(x,y,t) \\ z_N(x,y,t) \end{bmatrix} = \sum_{k,l \in I_N} \begin{bmatrix} \hat{u}_{kl}(t) \\ \hat{v}_{kl}(t) \\ \hat{z}_{kl}(t) \end{bmatrix} e^{i(kx+ly)}, \tag{2}$$

where

$$I_N = \left\{ (k,l) : -\frac{N}{2} + 1 \le k, l \le \frac{N}{2} \right\}. \tag{3}$$

The truncated expansion can be seen as the result of the *projection* P_N from the space H_{per} of functions given by double Fourier series onto $U_N = \text{span}\,\{e^{i(kx+ly)} : (k,l) \in I_N\}$. We first consider a second-order semi-implicit time discretization of (1), where the terms giving rise to the fast gravity waves (such as the geopotential gradient) will be treated implicitly, while the nonlinear terms will be discretized explicitly by a leap-frog type scheme. The complete time discretization is given by

$$\begin{aligned} u^{n+1} - \Delta t v^{n+1} + \Delta t z_x^{n+1} - \nu_0 \Delta t \Delta u^{n+1} &= r_{1a}^{n-1} + r_{1b}^{n} \\ v^{n+1} + \Delta t u^{n+1} + \Delta t z_y^{n+1} - \nu_0 \Delta t \Delta v^{n+1} &= r_{2a}^{n-1} + r_{2b}^{n} \\ z^{n+1} + z_0 \Delta t \left(u_x^{n+1} + v_y^{n+1} \right) - \kappa_0 \Delta t \Delta z^{n+1} &= r_{3a}^{n-1} + r_{3b}^{n}, \end{aligned} \tag{4}$$

where a superscript n refers to the variables at time $t_n = n\Delta t$, and

$$\begin{aligned} r_{1a}^{n-1} &= u^{n-1} + \Delta t v^{n-1} - \Delta t z_x^{n-1} + \nu_0 \Delta t \Delta u^{n-1} \\ r_{2a}^{n-1} &= v^{n-1} - \Delta t u^{n-1} - \Delta t z_y^{n-1} + \nu_0 \Delta t \Delta v^{n-1} \\ r_{3a}^{n-1} &= z^{n-1} - \Delta t z_0 \left(u_x^{n-1} + v_y^{n-1} \right) + \kappa_0 \Delta t \Delta z^{n-1} \end{aligned} \tag{5}$$

$$
\begin{aligned}
r_{1b}^n &= -2\Delta t\left(u^n u_x^n + v^n u_y^n\right)\\
r_{2b}^n &= -2\Delta t\left(u^n v_x^n + v^n v_y^n\right)\\
r_{3b}^n &= -2\Delta t((z^n u^n)_x + \left(z^n v^n\right)_y + F).
\end{aligned}
\tag{6}
$$

The Galerkin scheme is obtained through the projection of the time discretizated equations on U_N :

$$
\begin{aligned}
u_N^{n+1} - \Delta t v_N^{n+1} + \Delta t z_{N,x}^{n+1} - \nu_0 \Delta t \Delta u_N^{n+1} &= P_N\left(r_{1a}^{n-1} + r_{1b}^n\right)\\
v_N^{n+1} + \Delta t u_N^{n+1} + \Delta t z_{N,y}^{n+1} - \nu_0 \Delta t \Delta v_N^{n+1} &= P_N\left(r_{2a}^{n-1} + r_{2b}^n\right)\\
z_N^{n+1} + z_0 \Delta t\left(u_{N,x}^{n+1} + v_{N,y}^{n+1}\right) - \kappa_0 \Delta t \Delta z_N^{n+1} &= P_N\left(r_{3a}^{n-1} + r_{3b}^n\right).
\end{aligned}
\tag{7}
$$

Assuming for the moment that the projections in the right-hand side of (7) have been computed, this linear system can be decomposed for each spectral component $(k, l) \in I_N$ as

$$
\begin{bmatrix}
1+\nu_0\gamma_{kl} & -\Delta t & \mathbf{i}k\Delta t\\
\Delta t & 1+\nu_0\gamma_{kl} & \mathbf{i}l\Delta t\\
\mathbf{i}kz_0\Delta t & \mathbf{i}lz_0\Delta t & 1+\kappa_0\gamma_{kl}
\end{bmatrix}
\cdot
\begin{bmatrix}
\hat{u}_{kl}^{n+1}\\ \hat{v}_{kl}^{n+1}\\ \hat{z}_{kl}^{n+1}
\end{bmatrix}
=
\begin{bmatrix}
\hat{r}_{1a,kl}^{n-1} + \hat{r}_{1b,kl}^n\\
\hat{r}_{2a,kl}^{n-1} + \hat{r}_{2b,kl}^n\\
\hat{r}_{3a,kl}^{n-1} + \hat{r}_{3b,kl}^n
\end{bmatrix},
\tag{8}
$$

where $\gamma_{kl} = \Delta t \lambda_{kl}$ and $\lambda_{kl} = k^2 + l^2$ is the corresponding eigenvalue of minus the Laplacian.

The system (8) can be written as a system for $(\hat{u}_{kl}^{n+1}, \hat{v}_{kl}^{n+1})$ (depending on the $\hat{z}_{kl}^{n+1}$ variable)

$$
\begin{bmatrix}
(1+\nu_0\gamma_{kl}) & -\Delta t\\
\Delta t & (1+\nu_0\gamma_{kl})
\end{bmatrix}
\cdot
\begin{bmatrix}
\hat{u}_{kl}^{n+1}\\ \hat{v}_{kl}^{n+1}
\end{bmatrix}
=
\begin{bmatrix}
\hat{r}_{1a,kl}^{n-1} + \hat{r}_{1b,kl}^n - \mathbf{i}k\Delta t\hat{z}_{kl}^{n+1}\\
\hat{r}_{2a,kl}^{n-1} + \hat{r}_{2b,kl}^n - \mathbf{i}l\Delta t\hat{z}_{kl}^{n+1}
\end{bmatrix}
$$

whose solution is given by

$$
\begin{bmatrix}
\hat{u}_{kl}^{n+1}\\ \hat{v}_{kl}^{n+1}
\end{bmatrix}
= \frac{1}{\theta_{kl}}
\begin{bmatrix}
\hat{s}_{1,kl} - \mathbf{i}(k\Delta t(1+\nu_0\gamma_{kl}) + l\Delta^2 t)\hat{z}_{kl}^{n+1}\\
\hat{s}_{2,kl} + \mathbf{i}(k\Delta^2 t - l\Delta t(1+\nu_0\gamma_{kl}))\hat{z}_{kl}^{n+1}
\end{bmatrix},
\tag{9}
$$

where

$$
\begin{aligned}
\mathbf{i} &= \sqrt{-1}\\
\theta_{kl} &= (1+\nu_0\gamma_{kl})^2 + \Delta^2 t\\
\hat{s}_{1,kl} &= (1+\nu_0\gamma_{kl})\left(\hat{r}_{1a,kl}^{n-1} + \hat{r}_{1b,kl}^n\right) + \Delta t\left(\hat{r}_{2a,kl}^{n-1} + \hat{r}_{2b,kl}^n\right)\\
\hat{s}_{2,kl} &= -\Delta t\left(\hat{r}_{1a,kl}^{n-1} + \hat{r}_{1b,kl}^n\right) + (1+\nu_0\gamma_{kl})\left(\hat{r}_{2a,kl}^{n-1} + \hat{r}_{2b,kl}^n\right).
\end{aligned}
\tag{10}
$$

The combination of the two equations in (9) (equivalent to building the divergence of the new velocity field) provides the expression

$$
\mathbf{i}\left(k\hat{u}_{kl}^{n+1} + l\hat{v}_{kl}^{n+1}\right) = \frac{\left(\mathbf{i}(k\hat{s}_{1,kl} + l\hat{s}_{2,kl}) + \Delta t(1+\nu_0\gamma_{kl})\lambda_{kl}\hat{z}_{kl}^{n+1}\right)}{\theta_{kl}},
\tag{11}
$$

which when employed in the third equation of (8) leads to

$$\hat{z}_{kl}^{n+1} = \frac{1}{\alpha_{kl}} \left(\hat{r}_{3a,kl}^{n-1} + \hat{r}_{3b,kl}^{n} - \mathbf{i} \frac{z_0 \Delta t}{\theta_{kl}} (k\hat{s}_{1,kl} + l\hat{s}_{2,kl}) \right) \tag{12}$$

with

$$\alpha_{kl} = 1 + \frac{z_0 \Delta t \gamma_{kl}}{\theta_{kl}} (1 + \nu_0 \gamma_{kl}) + \kappa_0 \gamma_{kl}. \tag{13}$$

Altogether, we solve (8) by first deriving the value of $\hat{z}_{kl}^{n+1}$ from (12) and then using it in (9).

For the solution of (7) it remains to explain how to compute the projections of the right-hand side. Since they involve the nonlinear terms (including products of the variables), they cannot be computed directly from the spectral coefficients in a efficient way. Instead, we employ the so-called spectral transform method [8, 15], using an auxiliary grid

$$J_N = \left\{ (x_r, y_s) : x_r = \frac{2\pi(r-1)}{N}, \quad y_s = \frac{2\pi(s-1)}{N}, \quad r, s = 1, \ldots, N \right\}, \tag{14}$$

where we can evaluate the function and derivative values involved in the right-hand side of (7) (through Fourier transforms). On the grid, the products can be trivially formed and added. In order to project the right-hand side of (7) onto U_N a Fourier transform of the grid values is employed, leading to the spectral coeffcients. However, the product of two functions in U_N (such as u and u_x) lies in U_{2N}, and if it is evaluated on the grid J_N and transformed back to get the spectral representation, the high modes ($N+1$ to $2N$) will be aliased with the lower modes. This spurious transfer of energy from high to low modes is a potential source of (nonlinear) instability in the scheme. If we use the grid J_{2N} instead, we get the correct coefficients of the product term in U_{2N} and therefore the correct projection onto U_N. But for this purpose, it is sufficient to employ grid $J_{3N/2}$ where aliasing occurs only in the frequency range from $N+1$ to $3N/2$ and an alias free projection onto U_N will be computed (see [1]). This is the smallest grid to guarantee an alias-free computation and will be chosen in efficient computations. In summary, a complete time step will consist of the computation of the right-hand side of (7) on grid $J_{3N/2}$ amounting to Fourier transforms of u, v, z and their gradients, in a total of 9 fields transformed. The products are then computed on the grid, and the right-hand sides of (7) are transformed back. The system can then be solved and the time step completed. The computational costs are dominated by the 12 transforms per time step, on a grid of size $3N/2 \times 3N/2$. Therefore, the computational costs will be of the order $\vartheta(27N^2 \log_2(3N/2))$.

4. THE NONLINEAR GALERKIN METHOD

Nonlinear Galerkin methods are motivated by the theory of Approximate Inertial Manifolds (e.g. [9, 14, 12]). In principle, they can be used for equations of the form

$$\frac{dU}{dt} + AU + G(U) = f, \tag{15}$$

where A is a positive self-adjoint linear operator with compact inverse and G includes the nonlinearities. In our problem, A is the (diagonal) diffusion operator, $U = (u, v, z)$ and the

eigenfunctions of A (which form a complete set) are the Fourier modes. For $P_{N/2}$ defined as the projection (of H_{per}—the solution space) onto the space of low modes $U_{N/2}$ and $Q_N = I - P_{N/2}$ (I is the identity operator) Eq. (15) is equivalent to the following system,

$$\frac{dp}{dt} + Ap + P_{N/2}G(p+q) = P_{N/2}f \tag{16}$$

$$\frac{dq}{dt} + Aq + Q_N G(p+q) = Q_N f, \tag{17}$$

where $p = P_{N/2}U$ and $q = Q_N U$. Neglecting q (the projection onto the high modes) leads to the usual Galerkin scheme:

$$\frac{dp}{dt} + Ap + P_{N/2}G(p) = P_{N/2}f. \tag{18}$$

In the theory of inertial manifolds, q is also taken into account, being approximated as a function of the low modes $q = \Phi(p)$, such that $\Phi(p) = A^{-1}(Q_N f - Q_N G(p)$ is an approximate inertial solution of the equation for the high modes (17). The graphic of the equation $q = \Phi(p)$ is called the approximate inertial manifold. This diagnostic value of the high modes is employed in the formulation of the nonlinear Galerkin scheme:

$$\frac{dp}{dt} + Ap + P_{N/2}G(p + \Phi(p)) = P_{N/2}f. \tag{19}$$

For a practical and efficient method, some simplifications are necessary. First, instead of defining $Q_N = I - P_{N/2}$ we need to restrict it to a projection onto a finite dimensional subspace. We take $Q_N = P_N - P_{N/2}$. The second simplification we adopt is in the evaluation of $G(p + \Phi(p))$ in which we will not consider the interaction between high modes. Introducing the notation

$$\begin{aligned} u_N &= u_L + u_H, \quad (u_L = P_{N/2}(u), u_H = Q_N(u)) \\ v_N &= v_L + v_H, \quad (v_L = P_{N/2}(v), v_H = Q_N(v)) \\ z_N &= z_L + z_H, \quad (z_L = P_{N/2}(z), z_H = Q_N(z)) \\ F_N &= F_L + F_H, \quad (F_L = P_{N/2}F, F_H = Q_N F), \end{aligned} \tag{20}$$

we obtain the following systems, which define our nonlinear Galerkin method. We have

$$\begin{aligned} \frac{\partial u_L}{\partial t} - v_L + \frac{\partial z_L}{\partial x} - \nu_0 \Delta u_L &= s_1 \\ \frac{\partial v_L}{\partial t} + u_L + \frac{\partial z_L}{\partial y} - \nu_0 \Delta v_L &= s_2 \\ \frac{\partial z_L}{\partial t} + z_0\left(\frac{\partial u_L}{\partial x} + \frac{\partial v_L}{\partial y}\right) - \kappa_0 \Delta z_L &= s_3 \end{aligned} \tag{21}$$

$$\begin{aligned} -v_H + \frac{\partial z_H}{\partial x} - \nu_0 \Delta u_H &= s_4 \\ u_H + \frac{\partial z_H}{\partial y} - \nu_0 \Delta v_H &= s_5 \\ z_0\left(\frac{\partial u_H}{\partial x} + \frac{\partial v_H}{\partial y}\right) - \kappa_0 \Delta z_H &= s_6, \end{aligned} \tag{22}$$

where the right-hand sides are given by

$$
\begin{aligned}
s_1 &= P_{N/2}(-u_L(u_L+u_H)_x - u_H u_{L,x} - v_L(u_L+u_H)_y - v_H u_{L,y}) \\
s_2 &= P_{N/2}(-u_L(v_L+v_H)_x - u_H v_{L,x} - v_L(v_L+v_H)_y - v_H v_{L,y}) \qquad (23)\\
s_3 &= P_{N/2}(F - (z_L(u_L+u_H))_x - (z_H u_L)_x - (z_L(v_L+v_H))_y - (z_H v_L)_y)
\end{aligned}
$$

$$
\begin{aligned}
s_4 &= Q_N(-u_L u_{L,x} - v_L u_{L,y}) \\
s_5 &= Q_N(-u_L v_{L,x} - v_L v_{L,y}) \qquad (24)\\
s_6 &= Q_N(F - (z_L u_L)_x - (z_L v_L)_y).
\end{aligned}
$$

One could also keep the temporal derivatives in the equations for the high modes (22), using it in prognostic form. We choose the diagnostic form (22).

This nonlinear scheme is expected to lead to a better precision than the corresponding Galerkin method on $U_{N/2}$, since it includes the nonlinear interactions with the diagnostic values of the high modes. On the other hand, because of the definition of Q_N as the projection onto U_N, complementary to $P_{N/2}$, the method should not be more precise than the Galerkin scheme on U_N. The atractivity of the idea will be to achieve almost the same precision of the Galerkin scheme on U_N, with lower computational costs.

The numerical discretization of the nonlinear Galerkin method will follow the same lines of the Galerkin scheme. We employ a semi-implicit temporal discretization on a pseudo-spectral method, applying spectral transforms to compute nonlinear products. The projections on the high and low modes spaces will be computed exactly in an alias free scheme.

We have

$$
\begin{aligned}
u_L^{n+1} - \Delta t v_L^{n+1} + \Delta t z_{L,x}^{n+1} - \nu_0 \Delta t \Delta u_L^{n+1} &= S_1^{n-1} + 2\Delta t s_1^{n,n+1} \\
v_L^{n+1} + \Delta t u_L^{n+1} + \Delta t z_{L,y}^{n+1} - \nu_0 \Delta t \Delta v_L^{n+1} &= S_2^{n-1} + 2\Delta t s_2^{n,n+1} \qquad (25)\\
z_L^{n+1} + z_0 \Delta t \big(u_{L,x}^{n+1} + v_{L,y}^{n+1}\big) - \kappa_0 \Delta t \Delta z_L^{n+1} &= S_3^{n-1} + 2\Delta t s_3^{n,n+1}
\end{aligned}
$$

$$
\begin{aligned}
-v_H^{n+1} + z_{H,x}^{n+1} - \nu_0 \Delta u_H^{n+1} &= s_4^n = Q_N\big(-u_L^n u_{L,x}^n - v_L^n u_{L,y}^n\big) \\
u_H^{n+1} + z_{H,y}^{n+1} - \nu_0 \Delta v_H^{n+1} &= s_5^n = Q_N\big(-u_L^n v_{L,x}^n - v_L^n v_{L,y}^n\big) \qquad (26)\\
z_0\big(u_{H,x}^{n+1} + v_{H,y}^{n+1}\big) - \kappa_0 \Delta z_H^{n+1} &= s_6^n = Q_N\big(F - \big(z_L^n u_L^n\big)_x - \big(z_L^n v_L^n\big)y\big),
\end{aligned}
$$

with superscripts referring to time instants. The linear terms at time t^{n-1} in the equation for low modes are given by

$$
\begin{aligned}
S_1^{n-1} &= u_L^{n-1} + \Delta t v_L^{n-1} - \Delta t z_{L,x}^{n-1} + \nu_0 \Delta t \Delta u_L^{n-1} \\
S_2^{n-1} &= v_L^{n-1} - \Delta t u_L^{n-1} - \Delta t z_{L,y}^{n-1} + \nu_0 \Delta t \Delta v_L^{n-1} \qquad (27)\\
S_3^{n-1} &= z_L^{n-1} - \Delta t z_0\big(u_{L,x}^{n-1} + v_{L,y}^{n-1}\big) + \kappa_0 \Delta t \Delta z_L^{n-1}.
\end{aligned}
$$

The nonlinear terms are discretized as

$$
\begin{aligned}
s_1^{n,n+1} &= P_{N/2}\big(-u_L^n\big(u_L^n + u_H^{n+1}\big)_x - u_H^{n+1} u_{L,x}^n - v_L^n\big(u_L^n + u_H^{n+1}\big)_y - v_H^{n+1} u_{L,y}^n\big) \\
s_2^{n,n+1} &= P_{N/2}\big(-u_L^n\big(v_L^n + v_H^{n+1}\big)_x - u_H^{n+1} v_{L,x}^n - v_L^n\big(v_L^n + v_H^{n+1}\big)_y - v_H^{n+1} v_{L,y}^n\big) \qquad (28)\\
s_3^{n,n+1} &= P_{N/2}\big(F - \big(z_L^n\big(u_L^n + u_H^{n+1}\big)\big)_x - \big(z_H^{n+1} u_L^n\big)_x - \big(z_L^n\big(v_L^n + v_H^{n+1}\big)\big)_y - \big(z_H^{n+1} v_L^n\big)y\big).
\end{aligned}
$$

The equations are transformed into equations for the spectral coefficients of low modes ($(k, l) \in U_{N/2}$),

$$\begin{bmatrix} 1+\nu_0\gamma_{kl} & -\Delta t & \mathbf{i}k\Delta t \\ \Delta t & 1+\nu_0\gamma_{kl} & \mathbf{i}l\Delta t \\ \mathbf{i}kz_0\Delta t & \mathbf{i}lz_0\Delta t & 1+\kappa_0\gamma_{kl} \end{bmatrix} \cdot \begin{bmatrix} \hat{u}_{L,kl}^{n+1} \\ \hat{v}_{L,kl}^{n+1} \\ \hat{z}_{L,kl}^{n+1} \end{bmatrix} = \begin{bmatrix} \hat{S}_{1,kl}^{n-1} + 2\Delta t\hat{s}_{1,kl}^{n,n+1} \\ \hat{S}_{2,kl}^{n-1} + 2\Delta t\hat{s}_{2,kl}^{n,n+1} \\ \hat{S}_{3,kl}^{n-1} + 2\Delta t\hat{s}_{3,kl}^{n,n+1} \end{bmatrix}, \tag{29}$$

and for the high modes ($(k, l) \in W_N$),

$$\begin{bmatrix} \nu_0\lambda_{kl} & -1 & \mathbf{i}l \\ 1 & \nu_0\lambda_{kl} & \mathbf{i}k \\ \mathbf{i}lz_0 & \mathbf{i}kz_0 & \kappa_0\lambda_{kl} \end{bmatrix} \cdot \begin{bmatrix} \hat{u}_{H,kl}^{n+1} \\ \hat{v}_{H,kl}^{n+1} \\ \hat{z}_{H,kl}^{n+1} \end{bmatrix} = \begin{bmatrix} \hat{s}_{4,kl}^{n} \\ \hat{s}_{5,kl}^{n} \\ \hat{s}_{6,kl}^{n} \end{bmatrix}. \tag{30}$$

Assuming the right-hand side of (30) to be known, it can be solved as system (8) leading to the new values of $\hat{u}_H$, $\hat{v}_H$, and $\hat{z}_H$. These high modes are used in the right-hand side of (29), which can then be solved as (8) for each $(k, l) \in U_{N/2}$. The evaluation of the right-hand side of the systems for the low and high modes involves nonlinear terms. For this purpose, we employ spectral transforms, choosing appropriate auxiliary grids to guarantee alias-free results and an efficient scheme.

We need to compute products of low modes with either low or high modes, and then project the results onto $U_{N/2}$ or W_N . For functions ϕ, θ in $U_{N/2}$ the product $\phi\theta$ lies in U_N. With an $N \times N$ grid, a two-dimensional FFT will provide the correct values of all spectral coefficients of $\phi\theta$, and therefore the two projections $P_{N/2}(\phi\theta)$ and $Q_N(\phi\theta)$ are obtained alias free. For ϕ in $U_{N/2}$ and θ in W_N the product lies in $U_{3N/2}$. With an $N \times N$ grid, only the modes with wave number up to $N/2$ will be obtained alias free. But that's all we need to have $P_{N/2}(\phi\theta)$ computed correctly. Therefore, an $N \times N$ grid will be sufficient for obtaining all the right-hand sides with no aliasing.

In summary, a time step of the method will proceed as follows, assuming $\hat{u}_{L,kl}^{n-1}$, $\hat{v}_{L,kl}^{n-1}$, $\hat{z}_{L,kl}^{n-1}$ for $(k, l) \in I_{N/2}$ and the values of u_L^n, $u_{L,x}^n$, $u_{L,y}^n$, v_L^n, $v_{L,x}^n$, $v_{L,y}^n$, z_L^n, $z_{L,x}^n$, $z_{L,y}^n$, on a $N \times N$ grid to be known.

(a) Computation of high modes at t_{n+1}:

($a1$) Compute s_4^n, s_5^n, s_6^n on the grid and apply Fourier transforms to get $\hat{s}_{4,kl}^n$, $\hat{s}_{5,kl}^n$, $\hat{s}_{6,kl}^n$ for $(k, l) \in I_N$.

($a2$) Solve for $\hat{u}_H^{n+1}$, $\hat{v}_H^{n+1}$, $\hat{z}_H^{n+1} \in W_N$.

($a3$) Compute the values of u_H^{n+1}, $u_{H,x}^{n+1}$, $u_{H,y}^{n+1}$, v_H^{n+1}, $v_{H,x}^{n+1}$, $v_{H,y}^{n+1}$, z_H^{n+1}, $z_{H,x}^{n+1}$, $z_{H,y}^{n+1}$, on the $N \times N$ grid.

(b) Computation of the low modes at t_{n+1}:

($b1$) Using the values computed in step ($a3$) evaluate $s_i^{n,n+1}$, $i = 1, 2, 3$ on the $N \times N$ grid and use FFTs to get the corresponding spectral coefficients of their projection onto $U_{N/2}$.

($b2$) Complete the right-hand side with the linear terms S_i^{n-1}, $i = 1, 2, 3$ and solve the equations for $\hat{u}_L^{n+1}$, $\hat{v}_L^{n+1}$, and $\hat{z}_L^{n+1}$ in $U_{N/2}$.

($b3$) Generate the values of u_L^{n+1}, $u_{L,x}^{n+1}$, $u_{L,y}^{n+1}$, v_L^{n+1}, $v_{L,x}^{n+1}$, $v_{L,y}^{n+1}$, z_L^{n+1}, $z_{L,x}^{n+1}$, $z_{L,y}^{n+1}$, on the $N \times N$ grid, which shall be used in the following time step.

Altogether, 24 two-dimensional FFT's on $N \times N$ grids will be carried out, 12 when computing the high modes and 12 for the low modes. The total computational work will be of

the order of $\vartheta(24N^2\log_2(N^2))$. This work compares favorably to the $\vartheta(27N^2\log_2(3N^2/2))$ of the linear Galerkin method (on U_N), which needed a $3N/2 \times 3N/2$ grid for alias-free results.

We also consider the possibility of freezing the high modes coefficients for several time steps. In this case, the total computational cost consists essentially of the costs for the low modes equation, being reduced almost by a factor of two.

5. LINEAR STABILITY ANALYSIS

We present a linear stability analysis for the pseudo-spectral linear Galerkin method described in Section 2. The discrete equations are given by

$$\frac{\vec{V}_N^{n+1}-\vec{V}_N^{n-1}}{2\Delta t}+P_N\big((\vec{V}_N^{n}\cdot\nabla)\vec{V}_N^{n}\big)+\vec{k}\times\frac{\vec{V}_N^{n+1}+\vec{V}_N^{n-1}}{2}+\nabla\left(\frac{z_N^{n+1}+z_N^{n-1}}{2}\right)-\nu_0\Delta\left(\frac{\vec{V}_N^{n+1}+\vec{V}_N^{n-1}}{2}\right)=0 \tag{31}$$

$$\frac{z_N^{n+1}-z_N^{n-1}}{2\Delta t}+z_0\nabla\cdot\left(\frac{\vec{V}_N^{n+1}+\vec{V}_N^{n-1}}{2}\right)+P_N\big(\nabla\cdot\big(z_N^{n}\vec{V}_N^{n}\big)\big)-\kappa_0\Delta\left(\frac{z_N^{n+1}+z_N^{n-1}}{2}\right)=F_N,$$

with $\vec{V}_N=(u_N,v_N)\in U_N\times U_N$ and $z_N\in U_N$.
Linearizing system (31) around a state with constant velocity $\vec{U}$ leads to

$$\frac{\vec{V}_N^{n+1}-\vec{V}_N^{n-1}}{2\Delta t}+\big((\vec{U}\cdot\nabla)\vec{V}_N^{n}\big)+\vec{k}\times\frac{\vec{V}_N^{n+1}+\vec{V}_N^{n-1}}{2}+\nabla\left(\frac{z_N^{n+1}+z_N^{n-1}}{2}\right)-\nu_0\Delta\left(\frac{\vec{V}_N^{n+1}+\vec{V}_N^{n-1}}{2}\right)=0 \tag{32}$$

$$\frac{z_N^{n+1}-z_N^{n-1}}{2\Delta t}+z_0\nabla\cdot\left(\frac{\vec{V}_N^{n+1}+\vec{V}_N^{n-1}}{2}\right)+\vec{U}\cdot\nabla z_N^{n}-\kappa_0\Delta\left(\frac{z_N^{n+1}+z_N^{n-1}}{2}\right)=F_N.$$

We build the scalar product of the first equation in (32) with $z_0(\vec{V}_N^{n+1}+\vec{V}_N^{n-1})$ and add it to the scalar product of the second equation with $(z_N^{n+1}+z_N^{n-1})$. This, after simplifying the terms originating from the geopotential and of the divergence, in presence of the periodic boundary conditions, leads to the expression

$$\begin{aligned}&z_0\big|\vec{V}_N^{n+1}\big|^2+\big|z_N^{n+1}\big|^2+z_0\nu_0\Delta t\big\|\vec{V}_N^{n+1}+\vec{V}_N^{n-1}\big\|^2+\kappa_0\Delta t\big\|z_N^{n+1}+z_N^{n-1}\big\|^2\\&\quad=z_0\big|\vec{V}_N^{n-1}\big|^2+\big|z_N^{n-1}\big|^2-2z_0\Delta t\big\langle\vec{U}\cdot\nabla\vec{V}_N^{n},\vec{V}_N^{n+1}+\vec{V}_N^{n-1}\big\rangle\\&\qquad-2\Delta t\big\langle\vec{U}\cdot\nabla z_N^{n},z_N^{n+1}+z_N^{n-1}\big\rangle+2\Delta t\big\langle F_N,z_N^{n+1}+z_N^{n-1}\big\rangle,\end{aligned}$$

where we have $\vec{V}=(u,v)$, $\langle u,v\rangle=\int_\Omega uv$, $|\vec{V}|^2=|u|^2+|v|^2$, $|u|^2=\langle u,u\rangle$, $\|\vec{V}\|^2=\|u\|^2+\|v\|^2$, $\|\vec{u}\|^2=|u_x|^2+|u_y|^2$.
We add $z_0|\vec{V}_N^{n}|^2+|z_N^{n}|^2$ to both sides of last equation, in order to get

$$\begin{aligned}&G^{n+1}+z_0\nu_0\Delta t\big\|\vec{V}_N^{n+1}+\vec{V}_N^{n-1}\big\|^2+\kappa_0\Delta t\big\|z_N^{n+1}+z_N^{n-1}\big\|^2\\&\quad=G^{n}+NLT+2\Delta t\big\langle F_N,z_N^{n+1}+z_N^{n-1}\big\rangle\end{aligned} \tag{33}$$

where

$$G^n = z_0\left|\vec{V}_N^n\right|^2 + z_0\left|\vec{V}_N^{n-1}\right|^2 + \left|z_N^n\right|^2 + \left|z_N^{n-1}\right|^2, \tag{34}$$

$$NLT = -2z_0\Delta t\left\langle \vec{U}\ .\ \nabla\vec{V}_N^n, \vec{V}_N^{n+1} + \vec{V}_N^{n-1}\right\rangle - 2\Delta t\left\langle \vec{U}\ .\ \nabla z_N^n, z_N^{n+1} + z_N^{n-1}\right\rangle. \tag{35}$$

For estimating the term with the mass forcing, we first observe that the integration of (32) on the whole domain leads to $\int_\Omega z_N^{n+1} - z_N^{n-1} = 2\Delta t\int_\Omega F_N$. This provides a limitation of the constant part of the solution (denoted by $\overline{z}_N^n$) of the type

$$\left|\overline{z}_N^n\right| \le Cn\Delta t|F_N|,$$

where C is a positive constant. We now split $z_N^n = \overline{z}_N^n + \tilde{z}_N^n$, where $\tilde{z}_N^n$ has zero mean value. We then have

$$\begin{aligned}
2\Delta t\left\langle F_N, z_N^{n+1} + z_N^{n-1}\right\rangle &= 2\Delta t\left(\left\langle F_N, \overline{z}_N^{n+1} + \overline{z}_N^{n-1}\right\rangle + \left\langle F_N, \tilde{z}_N^{n+1} + \tilde{z}_N^{n-1}\right\rangle\right)\\
&\le 2\Delta t|F_N|\left(\left|\tilde{z}_N^{n+1} + \tilde{z}_N^{n-1}\right| + \left|\overline{z}_N^{n+1} + \overline{z}_N^{n-1}\right|\right)\\
&\le 2\Delta t|F_N|\left(2Cn\Delta t|F_N| + C_1\left\|z_N^{n+1} + z_N^{n-1}\right\|\right)\\
&\le \frac{1}{2}\kappa_0\Delta t\left\|z_N^{n+1} + z_N^{n-1}\right\|^2 + 2\Delta t\left(\frac{C_1^2}{\kappa_0} + 2Cn\Delta t\right)|F_N|^2.
\end{aligned} \tag{36}$$

In the last estimates we have used a Poincaré inequality (C_1 is a positive constant) and the algebraic inequality ($ab \le a^2/4 + b^2$). We also observe, using the periodic boundary conditions, that

$$-2z_0\Delta t\left\langle \vec{U}\ .\ \nabla\vec{V}_N^n, \vec{V}_N^{n-1}\right\rangle = 2z_0\Delta t\left\langle \vec{U}\ .\ \nabla\vec{V}_N^{n-1}, \vec{V}_N^n\right\rangle.$$

Similarly, $-2\Delta t\langle\vec{U}\ .\ \nabla z_N^n, z_N^{n-1}\rangle = 2\Delta t\langle\vec{U}\ .\ \nabla z_N^{n-1}, z_N^n\rangle$. Substituting in (35) we have

$$\begin{aligned}
NLT &= -2z_0\Delta t\left\langle \vec{U}\,.\nabla\vec{V}_N^n, \vec{V}_N^{n+1}\right\rangle - 2\Delta t\left\langle \vec{U}\,.\nabla z_N^n, z_N^{n+1}\right\rangle\\
&\quad + 2z_0\Delta t\left\langle \vec{U}\,.\nabla\vec{V}_N^{n-1}, \vec{V}_N^n\right\rangle + 2\Delta t\left\langle \vec{U}\,.\nabla z_N^{n-1}, z_N^n\right\rangle\\
&= H^{n+1} - H^n,
\end{aligned}$$

with

$$H^n = -2z_0\Delta t\left\langle \vec{U}\,.\,\Delta\vec{V}_N^{n-1}, \vec{V}_N^n\right\rangle - 2\Delta t\left\langle \vec{U}\ .\ \Delta z_N^{n-1}, z_N^n\right\rangle.$$

Using the last expressions in (33) we obtain

$$\begin{aligned}
G^{j+1} &\le G^{j+1} + z_0\nu_0\Delta t\left\|\vec{V}_N^{j+1} + \vec{V}_N^{j-1}\right\|^2 + \frac{1}{2}\kappa_0\Delta t\left\|z_N^{j+1} + z_N^{j-1}\right\|^2\\
&\le G^j + H^{j+1} - H^j + 2\Delta t\left(\frac{C_1^2}{\kappa_0} + 2C_n\Delta t\right)|F_N|^2.
\end{aligned} \tag{37}$$

Adding up for j from 1 to n results in

$$G^{n+1} \le G^1 + H^{n+1} - H^1 + 2\left(\frac{C_1^2}{\kappa_0} + 2C_n\Delta t\right)n\Delta t|F_N|^2. \tag{38}$$

We can now limit

$$\begin{aligned}|H^j| &\le 2z_0\Delta t|\vec{U}|_\infty\|\vec{V}_N^{j-1}\|\,|\vec{V}_N^j| + 2\Delta t|\vec{U}|_\infty\|z_N^{j-1}\|\,|z_N^j| \\ &\le 2z_0\sqrt{2}N\Delta t|\vec{U}|_\infty|\vec{V}_N^{j-1}|\,|\vec{V}_N^j| + 2\sqrt{2}N\Delta t|\vec{U}|_\infty|z_N^{j-1}|\,|z_N^j| \\ &\le \sqrt{2}N\Delta t|\vec{U}|_\infty\big(z_0|\vec{V}_N^{j-1}|^2 + z_0|\vec{V}_N^j|^2 + |z_N^{j-1}|^2 + |z_N^j|^2\big) \\ &\le (\sqrt{2}N\Delta t|\vec{U}|_\infty)G^j.\end{aligned}$$

It then follows from (38) that for any $T > 0$ fixed and $n\Delta t \le T$:

$$(1-\sqrt{2}N\Delta t|\vec{U}|_\infty)G^{n+1} \le (1+\sqrt{2}N\Delta t|\vec{U}|_\infty)G^1 + 2\left(\frac{C_1^2}{\kappa_0} + 2CT\right)T|F_N|^2.$$

So, if $\Delta t < \frac{1}{\sqrt{2}N|\vec{U}|_\infty}$ we have

$$G^{n+1} \le \frac{1+\sqrt{2}N\Delta t|\vec{U}|_\infty}{1-\sqrt{2}N\Delta t|\vec{U}|_\infty}G^1 + \frac{2\big(\kappa_0^{-1}C_1^2 + 2CT\big)T}{1-\sqrt{2}N\Delta t|\vec{U}|_\infty}|F_N|^2. \tag{39}$$

We have therefore proved:

PROPOSITION. *If Δt obeys the CFL condition $\Delta t < \frac{1}{\sqrt{2}N|\vec{U}|_\infty}$, then $G^n \le C_2G^1 + C_3T|F_N|^2$, where C_2 and C_3 are constants obtained from the inequality* (39), *T a fixed time, and G^n is given by* (34). *It follows that the scheme is linearly stable.*

Observation. The same proof above applies for the nonlinear galerkin method (with projections $P_{N/2}$ and Q_N). It will lead to the stability condition $\Delta t < \frac{1}{\sqrt{2}\frac{N}{2}|\vec{U}|_\infty}$, indicating that the CFL condition for the nonlinear galerkin method is given by the equation for the low modes. This agrees with the fact that the equation for the high modes is employed as a diagnostic equation.

An analysis in which the linearization is done around a spatially variable basic state and the interaction between low and high modes is present in the resulting system is presented in [2].

6. NUMERICAL RESULTS

We present in this section numerical results obtained with the nonlinear Galerkin method, which are compared with results from the pseudo-spectral Galerkin method described in section (4).

We first consider a smooth solution with initial state given by

$$\begin{aligned}u &= \frac{1}{100\nu_0 z_0}\sin y \\ v &= \frac{1}{100 z_0}\sin y \\ z &= \frac{1}{100\nu_0 z_0}\cos y.\end{aligned} \tag{40}$$

This initial state is nearly stationary for a forcing term $F = \cos y/100$. We adopt the values $z_0 = 8$, $\nu_0 = \kappa_0 = 1/48$ in the nondimensional model and integrate the equations with a

forcing term of the order of $1/10$. (Our choice of parameters correspond, on a domain with typical lenght $L = 1080$ Km, to heights of the order of 8 km and Coriolis factor $f^{-1} = 10800$ s.)

We first tested the method in different situations, also using the high-mode equations in prognostic form. This brings no advantage over the diagnostic form (22) described in the text. We have experimented with the possibility of freezing the high modes for several time steps, with different test cases, resolutions, and time steps. The number of steps the high modes can be kept unchanged (without significant loss in precision) varied from around 10 to more than 50. The results indicate that the high modes don't need to be updated as frequently as the low modes (this is in accordance with results of [4] for the Navier–Stokes equations).

In Fig. 1 we display results obtained with the nonlinear method (NLG) (with the high modes frozen for every 30 time steps) and with the pseudo-spectral Galerkin (GL) method at several resolutions. In these tests, a small time step ($\Delta t = 9$ min.) was employed, in order to keep the time truncation errors very small. In this way we can observe better the differences in spatial resolution of both methods. The results, after 200 time steps, show that the nonlinear method at a resolution N, has an accuracy close to that of the corresponding Galerkin method at same truncation and better than the accuracy of the Galerkin method at resolution $N/2$, as expected. The lower cost of the nonlinear Galerkin method (compared to the Galerkin method at same truncation) makes the scheme interesting. In Fig. 1 the geopotential field is shown for $N = 8$, 16, and 32. In each graphic, we compare the linear Galerkin method (GL) at truncation N either to the same method at truncation $N/2$ or to the nonlinear method (NLG) at truncation N.

The relative computational efficiency of the schemes can be seen in Fig. 2. For N from 16 to 128, we display the CPU times for the whole integration (200 time steps) with the Linear Galerkin (LG), nonlinear Galerkin (NLG), and nonlinear Galerkin with high modes frozen (NLGF). We observe that the nonlinear method (NLG) is faster than LG (for the same resolution) by around 12%, and when the high modes are frozen it is almost two times faster.

We also consider less smooth solutions by taking an extra forcing term of the form $F_1(x, y) = \delta_r(x)\delta_r(y)$ where

$$\delta_r = \begin{cases} \frac{r}{4}(1. + \cos(rs + \pi)), & |s| \le \pi(1 + r) \\ 0, & \text{otherwise;} \end{cases}$$

(see Fig. 3). This forcing approximates a local (dirac type) mass source at the center of the domain. The results for the geopotential after 200 time steps (departing from the same initial state as before) are displayed in Fig. 4. We present results for the linear Galerkin method and for the nonlinear method with frozen high modes for every 30 steps, for $N = 8$, 16, 32, and 64. We can observe the same qualitative behavior of the previous example also in this case, in which at least 32 modes are necessary to provide a good resolution of the solution. Again, the nonlinear method at truncation N leads to an intermediate accuracy, between the ones of the Galerkin method at resolutions $N/2$ and N, closer to the latter (at lower cost).

We carried out several stability tests, confirming the CFL-type condition for stability (dependent on the maximal flow velocities). The numerical experiments also confirmed that the nonlinear method at truncation N is as stable as the Galerkin method at resolution $N/2$ (being able to employ time steps two times larger than the Galerkin method at truncation N). This is a potential advantage of the nonlinear method.

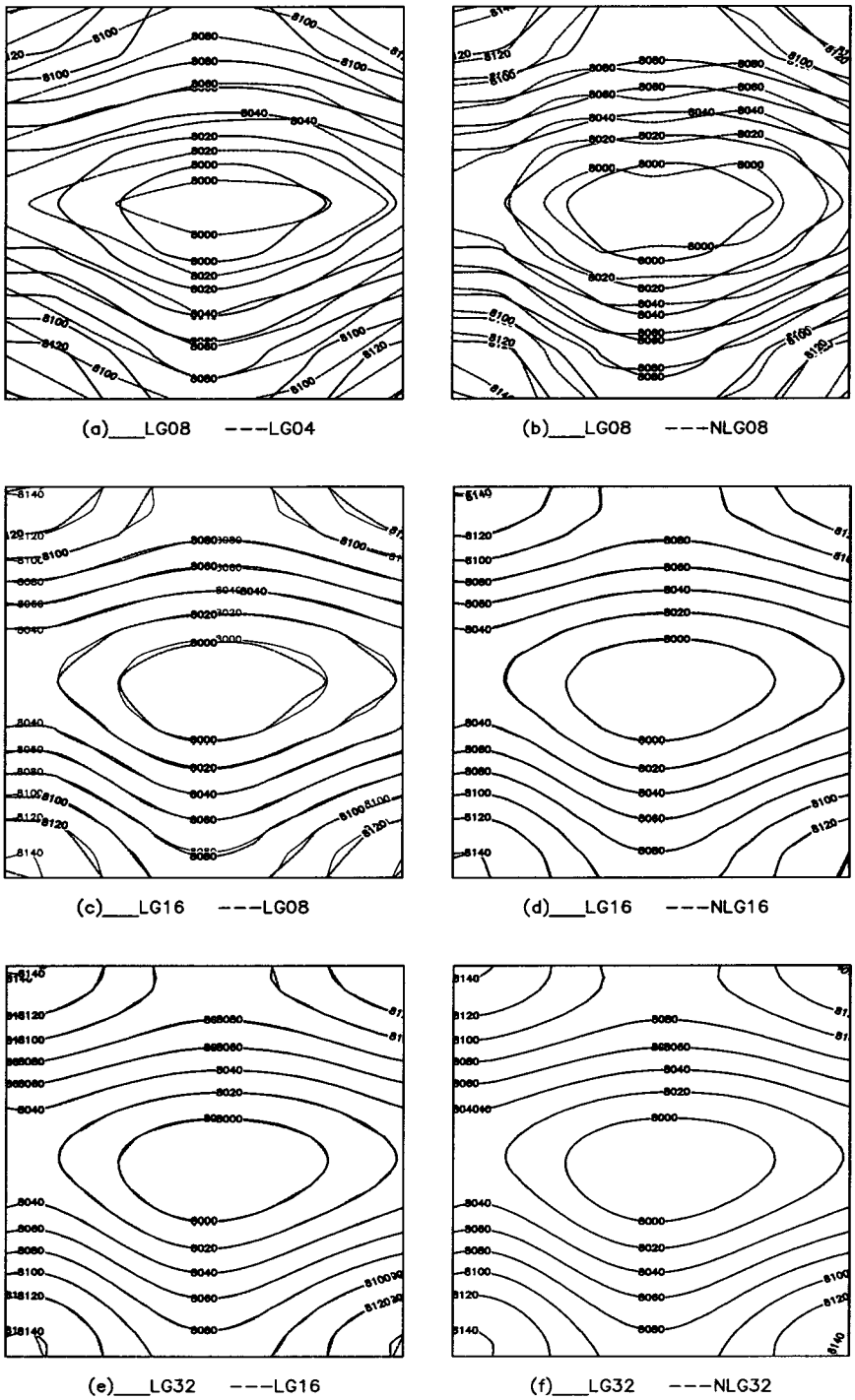

FIG. 1. Display of the geopotential field after 200 time steps for several resolutions. In the left (plots a, c, and e) we compare the results of the (linear) Galerkin method with truncations $N = 8$, 16, and 32 to the results of the same method at half resolution ($N = 4$, 8, and 16) respectively. In the right (plots b, d, and f) the Galerkin (GL) method (with truncations $N = 8$, 16, and 32) is compared to the non-linear method (NLG) at same resolution (truncations $N = 8$, 16, and 32, respectively).

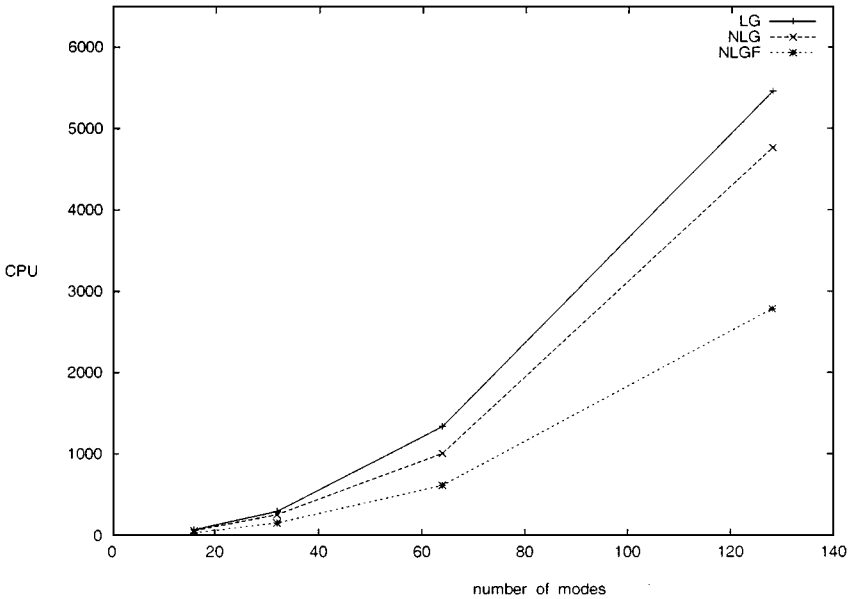

FIG. 2. CPU time.

7. A FEW REMARKS ON THE APPLICABILITY OF THE METHOD

The use of the shallow-water equations on a f-plane with periodic conditions and constant coefficients leads to some simplifications, which are not necessarily present in more realistic applications. We would like, therefore, to make some remarks pointing out how to handle more general situations. First, if the Coriolis factor is variable (as in the case of a β-plane or global models), its implicit treatment as done in this paper would not be feasible. The Coriolis terms should then be discretized explicitly (at time t^n). The products fu and fv would be evaluated on the grid and contribute to the right-hand sides of the equations. This modification has no significant impact either on the accuracy or on the stability of the method, and could have been used in the present model.

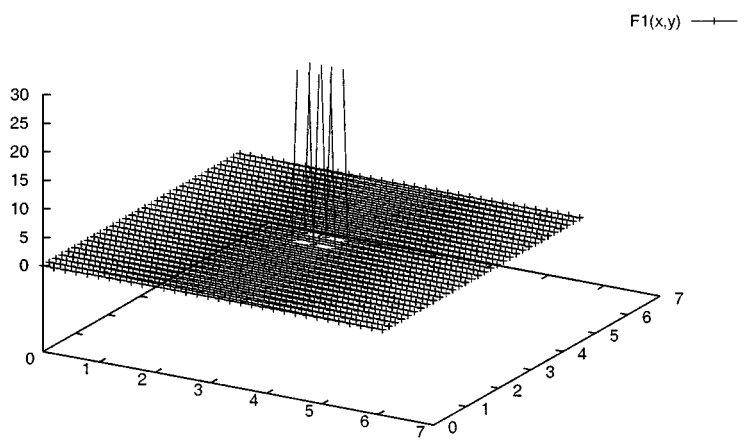

FIG. 3. Shallow water equations: forcing F_1.

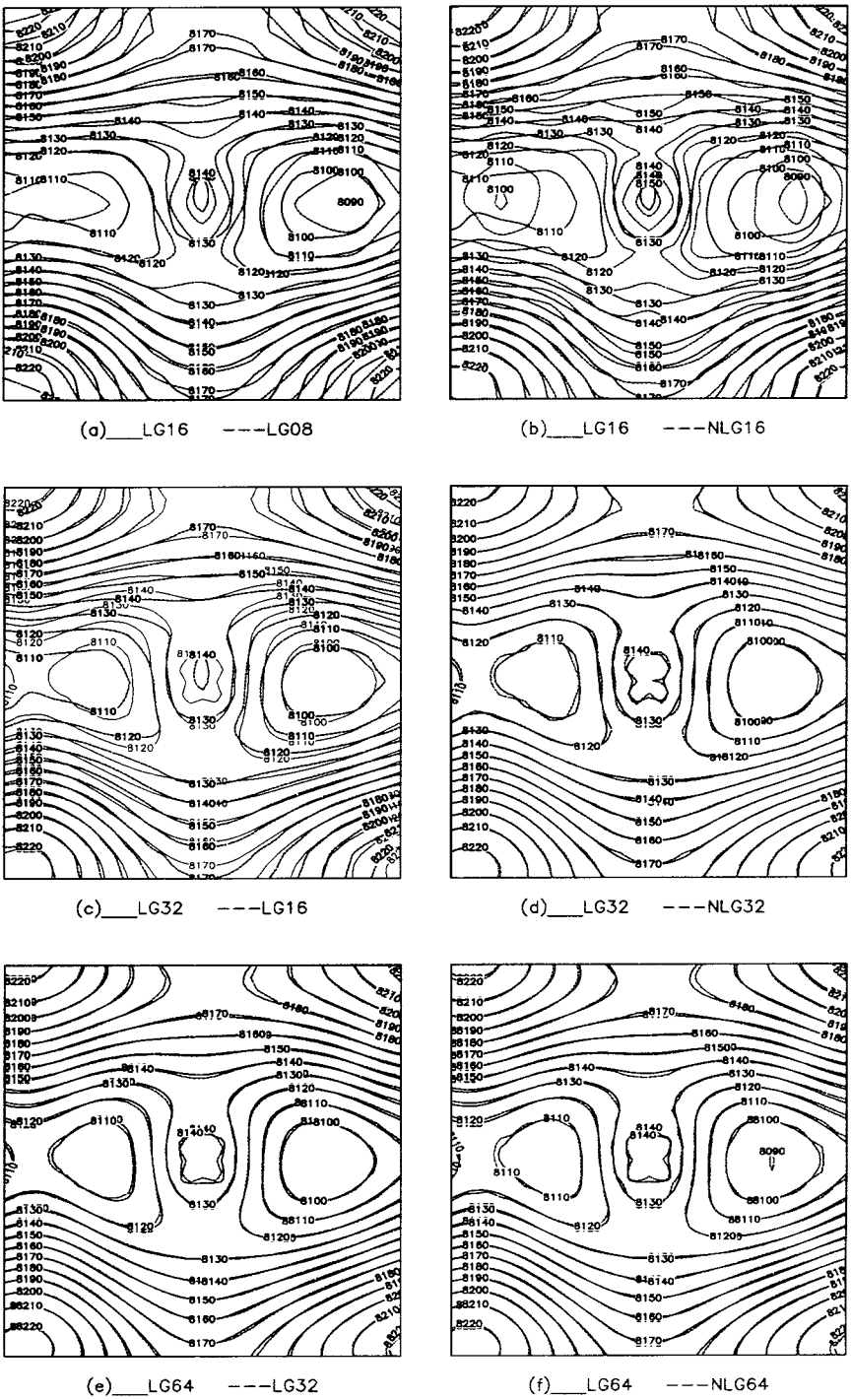

FIG. 4. Display of the geopotential field after 200 time steps with a local mass source. In the left (plots a, c, and e) we compare the results of the (linear) Galerkin method with truncations $N = 16$, 32, and 64 to the results of the same method at half resolution ($N = 8$, 16, and 32), respectively. In the right (plots b, d, and f) the Galerkin (LG) method (with truncations $N = 16$, 32, and 64) is compared to the nonlinear method (NLG) at same resolution (truncations $N = 16$, 32, and 64, respectively).

The presence of variable horizontal diffusion coefficients would also prevent the implicit treatment of the Laplacian terms, in a spectral framework as employed in this paper. The solution for this, is to discretize the diffusion terms explicitly (at time t^{n-1}, as in an Eulerian method with time step $2\Delta t$). This change, however, will change the stability properties of the schemes (both of the Galerkin and of the nonlinear Galerkin method). An explicit treatment of diffusion is usual in oceanic applications of shallow water flow (see e.g. [3]) and the stability conditions seem to be acceptable from a pratical point of view. In global weather forecasting models, however, it is a common practice to employ constant horizontal diffusion coefficients (see e.g. [16]), and in this case, the related terms can be treated implicitly as described here. Another aspect is the possible presence of a variable orography, instead of the flat bottom considered in this paper. The extra terms, involving the gradient of the orography, could be evaluated explicitly (at time t^n) on the grid, with the linear part of the mass divergence term still handled implicitly as in the present model. With these modifications, the scheme should present similar numerical properties, concerning accuracy and stability.

The application of the nonlinear Galerkin method to a global spectral shallow-water model, involving the ideas discussed in this section, is a subject of our ongoing research.

8. CONCLUSIONS

We present a nonlinear (pseudo-spectral) Galerkin method for the shallow-water equations on bidimensional periodic domains, and compare it to a pseudo-spectral Galerkin method. Both schemes employ a semi-implicit second-order accurate time discretization, and have a CFL stability restriction governed by the flow velocities, and not by the fast gravity wave modes of the shallow-water equations. This stability condition is more restrictive (by a factor two) for the Galerkin method than for the nonlinear scheme. This fact is verified numerically and supported by a linear stability analysis, in which the stability criteria are derived. Both schemes are derived to be free of aliasing resulting from nonlinear interaction. With the linear Galerkin method, employing double Fourier expansions with N modes in each direction, this is achieved at the cost of using a $3N/2 \times 3N/2$ auxiliary grid in the spectral transforms. In the nonlinear Galerkin method we developed, a $N \times N$ grid is sufficient for alias-free computations. In this way, every time step of the nonlinear Galerkin method is faster than the corresponding step of the linear Galerkin method at same resolution. In the nonlinear method we can also freeze the high modes for some time steps, therefore reducing the computational work by a significant amount. Our numerical results indicate that the nonlinear Galerkin method, even with the high modes frozen for many steps, still achieves an accuracy comparable to that of the linear scheme at same resolution (at a considerable lower cost). This fact makes the approach potentially interesting for applications. We are currently investigating its use on spectral schemes for global shallow-water models, with atmospheric applications in view.

REFERENCES

1. C. Canuto, Y. Husaini, A. Quarteroni, and T. Zang, *Spectral Methods in Fluid Dynamics* (Springer-Verlag, New York, 1988).
2. J. W. Cardenas, Variedades Inerciais Aproximadas e Métodos de Galerkin não linear para as equações de água rasa (Approximate inertial manifolds and non-linear Galerkin methods for the shallow-water equations), PhD thesis (Universidade de São Paulo, S. Paulo, Brasil, 1999).

3. V. Casulli and E. Cattani, Stability, accuracy and efficiency of a semi-implicit method for three-dimensional shallow water flow, Comput. Math. Appl. **27**(4), 99 (1994).
4. A. Debussche, T. Dubois, and R. Temam (1993). The nonlinear Galerkin method: A multi-scale method applied to the simulation of homogeneos turbulent flows, Icase Report No. 93-93.
5. L., Dettori, D. Gottlieb, and R. Temam, A nonlinear Galerkin method: The two-level Fourier-collocation case, *J. Sci. Comput*. **10**(4), 371 (1995).
6. B. di Martino and P. Orenga, Résolution des équations de shallow water par la méthode de Galerkin non linéaire, M^2 AN *Math. Model. Numer. Anal*. **32**, 451 (1998).
7. T. Dubois, F. Jabertau, and R. Temam, Solution of the incompressible Navier–Stokes equations by the Nonlinear Galerkin Methods, *J. Sci. Comput*. **8**(2), 167 (1993).
8. E. Eliasen, B. Machenhauer, and E. Rasmussen, On a numerical method for integration of the hydrodynamical equations with a spectral representation of the horizontal fields, Report Nr. 2 (Institut for Teoretisk Meteorologi, Københavns Universitet, Haraldsgade 6, DK-2200 Copenhagen N, Denmark, 1970).
9. C. Foias, O. P. Manley, and R. Temam, Modelization of the iteration of small and large eddies in two dimensional turbulent flows M^2 AN *Math. Model. Numer. Anal*. **22**, 93 (1988).
10. D. Gottileb and R. Temam, Implementation of the Nonlinear Galerkin Method with pseudospectral (collocation) discretizations, *Appl. Numer. Math*. **12**, 119 (1993).
11. F. Jabertau, C. Rosier, and R. Temam, A Nonlinear Galerkin Method for the Navier-Stokes Equation, *Computer Meth. Appl. Mech. Eng.* **80**, 245 (1990).
12. D. E. Jones, L. G. Margolin, and E. S. Titi, On the effectiveness of the approximate inertial manifold—a computational study, *Thoret. Comput. Fluid Dynamics* **7**, 243 (1995).
13. E. Lorenz, Attractor Sets and Quasi-Geostrophic Equilibrium, *J. Atmos. Sci*. **37**, 1685 (1980).
14. M. Marion and R. Temam, Nonlinear Galerkin Methods, *SIAM J. Num. Anal*. **26**, 1139 (1989).
15. S. A. Orzag, Transform method for calculation of vector–coupled sums: Application to the spectral form of the vorticity equation, *J. Atmos. Sci*. **27**, 890 (1970).
16. H. Ritchie, C. Temperton, A. J. Simmons, M. Hortal, T. Davies, D. Dent and M. Hamrud, Implementation of the semi-Lagrangian method in a high-resolution version of the ECMWF foracast model, *Mon. Wea. Rev*. **123**, 489 (1995).