QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY
Q. J. R. Meteorol. Soc. **135**: 000–000 (2009)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/qj.487

RMetS
Royal Meteorological Society

A conservative Fourier-finite-element method for solving partial differential equations on the whole sphere

T. Dubos*
IPSL/Laboratoire de Météorologie Dynamique, École Polytechnique, Palaiseau, France

ABSTRACT: Solving transport equations on the whole sphere using an explicit time stepping and an Eulerian formulation on a latitude–longitude grid is relatively straightforward but suffers from the pole problem: due to the increased zonal resolution near the pole, numerical stability requires unacceptably small time steps. Commonly used workarounds such as near-pole zonal filters affect the qualitative properties of the numerical method. Rigorous solutions based on spherical harmonics have a high computational cost.

The numerical method we propose to avoid this problem is based on a Galerkin formulation in a subspace of a Fourier-finite-element spatial discretization. The functional space we construct provides quasi-uniform resolution and high-order accuracy, while the Galerkin formalism guarantees the conservation of linear and quadratic invariants. For N^2 degrees of freedom, the computational cost is $\mathcal{O}(N^2 \log N)$, dominated by the zonal Fourier transforms. This is more than with a finite-difference or finite-volume method, which costs $\mathcal{O}(N^2)$, and less than with a spherical harmonics method, which costs $\mathcal{O}(N^3)$. Differential operators with latitude-dependent coefficients are inverted at a cost of $\mathcal{O}(N^2)$.

We present experimental results and standard benchmarks demonstrating the accuracy, stability and efficiency of the method applied to the advection of a scalar field by a prescribed velocity field and to the incompressible rotating Navier–Stokes equations. The steps required to extend the method towards compressible flows and the Saint-Venant equations are described. Copyright © 2009 Royal Meteorological Society

AQ1

KEY WORDS •

Received 31 January 2009; Revised 3 July 2009; Accepted 4 July 2009

1. Introduction

Global weather and climate modelling require the numerical solution of partial differential equations on the whole sphere. A difficult part of this task is the discretization of the dynamical core, dealing with the transport of mass, momentum and various species. Ideally a numerical scheme should be accurate, stable and computationally efficient. In the context of climate studies, a crucial additional requirement is that it be conservative: the discretized system should enforce the exact conservation of a discrete approximation of the total mass, momentum and, if possible, energy and enstrophy. For an in-depth review of the evolution of dynamical cores, the reader is referred to Williamson (2007). Although a single optimal scheme has not emerged yet, most dynamical cores in use today use one of two methods: the finite-difference method (Arakawa, 1966; Sadourny, 1975; Arakawa and Lamb, 1981) or the spectral-transform method (Orszag, 1970; Swarztrauber, 1996), both using a structured latitude–longitude grid.

The finite-difference method emphasizes efficiency. For $\sim N$ grid points along latitude and longitude circles, an optimal cost of $\mathcal{O}(N^2)$ is achieved by approximating derivatives locally through finite-difference formulae. Carefully crafted finite-difference formulae conserve mass and one or two quantities from amongst angular momentum, energy and enstrophy (Arakawa, 1966; Sadourny, 1975; Arakawa and Lamb, 1981). On the sphere, apparently only second-order accurate formulae have been found that are also conservative. Fourth-order formulae exist on planar uniform grids that provide good accuracy on spherical quasi-uniform grids (Rancic *et al.*, 2008). Staggered grids avoid numerical instabilities due to computational modes, but the temporal stability is limited by a Courant–Friedrichs–Lewy (CFL) condition. This implies that the maximum allowed time step is proportional to the smallest grid interval. Near the Poles, zonal grid intervals are $\sim a/N^2$ with a the Earth radius, much smaller than the grid interval $\sim a/N$ near the Equator. This so-called 'pole problem' can be overcome by applying latitude-dependent zonal Fourier filters near the Poles. Although widely adopted, this fix remains 'unsatisfying' (Williamson, 2007) and may introduce discretization errors (Purser, 1988). Furthermore, Fourier filtering incurs an asymptotic cost of $\mathcal{O}(N^2 \log N)$. Today, finite-difference methods tend to be replaced by finite-volume methods, which have the same efficiency but can be more flexible in terms of the underlying grid, allowing the use of non-structured grids with no singular points (Satoh *et al.*, 2008; Lee and MacDonald, 2009), and in terms of

*Correspondence to: T. Dubos, IPSL/Laboratoire de Météorologie Dynamique, École Polytechnique, Palaiseau, France.
E-mail: dubos@lmd.polytechnique.fr

qualitative properties of the method, especially in dealing with discontinuities. The main differences between our method and finite-volume methods on the sphere are discussed in section 5.

The spectral-transform method emphasizes stability. By representing the solution as a combination of spherical harmonics, it elegantly removes the singularity at the pole introduced by spherical coordinates. With adequate truncation, aliasing and nonlinear instabilities are avoided too. This strong stability comes at a fairly high computational cost of $\mathcal{O}(N^3)$, dominated by the Legendre transforms associated with the spherical basis. In principle, the spectral transform provides higher spatial accuracy than any fixed-order method, provided the solution is sufficiently smooth. However, since temporal integration is usually of low order, the overall accuracy is typically no better than third order.

While the conservation properties of finite-difference methods are fairly ad hoc, the conservation properties of the spectral-transform method are generic and common to the larger class of Galerkin methods. Consider, for instance, the incompressible Euler equations in vorticity–stream function form:

$$\partial_t \zeta + J(\psi, \zeta) = 0 \qquad \text{where } \nabla^2 \psi = \zeta. \tag{1}$$

J is the Jacobian operator and ∇^2 is the Laplacian operator. In a Galerkin formulation, the stream function ψ and vorticity ζ are approximated within a finite-dimensional function space $\mathcal{S}$. For a spectral-transform method, $\mathcal{S}$ is spanned by spherical harmonics and is finite-dimensional because of (typically triangular) truncation. In the Galerkin framework, the discrete version of (1) is obtained by requiring that

$$\forall g \in \mathcal{S} \qquad \langle g\ \nabla^2 \psi \rangle - \langle g\ \zeta \rangle = 0, \tag{2}$$

$$\forall g \in \mathcal{S} \qquad \langle g\ \partial_t \zeta \rangle + \langle g\ J(\psi, \zeta) \rangle = 0, \tag{3}$$

where $\langle f \rangle$ is the mean value of f over the sphere. To achieve a given accuracy, the functional space $\mathcal{S}$ must have a sufficient approximating power and the integrals $\langle g\ \nabla^2 \psi \rangle$, $\langle g \partial_t \omega \rangle$ and $\langle g J(\psi, \omega) \rangle$ must be computed with that same accuracy. In the spectral-transform method, $\mathcal{S}$ has spectral approximating power and the integrals can be computed exactly. In this case, conservation properties hold. For instance, letting $g = \omega$ in (3) proves the conservation of enstrophy (see subsection 4.2 for more details). These exact conservation properties come 'for free' provided only that one adheres to the Galerkin framework and that the integrals can be computed exactly.

Much current research on numerical schemes for dynamical cores has abandoned the latitude–longitude grid and focuses on the use of quasi-uniform grids with less severe singularities (Williamson, 2007; Rancic *et al.*, 2008). In the present work, we instead keep the familiar latitude–longitude grid and design a new numerical method that is more accurate than finite differences and more efficient than the spectral transform, borrowing from the two approaches to achieve comparable stability and conservation properties.

In order to conserve linear and quadratic invariants, we adhere to the Galerkin framework throughout. Therefore, designing the scheme boils down to designing the functional space used for the approximation of the dynamical fields. In section 2 we show that exact quadrature is possible within a Fourier-finite-element space $\mathcal{S}$. Zonal Fourier discretization brings spectral accuracy, zonal invariance and fast transformation. Latitudinal finite elements provide adjustable accuracy, optimal efficiency and spatial locality. In section 3 the pole problem is addressed. Using the local support of the latitudinal finite elements, we construct a functional subspace $\mathcal{S}' \subset \mathcal{S}$ with quasi-uniform resolution. Our approach is based on the relationship between the effective grid size entering the CFL condition and the largest eigenvalue of the Laplacian operator. In section 4 we derive the formal properties of conservation and stability of our approach applied to two two-dimensional prototype problems involving only scalar fields: advection–diffusion of a scalar by a non-divergent flow and incompressible, rotating Navier–Stokes dynamics in vorticity–stream function formulation. In section 5, experimental results are presented demonstrating the accuracy, stability and conservation properties of the method, as well as its expected deficiencies when dealing with discontinuous fields. In section 6 we discuss the relationship of our method to zonal filters and other numerical methods recently developed for the sphere. An extension to compressible flows and the Saint-Venant equations is also considered. Our implementation is described in an appendix, where its computational cost is evaluated.

2. Approximation of scalar fields by Fourier-Finite elements

2.1. Behaviour of smooth scalar fields near the pole

Let (x, y, z) be a set of Cartesian coordinates, and (λ, ϕ, r) the associated longitude–latitude–radius coordinates, i.e. $x = r\ \cos\ \lambda\ \cos\ \phi$, $y = r\ \sin\ \lambda\ \cos\ \phi$ and $z = r\ \sin\ \phi$. Here $\lambda \in [-\pi, \pi]$ is the longitude and $\phi \in [-\pi/2, \pi/2]$ is the latitude. We note that $(\mathbf{e}_\lambda, \mathbf{e}_\phi, \mathbf{e}_r)$ is the local orthonormal basis associated with the longitude–latitude–radius coordinates.

Consider a k times continuously differentiable scalar function f defined on the sphere $r = 1$. We can extend f to the whole three-dimensional space except the origin by letting $f_{3\mathrm{D}}(x, y, z) = f(x/r, y/r, z/r)$. $f_{3\mathrm{D}}$ is also k times continuously differentiable in (x, y, z). A Taylor expansion of $f_{3\mathrm{D}}$ near a pole ($\phi = \pm\pi/2$) then approximates f by a polynomial in x, y, z of degree k. Considering that if $r = 1$ then

$$(x + \mathrm{i}y)^m (x - \mathrm{i}y)^n z^p = \mathrm{e}^{\mathrm{i}(m-n)\lambda} \cos^{m+n} \phi \sin^p \phi, \tag{4}$$

DOI: 10.1002/qj

it follows that the zonal Fourier mode $\hat{f}_m(\phi)$ of f for $m \leq k$ decays like $\cos^{|m|}\phi$ near the poles:

$$\hat{f}_m(\phi) = \frac{1}{2\pi}\int f(\lambda,\phi)\mathrm{e}^{-\mathrm{i}m\lambda}\mathrm{d}\lambda \sim \cos^{|m|}\phi, \quad \phi \longrightarrow \pm\pi/2. \tag{5}$$

For the approximation of infinitely differentiable functions, the decay (5) must be satisfied for any zonal mode m, as the spherical harmonics require. This behaviour of smooth functions near the poles serves us as a guide to construct suitable functional spaces in the next subsections.

2.2. Exact quadrature in latitude–longitude coordinates

Discretizing a continuous dynamical system by the Galerkin method involves the computation of integrals of the form $\int f\mathrm{d}\Omega$ where f is the product of basis functions and their derivatives and $\mathrm{d}\Omega = \cos\phi\,\mathrm{d}\phi\,\mathrm{d}\lambda$ is the element of integration on the sphere. The efficient way to compute such integrals is via quadrature rules, which are weighted sums of the values of f at well-chosen quadrature points. We now construct a space $\mathcal{S}$ of scalar functions for which exact quadrature is possible.

In the following we let $r = 1$, hence $z = \sin\phi$, $\mathrm{d}z = \cos\phi\,\mathrm{d}\phi$ and $\partial_\phi = \cos\phi\,\mathrm{d}z$.

We split the interval $[-1, 1]$ into N subintervals bounded by $N+1$ nodes $-1 = z_0 < z_1 < \cdots < z_N = 1$ and define $\mathcal{S}$ as the set of functions $f(\lambda, z = \sin\phi)$ such that

- the zonal Fourier mode $\hat{f}_m$ is of the form

$$\hat{f}_m = g_m(z) \qquad m \text{ even}, \tag{6}$$

$$\hat{f}_m = \cos\phi\, g_m(z) \qquad m \text{ odd}, \tag{7}$$

  and
- each $g_m(z)$ is a polynomial in z over each subinterval $[z_k, z_{k+1}]$.

In conclusion, we say that $f \in \mathcal{S}$ is admissible. The integral

$$\int f\mathrm{d}\Omega = \int f\mathrm{d}\lambda\cos\phi\mathrm{d}\phi = 2\pi\int \hat{f}_0\,\mathrm{d}z \tag{8}$$

can be computed in two steps from the values of f on a regular latitude–longitude grid. First we consider N_{lon} equally spaced longitudes λ_i, where N_{lon} is larger than the zonal bandwidth of f. Next we use Gauss quadrature formulae on each subinterval $[z_k, z_{k+1}]$, with enough points for exact integration up to the degree of the piece-wise polynomial $\hat{f}_0$. With q quadrature points in each subinterval, we have a total of $N_{\mathrm{lat}} = qN$ quadrature points z_j, corresponding to latitudes $\phi_j = \arcsin z_j$. The N_{lat} quadrature weights are normalized by $\sum_j w_j = 1$. Then the value of the zonal mean $\hat{f}_0$ at latitudes ϕ_j is exactly

$$\hat{f}_0(\phi_j) = \frac{1}{N_{\mathrm{lon}}}\sum_i f(\lambda_i, \phi_j), \tag{9}$$

and latitudinal quadrature formulae yield

$$\frac{1}{4\pi}\int f\mathrm{d}\Omega = \frac{1}{2}\sum_k \int_{z_k}^{z_{k+1}} \hat{f}_0\,\mathrm{d}z = \sum_j w_j \hat{f}_0(\phi_j) = \frac{1}{N_{\mathrm{lon}}}\sum_{i,j} w_j f(\lambda_i, \phi_j). \tag{10}$$

Notice that $\mathcal{S}$ is in fact algebraic: the product fg of two admissible functions is admissible. Indeed, it is enough to consider that f and g are pure zonal modes m and m'. If both m and m' are even, fg obviously satisfies (6). If both are odd, (7) is still satisfied because $\cos^2\phi = 1 - z^2$. If either m or m' is odd, (7) is satisfied. This is a required property, since we shall eventually integrate double or triple products of admissible functions.

The nodes z_k can be chosen arbitrarily. In the following they are defined by $z_k = \cos(\pi k/N)$, corresponding to equally spaced latitudes. We have not tried to improve the accuracy or stability of the numerical method by adjusting the nodes z_k. Once the nodes z_k are chosen, the number N_{lat}, the latitudes ϕ_j of the quadrature points and their weights w_j are determined by the quadrature rule used on each subinterval $[z_k, z_{k+1}]$. For a q-point Gauss–Legendre quadrature rule, $N_{\mathrm{lat}} = Nq$ and polynomials of degree up to $2q-1$ can be integrated exactly.

2.3. A Fourier-finite-element functional space

We have not yet imposed latitudinal smoothness conditions, not even continuity, on admissible functions. Hence a finite-dimensional subspace $\mathcal{S}' \subset \mathcal{S}$ suitable to the Galerkin discretization of partial differential equations (PDEs) will enforce additional smoothness conditions as well as near-pole decay reflecting the order of the PDE.

Consider the advection–diffusion equation with non-divergent flow:

$$\partial_t f + J(\psi, f) = \kappa\nabla^2 f, \tag{11}$$

where ψ is a prescribed stream function, J is the Jacobian and ∇^2 is the Laplacian operator. Upon multiplication of (11) by an arbitrary test function $g \in \mathcal{S}'$ and integration by parts, the PDE (11) is discretized into

$$\forall g \in \mathcal{S}' \qquad M(g, \partial_t f) + T(g, f) = -\kappa S(g, f), \tag{12}$$

where the quadratic forms M, T and S are defined by

$$M(g, f) = \langle g^* f\rangle, \tag{13}$$

$$T(g, f) = \langle g^*\, J(\psi, f)\rangle, \tag{14}$$

$$S(g, f) = \langle \nabla g^* \cdot \nabla f\rangle, \tag{15}$$

where g^* is the complex conjugate of g and $\langle f\rangle = [1/(4\pi)]\int f\mathrm{d}\Omega$.

In the following we shall use the definition

$$||S||_{\mathcal{S}'} = \sup_{g\in\mathcal{S}'} \frac{\langle \nabla g^* \cdot \nabla g\rangle}{||g||^2}, \tag{16}$$

DOI: 10.1002/qj

where $||g||^2 = \langle |g|^2 \rangle$ and $\mathcal{S}'$ is any subspace of $\mathcal{S}$. Differentiating $S(g,g)/M(g,g)$ with respect to g, one finds that $||S||_{\mathcal{S}'}$ is also the largest solution σ of the generalized eigenvalue problem:

$$\forall g \in \mathcal{S}' \qquad S(g,f) = \sigma M(g,f), \tag{17}$$

where $f \in \mathcal{S}'$ is an unknown eigenvector.

The quadratic forms M, T and S are well-defined if, for any $f \in \mathcal{S}'$,

$$\nabla f = \frac{1}{\cos\phi}\partial_\lambda f \mathbf{e}_\lambda + \cos\phi \partial_z f \ \mathbf{e}_\phi \tag{18}$$

is square-integrable. This requires only that $\hat{f}_m$ be continuous across nodes z_k, and that it decay like $\cos(\phi)$ near the poles for $m \geq 1$. This is equivalent to imposing the boundary conditions

$$g_m(z = \pm 1) = 0 \qquad |m| \geq 2. \tag{19}$$

Although zonal and latitudinal components of ∇f are not smooth scalar fields, the derivatives $\partial_z f$ and $\partial_\lambda f$ are admissible, hence the Jacobian $J(f,g) = \partial_\lambda f \partial_z g - \partial_\lambda g \partial_z f$ is admissible. Assuming $\psi \in \mathcal{S}$, the quadratic forms M and T can then be computed exactly by quadrature (see the appendix for practical details).

The dimensionality of $\mathcal{S}'$ is kept finite by truncating frequencies to $-M \leq m \leq M$ for some M and limiting the polynomials g_m to a finite degree. In the following we use B-splines of degree d. B-splines of degree d are $d-1$ times continuously differentiable, provide $(d+1)$th order accuracy and are generated by a basis of $N+d$ piecewise polynomials $B_0(z), \ldots, B_{N+d-1}(z)$, support for which consists of at most $d+1$ successive subintervals. Figure 1 shows the shape of the basis functions B_l for $N = 18$ and $d = 1, 2, 3$. Functions $B_d, \ldots, B_N - 1$ have $d-1$ vanishing derivatives at $z = \pm 1$, while the first d and last d basis functions control the boundary conditions. Therefore boundary conditions (19) are easily enforced by discarding B_0 and B_{N+d-1}. To summarize, functions $f \in \mathcal{S}'$ are of the form

$$f = \sum_{m=-1}^{1} \sum_{0}^{N+d-1} \alpha_{ml} F_{ml} + \sum_{1<|m|\leq M} \sum_{1}^{N+d-2} \alpha_{ml} F_{ml},$$

$$m \text{ even} \qquad F_{ml} = B_l(\sin\phi)\ \mathrm{e}^{\mathrm{i}m\lambda},$$

$$m \text{ odd} \qquad F_{ml} = \cos\phi\ B_l(\sin\phi)\ \mathrm{e}^{\mathrm{i}m\lambda}.$$

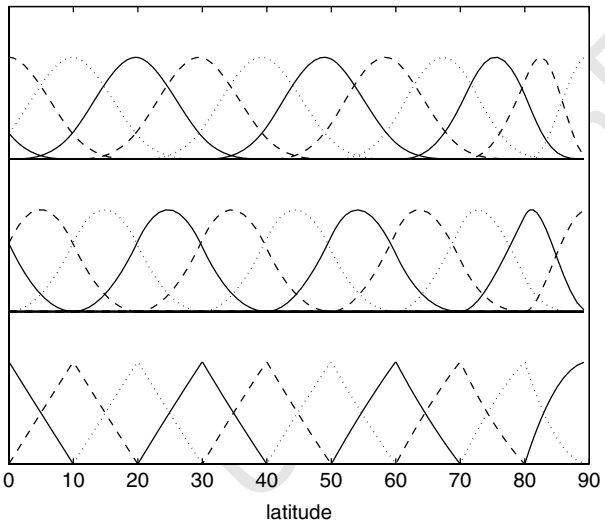

Figure 1. B-spline basis $B_l(z)$ (where $z = \sin\phi$ and ϕ is the latitude) for $N = 18$ and $d = 1, 2, 3$ (from bottom to top). The support of each B_l consists of at most $d+1$ consecutive latitude bands, each with a width of $180/N = 10^\circ$. Each B_l is a polynomial in z of degree d within each latitude band, and is globally $d-1$ times continuously differentiable. Discarding the rightmost basis function enforces the boundary conditions (19).

3. Subspaces providing quasi-uniform resolution

3.1. The pole problem

The pole problem occurs when solving (12) with an explicit time-stepping scheme. We consider here the case $\kappa = 0$ for simplicity. The stability of an explicit scheme is limited by the largest proper frequency ω of the transport operator, defined as the largest solution ω of the generalized eigenproblem:

$$\forall g \in \mathcal{S}' \qquad T(g,f) = \mathrm{i}\omega M(g,f), \tag{20}$$

where $f \in \mathcal{S}'$ and ω are an unknown eigenvector and proper frequency. More precisely, the time step τ must satisfy $\omega\tau \leq c$, where the non-dimensional constant c depends only on the time-stepping scheme (leapfrog etc.).

For the transport operator, one finds that ω is bounded by

$$|\omega| \leq U/\delta, \tag{21}$$

where $U = \max|\nabla\psi|$ is the maximum velocity and

$$\delta^{-2} = ||S||_{\mathcal{S}'}. \tag{22}$$

Indeed,

$$|\omega| \leq \sup_{f,g\in\mathcal{S}'} \frac{T(g,f)}{||f||\ ||g||},$$

$$|T(g,f)| \leq U||f||\ ||\nabla g||$$

and definition (16) implies that $||\nabla g|| \leq ||g||/\delta$. Therefore stability is guaranteed if

$$\tau U \leq c\delta. \tag{23}$$

The effective grid scale δ entering the CFL criterion (23) and controlling the time step is therefore defined from the largest eigenvalue of the Laplacian operator, as expressed by (22).

As it stands, the functional space $\mathcal{S}'$ suffers from the pole problem to the same extent as finite differences or

DOI: 10.1002/qj

double Fourier series on a latitude–longitude grid: δ is controlled by the near-pole zonal resolution, much finer than that near the Equator. This fine resolution is wasted since the discretization error arising near the Equator eventually propagates to the whole sphere under the effect of advection. In the next subsection, we remove this excess resolution by restricting the Galerkin formulation to a subspace $\mathcal{S}''$ of $\mathcal{S}'$.

3.2. Quasi-uniform resolution

We define $\mathcal{S}''$ as the space generated by the basis functions F_{ml}, but for only a subset K of indices l, m. For this, let us define $\mathcal{S}_m^L$ as the space spanned by the basis functions (F_{ml}) with $L \leq l < N + d - L$. With this definition, functions in $\mathcal{S}_m^L$ are zonally monochromatic and we have discarded L degrees of freedom near each pole. Our goal is to define for each m an increasing number $L(m)$ of near-pole degrees of freedom to discard (with $L(-m) = L(m)$), and to define $\mathcal{S}'' = \bigoplus_{m=-M}^{M} \mathcal{S}_m^{L(m)}$, e.g. the space spanned by the $(F_{ml})_{(m,l)\in K}$ with $K = \{-M \leq m \leq M \text{ and } L(m) \leq l < N + d - L(m)\}$.

There are certainly several strategies to determine $L(m)$ and ensure a quasi-uniform resolution. Our approach is to set an upper bound on the largest eigenvalue of the Laplacian operator, which will provide an a priori control of the CFL stability criterion. First we use the fact that the Laplacian does not couple the different zonal modes. Hence

$$||S||_{\mathcal{S}''} = \max_{0\leq m\leq M} ||S||_{\mathcal{S}_m^{L(m)}},$$

where $||S||_{\mathcal{S}_m^L}$ is defined as in (22). Furthermore, for a given L, $||S||_{\mathcal{S}_m^L}$ is obtained as the largest eigenvalue λ of the generalized eigenvalue problem:

$$S_{kl}^{(m)}\alpha_l = \lambda M_{kl}^{(m)}\alpha_l,$$

where

$$S_{kl}^{(m)} = \langle \nabla F_{km}^* \cdot \nabla F_{lm} \rangle,$$
$$M_{kl}^{(m)} = \langle F_{km}^* F_{lm} \rangle,$$

and implicit summation ranges over the index $l \in [L,\ 2N + 1 - L]$. The matrices $\mathbf{S}^{(m)}$ and $\mathbf{M}^{(m)}$ are real, symmetric, positive and $(2d + 1)$-diagonal due to the finite support of the basis functions B_l.

Notice that we have no reason to discard near-pole degrees of freedom for $-1 \leq m \leq 1$. Indeed, the corresponding functions have the correct near-pole decay, as $\cos^m \phi$. Hence we decide that $L(m) = 0$ for $-1 \leq m \leq 1$ and that

$$\delta^{-2} = \max\ (||S||_{\mathcal{S}_0^0},\ ||S||_{\mathcal{S}_1^0}). \tag{24}$$

This resolution δ is entirely determined by the latitudinal resolution, e.g. by the number N of latitudinal intervals and by the positions of the nodes z_k. We then define $L(m)$ for $l > 1$ as the smallest number L such that $||S||_{\mathcal{S}_m^L} \leq \delta^{-2}$. This guarantees that the CFL criterion indeed involves the latitudinal resolution, and not a very small near-pole zonal resolution. Notice that $||S||_{\mathcal{S}_m^L} \geq m^2$, which also bounds the number M of zonal modes for a given N.

In the following we set $N = M$. By analogy with spectral truncation, a specific choice of M is called in the following the 'truncation M', for instance T42 in the case $M = N = 42$. We have computed the effective grid size δ defined by (24) for truncations ranging from $T16$ to $T341$. The values (in degrees) are summarized in Table I together with the numbers M and N, the number N_{lon} of zonal grid points and the zonal grid spacing at the Equator, $360/N_{\text{lon}}$. The effective grid size, limiting the length of the time step through the CFL condition, is seen to be roughly a third of the zonal grid size at the Equator.

We now turn to the properties of the Galerkin method based on the functional space $\mathcal{S}'' \subset \mathcal{S}'$. We present formal and experimental results demonstrating that the method is conservative and stable, and that the truncation strategy we have adopted does not degrade the approximating power of the functional space $\mathcal{S}''$.

4. Formal properties

4.1. Advection

Due to exact quadrature, the transport operator T defined by Equation (14) is anti-Hermitian provided $\psi \in \mathcal{S}'$:

$$T(f, g) = \langle \psi J(f^*, g) \rangle = -\ T(g, f)^*. \tag{25}$$

Table I. Effective grid size in degrees for truncations ranging from $T16$ to $T341$ together with the numbers M and N, the number N_{lon} of zonal grid points and the zonal grid spacing at the Equator.

	$T15$	$T21$	$T31$	$T42$	T63	T85	T127	T170	T255	T341
$M = N$	15	21	31	42	63	85	127	170	255	341
N_{lon}	48	64	96	128	192	256	384	512	768	1024
$360/N_{\text{lon}}$	7.5	5.6	3.8	2.8	1.9	1.4	0.94	0.70	0.47	0.35
$180/\pi\,\delta$ $(d = 1)$	2.6	1.9	1.3	0.94	0.63	0.46	0.31	0.23	0.15	0.12
$180/\pi\,\delta$ $(d = 2)$	2.6	1.8	1.2	0.91	0.61	0.45	0.30	0.23	0.15	0.11
$180/\pi\,\delta$ $(d = 3)$	2.3	1.65	1.1	0.83	0.55	0.41	0.27	0.20	0.14	0.10

DOI: 10.1002/qj

We have checked numerically that (25) is satisfied within round-off error. Therefore the spatially discretized, continuous-time system (12) with $\kappa = 0$ exactly preserves the mean value and total variance of f:

$$\partial_t \langle f \rangle = T(1, f) = 0, \tag{26}$$

$$\partial_t \langle f^* f \rangle = M(f, \partial_t f) + M(\partial_t f, f) = T(f, f) + T(f, f)^* = 0. \tag{27}$$

Conservation of the total variance is usually not preserved once time is discretized too, by a leap-frog scheme for example. Nevertheless, since T is anti-Hermitian, it has purely imaginary eigenvalues $\mathrm{i}\omega$ (Equation (17)). As a result, if the CFL criterion is satisfied, both the physical and the computational modes of the leap-frog scheme are exactly neutral. This implies that the variance $\langle f^* f \rangle$, although not strictly conserved, can only oscillate around a mean value with an amplitude set by the strength of the computational mode. The latter depends on the method used to perform the first step of temporal integration. Furthermore, the scheme need not be stabilized by diffusion, an Asselin filter or another method.

4.2. Incompressible rotating Navier–Stokes equations

We now consider the (barotropic) incompressible rotating Navier–Stokes equations:

$$\partial_t \mathbf{u} + (\zeta + 2\Omega z)\mathbf{e}_r \times \mathbf{u} + \nabla p = \nu(\nabla^2 \mathbf{u} + 2\mathbf{u}), \tag{28}$$

where ν is the kinematic viscosity, Ω the rotation rate of the sphere and p the dynamic pressure. This equation models a thin, incompressible, barotropic atmosphere of constant thickness. Barotropy means the absence of vertical shear, hence a three-dimensional motion $\mathbf{u}_{3D} = r\mathbf{u}$ within each atmospheric column. Evaluating at $r = 1$, the usual three-dimensional viscous term $\nu\nabla^2(r\mathbf{u}) = \nu(r\nabla^2\mathbf{u} + (2/r)\mathbf{u})$ yields the right-hand side of (28).

The vorticity–stream function form of (28) is obtained by letting $\mathbf{u} = \mathbf{e}_r \times \nabla\psi$, where ψ is the stream function, and by taking the curl of (28). This yields

$$\partial_t \zeta + J(\psi, \zeta + 2\Omega z) = \nu(\nabla^2 \zeta + 2\zeta) \tag{29}$$

where

$$\Delta\psi = \zeta. \tag{30}$$

Here the apparently peculiar form of the viscous term is seen to guarantee that any solid-body rotation is an exact steady solution of (29): for solid-body rotation, $\nabla^2\zeta = -2\zeta$. If the viscous term were simply $\nu\nabla^2\zeta$, solid-body rotations would be damped and total angular momentum would decay instead of being conserved, as is physically required.

The stream function ψ exists only if the relative vorticity ζ has zero average. To ensure the unity of ψ, we require that $\psi(z = -1) = 0$. The spatially discretized version of (29) reads

$$\langle \nabla g^* \cdot \nabla\psi \rangle + \langle g^* \zeta \rangle = 0 \qquad \forall g \in \mathcal{S}'', \tag{31}$$

$$\langle g^* \partial_t \zeta \rangle + \langle g^* J(\psi, \zeta + 2\Omega z) \rangle + \nu\langle \nabla g^* \cdot \nabla\zeta - 2g^*\zeta \rangle = 0 \quad \forall g \in \mathcal{S}''. \tag{32}$$

In (31), the stream function $\psi \in \mathcal{S}''$ is computed given the relative vorticity field $\zeta \in \mathcal{S}''$. Advection–diffusion of absolute vorticity $\zeta + 2\Omega z$ is expressed by (32), from which $\partial_t\zeta$ is obtained. It results from the previous subsection that the spatial average of ζ remains zero provided it is initially zero. We now show that the axial angular momentum is conserved, and that enstrophy and energy are conserved when $\nu = 0$. In this special case, (28) reduces to the Euler equations and (29) to their vorticity–stream function formulation.

Conservation of axial angular momentum $L_z = \langle z\zeta \rangle$ results from the fact that $z \in \mathcal{S}''$. We can therefore let $g = z$ in (32):

$$\partial_t L_z - \langle \psi \; J(z, \zeta + 2\Omega z) \rangle + \nu\langle \cos\phi \partial_\phi \zeta - 2z\zeta \rangle = 0. \tag{33}$$

Now $J(z, 2\Omega z) = 0$ and $\langle \cos\phi \partial_\phi \zeta \rangle = \langle (1 - z^2)\partial_z \zeta \rangle = \langle 2z\zeta \rangle$ after integration by parts. Furthermore,

$$\langle \psi \; J(\zeta, z) \rangle = \langle \psi \; \partial_\lambda \zeta \rangle = 0, \tag{34}$$

because ψ and ζ are related by the zonally symmetric, Hermitian relationship (31). Hence $\partial_t L_z = 0$. Notice that we prove the conservation of L_z in the viscous case, but that all simulations presented in the following are inviscid.

Conservation of enstrophy when $\nu = 0$ is analogous to the conservation of scalar variance, and is proven by taking $g = \zeta$. We now prove the conservation of the energy defined by $E = -\langle \psi\zeta \rangle/2$. Notice that E is a quadratic function of ζ, hence $\partial_t E = -\langle \psi \partial_t \zeta \rangle$. Now $\psi \in \mathcal{S}''$, hence we can let $g^* = \psi$ in (32), yielding $\langle \psi \partial_t \zeta \rangle = \langle (\zeta + 2\Omega z) J(\psi, \psi) \rangle = 0$. We have checked numerically that $\langle z\partial_t\zeta \rangle = 0$, $\langle \zeta \partial_t \zeta \rangle = 0$ and $\langle \psi \partial_t \zeta \rangle = 0$ within round-off error at all truncations.

What precedes substantiates our claim (see the appendix) that the conservation properties of our scheme are not affected if inexact quadrature formulae are used to compute the stiffness matrix of the Laplacian operator (Equations (A3) and (A9)). Exact quadrature is required only for the nonlinear term, so that

$$\langle g^* \; J(\psi, \zeta + 2\Omega z) \rangle = \langle \psi \; J(\zeta + 2\Omega z, g^*) \rangle \tag{35}$$

$$= \langle (\zeta + 2\Omega z) \; J(g^*, \psi) \rangle. \tag{36}$$

It is then sufficient that $\psi \in \mathcal{S}''$ be any Hermitian function of ζ for the energy $-\langle \psi\zeta \rangle/2$ to be conserved. The additional requirement that the relationship between ζ and ψ be zonally invariant guarantees the conservation of axial angular momentum.

5. Experimental results

5.1. Truncation error

We now check that, despite near-pole zonal truncation, the functional space $\mathcal{S}''$ provides an approximation of the same order as $\mathcal{S}'$, i.e. of order $d + 1$. For this we pick

DOI: 10.1002/qj

a point (λ_0, ϕ_0) on the sphere and consider two scalar functions, a Gaussian

$$f(\lambda, \phi) = \exp \frac{\cos\alpha - 1}{\alpha_c^2} \tag{37}$$

and a cosine bell

$$f(\lambda, \phi) = 1 + \cos\{\pi \ \min(1, \alpha/\alpha_c)\}, \tag{38}$$

where α is the geodesic angle between (λ, ϕ) and (λ_0, ϕ_0) and $\alpha_c = \pi/8$. We compute an approximation $\tilde{f} \in \mathcal{S}''$ from the value of f at the quadrature points; then the maximal pointwise error, as well as the largest pointwise error in the approximation of the gradient, can be obtained from

$$\epsilon(N, d) = \max |f(\lambda_i, \phi_j) - \tilde{f}(\lambda_i, \phi_j)|, \tag{39}$$

$$\epsilon_\nabla(N, d) = \max ||\nabla f(\lambda_i, \phi_j) - \nabla \tilde{f}(\lambda_i, \phi_j)||. \tag{40}$$

We repeat the process for 100 random values of (λ_0, ϕ_0) and retain the largest errors. Figure 2 displays $\epsilon(N, d)$ as a function of the zonal grid size at the Equator, $360/N$, for finite elements of degree $d = 1, 2, 3$ (circles, crosses, triangles), for the Gaussian (solid line) and for the cosine bell (dashed line). For the cosine bell, the error scales like $\epsilon \sim N^{-2}$, indicating second-order accuracy. This is consistent with the cosine bell being only continuously differentiable with a bounded second derivative. Quadratic finite elements (crosses) provide slightly better accuracy than linear finite elements (circles), and cubic finite elements (triangles) do not improve on them. For the Gaussian however, the error scales like $\epsilon \sim N^{-(d+1)}$, demonstrating that the formal order of accuracy is indeed achieved in practice.

Figure 3 displays $\epsilon_\nabla(N, d)$ as a function of the zonal grid size at the Equator, $360/N$. For the cosine bell the error scales like $\epsilon_\nabla \sim N^{-1}$, while for the Gaussian the error scales like $\epsilon_\nabla \sim N^{-d}$. Hence the order of the approximation of the gradients is one less than the order of the pointwise approximation. This is the expected behaviour for a finite-element method.

5.2. Advection

We have checked the conservative and stability properties of our numerical scheme in the problem of advection of a cosine bell, first by a solid-body rotation field and then by an arbitrary stream function. We use a leap-frog temporal scheme. In the first case the stream function ψ is linear in the coordinates x, y, z, hence $\psi \in \mathcal{S}'$ naturally. In the second case we generate a stream function whose Laplacian (vorticity) is a spatially white noise. As argued above, the model remains stable for long times (hundreds of solid-body rotations or turnover times) without diffusion or temporal filtering. The relative amplitude of the oscillations of the variance ranges from 10^{-3} at $T15$ to 10^{-6} at $T170$.

Figure 4 displays the results of the simulation of advection of a cosine bell by a solid-body rotation with unit angular velocity and an axis making an angle $\pi/2 - 0.05$ with the z axis. The resolution is $T42$ with quadratic B-splines, and we use a leap-frog time scheme. The time step is set to $\delta/2$, half the maximum time step allowed by the CFL criterion. The initial condition (top panel) is the best approximation of the cosine-bell function within the space $\mathcal{S}''$. We let it sit across the Equator (parameter $\phi_0 = 0$ in the preceding subsection). Slightly negative values due to Gibbs oscillations are visible in the region where the exact cosine-bell function is zero. We run the simulation during one period of solid rotation and compare the final field with the initial field by taking the difference between the two (bottom panel). The largest pointwise error is about 0.02, or 1% of the maximum value of the cosine bell.

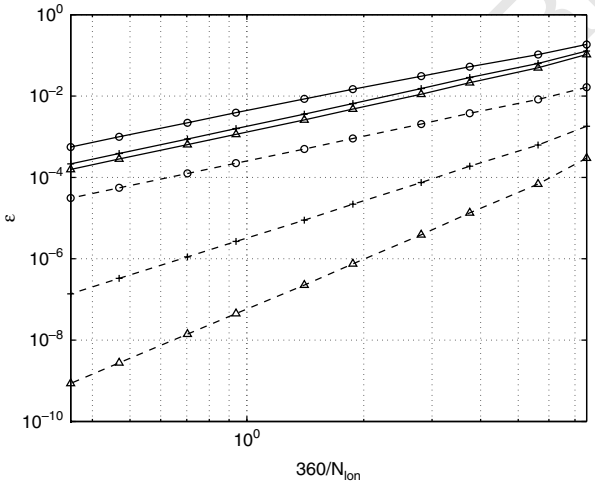

Figure 2. Pointwise discretization error $\epsilon(N, d) = \max |f(\lambda_i, \phi_j) - \tilde{f}(\lambda_i, \phi_j)|$ as a function of the zonal grid size $360/N_{\text{lon}}$ (°) for finite elements of degree $d = 1, 2, 3$ (circles, crosses, triangles), a Gaussian (dashed line) and a cosine bell (solid line).

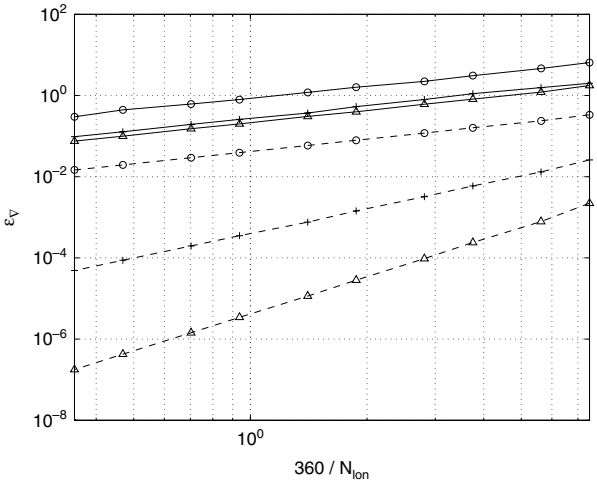

Figure 3. Pointwise discretization error of the gradient $\epsilon_\nabla(N, d) = \max \left\| \nabla f(\lambda_i, \phi_j) - \nabla \tilde{f}(\lambda_i, \phi_j) \right\|$ as a function of the zonal grid size $360/N_{\text{lon}}$ (°) for finite elements of degree $d = 1, 2, 3$ (circles, crosses, triangles), a Gaussian (dashed line) and a cosine bell (solid line).

DOI: 10.1002/qj

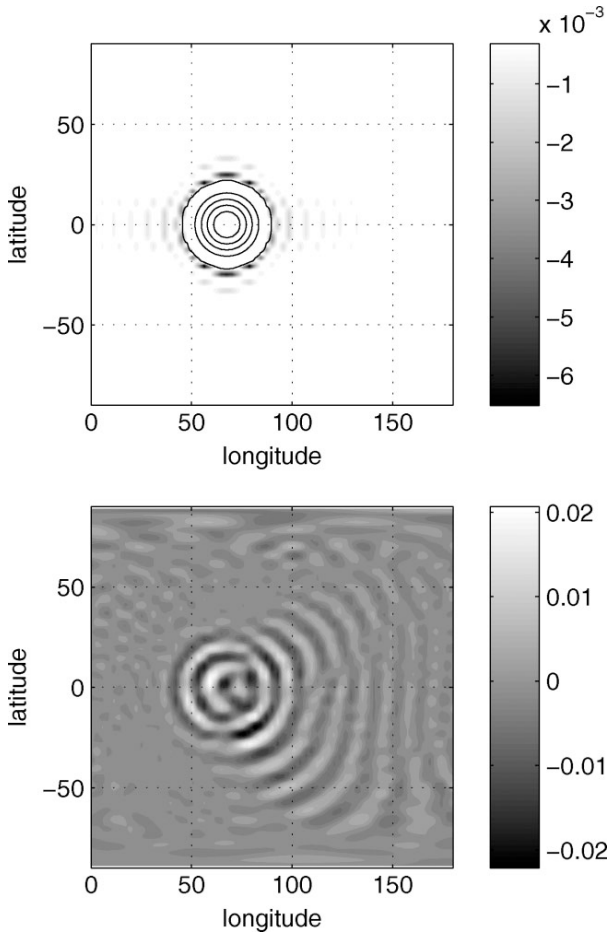

Figure 4. Advection of a cosine bell by a solid-body rotation at resolution $T42$ using quadratic B-splines. Top: values of the initial condition f_0. Solid line: contours $f_0 = 0, 0.4, 0.8, 1.2, 1.6$. Grey shading: negative values of f_0 due to Gibbs oscillations. Bottom: difference between the simulated values of f after one complete revolution and the initial values.

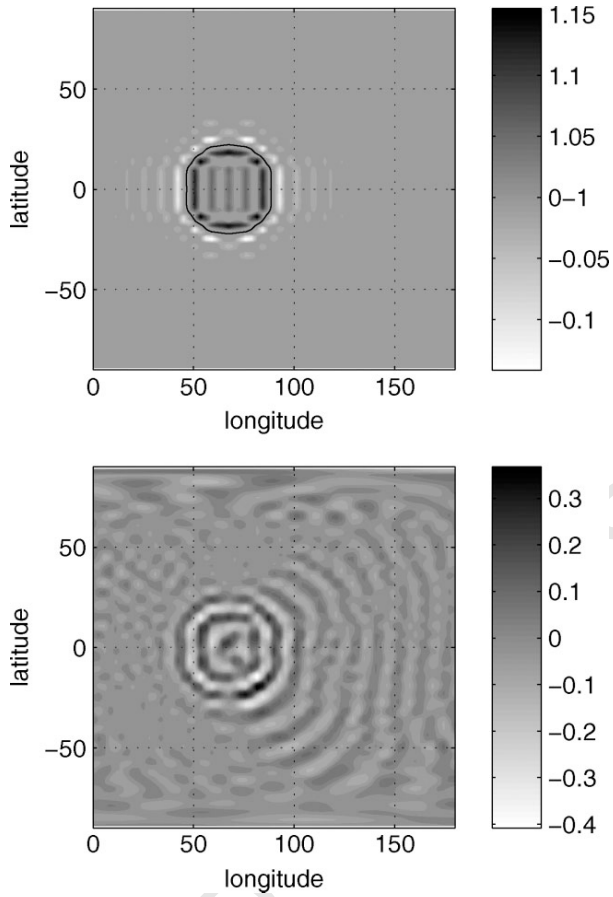

Figure 5. Advection of a uniform circular patch by a solid-body rotation at resolution $T42$ using quadratic B-splines. Top: values of the initial condition f_0. Solid line: contour $f_0 = 0.5$. The grey-scale emphasizes Gibbs oscillations outside the range [0 1]. Bottom: difference between the simulated values of f after one complete revolution and the initial values.

In order to give some indications of the performance of our method with discontinuous functions, which we expect to be poor, we also compute the advection of a uniform circular patch by the same solid-body rotation (Figure 5). The initial condition is defined as 0 where the cosine bell is 0, and 1 inside the disc where it is >0. The approximation of the initial condition displays significant zonal Gibbs oscillations, as expected when a discontinuous function is approximated using Fourier series (top panel). Again, we run the simulation during one period of solid rotation and compare the final field with the initial field by taking the difference between the two (bottom panel). Gibbs oscillations are seen to propagate to, and significantly pollute, the whole sphere.

5.3. Propagation of a Rossby–Haurwitz wave

We finally use our method to simulate the propagation of a Rossby–Haurwitz wave, a classical test case (Cheong, 2000). We introduce a rotated coordinate system (x', y', z') such that the z and z' axes make an angle α relative to each other. We solve the vorticity equation (29), where $\zeta + 2\Omega z$ is replaced by $\zeta + 2\Omega z'$: the rotation axis z' of the sphere differs from the numerical axis z. We use as an initial condition a wave with $(l, m) = (5, 4)$:

$$\zeta_0 = \rho l(l+1)\cos^4\theta' \sin\theta' \cos\ 4\lambda', \tag{41}$$

$$\psi_0 = -\rho \cos^4\theta' \sin\theta' \cos\ 4\lambda', \tag{42}$$

where $\rho = 2/\{l(l+1) - 2\}$ is the wave amplitude and (λ', θ') are the longitude and latitude relative to (x', y', z'). This wave has a period

$$T = \frac{l(l+1)}{2m} T_0, \tag{43}$$

where $T_0 = 1$ day. We set $\alpha = \pi/2 - 0.05$, which is considered a more stringent test (Cheong, 2000). We perform several inviscid simulations ($\nu = 0$) lasting 500 wave periods (1925 days) and record the ψ and ζ fields at every integer multiple of the wave period. We monitor the evolution of energy E and enstrophy Z, which should be conserved.

While the initial condition ψ_0 results in an exactly time-periodic solution of the continuous equations (29),

DOI: 10.1002/qj

this may not be so after discretization, due to small discretization errors in space and time. This leads to a non-zero departure from the initial condition:

$$\delta\psi = \psi - \psi_0. \tag{44}$$

We compute the departure $\delta\psi$ from the initial condition at every integer multiple of the wave period T and monitor the evolution of its of energy and enstrophy:

$$E_\delta = -\frac{1}{2}\langle \delta\psi \ \nabla^2\delta\psi\rangle, \tag{45}$$

$$Z_\delta = \frac{1}{2}\langle(\nabla^2\delta\psi)^2\rangle. \tag{46}$$

We first perform a simulation at resolution $T21$ with quadratic B-splines ($d = 2$) using a leap-frog temporal scheme. The time step of the leap-frog scheme is set to $T/64$ (simulation TW_T21_64). We prevent the leap-frog instability by applying a forward Euler step every $N_{\text{Euler}} = 512$ time steps (i.e. every 8 wave periods, or 30 days). The growth of δE and δZ is displayed in Figures 6 and 7 with solid lines. Figure 6 focuses on the first 500 days of simulation. The relative departures $\delta E/E_0$ and $\delta Z/Z_0$ happen to be virtually identical. Two phases are present: in a first phase until $t \simeq 200$ days, small errors accumulate linearly in time; after this initial phase δE and δZ soon grow exponentially. At later stages, δE and δZ saturate at an amplitude as large as the initial energy E_0 and enstrophy Z_0. The exponential growth of δE and δZ suggests that the wave is unstable. Indeed, energy and enstrophy are conserved throughout the simulation to an accuracy better than 1% (Figure 7), indicating a physical, rather than numerical, instability. Such an instability of the (4,5) Rossby–Haurwitz wave has been reported in previous studies and attributed to the interaction of resonant triads (Lynch, 2009).

To confirm this, we perform a second simulation with a time step set to $T/128$ and $N_{\text{Euler}} = 1024$ (simulation TW_T21_128, dashed lines). During the initial phase, the numerical errors accumulate with a rate diminished by a factor 1/4, consistent with the accumulation of temporal errors produced by a second-order scheme. The delay before the exponential growth occurs is roughly the same, which confirms the presence of a genuine, dynamical instability. Indeed, the growth rate of a numerical instability would have decreased with a decreased time step. Energy and enstrophy also remain constant within a few per cent (Figure 7). We finally perform a third simulation at a higher resolution $T42$ with the same time step $T/128$ and $N_{\text{Euler}} = 1024$ (TW_T42_128, dotted lines). Notice that the initial accumulation of error is not reduced, as expected for temporal errors, since the time step is identical to TW_T21_128. The conservation of energy is as good as with TW_T21_128, while the enstrophy is even better conserved. We conclude that our spatial discretization allows long inviscid simulations with very good conservation of energy and enstrophy and no numerical instability.

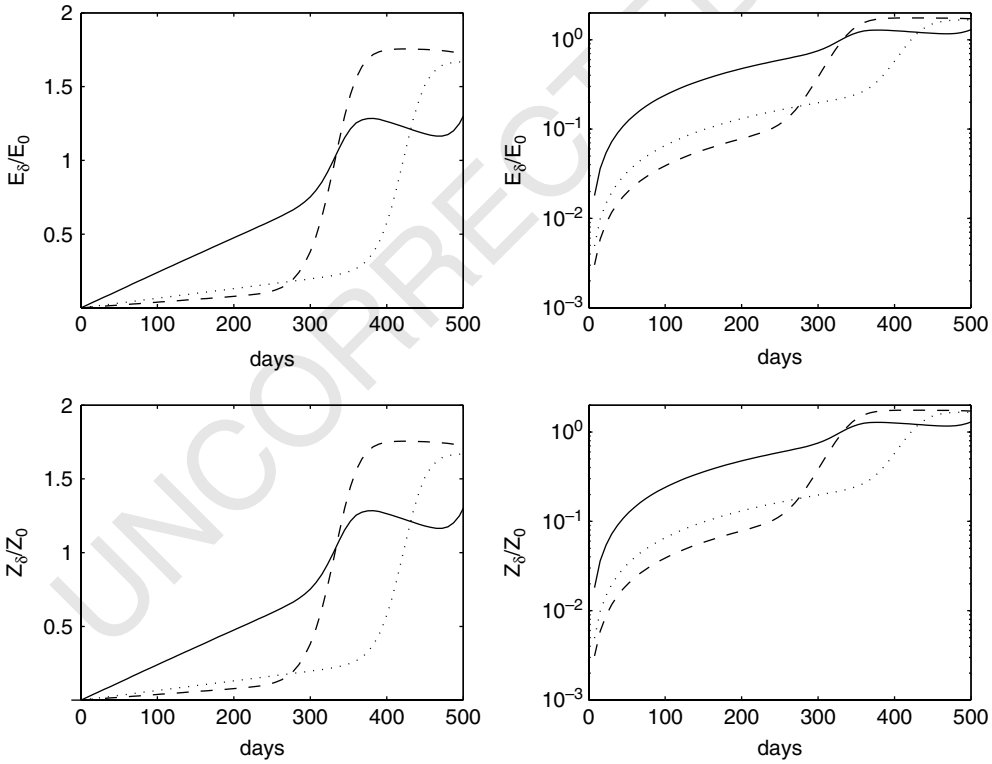

Figure 6. Inviscid simulations TW_T21_64 (solid), TW_T21_128 (dashed) and TW_T42_128 (dotted) of a travelling Rossby wave: growth of the departure $\delta\psi = \psi - \psi_0$ from initial conditions during the first 500 days. Upper panels: energy E_δ of $\delta\psi$ normalized by E_0, the energy of the initial condition ψ_0; lower panels: enstrophy Z_δ of $\delta\psi$, normalized by Z_0, the enstrophy of the initial condition.

DOI: 10.1002/qj

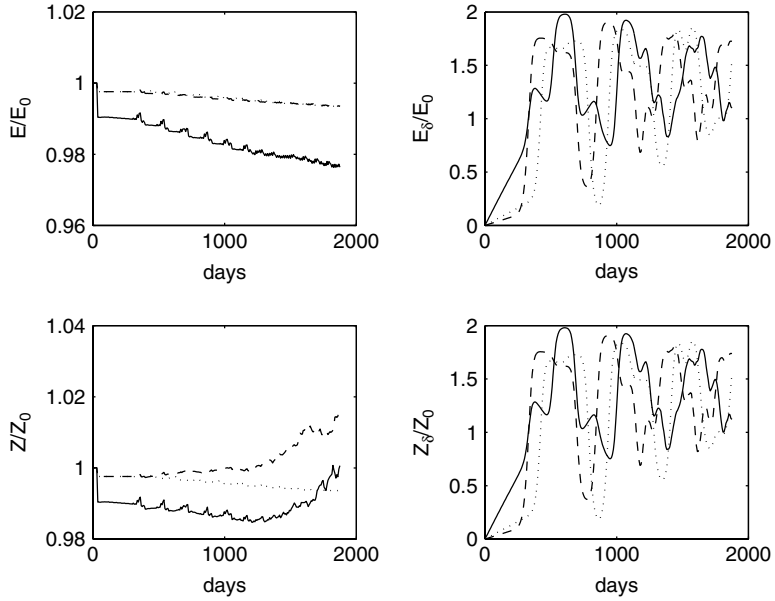

Figure 7. Inviscid simulations TW_T21_64 (solid), TW_T21_128 (dashed) and TW_T42_128 (dotted) of a travelling Rossby wave: evolution of energy and enstrophy (left panels) and growth of the departure $\delta\psi = \psi - \psi_0$ of the flow from its initial condition (right panels). Upper panels: total energy E and energy E_δ of $\delta\psi$, both normalized by the initial value E_0 of E; lower panels: flow enstrophy Z and enstrophy Z_δ of $\delta\psi$, both normalized by the initial value Z_0 of Z.

6. Discussion

We have designed a conservative, accurate and stable method for the solution of scalar PDEs on the whole sphere. Attractive features of the method are its simplicity, associated with the latitude–longitude grid, and its efficiency, intermediate between finite-difference and spectral methods. For N^2 degrees of freedom, the computational cost is $\mathcal{O}(N^2 \log N)$, dominated by the zonal Fourier transforms and independent of the order of the method. This is more than with a finite-difference method, which costs $\mathcal{O}(N^2)$, and asymptotically much lower than with a spherical harmonics spectral method, which costs $\mathcal{O}(N^3)$. Differential operators with zonally symmetric coefficients are inverted at a small cost of $\mathcal{O}(N^2)$. Two key constraints led to the choices we made: the method had to be conservative, and the pole problem had to be overcome. Conservativeness of the method relies on the Galerkin framework and the exact quadrature of nonlinear terms. The pole problem is overcome by varying the zonal resolution near the poles.

In the spectral-transform method, the zonal resolution is effectively reduced near the poles because spherical harmonics with zonal wavenumber m decay like $\cos^m \phi$. However spherical harmonics have global support, which results in a high computational cost. Here we achieve latitude-dependant zonal resolution and computational efficiency through the basis functions $F_{ml}(\lambda, \phi)$, which have local support in latitude and in zonal frequency. While latitudinal local support is straightforwardly provided by piecewise polynomials in the ϕ or $z = \sin\phi$ variables, special care has been necessary to achieve exact quadrature. A contribution of this work is to recognize that exact quadrature is possible within the algebra $\mathcal{S}$, which deals with even and odd zonal modes separately. The desired order, smoothness and zonal resolution can then be attained by considering the adequate subspace $\mathcal{S}''$ of $\mathcal{S}$. Flexibility in the choice of $\mathcal{S}''$ results in a family of numerical methods where accuracy and smoothness can be adjusted depending on the specific problem to be solved. A second contribution of this work is to base the zonal truncation on an explicit bound of the CFL number, thus providing a priori control of the numerical stability of the method. Again the Galerkin framework was instrumental in establishing this bound.

Many other methods have been proposed to solve partial differential equations on the sphere, several of which can be usefully compared with ours. The present work presents only a few similarities with the Cote and Staniforth (1990) finite-element method. Cote and Staniforth (1990) use finite elements, essentially for efficiency reasons, within a semi-Lagrangian method. The semi-Lagrangian method solves the pole problem but enforces only mass conservation. Accordingly they use inexact second-order quadrature, and no zonal filters.

The order of accuracy of our method can be made as high as desired. An alternative to doing this is the spectral-element method. The latter has been implemented on the sphere using a tiling of the sphere with quadrangular elements (Taylor, 1997; Baer *et al.*, 2006). On such a tiling, accurate but not exact quadrature can be used. The resulting method is spectrally accurate but not de-aliased. Therefore viscosity or filtering is required to remove small-scale noise (Taylor, 1997). Strictly speaking, integral invariants are not guaranteed to be conserved, but this may not be significant if sufficiently high-order elements are used. By contrast, our method is conservative and fully de-aliased, and long inviscid runs can be performed.

DOI: 10.1002/qj

The latitude-dependent zonal truncation that we define presents obvious similarities with the practice of zonal filters in association with methods based on finite differences (Purser, 1988) or double Fourier series (Cheong, 2000). However a significant difference is that zonal filters are used in that context as an a posteriori fix, while in our case they are embedded from the start in the construction of the functional space $\mathcal{S}''$. As a result, basis functions spread the latitudinal jumps in zonal bandwidth over several latitudes. This preserves the accuracy of the approximation of the latitudinal gradients, in contrast with what happens with finite differences (Purser, 1988). Furthermore, the Galerkin framework provides nonlinear stability while the pseudo-spectral methods based on truncated double Fourier series require the periodic application of a spherical harmonics projector (Cheong, 2000).

Weather patterns like fronts and natural forcings like orography tend to present sharp horizontal gradients, which will look like discontinuities from the point of view of a coarsely resolved climate model. Since our basis functions are infinitely differentiable in the zonal direction, they approximate discontinuous fields poorly and produce Gibbs oscillations. This behaviour is well-known for spectral methods and is especially undesirable for constituents such as water vapour (Rasch and Williamson, 1990). With our method, as in spectral models, one can consider damping spurious oscillations using explicit diffusion, like an iterated Laplacian operator. It is easy to include such a diffusion in our method, provided the basis functions have sufficient latitudinal smoothness. However it is probably preferable not to use our method when positivity is crucial, since it suffers from the same drawbacks as the spectral method.

We have not tried to design a shock-resolving method, i.e. one that would accommodate for discontinuities in the scalar fields. In such methods, like the finite-volume method or the discontinuous Galerkin method, there are several consistent ways to compute fluxes across discontinuities. This extra flexibility can be exploited to guarantee the positivity of the transported quantities, usually through some form of upwinding (Lin and Rood, 1996; Hourdin and Armengaud, 1999 and references therein). Such strategies could be applied in the latitudinal direction, using a subspace of $\mathcal{S}$ not enforcing latitudinal continuity. However, the resulting method would be shock-resolving only for discontinuities parallel to the Equator. For a general discontinuity, we still expect Gibbs oscillations to occur due to the Fourier-based discretization in the zonal direction. Control of these oscillations would require explicit dissipation, thus annihilating much of the appeal of shock-resolving methods, which can run without explicit dissipation. It should be feasible, though not straightforward, to design a truly shock-resolving method by dividing the zonal interval into elements, and replacing the zonal Fourier basis by a basis of functions with support limited to the zonal elements. It may be interesting to explore this avenue in future.

We have restricted the present work to scalar fields and non-divergent flows, while the minimal test-bed for atmospheric applications is the compressible Saint-Venant model. This restriction is motivated by the fact that exact conservation of linear and quadratic integral invariants can be generically obtained within the Galerkin framework. We therefore focused on problems with such invariants as a first step. The energy and enstrophy of the Saint-Venant model are not quadratic invariants, and will not be exactly conserved even with exact quadrature. Nevertheless, the method developed here can be readily extended to compressible flows. A straightforward possibility is to use the stream function–velocity potential representation, as in spectral models. Continuously differentiable scalar fields would be needed in order to represent a continuous wind field. Another possibility would be to design algebras $\mathcal{S}_u$ and $\mathcal{S}_v$ similar to $\mathcal{S}$ but suitable to represent the zonal and latitudinal wind components, with their specific near-pole behaviour. This should be slightly more economical since the required smoothness, and hence polynomial degree, is less. Work is under way to explore both possibilities (Dubos 2009).

Acknowledgements

The author thanks two anonymous referees for useful criticism and suggestions.

This work was initiated during a research visit to the Applied Physics Laboratory, University of Washington, Seattle (USA) supported by the Délégation Générale à l'Armement, grant ERE 0760027.

Appendix

Implementation of the method

We consider the advection–diffusion problem (11), spatially discretized following the Galerkin method associated with the space $\mathcal{S}''$, i.e. we look for a time-dependent

$$f = \sum_{(m,l)\in K} a_{ml}(t) F_{ml}(\lambda, \phi) \tag{A.1}$$

such that

$$\forall g \in \mathcal{S}'' \; M\ (g,\ \partial_t f) + T\ (g,\ f) = -\kappa S\ (g,\ f). \tag{A.2}$$

This yields the coupled system of ODEs satisfied by the set of complex coefficients $\mathbf{a} = (a_{ml})$:

$$M^{(m)}_{ll'}\ \partial_t a_{ml'} + T_{ml}(\mathbf{a}) + \kappa S^{(m)}_{ll'} a_{ml'} = 0, \tag{A.3}$$

where $(m,\ l) \in K$, $(m,\ l') \in K$, summation over the index l' is implied, and

$$T_{ml}(\mathbf{a}) = T(F_{ml},\ F_{m'l'}) a_{m'l'}, \tag{A.4}$$

where summation over indexes $(m', l') \in K$ is implied.

DOI: 10.1002/qj

Notice that the mass matrices $\mathbf{M}^{(m)} = (M_{ll'}^{(m)})$ depend only on whether m is even or odd:

$$M_{ll'}^{(2m)} = M_{ll'}^{\text{even}}, \qquad M_{ll'}^{(2m+1)} = M_{ll'}^{\text{odd}}.$$

The matrices $\mathbf{M}^{\text{even}}$ and $\mathbf{M}^{\text{odd}}$ are computed by quadrature as follows. Consider the matrix $\mathbf{B} = (B_l(z_j))$ containing the values of the finite elements B_{ml} at the qN quadrature points $z_j = \sin\phi_j$. Then

$$M_{ll'}^{\text{even}} = \int H_l(z) H_{l'}(z)\ \mathrm{d}z = \sum_j w_j H_l(z_j) H_{l'}(z_j), \tag{A.5}$$

$$M_{ll'}^{\text{odd}} = \int \cos\phi H_l(z)\ \cos\phi H_{l'}(z)\ \mathrm{d}z = \sum_j (1-z_j)^2 w_j H_l(z_j) H_{l'}(z_j), \tag{A.6}$$

where w_j are the qN Gaussian quadrature weights. Hence

$$\mathbf{M}^{\text{even/odd}} = \mathbf{H}^* \cdot \mathbf{W}^{(0/2)} \cdot \mathbf{H}, \tag{A.7}$$

$$W_{jj}^{(2m)} = (1-z_j^2)^m w_j \phi_j, \tag{A.8}$$

where $\mathbf{W}^{(2m)}$ is a $qN \times qN$ diagonal matrix. Also, one finds that

$$\mathbf{S}^{(m)} = m^2 \mathbf{N}^{\text{even/odd}} + \mathbf{S}'^{\text{even/odd}}, \tag{A.9}$$

where

$$\mathbf{N}^{\text{even/odd}} = \mathbf{H}^* \cdot \mathbf{W}^{(-2/0)} \cdot \mathbf{H}, \tag{A.10}$$

$$\mathbf{S}'^{\text{even/odd}} = \mathbf{G}^* \cdot \mathbf{W}^{(0/2)} \cdot \mathbf{G}, \tag{A.11}$$

$$G_{jl} = \frac{\mathrm{d}H_l}{\mathrm{d}z}(z_j). \tag{A.12}$$

Elements of the matrices $\mathbf{M}^{\text{even/odd}}$ and $\mathbf{S}^{\text{even/odd}}$ are integrals of piecewise polynomials of degree at most $2d+2$, since $\cos^2\ \phi = 1 - z^2$. Therefore exact quadrature is achieved with $d+1$ Gauss quadrature points in each latitudinal subinterval. Notice that while $\mathbf{N}^{\text{odd}} = \mathbf{M}^{\text{even}}$, the quadrature formula used to compute $\mathbf{N}^{\text{even}}$ is not exact due to the non-polynomial weight $\cos^{-2}\ \phi$. This does not affect the conservation properties of the scheme, as discussed in subsection 4.2. Only the matrices $\mathbf{H}$, $\mathbf{G}$, $\mathbf{M}^{\text{even/odd}}$, $\mathbf{S}^{\text{even/odd}}$ and $\mathbf{N}^{\text{even}}$ need be computed, once for all. Due to the local support of the finite elements, $\mathbf{M}^{\text{even/odd}}$, $\mathbf{S}^{\text{even/odd}}$ and $\mathbf{N}^{\text{even}}$ are banded with $(2d+1)$ diagonals; also, $\mathbf{F}$ and $\mathbf{G}$ are qd-diagonal. The cost for computing and storing these matrices is $\mathcal{O}(Nd^2)$ and $\mathcal{O}(Nd)$ respectively.

The computation of the transport term $T_{ml}(\mathbf{a})$ takes place in three steps.

(1) The values $f(\lambda_i,\ \phi_j)$ at the quadrature points are first obtained from the complex coefficients (α_{lm}). For this, the even and odd coefficients α_{lm} are Fourier-transformed for each $l = 0 \ldots N+d-1$, producing sets of values $a_l^{\text{even}}(\lambda_i)$ and $a_l^{\text{odd}}(\lambda_i)$ at zonal quadrature points. This costs $\mathcal{O}(N^2 \ln(N))$. The values $f(\lambda_i, \phi_j)$ are then obtained as

$$f(\lambda_i,\ \phi_j) = a_l^{\text{even}}(\lambda_i) H_l(\phi_j) + a_l^{\text{odd}}(\lambda_i) \cos(\phi_j) H_l(\phi_j),$$

where summation over the index l is implied. This costs $\mathcal{O}\ ((d+1)qN^2)$, again because the matrix $\mathbf{B} = (B_l(\phi_j))$ has only $\mathcal{O}\ ((d+1)qN)$ non-zero entries. The gradients $(\partial_\lambda f/\cos\phi,\ \partial f)$ and $(\partial_\lambda \psi/\cos\phi,\ \partial\psi)$ are computed at the quadrature points by an analogous operation also involving the matrix $\mathbf{G} = (\mathrm{d}B_l/\mathrm{d}z_j)$. The boundary conditions (19) guarantee finite gradients near the poles.

(2) The Jacobian $J(\psi, f)$ is then computed pointwise at the quadrature points. This costs $\mathcal{O}\ (qN^2)$.

(3) The integrals

$$T_{ml}(\mathbf{a}) = \langle F_{ml}^* J(\psi,\ f)\rangle \tag{A.13}$$

are finally computed using the quadrature rules:

$$\begin{aligned}
&\langle F_{ml}^*\ J(\psi,\ f)\rangle \\
&= \int \cos^p\ \phi H_l(\phi) \mathrm{e}^{-\mathrm{i}m\lambda}\ J(\phi,\lambda)\ \frac{\cos\phi \mathrm{d}\phi \mathrm{d}\lambda}{4\pi} \\
&= \int \cos^p\ \phi H_l(\phi)\ \hat{J}_m(\phi) \frac{\mathrm{d}z}{2} \\
&= \sum_j w_j\ \cos^p\ \phi_j H_l(\phi_j)\ \hat{J}_m(\phi_j),
\end{aligned}$$

where $\hat{J}$ is the discrete zonal Fourier transform of J and $p = 0$ (resp. $p = 1$) if m is even (resp. odd). The columns of the matrix $T_{ml}(\mathbf{a})$ can therefore be obtained by first Fourier-transforming into $\hat{\mathbf{J}}$ the matrix $\mathbf{J}$ whose columns contain the values of the flux J along each meridian, then separating even and odd modes and finally left-multiplying by $\mathbf{B}^*$ and $\mathbf{B}^*\mathbf{W}^{(1)}$ respectively. The associated costs are $\mathcal{O}(qN^2 \log(N))$ and $\mathcal{O}\ ((d+1)qN^2)$. A more efficient approach is to first compute the 2 matrices $\mathbf{Q_p} = \mathbf{H}^*\mathbf{W}^{(\mathbf{p})}\mathbf{J}$ $(p = 0,\ 1)$ and then their Fourier transforms $\hat{\mathbf{Q}}_\mathbf{p}$. The values $T_{ml}(\mathbf{a})$ are then obtained by combining the columns of $\hat{\mathbf{Q}}_\mathbf{o}$ (m even), and $\hat{\mathbf{Q}}_\mathbf{1}$ (m odd). The associated costs are $\mathcal{O}\ ((d+1)qN^2)$ and $\mathcal{O}\ (N^2 \log N)$.

As $\log(N) \gg (d+1)q$, the cost is dominated by the zonal Fourier transforms and scales like $N^2 \log(N)$, independently of the number q of quadrature points and the degree d of the B-splines. The quadrature rules are applied to triple products of the form $\langle \partial_z F_{ml}^*\ F_{m'l'}\ F_{m''l''}\rangle$. Therefore the usual 2/3 truncation rule applies to the zonal Fourier transforms. Furthermore, there are at most two odd numbers among m, m', m''. The degree of the piecewise polynomials to be integrated therefore cannot exceed $3d+1$ and $q = 3$ (resp. $q = 4,\ 6$). Gauss–Legendre

DOI: 10.1002/qj

quadrature points provide exact quadrature for $d = 1$ (resp. $d = 2$, 3) respectively.

References

Arakawa A. 1966. Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. Part I. *J. Comput Phys.* **1**: 119–143.

Arakawa A, Lamb VR. 1981. A potential enstrophy and energy conserving scheme for the shallow water equations. *Mon. Weather Rev.* **109**: 18–36.

Baer F, Wang H, Tribbia JJ, Fournier A. 2006. Climate modeling with spectral elements. *Mon. Weather Rev.* **134**: 3610–3624.

Cheong HB. 2000. Double Fourier series on a sphere: Applications to elliptic and vorticity equations. *J. Comput Phys.* **157**: 327–349.

Cote J, Staniforth A. 1990. An accurate and efficient finite-element global model of the shallow-water equations. *Mon. Weather Rev.* **118**: 2707–2717.

Dubos T. 2009. 'High-order quasi-uniform approximation on the sphere using Fourier-finite-elements'. In *International Conference on High-Order Methods*.

Hourdin F, Armengaud A. 1999. The use of finite-volume methods for atmospheric advection of trace species. Part I: Test of various formulations in a general circulation model. *Mon. Weather Rev.* **127**: 822–837.

Lee J-L, MacDonald AE. 2009. A finite-volume icosahedral shallow-water model on a local coordinate. *Mon. Weather Rev.* **137**: 1–0000.

Lin S-J, Rood RB. 1996. Multidimensional flux-form semi-Lagrangian transport schemes. *Mon. Weather Rev.* **124**: 2046–2070.

Lynch P. 2009. On resonant Rossby–Haurwitz triads. *Tellus A* **61**: 438–445.

Orszag SA. 1970. Transform method for the calculation of vector-coupled sums: Application to the spectral form of the vorticity equation. *J. Atmos. Sci.* **27**: 890–895.

Purser RJ. 1988. Degradation of numerical differencing caused by Fourier filtering at high-latitudes. *Mon, Weather Rev.* **116**: 1057–1066.

Rancic M, Zhang H, Savic-Jovcic V. 2008. Nonlinear advection schemes on the octagonal grid. *Mon. Weather Rev.* **136**: 4668–4686.

Rasch PJ, Williamson DL. 1990. Computational aspects of moisture transport in global models of the atmosphere. *Q. J. R. Meteorol. Soc.* **116**: 1071–1090.

Sadourny R. 1975. Compressible model flows on the sphere. *J. Atmos. Sci.* **32**: 2103–2110.

Satoh M, Matsuno T, Tomita H, Miura H, Nasuno T, Iga S. 2008. Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations. *J. Comput. Phys.* **227**: 3486–3514.

Swarztrauber PN. 1996. Spectral transform methods for solving the shallow-water equations on the sphere. *Mon. Weather Rev.* **124**: 730–744.

Taylor M. 1997. The spectral element method for the shallow water equations on the sphere. *J. Comput. Phys.* **130**: 92–108.

Williamson DL. 2007. The evolution of dynamical cores for global atmospheric models. *J. Meteorol. Soc. Jpn* **85B**: 241–269.

DOI: 10.1002/qj