RMetS
Royal Meteorological Society

Horizontal grids for global weather and climate prediction models: a review

Andrew Staniforth[a*†] and John Thuburn[b]
[a]*Met Office, Exeter, UK*
[b]*University of Exeter, UK*
*Correspondence to: A. Staniforth, Met Office, FitzRoy Road, Exeter EX1 3PB, UK.
E-mail: Andrew.Staniforth@metoffice.gov.uk
†The contribution of this author was written in the course of his employment at the Met Office, UK, and is published with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

A latitude–longitude grid is used by almost all operational atmospheric forecast models, and many research models. However, it is expected that the advantages of a latitude–longitude grid will become outweighed on massively parallel computers by data-communication bottlenecks. There is therefore renewed interest in quasi-uniform alternatives. This review surveys and assesses previously proposed horizontal grids for modelling the atmosphere over the sphere. Aspects of numerical accuracy likely to be affected by grid structure are discussed; particular attention is paid to computational modes and grid imprinting. Computational modes are potentially very serious, since they may be excited in realistic applications by boundary conditions, nonlinearity, physical forcing, and data assimilation. The geometry of polyhedra is reviewed due to its relation to numerical degrees of freedom, and hence to numerical wave dispersion and the possible existence of computational modes.

All grids proposed to date have known problems or issues that merit further investigation. Orthogonal logically rectangular grids may be generated using conformal maps, but these suffer from singularities and resolution clustering. Resolution clustering may be avoided by using overset grids, but there are potential issues associated with the overlap regions. Alternatively, resolution clustering may be avoided, whilst retaining a logically rectangular grid, by giving up orthogonality; however, existing numerical schemes exploit orthogonality to obtain various properties thought to be important for accuracy, and it is not yet known whether these can also be obtained on non-orthogonal grids. Quasi-uniformity and orthogonality can be obtained without resolution clustering or overlaps by using non-quadrilateral grid cells, such as triangles, or pentagons and hexagons. However, when a staggered placement of variables is used to minimise dispersion errors for fast waves, non-quadrilateral grids support computational modes.

In view of the lack of a single ideal grid, several topics meriting further investigation are identified. **Copyright © 2011 Royal Meteorological Society and British Crown Copyright, the Met Office**

Key Words: computational mode; conformal mapping; grid imprinting; numerical dispersion; orthogonality; polyhedra

Received 29 June 2011; Revised 26 September 2011; Accepted 28 September 2011; Published online in Wiley Online Library 14 November 2011

Citation: Staniforth A, Thuburn J. 2011. Horizontal grids for global weather and climate prediction models: a review. *Q. J. R. Meteorol. Soc.* **138**: 1–26. DOI:10.1002/qj.958

1. Introduction

Many operational weather and climate models are based on a latitude–longitude (lat–lon) spherical grid. Its logically rectangular structure, orthogonality, and symmetry properties, make it relatively straightforward to obtain various desirable, accuracy-related, properties (section 1.2). However, due to the convergence of the meridians at the poles, the lat–lon grid is expected to lead to a severe scalability bottleneck on the massively parallel computing platforms that are becoming available, and many modelling groups are now investigating alternatives. It is therefore timely to review the spherical grids that have been proposed in the literature. However, when considering the choice of grid, it is important not to focus exclusively on the massively parallel scaling issue since it is only one, albeit an important one, of a number of issues that require examination *in toto*. The geometrical properties of a grid, in combination with the numerical methods used, have direct consequences for the accuracy of numerical simulations. In particular, it is possible that the structure of the underlying grid becomes visible in the solution in the form of noise or systematic errors ('grid imprinting'), or the numerical model might support spurious dynamical modes ('computational modes') that have no physical counterparts and degrade the fidelity of the numerical solution. Perfect scalability of a numerical model on a massively parallel computer is of no benefit if it is achieved by unacceptably compromising accuracy or stability. Also, the grid structure can constrain the types of numerical scheme that are feasible. This review surveys the various approaches that have been proposed for gridding the sphere, and some of the likely implications of grid structure for numerical accuracy.

1.1. Limitations of the full lat–lon grid

On a full lat–lon grid, the convergence of the meridians leads to resolution clustering at the poles. (Here we use the term 'full' to signify the standard lat–lon grid, and thereby distinguish it from the 'reduced' lat–lon grids discussed in section 3.1.) For an explicit time integration scheme with Eulerian advection, this convergence places an unbearable restriction on the size of time step as resolution is increased. This restriction, and other problems related to the poles, stimulated some early research into quasi-uniform spherical grids (Sadourny *et al.*, 1968; Williamson, 1968, 1969, 1970, 1971; Sadourny, 1972; Cullen, 1974). However, the development of semi-implicit time integration schemes in combination with semi-Lagrangian advection (e.g. Staniforth and Côté, 1991; Staniforth, 1997; Williamson, 2007), allowed the severe time-step restriction to be bypassed whilst retaining the lat–lon grid, thereby greatly enhancing the viability of lat–lon grids. Such schemes have proven very popular in operational and research models worldwide (e.g. Williamson, 1997; Côté *et al.*, 1998; Tolstykh, 2003; Davies *et al.*, 2005; Zhang and Shen, 2008).

However, both semi-implicit time integration (which requires the solution of a global elliptic problem at each time step) and semi-Lagrangian advection require significant data communication among the grid points clustered around the poles. When the grid is decomposed across a number of processors on a parallel computer, these schemes then require significant data communication between processors. Communication between processors is slow and can leave the processors waiting for data; this then becomes the bottleneck to performance, and to the reduction of computing time with increasing number of processors (scalability). Current thinking on how computer architectures will evolve is that increased computer power will be achieved by harnessing hundreds of thousands of processors which are individually little or no faster than current ones. If true, it is anticipated that the resolution-clustering property of the lat–lon grid around its poles will ultimately make it unfeasible for high-resolution modelling, since increased resolution will only worsen the scaling problem. There is therefore renewed impetus to investigate alternative, quasi-uniform spherical grids as a basis for global weather and climate models (Williamson, 2007). In this regard, a valuable survey of recent developments in numerical techniques for global atmospheric models, based upon lectures given at a symposium held in 2008 at the National Center for Atmospheric Research, Boulder, Colorado, may be found in Lauritzen *et al.* (2011).

1.2. Essential and desirable properties of a dynamical core

The *dynamical core* may be loosely defined as the component of a numerical model that solves the adiabatic and frictionless governing equations on resolved scales. (There are some subtleties, such as whether scale-selective dissipation terms should be considered part of the dynamical core or a parametrisation of subgrid-scale processes (Williamson *et al.*, 1992; Thuburn, 2008b); these need not concern us here.) Certain properties are considered essential, or at least highly desirable, for a dynamical core. They affect the stability and accuracy of the numerical solution, particularly the behaviour of marginally resolved scales. The ease with which we can obtain these properties is strongly influenced by the geometric properties of the chosen grid.

1. Mass conservation.

Conservation of mass of dry air and mass of trace species is considered highly desirable for climate research and climate prediction. (For the much shorter integration times used in weather prediction, however, conservation is not usually considered essential.) Mass conservation is straightforward to achieve with an Eulerian finite-volume discretisation of the mass continuity equation on any grid structure. However, mass-conserving semi-Lagrangian discretisations are more complex (Zerroukat *et al.*, 2004; Lauritzen *et al.*, 2010b) and typically depend on being able to exploit some underlying grid structure, such as a logically-rectangular structure, to avoid being excessively expensive. Because of the complexity of conservatively mapping from the Eulerian grid cells to arbitrary Lagrangian departure cells, it seems unlikely that mass-conserving semi-Lagrangian schemes can be competitive on more general kinds of grid.

2. Accurate representation of balanced flow and adjustment.

The large-scale atmospheric flow is dominated energetically by motions close to hydrostatic and geostrophic balance. It is crucial, therefore, for a dynamical core to accurately represent the slow, balanced component of the flow: the Rossby waves and nonlinear vortical motions. It is also crucial for a dynamical core to represent the *adjustment* process by which the flow returns towards balance after perturbation away from balance. Adjustment occurs

through the radiation, dispersion, and ultimate dissipation of fast acoustic and inertio-gravity waves. To capture the dispersion of the full spectrum of fast waves sufficiently accurately, three approaches are currently known: (i) a *C*-grid or *Z*-grid placement of the predicted variables (section 5) with centred-difference approximations to the relevant spatial derivatives, (ii) a spectral representation of the predicted variables with spatial derivatives calculated directly from the basis functions, or (iii) a collocated placement of predicted variables with a Riemann solver based numerical method, e.g. Ullrich *et al.* (2010). Spectral methods require global rearrangement of data in transforming between the spectral and grid-space representation of model fields at each time step; therefore there is growing concern regarding the parallel efficiency of spectral methods on future exascale supercomputing architectures. Riemann solver methods are at an early stage of development for atmospheric modelling; they are not yet widely used and, to date and to our knowledge, are exclusively based on directional splitting, restricting their use to grids with a local logically-rectangular structure. The *Z*-grid requires the solution of Poisson equations on each model level at each time step to recover the wind field from the vertical vorticity and horizontal divergence; its use, therefore, is contingent on the availablity of a sufficiently fast and scalable Poisson solver. For these reasons we concentrate mainly on *C*-grid discretisations in the rest of this review.

3. Computational modes should be absent or well controlled.

When linearised about some simple basic state, such as a state of rest, the discretisation may be expected to support a spectrum of wave modes that are analogous to those supported by the continuous governing equations. In some cases, however, the discrete modes may have no continuous analogues, or they may exhibit unphysical behaviour such as failure to propagate (zero frequency); they are then called *computational modes*. Some schemes have one or more families of computational modes in their discrete dispersion relation (section 5). Computational modes often have spatial structure close to the grid scale. Their existence can lead to unphysical behaviour in response to forcing by boundary conditions, physical parametrisations, or nonlinearity, or to the insertion of data through data assimilation. The problem often manifests itself as near-grid-scale noise in one or more fields, but it can also appear, for example, when a necessary condition for a physical instability is spuriously satisfied by the numerics (Arakawa and Moorthi, 1988). For these reasons, computational modes are considered highly undesirable. If a scheme supports them, then a mechanism for controlling them is essential. The existence of (branches of) computational modes in the discrete dispersion relation is related to the relative numbers of degrees of freedom in the mass and velocity fields. The number of degrees of freedom is in turn related to the chosen grid staggering, and to the associated number of vertices, faces, and edges. Much of the discussion below therefore focuses on these numbers of degrees of freedom for different grids, assuming a *C*-grid placement of variables.

4. The geopotential gradient and pressure gradient should produce no unphysical source of vorticity.
5. Terms involving the pressure should be energy conserving.
6. Coriolis terms should be energy conserving.
7. There should be no spurious fast propagation of Rossby modes; geostrophic balance should not spontaneously break down.
8. Axial angular momentum should be conserved.

These last five properties are all related to the ability of the discretisation to mimic basic geometrical or mathematical properties of the continuous equations. Schemes with such properties are sometimes called *mimetic*, e.g. Hyman and Shashkov (1997) and Taylor and Fournier (2010).

The geopotential-gradient and pressure-gradient terms are the two largest terms in the governing equations, so it is crucial that these terms are approximated accurately. It is especially important that there should be no spurious forcing of vortical motions that could project onto the energetically-dominant meteorological signal. The requirement here is that the discretisation should possess a discrete analogue of the vector calculus identity $\nabla \times \nabla\Psi \equiv 0$ for any scalar field Ψ.

The adiabatic frictionless continuous governing equations conserve energy. However, they do permit transfer of energy between scales, and there is evidence of a small but systematic nonlinear downscale energy transfer in the real atmosphere at scales near the resolution limit of typical weather and climate models. There is room for debate, therefore, over whether a dynamical core should conserve energy on resolved scales or permit a transfer to unresolved scales (Thuburn, 2008b). However, there is a strong argument for requiring the large, and essentially linear, pressure-gradient and Coriolis terms to be individually energy conserving. Energy conservation by the terms involving pressure typically reduces to the requirement that the discretisation have a discrete analogue of the vector calculus identity $\mathbf{v} \cdot \nabla\Psi + \Psi\nabla \cdot \mathbf{v} \equiv \nabla \cdot (\Psi\mathbf{v})$ for any scalar field Ψ and vector field $\mathbf{v}$. Energy conservation by the Coriolis terms can be achieved straightforwardly when all components of the velocity vector are stored at the same location. However, for a *C*-grid staggering it becomes non-trivial, even on a lat–lon grid (Arakawa and Lamb, 1977).

As discussed above, the discretisation should support a spectrum of waves analogous to those supported by the continuous equations. The Rossby waves should be slowly propagating. In the limit of constant Coriolis parameter f, their frequency should become zero, and for non-constant f their frequency should be of order β, the gradient of f. Failure to satisfy these properties means that flows that ought to be geostrophically balanced spontaneously break down on a short timescale. The first *C*-grid scheme to be analysed for a hexagonal grid (Ničković *et al.*, 2002) suffered from this spontaneous breakdown of balance, though a solution was subsequently found (Thuburn, 2008a; Ringler *et al.*, 2010). For linear shallow-water flow with constant f, the solution requires the scheme to possess a discrete analogue of the vorticity equation $\partial_t\zeta + f\nabla \cdot \mathbf{v} = 0$, so that the vorticity ζ is steady when the divergence $\nabla \cdot \mathbf{v}$ vanishes.

Conservation of axial angular momentum is desirable for an accurate representation of the zonal jets that dominate the large-scale circulation of the troposphere and middle atmosphere. However, as far as we are aware, to date this has only been achieved for discretisations on a lat–lon grid (in this context, the Gaussian grid of a spectral model is considered to be a lat–lon grid). To achieve it on a grid other than a lat–lon grid appears to be particularly challenging.

It is far from trivial to obtain all of the properties 1 to 8 in a single scheme. One of the reasons for the sustained

popularity of the lat–lon grid is that its logically rectangular structure, orthogonality and symmetry can be exploited to obtain all eight properties. Obtaining the same properties on alternative grid structures is a challenging current area of research. Some recent progress has been made for horizontal finite-difference *C*-grid approximations (Thuburn *et al.*, 2009; Ringler *et al.*, 2010) provided the grid has a dual that is orthogonal to it–for example a Voronoi diagram and its associated Delaunay triangulation (Augenbaum and Peskin, 1985). This scheme gives properties 1, 2, and 4–7; orthogonality is relied upon to obtain 4–7. Whether or not property 3 is obtained depends on the numbers of mass and velocity degrees of freedom.

9. Accuracy approaching second order.
10. Minimal grid imprinting.

Only five truly homogeneous spherical grids are possible, and these are all far too coarse for practical use (section 2). Any practical grid will be, at best, quasi-uniform, and will have a number of special points, or special lines, where the grid structure is locally different from the structure elsewhere. It may be difficult to achieve a discretisation that is second-order accurate near these special points and lines. Even if second-order accuracy can be achieved, the dependence of the truncation error on the local grid structure is likely to leave some signal of the grid structure in the numerical solution (section 6). It is desirable to choose the grid and discretisation to minimise such grid imprinting.

1.3. Scope of the review and outline

Because grid geometry is so important, some of the mathematics of polyhedra is reviewed and discussed in section 2. This provides some important information on the relative number of degrees of freedom that result from where one places dynamical variables on the differently-shaped polygonal faces of a polyhedron.

Many quadrilateral grids have been proposed in the past, and these are reviewed in section 3. A brief overview of triangular and hexagonal/pentagonal grids that have been used for atmospheric models–there are relatively few such grids–is presented in section 4. Atmospheric models have also been developed using unstructured triangular and pentagonal-hexagonal grids (e.g. Bacon *et al.*, 2000; Gopalakrishnan *et al.*, 2002; Ford *et al.*, 2004a,b; Pain *et al.*, 2005; Piggott *et al.*, 2005; Läuter *et al.*, 2007; Weller and Weller, 2008; Bernard *et al.*, 2009b; Weller *et al.*, 2009, 2010; Ringler *et al.*, 2011) and are similar in many ways to those for oceanic applications. An important virtue of unstructured grids is their ability to match a grid to awkward-shaped boundaries, such as the land boundaries of oceanic models. This is however much less important for atmospheric modelling since the lower boundary can generally be handled by terrain-following coordinates. A further advantage of unstructured grids is for static or dynamically adaptive mesh refinement. Although very attractive in principle, in practice there are problems with this approach, principally due to spurious dispersion issues; Slingo *et al.* (2009) provide a recent review of this area of research. Any model that uses an unstructured grid must, as a minimum, perform well for the special case when the grid is configured to have structure. For this reason, and in the present context, unstructured grid methods are only considered peripherally.

Analyses of Rossby-wave dispersion and geostrophic adjustment for discretisations of the shallow-water equations are discussed in section 5. This kind of analysis is a powerful tool for exposing the existence of spurious or poorly represented wave modes for any proposed grid and discretisation, and hence for predicting possible poor behaviour in more complete models. Previous investigations into grid imprinting are reviewed in section 6.

Conclusions are drawn in section 7, and some open questions that merit further investigation are identified.

2. Polyhedra

2.1. Euler's formula for convex polyhedra

Euler's formula for convex polyhedra states that the number of vertices plus the number of faces equals the number of edges plus two. Thus

$$V + F = E + 2, \tag{1}$$

where V, F, and E are the numbers of vertices, faces, and edges, respectively. For a proof of (1), and a fun discussion of Platonic and Archimedian polyhedra, the interested reader is referred to White (1997).

2.2. The Platonic solids

A Platonic solid is a regular convex polyhedron, i.e. its faces are congruent regular polygons, with the same number of faces meeting at each vertex, so that all its edges are congruent, as are all its vertices and angles. Only five such solids exist (Figure 1).

2.2.1. Constraints for regular convex polyhedra

For a regular convex polyhedron with $F \geq 3$ congruent faces, let $m \geq 3$ be the number of edges of each polygonal face. Since there are $n \geq 3$ edges that meet at each vertex, and there are V vertices, this gives nV associated edges. But each edge has been double counted since it is associated with, and joins, exactly two vertices. Thus

$$E = \frac{nV}{2}. \tag{2}$$

Furthermore, there are n faces that meet at each vertex, and there are V vertices, giving nV faces. But each face has $m \geq 3$ edges and is associated with m vertices, so this has led to an overcounting of faces by a factor of m. Taking this into consideration gives

$$F = \frac{nV}{m}. \tag{3}$$

Solving (1)–(3) for V, F and E (all positive integers) in terms of $m \geq 3$ and $n \geq 3$, both integers, then yields

$$\left.\begin{aligned} V &= \frac{4m}{2m - (m-2)\,n}, \\ F &= \frac{4n}{2m - (m-2)\,n}, \\ E &= \frac{2mn}{2m - (m-2)\,n}. \end{aligned}\right\} \tag{4}$$

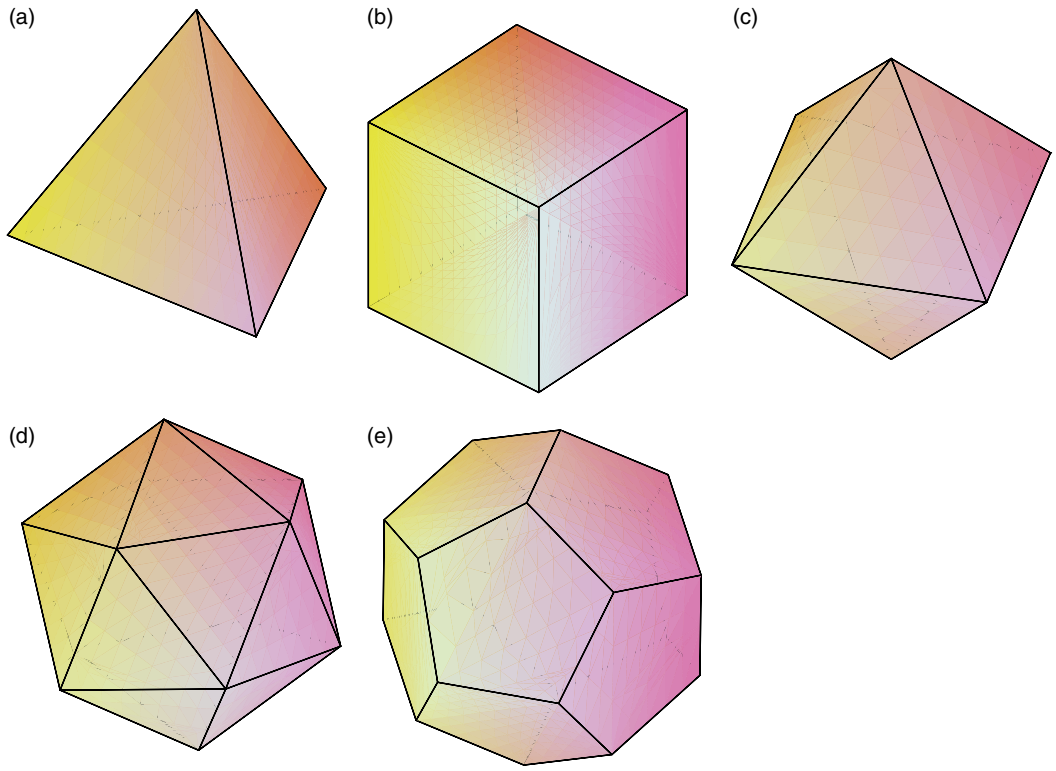

Figure 1. The five Platonic solids: (a) tetrahedron, (b) hexahedron (cube), (c) octahedron, (d) icosahedron, and (e) dodecahedron.

Now the interior angle of an m-sided polygon is $(m-2)\,\pi/m$. Since the number of polygons meeting at a vertex is n, the total angle subtended by the polygons meeting there is $n\,(m-2)\,\pi/m$, and this has to be strictly less than 2π for a convex polyhedron. Thus

$$n\,(m-2)\,\pi/m < 2\pi \quad \Rightarrow \quad 2m-(m-2)\,n > 0, \qquad (5)$$

i.e. the denominator in (4) is strictly positive, and therefore guaranteed to be non-singular.

2.2.2. Regular convex polyhedra composed entirely of equilateral triangles

For a regular convex polyhedron composed entirely of equilateral *triangles*, the number of edges of each triangular face is $m = 3$. Since the sum of the angles of the (equilateral) triangles that meet at a vertex has to be strictly less than 2π, so $3 \leq n \leq 5$. Setting $m = 3$ in (4), and letting n successively vary from 3 to 5, then gives:

1. a regular *tetrahedron* (Figure 1(a)), where

$$\{m,n\} = \{3,3\}\,, \quad V = 4,\ E = 6,\ F = 4, \qquad (6)$$

   and $\{m,n\}$ is termed the Schläfli symbol (Coxeter, 1973);
2. a regular *octahedron* (Figure 1(c)), where

$$\{m,n\} = \{3,4\}\,, \quad V = 6,\ E = 12,\ F = 8; \qquad (7)$$

3. a regular *icosahedron* (Figure 1(d)), where

$$\{m,n\} = \{3,5\}\,, \quad F = 20,\ V = 12,\ E = 30. \qquad (8)$$

2.2.3. Regular convex polyhedron composed entirely of squares

For a regular convex polyhedron composed entirely of *squares*, the number of edges of each square is $m = 4$. Since the sum of the angles of the squares that meet at a vertex has to be strictly less than 2π, so $n = 3$. Setting $m = 4, n = 3$, in (4) then gives a *cube* (Figure 1(b)), which is a special case of a *hexahedron* (i.e. a solid composed of six plane faces), where

$$\{m,n\} = \{4,3\}\,, \quad V = 8,\ E = 12,\ F = 6. \qquad (9)$$

2.2.4. Regular convex polyhedron composed entirely of pentagons

For a regular convex polyhedron composed entirely of pentagons, the number of edges of each pentagon is $m = 5$. Since the sum of the angles of the pentagons that meet at a vertex has to be strictly less than 2π, so $n = 3$. Setting $m = 5, n = 3$ in (4) then gives a regular *dodecahedron* (Figure 1(e)), where

$$\{m,n\} = \{5,3\}\,, \quad V = 20,\ F = 12,\ E = 30. \qquad (10)$$

2.2.5. Other regular convex polyhedra

Other regular convex polyhedra are not possible since (4)–(5) cannot be simultaneously satisfied for $m \geq 6$ and $n \geq 3$. For example, if $m = 6$, i.e. the faces are all regular hexagons, then (5) implies that $n < 3$, in contradiction of the constraint $n \geq 3$.

2.2.6. Summary table

Only five regular convex polyhedra exist, termed the Platonic solids, and their properties are summarised in Table I. All

Table I. Number of vertices (V), faces (F), and edges (E), for regular polyhedra, as constrained by Euler's formula. $\{m, n\}$ is the Schläfli symbol, where $m \geq 3$ is the number of edges of each regular polygonal face, and $n \geq 3$ is the number of edges or, equivalently, polygonal faces, that meet at a vertex. When a regular polyhedron is inscribed in the unit sphere, L is the length of any of its edges.

Polyhedron	V	+	F	=	E	+	2	=	Total	$\{m, n\}$	L
Tetrahedron	4	+	4	=	6	+	2	=	8	$\{3, 3\}$	$2\sqrt{2/3}$
Octahedron	6	+	8	=	12	+	2	=	14	$\{3, 4\}$	$\sqrt{2}$
Cube	8	+	6	=	12	+	2	=	14	$\{4, 3\}$	$2/\sqrt{3}$
Icosahedron	12	+	20	=	30	+	2	=	32	$\{3, 5\}$	$\{\sin(2\pi/5)\}^{-1}$
Dodecahedron	20	+	12	=	30	+	2	=	32	$\{5, 3\}$	$(\sqrt{5}-1)/\sqrt{3}$

five of them can be inscribed in a sphere of unit radius. When this is done, repeated use of the Pythagoras theorem leads to the length L of their edges taking the values given in the last column of Table I.

2.2.7. *Duality of the Platonic solids*

A Platonic solid can be classified by its Schläfli symbol $\{m, n\}$, where $m \geq 3$ is the number of edges of each polygonal face, and $n \geq 3$ is the number of edges/ polygonal faces meeting at each vertex (Table I).

The Platonic solids come in natural pairs, such that the number of vertices of one polyhedron correspond to the number of faces of the other, and *vice versa*, with each twin being dual to the other:

- the tetrahedron is self dual, i.e. it is its own dual;
- the cube and octahedron are dual to each other; and, similarly,
- the icosahedron and dodecahedron are dual to each other.

The Schläfli symbol of the dual is just the original written backwards, for example the dual of {5, 3}, i.e. the dodecahedron, is {3, 5}, i.e. the icosahedron.

2.3. *Uniform tiling of the sphere*

There are only five ways of uniformly tiling the surface of the sphere with congruent polygonal surfaces. This is seen by examining the gnomonic projection of the regular convex polyhedra, i.e. the Platonic solids, inscribed within a sphere, onto its surface. (The *gnomonic* projection projects along straight lines drawn from the centre of the sphere. The projection of an edge of a regular polyhedron onto the sphere is thus the arc of a great circle, and the projected edges of a regular convex polyhedron are thus all arcs of great circles.) This means (Table I) that it is impossible to tile the sphere perfectly uniformly with more than twenty identical faces, i.e. the number of (triangular) faces of the regular icosahedron. Each face of a regular polyhedron can however be subdivided into sub-areas of equal area, e.g. each face of the regular icosahedron can be divided into p^2 equilateral triangles for $p = 2, 3, \ldots$. However, when these faces are gnomonically projected onto the sphere, they are no longer of equal area. Also a different number of faces meet at some vertices than at others.

2.4. *Vertex constraints for general convex polyhedra*

Let V_n be the number of vertices where n faces meet, for example V_3 is the number of vertices where three faces meet. The total number of vertices can then be written as

$$V = V_3 + V_4 + V_5 + V_6 + V_7 + \ldots \tag{11}$$

At each vertex associated with V_n, exactly n edges radiate away from it. There are therefore a total of nV_n edges associated with the vertices where n faces come together. Summing all such vertices for $n = 3, 4, 5, \ldots$, then gives $3V_3 + 4V_4 + 5V_5 + 6V_6 + 7V_7 + \ldots$ associated edges. But summing this way, each edge is counted twice, since it connects exactly two vertices. Thus

$$2E = 3V_3 + 4V_4 + 5V_5 + 6V_6 + 7V_7 + \ldots \tag{12}$$

2.5. *Polyhedra composed entirely of triangles*

For a convex polyhedron, not necessarily regular, composed entirely of triangles, each face has three edges, but each of these edges is shared across two adjoining faces. Thus

$$3F = 2E, \tag{13}$$

i.e. for every two faces, there are exactly three edges. Eliminating E, F and V from (1), (11)–(12) and (13) then gives

$$3V_3 + 2V_4 + V_5 = 12 + V_7 + 2V_8 + \ldots \tag{14}$$

Also

$$E = 3(V-2)\,, \quad F = 2\,(V-2)\,, \quad V = \frac{1}{3}E+2 = \frac{1}{2}F+2. \tag{15}$$

From (15) it is seen that at high-enough asymptotic resolution (but not at low resolution), for every vertex there are two faces and three edges. Also (14) imposes no constraint on V_6, i.e. on the number of vertices where six triangles meet.

Setting $0 = V_4 = V_5 = V_7 = V_8 = \ldots$, i.e. allowing exactly three triangles to meet at a vertex, and no more, reduces (14)–(15) to

$$V = V_3 = 4, \quad E = 6, \quad F = 4, \tag{16}$$

which corresponds to a *tetrahedron*, regular or otherwise. Similarly, setting $0 = V_3 = V_5 = V_7 = V_8 = ...$, i.e. allowing exactly four triangles to meet at a vertex, reduces (14)–(15) to

$$V = V_4 = 6, \quad E = 12, \quad F = 8, \tag{17}$$

which corresponds to an *octahedron*; and setting $0 = V_3 = V_4 = V_7 = V_8 = ...$, i.e. allowing exactly five triangles to meet at a vertex, reduces (14)–(15) to

$$V = V_5 = 12, \quad E = 30, \quad F = 20, \tag{18}$$

which corresponds to an *icosahedron*.

Equations (16)–(18) correspond to (6)–(8) but are more general inasmuch as the triangles no longer have to be congruent.

Are there other polyhedra composed entirely of triangles? Yes, there are lots of possibilities: one only has to choose $V_3, V_4, V_5, ...$, to satisfy (11) and (14), and use (15) to determine the number of edges (E) and faces (F). Note though, from (14), that V_3, V_4 and V_5 cannot all be simultaneously zero. Setting $0 = V_3 = V_4 = V_7 = V_8...$, $V_5 = 12$, and $V_6 = p^2$, gives the example of section 2.3, where each face of an icosahedron is subdivided into p^2 triangles before gnomonic projection to the sphere.

2.6. *Polyhedra composed entirely of quadrilaterals*

For a convex polyhedron, not necessarily regular, composed entirely of quadrilaterals, each face has four edges, but each of these edges is shared across two adjoining faces. There are thus exactly twice as many edges as there are faces, i.e.

$$E = 2F. \tag{19}$$

Eliminating E, F and V from (1), (11)–(12) and (19) then gives

$$V_3 = 8 + V_5 + 2V_6 + 3V_7 + ... \tag{20}$$

Also

$$E = 2\,(V-2)\,, \quad F = V-2, \quad V = \frac{1}{2}E + 2 = F + 2. \tag{21}$$

Equation (21) imposes no constraint on V_4, i.e. on the number of vertices where four quadrilaterals meet. Examination of (20) shows that, for polyhedra composed entirely of quadrilaterals, there must be at least eight vertices where only three quadrilaterals meet.

Since $V_n \geq 0, \forall n$, the right-hand side of (20) is a minimum when $0 = V_5 = V_6 = V_7...$, and then, using (21) and assuming that V_4 is also zero, i.e. there are no vertices where four quadrilaterals meet,

$$V = V_3 = 8, \quad E = 12, \quad F = 6, \tag{22}$$

which is a *hexahedron* composed entirely of quadrilateral faces. When the faces are congruent squares, (22) corresponds to (9), i.e. to a *cube*.

A further example of a polyhedron composed entirely of quadrilaterals is the truncated square dipyramid (Figure 2(a)), which has

$$E = 28, \quad F = 14, \quad V_3 = V_4 = 8, \quad V = V_3 + V_4 = 16. \tag{23}$$

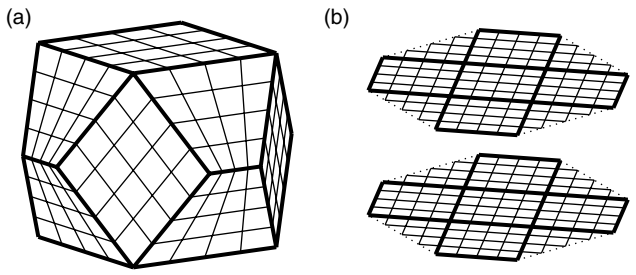

Figure 2. (a) Truncated square dipyramid with: two horizontal square faces (top and bottom), eight inclined trapezium faces (four per hemisphere), and four vertical rhomboidal faces (shared between the two hemispheres). (b) The two faces (top and bottom) of the octagonal dihedron.

To construct it, first take two square pyramids and glue their bases together. Next, horizontally chop off their two apexes to give a square top and a square bottom (and obtain a glued pair of quadrilateral frustrums). Finally, vertically chop off the four corners to get an octagon in the equatorial plane. This creates a rhombus in a vertical plane at each of the four corners, plus eight inclined trapeziums, which are all that remains of the faces of the original dipyramid. (Although one might call this solid a truncated octahedron, this term already denotes a different solid.)

The truncated square dipyramid is (after further subdivision) topologically equivalent to the the conformal octagonal grid of Purser and Rančić (1997)–section 3.2.2. Variants on this idea are also possible, for example Figure 1b of Purser and Rančić (1997) is topologically equivalent to a polyhedron such that

$$E = 56, \quad F = 28, \quad V_3 = 8, \quad V_4 = 22, \quad V = V_3 + V_4 = 32. \tag{24}$$

For both polyhedra described above, there are eight vertices where three quadrilaterals meet, with four quadrilaterals meeting at each of the remaining vertices.

2.7. *Polyhedra composed entirely of pentagons*

For a convex polyhedron, not necessarily regular, composed entirely of pentagons, each face has five edges, but each of these edges is shared across two adjoining faces. Thus

$$5F = 2E. \tag{25}$$

Eliminating E, F and V from (1), (11)–(12) and (25) then gives

$$V_3 = 20 + 2V_4 + 5V_5 + 8V_6 + 11V_7 + ... \tag{26}$$

Also

$$\left.\begin{aligned} E &= \frac{5}{3}\,(V-2)\,, \quad F = \frac{2}{3}\,(V-2)\,, \quad 2E = 5F, \\ V &= \frac{3}{5}E + 2 = \frac{3}{2}F + 2. \end{aligned}\right\} \tag{27}$$

Since $V_n \geq 0, \forall n$, the right-hand side of (26) is a minimum when $0 = V_4 = V_5 = V_6 = V_7...$, and then (using (27))

$$V = V_3 = 20, \quad E = 30, \quad F = 12, \tag{28}$$

which corresponds to a *dodecahedron*, cf. (10).

2.8. *Polyhedra composed entirely of hexagons*

For a convex polyhedron composed entirely of hexagons, each face must have six edges, but each of these edges is shared across two adjoining faces. Thus there must be exactly thrice as many edges as there are faces, i.e.

$$E = 3F. \tag{29}$$

Eliminating E, F and V from (1), (11)–(12) and (29) then gives

$$6 + V_4 + 2V_5 + 3V_6 + 4V_7 + \ldots = 0. \tag{30}$$

Since $V_n \geq 0, \forall n$, it is impossible to satisfy (30), and therefore *it is impossible to form a convex polyhedron composed entirely of hexagonal faces*, regular or otherwise, consistent with the more-limited conclusion given in section 2.2.5.

2.9. *Polyhedra composed of a mix of triangles and quadrilaterals*

Consider now an arbitrary combination of triangular and quadrilateral faces. Let F^{T} and F^{Q} be the number of triangular and quadrilateral faces, respectively, so

$$F = F^{\mathrm{T}} + F^{\mathrm{Q}}. \tag{31}$$

Each triangular face has three associated edges, but each of these edges is shared across two adjoining faces. Similarly, each quadrilateral face has four associated edges, each of which is again shared across two adjoining faces. Thus

$$3F^{\mathrm{T}} + 4F^{\mathrm{Q}} = 2E. \tag{32}$$

Eliminating E and V from (1) and (11)–(12) gives

$$\begin{aligned} 2F &= 2F^{\mathrm{T}} + 2F^{\mathrm{Q}} \\ &= 4 + V_3 + 2V_4 + 3V_5 + 4V_6 + 5V_7 + \ldots \end{aligned} \tag{33}$$

Eliminating E, F and V from (1), (11)–(12) and (31)–(32) yields

$$F^{\mathrm{T}} + V_3 = 8 + V_5 + 2V_6 + 3V_7 + \ldots, \tag{34}$$

$$3V_3 + 2V_4 + V_5 = 2F^{\mathrm{Q}} + 12 + V_7 + 2V_8 + \ldots \tag{35}$$

From (35), V_3, V_4 and V_5 cannot all be simultaneously zero.

A simple example is the *cuboctahedron* (Figure 3), which is an Archimedian solid where the corners of a cube are cut off by joining the midpoints of adjacent edges. This is obtained by setting $V_4 = 12$ and $0 = V_3 = V_5 = V_6 = V_7 = \ldots$ in (31)–(35), leading to

$$E{=}24, \quad F^{\mathrm{T}}{=}8, \quad F^{\mathrm{Q}}{=}6, \quad F{=}14, \quad V{=}12. \tag{36}$$

Motivated by the discussion in section 5, one may wish to impose the constraint that the number of edges is exactly equal to twice the number of faces, i.e. $E = 2F$. The above results may be used to find what constraint this implies for the vertices. Imposing $E = 2F$ on (31)–(35) gives

$$\left.\begin{aligned} &F^{\mathrm{T}} = 0, \\ &F = F^{\mathrm{Q}} = 6 + V_4 + 2V_5 + 3V_6 + 4V_7 + \ldots, \\ &E = 2F, \end{aligned}\right\} \tag{37}$$

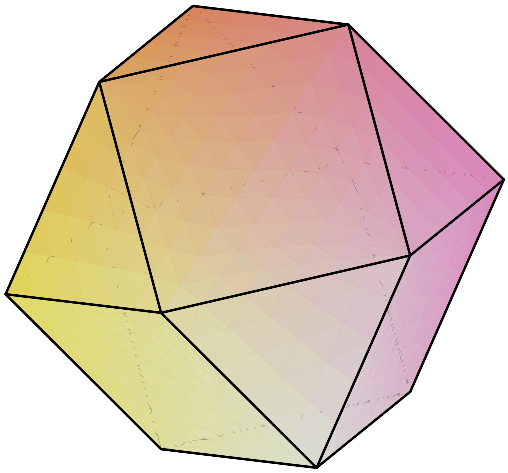

Figure 3. Cuboctahedron.

$$V_3 = 8 + V_5 + 2V_6 + 3V_7 + \ldots\ldots \tag{38}$$

Equation (37) means that if one wishes to have exactly twice the number of edges as faces with a mix of triangular and quadrilateral faces, then this can only be accomplished by exclusively using quadrilateral faces, with no triangular ones allowed. This case then reduces to that discussed in section 2.6, and (38) agrees with (20), as it should.

A further example of exploiting the above constraints is to answer the question: if one wishes to impose the constraint that the number of edges is exactly equal to twice the number of vertices, i.e. $E = 2V$, what constraint does this imply? Imposing $E = 2V$ on (31)–(35) gives

$$\left.\begin{aligned} &F^{\mathrm{T}}{=}8, \\ &F{=}8{+}F^{\mathrm{Q}}{=}2 + V_4{+}\,2V_5{+}\,3V_6{+}\,4V_7{+}\,\ldots, \\ &E{=}2V, \end{aligned}\right\} \tag{39}$$

$$V_3 = V_5 + 2V_6 + 3V_7 + \ldots. \tag{40}$$

Equation (39) means that if one wishes to have exactly twice the number of edges as vertices with a mix of triangular and quadrilateral faces, then this can only be accomplished by using exactly eight triangular faces (no more, nor no less) with all remaining faces being quadrilaterals. This can, for example, be accomplished by taking the *cuboctahedron*, which ((36)) has eight triangular faces, and subdividing each of its quadrilateral faces into a $p \times p$ grid of quadrilaterals, before projection to the sphere.

2.10. *Polyhedra composed of a mix of pentagons and hexagons*

Consider now an arbitrary combination of pentagonal and hexagonal faces. Let F^{P} and F^{H} be the number of pentagonal and hexagonal faces, respectively, so

$$F = F^{\mathrm{P}} + F^{\mathrm{H}}. \tag{41}$$

Each pentagonal face has five associated edges, but each of these edges is shared across two adjoining faces. Similarly, each hexagonal face has six associated edges, each of which is again shared across two adjoining faces. Thus

$$5F^{\mathrm{P}} + 6F^{\mathrm{H}} = 2E. \tag{42}$$

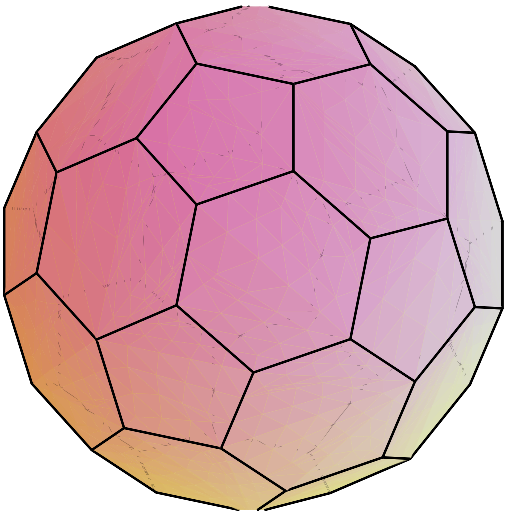

Figure 4. Truncated icosahedron.

Eliminating E and V from (1) and (11)–(12), and using (41), gives

$$\begin{aligned} 2F &= 2F^{\mathrm{P}} + 2F^{\mathrm{H}} \\ &= 4 + V_3 + 2V_4 + 3V_5 + 4V_6 + 5V_7 + \ldots \end{aligned} \tag{43}$$

Eliminating E, F and V from (1), (11)–(12) and (41)–(42) yields

$$F^{\mathrm{P}} = 12 + 2V_4 + 4V_5 + 6V_6 + 8V_7 + \ldots, \tag{44}$$

$$V_3 = 2F^{\mathrm{H}} + 20 + 2V_4 + 5V_5 + 8V_6 + 11V_7 + \ldots \tag{45}$$

A simple example is provided by a football (soccer ball). This is a gnomonic projection of the *truncated icosahedron* (Figure 4), which is an Archimedian solid composed of $F^{\mathrm{P}} = 12$ identical pentagonal panels plus $F^{\mathrm{H}} = 20$ identical hexagonal ones, for a total of $F = 32$ panels. There are $V = V_3 = 60$ vertices and $E = 90$ edges.

3. Global quadrilateral grids

Over the years, many quadrilateral grids have been proposed to tile the sphere. They have the advantage that the number of edges is equal to twice the number of faces, thus satisfying a necessary condition on the ratio of velocity to mass degrees of freedom on a C-grid to avoid computational modes (section 5). In many cases, the grid has a local logically rectangular structure, which simplifies coding and facilitates the use of a semi-Lagrangian advection scheme. This section provides an overview of quadrilateral grids.

3.1. *Latitude–longitude*

The traditional lat–lon grid has been the most popular grid, being used, from the 1960s to the present, in many numerical weather and climate prediction models. As mentioned in the introduction, the principal difficulty with this grid is due to the convergence of the meridians at its two singular points, the poles of the underlying spherical polar coordinate system. (Depending on the numerical scheme, the row of grid cells nearest each pole may be treated either as triangles, or as quadrilaterals with one edge of zero length.) Although the difficulties associated with convergence of the meridians have been addressed by modern numerical methods (Williamson, 2007), there remains the issue of poor scaling on massively parallel computer architectures: this is primarily due to inter-processor communication problems over and around the poles, but also to a loss of computational efficiency from anisotropic resolution.

To try to address anisotropy issues, and in particular the very small time steps associated with an Eulerian discretisation of the governing equations, Kurihara (1965) proposed a reduced version of the lat–lon grid where fewer gridpoints are used at each latitude moving poleward from the Equator. (More recently Leopardi (2006) has proposed a version of the reduced grid with uniform cell area.) However, it has been found (discussion in Williamson, 2007) that this results in differential phase errors and spurious meridional transport, so this approach has been largely abandoned. Although it is not impossible that a numerical method might be devised to overcome the identified difficulties, it does seem unlikely: reducing the inhomogeneity in east–west gridlength comes at the expense of introducing one in the north–south grid structure.

The spectral method uses essentially a lat–lon grid, though with the latitudes slightly irregularly spaced at the Gaussian quadrature points, and avoids the problems associated with polar convergence of the meridians. Spectral models can be designed to conserve mass (Bourke, 1972), though in practice most spectral models predict the logarithm of the mass variable (e.g. logarithm of surface pressure in the hydrostatic case) and so do not conserve mass exactly. In principle, if mass is conserved then conservation of energy (in the absence of time truncation errors) is also possible provided the relevant terms are calculated without aliasing, for example by the use of a sufficiently high resolution transform grid (Hoskins and Simmons, 1975). In practice, however, various compromises mean that energy conservation is not achieved; these include non-conservation of mass, the use of an alias-permitting transform grid, and the use of semi-Lagrangian advection. The spectral method can be adapted to use a reduced Gaussian grid (Temperton *et al.*, 2001) whilst avoiding most of the issues of reduced lat–lon grids mentioned above. The recent advent of a fast Legendre transform reduces the asymptotic scaling problem of the traditional Legendre transforms. However, there remain the problems of data communication required for the spectral transforms and the spectral-ringing problem due to unphysical spectral overshoots and undershoots.

3.2. *Grids generated by conformal maps*

One way to generate an orthogonal coordinate system on the surface of the sphere is through a conformal map from some plane map space to the surface of the sphere (Rančić *et al.*, 1996). A conformal map is one that preserves angles; in particular, the right angles between Cartesian coordinate lines in the map space map to right angles between the corresponding coordinate lines on the surface of the sphere. Conformal maps may be chosen that avoid the convergence of coordinate lines that occurs at the poles of the lat–lon grid. Unfortunately, there is no conformal map from (a subset of) the plane to the surface of the sphere that is completely free of singularities. Orthogonality breaks down at the singularities, and some resolution clustering associated with these singularities is unavoidable. Although

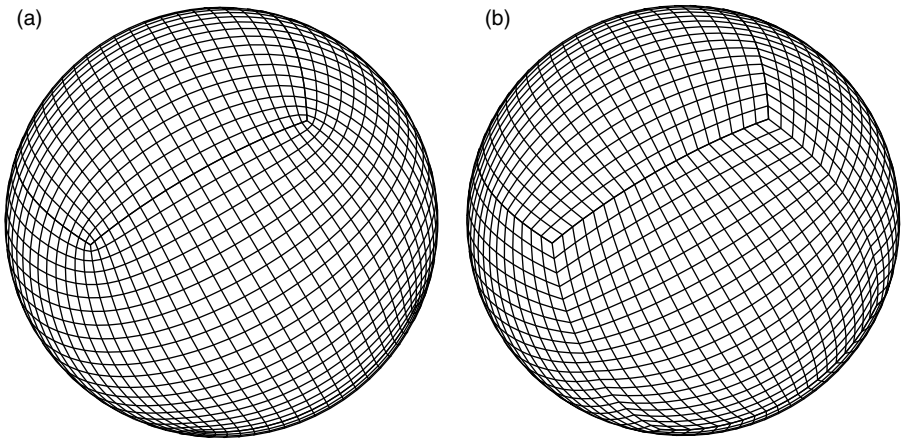

Figure 5. Cubed sphere: (a) conformally projected with isolines equally spaced in map coordinates, and (b) gnomonically projected with equal angles at an axis of the sphere.

it may be milder than that occuring at the poles of the lat–lon grid, it will nevertheless have some implications for parallel scaling.

3.2.1. *Conformally projected cube*

Cubed spheres come in two basic flavours (plus various enhancements): conformally projected and gnomonically projected. The conformally projected cube is discussed here, and the gnomonically projected cube in section 3.4.1.

Regional grids based on conformal maps were first used in the 1950s, e.g. Charney *et al.* (1950). A global conformal grid, the conformally projected cube, was proposed by Rančić *et al.* (1996); Figure 5(a) shows an example of what this gives. Their approach was subsequently employed in three-dimensional (3D) models by McGregor (1996, 1997) and Adcroft *et al.* (2004). To focus resolution on a region of the globe, McGregor (2005) suggests combining this with the use of the Schmidt (1977) conformal transformation, but Geleyn and Caian (1997) discuss the limits on how far the focusing can be pushed with this approach.

At the heart of the conformally projected cubed sphere is the conformal mapping between a circle and a square; Figure 2 of Rančić *et al.* (1996) shows the result of conformally mapping a uniform square mesh to the circle. Because conformality implies orthogonality (however the converse is not true), the six local coordinate systems (one for each face) are not only free of coordinate singularities, but are orthogonal, and the directions of coordinate isolines are continuous across face boundaries. However the eight vertices of the cube are still singular points. There is also a highly significant clustering of resolution around the eight vertices of the cube, which gets progressively worse as a function of increasing resolution, with a corresponding increase in the ratio of maximum to minimum gridlength.

To partially address the resolution-clustering behaviour of the conformally projected cubed sphere, Adcroft *et al.* (2004) proposed a rescaling of the local face coordinates (this amounts to choosing a different spacing of the coordinate isolines, i.e. to using a different set of values for the isolines). This is illustrated in their Figure 7, which shows (a) the unmodified Rančić *et al.* (1996) conformally projected grid, and (b)–(d) three modifications, by different rescalings, of this grid. For the modified grids there is less clustering of resolution in the immediate vicinity of the four corner points but, to achieve this, resolution is less uniform normal to face edges, and near to them. Although this tweaking of the conformal cube approach helps mitigate the resolution clustering problem, it does not really address it at source, but rather postpones the seriousness of the problem to higher resolution.

3.2.2. *Conformal octagonal grid*

Purser and Rančić (1997) proposed what they termed the conformal octagonal grid as another conformal mapping strategy, similar to the conformally projected cube, but with an emphasis on regional forecasting using the Schmidt (1977) transformation. Their Figure 1 depicts the double-octagon geometry, and two views of the resulting symmetric octagonal grid are shown in their Figure 2. Purser and Rančić (1997) describe the conformal octagonal geometry in the following way:

'*Each octagon is formed by snipping off the corners of a square in a way that preserves the symmetries of the square. We see* (their Figure 1(a)) *that, apart from the eight corner singularities, the domain formed by placing two such octagons back-to-back allows a continuous square gridding when the eight corners are located at rational proportions of the span of the sides of the original square. In the simplest case, we form each octagon from the square divided in both directions into three equal strips (that is, nine squares) and simply bisect the four corner squares diagonally. By joining two such octagons, the corner squares recover their integrity to give 14 subdomains in all. Further rational subdivisions of these squares allow* $14M^2$ *equal square subdomains to be formed. Alternatively, we might start by dividing the original square into four* equal *strips in both directions, snipping the corners to yield a decomposition into 28 squares (or* $28M^2$ *after further subdivision), but* (their Figure 1(b)) *the overall octagonal shape is unaltered–only rotated by 45°.*'

Purser and Rančić (1997)'s conformal octagonal grid is topologically equivalent to the truncated square dipyramid described in section 2.6 (Figure 2(a)). Note, however, that the conformal aspect of the conformal octagonal grid is most simply obtained by defining a square mesh on an octagonal dihedron (a dihedron is a two-faced polyhedron composed of two polygonal faces with shared edges)–Figure 2(b)–and then conformally mapping it to the sphere. Some other

grids may be viewed as being derivable from a dihedron–cf. section 3.5.1. In Purser and Rančić (1997)'s Figure 2(b), there is a noticeable clustering of resolution at the singular points of the transformation. An example of an octagon grid focused over North America using the Schmidt (1977) transformation, is displayed in their Figure 3, and the resolution clustering is even more apparent.

3.2.3. *Lanser et al.'s composite twin polar-stereographic with lat–lon*

Lanser *et al.* (2000), inspired by Phillips (1957) (section 3.3.1 below), proposed a grid that uses three regions covered by three orthogonal coordinate systems (their Figure 4). In the polar regions, square Cartesian meshes are projected stereographically onto the sphere. For the tropical belt region, whereas Phillips (1957) used a Mercator grid, Lanser *et al.* (2000) argued that, since spherical coordinates are natural, in widespread use, and easily implemented in regions removed from the pole, they prefer to use a lat–lon grid there (although orthogonal, the lat–lon grid is not, in fact, conformal).

The polar-cap grids do not overlap the tropical belt grid (in contrast to the approach of Phillips, 1957). Instead, the polar-cap grids are joined by radial coordinate lines to the most poleward latitude circles of the tropical belt–their Figure 4 gives two different views of this. The coordinate lines are not orthogonal where the two coordinate systems join. Also, the highly anisotropic grid boxes surrounding the polar boxes are then fine in one direction, but very coarse in the other, and this must have a serious impact on accuracy for realistic flows as resolution is increased. In particular, the north–south meshlength along the meridian that intersects the polar grid midway along an edge asymptotes to a constant, rather than to zero. This is a very serious limitation of their approach.

3.2.4. *Other conformal transformations*

Other conformal transformations are possible (e.g. Lee, 1976; Murray, 1996). For example, Figure 11 of Murray (1996) shows a polar-cap grid generated from a focal arc. However, note the undesirable clustering of resolution about the two foci.

3.3. *Overset composite grids*

Global orthogonal quadrilateral grids can also be constructed from a number of partially overlapping regional grids. The singularities of the conformal-map-based grids are avoided, but at the expense of introducing the overlaps. The need to transfer information between the regional grids at the overlaps makes it significantly more complicated to obtain mass conservation (Peng *et al.*, 2006) and to solve the Helmholtz problem required for semi-implicit time integration (Qaddouri *et al.*, 2008). This also has the potential to impact scalability because of the associated data communication burden. It is likely that geostrophic adjustment, balance, and energy conservation are compromised at the overlaps, though we are not aware of any studies that quantify these effects. Three variants of overset composite grids are now discussed.

3.3.1. *Twin polar-sterographic with Mercator*

The first overset composite grid was proposed by Phillips (1957) – his Figure 1 and (the incorrectly ordered/labelled) Figure 2b of Lanser *et al.* (2000). It is composed of three conformally mapped (from the sphere) grids: a rectangular Mercator grid covering a tropical belt bounded by the latitudes $\pm\Phi_{\text{bound}}$, which is partially overlapped by two square polar-stereographic grids, each of which is centred on a pole, and tangent to it. The equatorward boundaries of each polar-sterographic grid (Figure 1 of Phillips, 1957) are fixed to be tangent to the bounding poleward circle of the Mercator grid, with four points common to each polar-sterographic grid and the Mercator grid. Phillips (1957) argued that the bounding latitudes of the Mercator grid (and thus the location of the square boundaries of the polar-stereographic domains), should be chosen to minimise the ratio of maximum to minimum meshlength over the sphere. This then led to the value of $\Phi_{\text{bound}} = 43.27^\circ$, and to a ratio of 1.373. (If there were no need for any overlap between the grids, then it can be shown that this factor would reduce to a mere 1.2, with $\Phi_{\text{bound}} = \sin^{-1}(3/5) \approx \pm 36.87^\circ$.) Phillips (1959) then integrated a shallow-water model over a hemisphere using this technique, but found evidence of a 'noise' issue. In Phillips (1962), this was attributed to the existence of computational modes. By using a different numerical technique, good results were then achieved, albeit with an unrealistically large temporal oscillation of an inertia-gravity mode, attributable to a lack of balance in the initial conditions.

3.3.2. *Twin polar-stereographic*

Browning *et al.* (1989)'s assessment of Phillips' work was: '*Unfortunately, the decision to butt the coordinate systems together at a common latitude and to couple them with interpolation led to an unstable method, so the concept was abandoned*'. They then argued that the difficulties observed by Phillips (1959) are not due to any fundamental flaw in the concept of overset grids. Rather, they are due to details of the numerical discretisation of the governing equations on these grids, and on how they are coupled together.

For the overset-composite-grid method, Browning *et al.* (1989) then argued that:

(a) it is essential to ensure that *all* boundary points of a region covered by a grid are interior points (tangency is insufficient) of a region covered by another grid (Starius, 1980); and

(b) provided this is done, and the equations are discretised in an appropriate manner, then the Phillips (1957) approach is indeed viable.

Using a simplified form of the Phillips (1957) approach, which amounts to dispensing with the Mercator grid altogether, and extending the bounds of each polar-stereographic grid to include an entire hemisphere (Figure 1 of Browning *et al.*, 1989, and Figure 2 of Williamson, 2007), they then built, and successfully integrated, a shallow-water model. As Browning *et al.* (1989) noted, their particular proof-of-concept implementation is sub-optimal in the sense that the ratio of maximum to minimum grid length for their two coupled polar-sterographic grids is a little more than two, which is substantially greater than the corresponding factor of 1.373 for the three-mesh combination proposed by Phillips (1957).

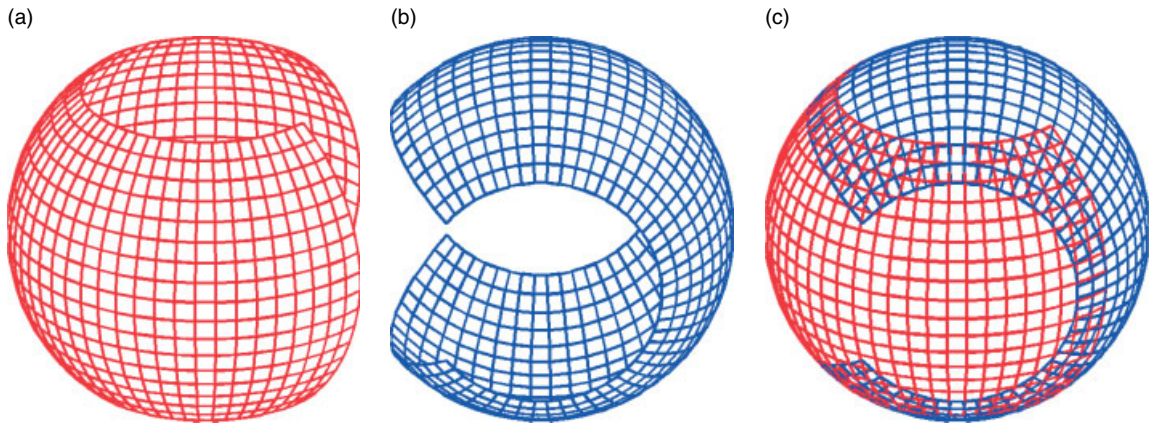

Figure 6. A Yin-Yang grid: (a) Yin grid (in red), (b) Yang grid (in blue), and (c) Yin-Yang merged grid. (Figures kindly provided by Abdessamad Qaddouri.)

In a further development, Dudhia and Bresch (2002) adopted and adapted the Browning *et al.* (1989) approach to build and demonstrate a global non-hydrostatic model that also uses two coupled polar-sterographic grids. They conceded that this is sub-optimal from an efficiency perspective, but argued that they were primarily interested in extending the domain of an existing regional model to the entire globe, thereby greatly enhancing the model's versatility.

3.3.3. *Yin-Yang*

The most recent overset composite grid to be proposed is the so-called Yin-Yang grid of Kageyama and Sato (2004), Kageyama (2005), Qaddouri *et al.* (2008), Li *et al.* (2008), Baba *et al.* (2010), Qaddouri (2011), Qaddouri *et al.* (2011), and Qaddouri and Lee (2011), a variant of which is shown schematically in Figure 6. This is a combination of two lat–lon grids with a small overlap between them. The geographical extent of each of these partial lat–lon grids is a little larger than 90° in latitude (with the bounding latitudes located slightly poleward of ±45°), and a little larger than 270° in longitude (with the bounding meridians, located at approximately 0° Greenwich, and 270° Greenwich in either direction).

The strengths of the Yin-Yang grid are:

1. its two components are traditional, orthogonal, lat–lon grids;
2. there are no singular points;
3. the metric factors of each orthogonal coordinate system are known analytically;
4. the grid spacing is quasi-uniform, with a ratio of maximum to minimum gridlength a little larger than $1/\cos(45^\circ) = \sqrt{2} \approx 1.414$;
5. approximately 20–25% fewer points are needed for quasi-uniform resolution than for the traditional lat–lon grid; and
6. existing lat–lon discretisations and codes can be adapted to the Yin-Yang grids.

The main limitations are due to the overlap between the Yin and Yang grids. This causes issues with the coupling between the two grids, particularly with respect to conservation. Mass conservation can be achieved by using a suitable mass-conserving remapping algorithm (Peng *et al.*, 2006), albeit at some computational cost. However, *a priori* conservation of momentum and energy would appear to be very challenging, although *a posteriori* 'fixers' are always possible, e.g. Sasaki (1976)'s variational technique. If a semi-implicit method is used, then careful consideration needs to be given to choosing an iterative solution algorithm for the resulting elliptic-boundary-value problem that has an acceptable rate of convergence–e.g. Qaddouri *et al.* (2008), Loisel *et al.* (2010) and Qaddouri (2011) give domain-decomposition methods suited to parallel computer architectures. Another concern is the possibility of decoupling between the solutions on the two component grids in the overlap region, particularly with the inclusion of physical parametrisation schemes whose output may depend very sensitively on their input. For example, deep convection may be triggered on one grid component but not on the other.

3.3.4. *Relative efficiency of the three overset composite grids*

All other things being equal, the twin polar-sterographic approach of section 3.3.2 is the least efficient of the three overset composite grids, since the ratio of maximum to minimum gridlength is greater than 2. For the Yin-Yang approach of section 3.3.3, because the bounding latitudes are at approximately ±45°, this ratio reduces to somewhat greater than $1/\cos(45^\circ) = \sqrt{2} \approx 1.414$. The most efficient overset grid, is the one proposed by Phillips (1957), and discussed in section 3.3.1, with a ratio of 1.373.

Although the Phillips (1957) grid is most efficient, the Yin-Yang grid is almost as efficient. For centres having an existing lat–lon grid model (with associated data assimilation system, postprocessing, etc.), a Yin-Yang grid model can be developed with relatively little modification of existing code, and is therefore an attractive proposition.

3.4. *Non-orthogonal quadrilateral grids*

A quadrilateral gridding of the sphere can be retained, without resolution clustering or overlaps, by sacrificing orthogonality. By avoiding both resolution clustering and overlaps, such grids avoid the main barriers to scalability. However, giving up orthogonality makes it significantly more complicated to obtain the mimetic properties 4–7 discussed in the introduction, and typically requires both the contravariant and covariant velocity components to enter the calculation.

3.4.1. *Gnomonically projected cube*

The first person to propose integrating a model on what is now popularly known as the 'cubed sphere' (Figure 5) appears to have been Sadourny (1972). The idea lay dormant for many years but has enjoyed a significant renaissance in recent years (Ronchi *et al.*, 1996; Taylor *et al.*, 1997; Thomas and Loft, 2000, 2002, 2005; Giraldo *et al.*, 2003; Fournier *et al.*, 2004; Giraldo and Rosmond, 2004; Dennis *et al.*, 2005; Nair *et al.*, 2005a,b; Rossmanith, 2006; Putman and Lin, 2007; Wang *et al.*, 2007; Chen and Xiao, 2008; Lauritzen and Nair, 2008; St-Cyr *et al.*, 2008; Ullrich *et al.*, 2009, 2010; Yang *et al.*, 2010; Lauritzen *et al.*, 2010b; Taylor and Fournier, 2010; Harris *et al.*, 2011).

From a geometric perspective, the gnomonically projected cube is arguably the most natural. Simply define a Cartesian mesh on each face of the cube, and then project these meshes from the centre of the Earth onto the sphere to obtain a spherical grid of curvilinear quadrilaterals–Figure 5(b). Each of the coordinate isolines is then a great-circle arc. The top and bottom faces of the unprojected cube lie in the latitudinal planes $\Phi = \sin^{-1}\left(1/\sqrt{3}\right) = 35.26^{\circ}$. (More generally, one can use a cuboid, i.e. a hexahedron composed of three pairs of rectangles, or, in principle but less usefully, a hexahedron composed of arbitrary quadrilaterals. Using a cuboid, for example, is one way of focusing resolution over a region of the globe.) The Cartesian mesh can be chosen to be uniform (leading to what is known as *equi-distant* projection), but this gives a less uniform grid on the sphere after gnomonic projection than using an *equi-angular* projection: asymptotically, the ratio of maximum to minimum gridlength for the *equi-distant* and *equi-angular* projections is approximately 5.2 and 1.3, respectively (Rančić *et al.*, 1996). (The equi-angular spacing can be considered to be doing things the opposite way around by uniformly spacing, in an angular sense, the meridians in each direction of a projected face, and then gnomonically projecting the spherical face to the cube to obtain the corresponding non-uniformly spaced Cartesian grid on the corresponding cube face.) The ratio of maximum to minimum gridlength of 1.3 for equi-angular spacing for the gnomonically projected cube is a bit better than the value of 1.373 for the three-mesh combination proposed by Phillips (1957), but not as good as the value of 1.2 if his proposed meshes were merged with almost no overlap.

For the gnomonically projected cube, the six local coordinate systems (one for each face) are free of coordinate singularities but are non-orthogonal. The eight corners of the sphere are singular points, and the direction of coordinate isolines changes abruptly when crossing face boundaries (Figure 5(b)). The lack of orthogonality needs to be taken into account when discretising the governing equations for atmospheric flow. In particular, attention needs to be paid to covariant versus contravariant components (e.g. Rančić *et al.*, 1996, and appendices of Adcroft *et al.*, 2004), and to the non-orthogonality of unit vectors.

3.4.2. *Smooth quasi-homogeneous*

In Purser and Rančić (1998), it is argued that, whilst the conformal projection introduced in Rančić *et al.* (1996) addresses the cross-boundary gradient continuity issues of the gnomonically projected cubed sphere, this is unfortunately achieved at the cost of resolution clustering around the eight singular points of the cube, with a ratio of maximum to minimum gridlength that increases with resolution, and a consequent non-negligible limitation on the time step of Eulerian discretisations. Purser and Rančić (1998) therefore proposed a compromise strategy which they termed smooth quasi-homogeneous gridding of the sphere, which is achieved via a variational principle. The compromise is to forgo conformality (and also orthogonality, particularly near singular points), in order to *simultaneously* obtain both quasi-homogeneity of the grid, and cross-boundary smoothness of coordinate lines. Provided one is prepared to live with non-orthogonal grids, and the ensuing discretisation consequences, the Purser and Rančić (1998) approach is quite attractive. Before and after examples for the conformal cube and conformal octagon grids are shown in their Figures 1 and 2, respectively. Purser and Rančić (1998) also note that the conformal cube grids give better results than the conformal octagon grids, and that '*in a semi-Lagrangian model, where regional inhomogeneities of grid resolution matter less, conformal grids remain more appropriate*'.

3.4.3. *Global logically rectangular*

Using a non-conformal mapping of a circle to and from a square, Calhoun *et al.* (2008) and Berger *et al.* (2009) propose mapping each hemisphere to a square region, and then combining these together to obtain a logically rectangular mapping of the whole sphere, with no overlapping regions e.g. Figure 5.1 of Calhoun *et al.* (2008) and Figure 2 of Berger *et al.* (2009). The resulting grid is then highly non-orthogonal, and this has to be taken into account when discretising the governing equations. With their simplest grid, the ratio of maximum to minimum gridlength is around 1.68, but with this grid the non-orthogonality and lack of smoothness are quite serious. Calhoun *et al.* (2008) do propose a way of making the grid smoother, but this is at the price of a larger ratio of maximum to minimum gridlength, and a consequent reduction in efficiency. Berger *et al.* (2009) propose a slightly smoother mapping than the one presented in Calhoun *et al.* (2008), and also one where one of the sets of grid lines are lines of constant latitude (although this is achieved at the price of being less smooth near the poles).

In principle, their non-conformal mapping between the circle and the square could be replaced by a conformal one, thereby leading to orthogonality (Figure 7(b)). However, this would then have the undesirable side effect of leading to the problem of resolution clustering at the four singular points of the transformation (in a similar manner to that discussed in section 3.2.1 for the cubed sphere, but amplified since the projected area, the surface of a hemisphere, is larger than that of the projected face of the cube).

A further wrinkle to the Calhoun *et al.* (2008) proposal would be to apply it over a spherical cap, rather than over a complete hemisphere. For each polar cap, this would give four singular points around the bounding circle of latitude, regardless of whether the circle-to-square mapping were to be performed as they propose, or conformally. The Calhoun *et al.* (2008) grid then amounts to constructing two such polar caps, each with four singular points, and then enlarging the polar caps to make them hemispheric in extent. This leads to coalescence, in pairs, of the singular points, reducing them from eight to just four.

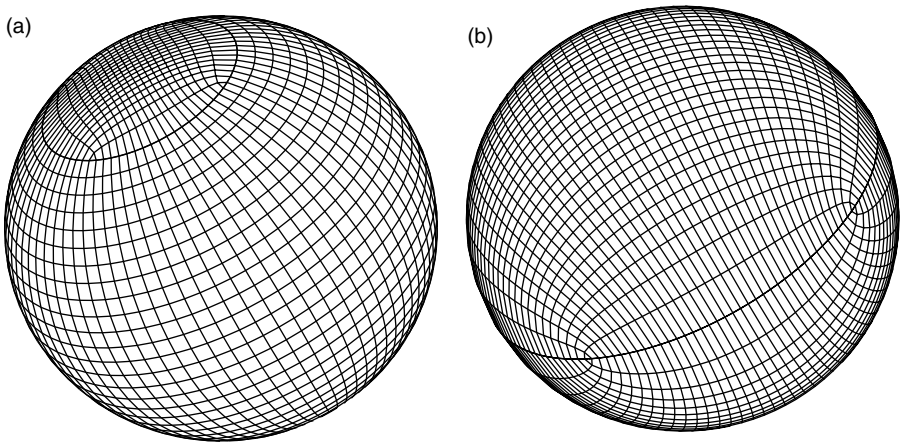

Figure 7. Composite meshes composed of a lat–lon tropical belt and twin polar-cap meshes: (a) where the meshes are joined at latitudes $\pm 60^\circ$, and (b) the limiting case where the lat–lon tropical belt shrinks to zero, and the twin polar-cap meshes are hemispheric and join at the Equator. The polar-cap meshes are obtained by conformally mapping a square Cartesian mesh to a circle before stereographically projecting onto the polar caps.

3.4.4. Fibonacci grids

Swinbank and Purser (2006) proposed a Fibonacci grid based on patterns found in nature in seeds and fruit; their Figures 1 and 2 show examples in plane and spherical geometry, respectively. Locally, a 2D but non-orthogonal coordinate system may be fitted through the grid points and exploited in any discretisation. A variable-resolution Fibonacci grid, in which the density of points is greater at the Equator than at high latitudes, is also possible (their Figure 7). Whilst Fibonacci grids are quite homogeneous, discretisation of equations on them is very difficult, and Swinbank and Purser (2006) found that computational stability was a significant issue when integrating a shallow-water model on such a grid.

3.5. Some further alternative quadrilateral grids

The analyses reported in section 5 show that, for *C*-grids, a logically rectangular arrangement of quadrilaterals generally leads to spatial discretisations that avoid branches of computational modes in the numerical dispersion relation. Other arrangements of quadrilaterals have twice as many velocity degrees of freedom as mass degrees of freedom and thus satisfy a necessary condition for avoiding branches of computational modes, though we are not aware of any published calculations of dispersion relations for such grids. This section explores the possibility of devising alternative quadrilateral-based tilings of the sphere that might conceivably lead to a good dynamical core.

3.5.1. Conformal twin polar-stereographic with lat–lon

The non-orthogonality of the Lanser *et al.* (2000) grid (section 3.2.3), where the tropical and polar coordinate systems join, can be avoided by conformally mapping a square Cartesian mesh to a circle before stereographically projecting onto the polar region. The coordinate lines in the polar region can then be made to join smoothly to those in the tropical belt region (Figure 7). All mesh lengths now asymptote to zero as resolution is increased, although the north–south resolution still varies abruptly across the joining latitude. Also, there remain coordinate singularities at the 'corners' of the polar coordinate system, and, associated with these, the ratio of maximum to minimum grid length increases as resolution is increased. Figure 7(b) shows the limiting case of this grid in which the joining latitude is the Equator. This grid may be generated from a rectangular gridding of a square dihedron, conformally mapped to the sphere.

A further disadvantage of this composite mapping, which is based on the conformal mapping between a circle and a square, compared with Rančić *et al.* (1996)'s conformal cube, is that the corner singularities are more severe: the proposed map has to map (distort) 180° to 90°, whereas the Rančić *et al.* (1996) conformal cube map has only to distort 120° to 90°.

3.5.2. Modified Yin-Yang

The Yin-Yang grid discussed in section 3.3.3 can be modified in the following manner. First, extend the Yin grid of Figure 6(a) by 90° in longitude (i.e. by one third of its geographical extent) to obtain the extended Yin grid (a periodic lat–lon tropical belt) depicted in Figure 8(a). Next, remove the middle one third of the Yang grid of Figure 6(b) to obtain the reduced Yang grid depicted in Figure 8(b). The merging of the extended Ying and reduced Yang grids then gives the modified Yin-Yang grid of Figure 8(c).

The modified Yin-Yang grid has the same number of degrees of freedom as the original Yin-Yang grid (with comparable efficiency), and shares with it the six strengths listed in section 3.3.3. However it also has some additional symmetry properties. The extended Yin grid (Figure 8(a)) is invariant under rotation about the polar axis by any integral number of longitudes, in contradistinction to the original Yin grid (Figure 6(a)). The two subgrids of the reduced Yang grid are invariant under rotation by 180° about an equatorial axis, whereas the original Yang grid is not. Because of the additional symmetry properties, it may be expected to be a little less vulnerable to grid imprinting than the original Yin-Yang grid. For example, it may better represent solid-body rotation about the polar axis.

Note, however, that Phillips (1957)'s twin-polar-stereographic with Mercator grid, discussed in section 3.3.2, is not only invariant under rotation by 180° about an equatorial axis of its polar-sterographic grids, but also by 90° and 270°, thereby possibly further reducing the intensity of grid

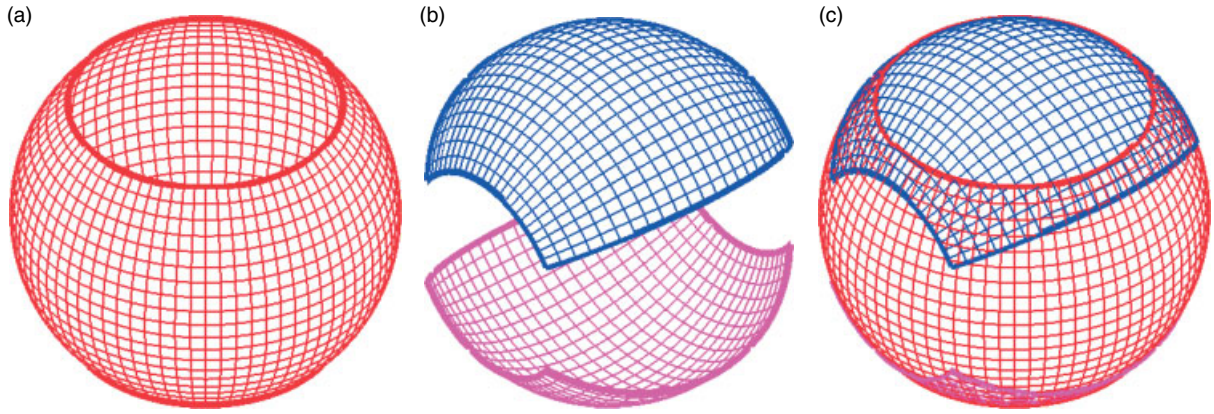

Figure 8. A modified Yin-Yang grid: (a) extended Yin grid (in red), (b) reduced Yang grid (in blue and magenta), and (c) modified Yin-Yang grid after merge of (a) with (b). (Figures kindly provided by Mohamed Zerroukat.)

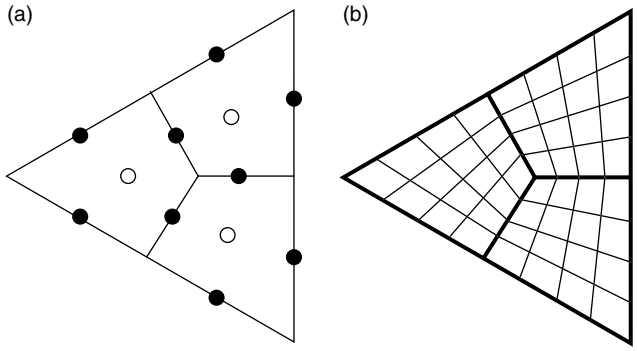

Figure 9. Sub-divided triangles: (a) an equilateral triangle divided into three quadrilaterals, with quadrilateral centroids denoted by open circles (○), and quadrilateral midpoint edges by filled circles (●); and (b) an equilateral triangle divided into $3 \times 4^2 = 48$ quadrilaterals.

imprinting. This, coupled with its uniformity-of-resolution property (section 3.3.4), suggest that it is likely to be a better overset grid than both the original and modified Yin-Yang grids.

3.5.3. *Triangles merged pairwise into quadrilaterals*

The generation of quadrilateral meshes has received much attention in the computer-aided design and engineering literature. Quadrilateral meshes may be generated from triangular meshes by merging pairs of triangles; there is a great deal of flexibility in how to do this in general, though optimising certain measures of mesh quality, and avoiding the occurrence of isolated triangles, can provide some constraints (e.g. Lau *et al.*, 1997; Itoh and Shimada, 2002, and references therein).

3.5.4. *Polygons subdivided into quadrilaterals (balanced polygons)*

Arbitrary polygonal regions can be subdivided into quadrilaterals. Again there is considerable freedom in how this is done, and again optimising mesh quality can provide constraints (e.g. Joe, 1995, and references therein). For global atmospheric modelling, the most simple and symmetrical such grids are most likely to be useful, and we consider some examples here.

A triangle may be subdivided into three quadrilaterals by joining the three edge midpoints to the centroid (Figure 9(a) for an equilateral triangle). Each of these quadrilaterals can then be subdivided into further quadrilaterals (Figure 9(b)).

Decomposition of a triangle into quadrilaterals is used in Giraldo (1998, 2001), Giraldo *et al.* (2002) and Giraldo and Rosmond (2004) to define quadrilateral spectral elements on the twenty triangular faces of an icosahedron; Figure 3 of Giraldo and Rosmond (2004) shows examples of this with two different resolutions. The non-uniform spacing of the quadrilaterals within an element is due to their use of numerical quadrature to perform certain integrals of the spectral-element discretisation.

With this strategy, if a domain can be triangulated, then it can be decomposed into quadrilaterals by simply applying the strategy to every triangle. (This might be a good approach for ocean modelling, whereby rectangles are used for most of the ocean, with triangles used to fit coastlines before decomposition into three quadrilaterals.)

An example of such a grid, obtained by subdividing an icosahedron into smaller triangles, then subdividing the triangles into quadrilaterals, is shown in Figure 10. The individual grid cells are kite-shaped quadrilaterals. For this particular example, although the grid is not a Voronoi grid, the nominal cell 'centres' can be chosen so that there is a dual mesh orthogonal to the primal mesh (Colin Cotter, personal communication).

Other polygons can also be subdivided into quadrilaterals by joining midpoint edges to the centroid; Figures 11(a, b) show the natural extension to pentagons and hexagons of the decomposition of triangles into quadrilaterals depicted in Figure 9. Thus, for example, a dodecahedron can be decomposed into quadrilaterals using the decomposition of Figure 11(a), with decomposition of quadrilaterals into further quadrilaterals also possible. An alternative division of a hexagon, this time into three quadrilaterals, is shown in Figure 11(c). Many other divisions of polygonal meshes into quadrilaterals are also possible.

Global *C*-grids based on this approach have exactly twice as many velocity as mass degrees of freedom, a necessary condition for avoiding computational modes in their numerical dispersion relation. However, their connectivity is different from that of a global logically rectangular mesh. It is not clear, therefore, whether they do in fact avoid computational modes. It would be valuable to have an analytical or numerical dispersion analysis for some examples of such grids to settle this question. Another potential limitation of such grids is that they will,

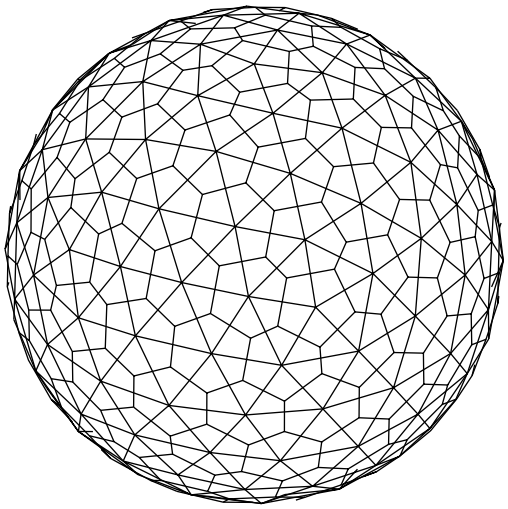

Figure 10. Kite grid: constructed from an icosahedral grid by subdividing each triangular face into subtriangles, followed by subdivision of each triangle into kite-shaped quadrilaterals.

in general, neither be orthogonal nor have an orthogonal dual (though particular examples, like that in Figure 10, may do so). It is currently not known whether a scheme may be constructed having all of the desirable properties listed in the introduction for a non-orthogonal *C*-grid.

4. Triangular and pentagonal-hexagonal grids

Spherical geodesic grids for global atmospheric modelling are generally based on either icosahedral grids, or on their pentagonal-hexagonal dual grid. The construction of such a grid by subdivision of a regular icosahedron is illustrated in Figure 12. The regular icosahedron is displayed in Figure 12(a) before subdivision of its triangular faces, and in Figure 12(b) and (c) after successive sub-division of each of its triangular faces into 4, and 4^2, equal subtriangles, respectively. Proceeding sequentially in such a manner, the asymptotic ratio of maximum to minimum gridlength is about 1.2 (e.g. Table 1 of Thuburn (1997)). Although two other triangle-faced Platonic solids exist, viz. the tetrahedron and octahedron (Figures 1(a) and (c)), subdivision of their faces results in larger ratios of maximum to minimum gridlength than for the icosahedron. Thus icosahedral-based grids are generally preferred, with discretisations either performed directly on triangulations of them, or on their pentagonal-hexagonal duals (Figure 12(d)).

After the early pioneering work of: Sadourny *et al.* (1968) and Williamson (1968) for the barotropic vorticity equation; Williamson (1970, 1971) and Cullen (1974) for the shallow-water equations; and Cullen and Hall (1979) for the hydrostatic primitive equations; interest in the use of these grids dwindled due mainly to noise and accuracy problems. There then followed a gradual renaissance of interest, starting with Masuda and Ohnishi (1986), and followed thereafter, for example, by Ničković (1994), Heikes and Randall (1995a,b), Giraldo (1997, 2000, 2005, 2006), Thuburn (1997), Stuhne and Peltier (1999), Ringler *et al.* (2000, 2010), Tomita *et al.* (2001), Ringler and Randall (2002a), Majewski *et al.* (2002), Randall *et al.* (2002), Tomita and Satoh (2004), Giraldo and Warburton (2005), Miura and Kimoto (2005), Satoh *et al.* (2008), Giraldo *et al.* (2008), Läuter *et al.* (2008), Walko and Avissar (2008a,b), Comblen *et al.* (2009), Lee and MacDonald (2009), and Ii and Xiao (2010).

Although the grids used by the above authors are generally based on a uniform division of an icosahedral face, they have often been 'tweaked' to facilitate application of particular numerical discretisations. For example, Tomita *et al.* (2001, 2002) used 'spring dynamics', a numerical smoothing technique, to increase the smoothness of their grid, whilst Heikes and Randall (1995b) modified their grid to improve the accuracy of the simple compact finite-difference approximation to the horizontal Laplacian that they employed. If the grid has a dual whose edges are orthogonal to the edges of the original grid, as is the case for a grid generated by Delaunay triangulation and Voronoi tessalation, then this orthogonality can be exploited to obtain desirable properties in the discretisation such as the mimetic properties 4–7 discussed in the introduction (Thuburn *et al.*, 2009; Ringler *et al.*, 2010). By avoiding resolution clustering and overlaps, such grids avoid the main barriers to parallel scalability.

A significant disadvantage of triangular and pentagonal-hexagonal *C*-grids is that their numerical dispersion relations support branches of computational modes. This is discussed further in section 5. Care must be taken to ensure that such modes are well controlled, otherwise they may lead to noisy and inaccurate solutions. A further disadvantage of triangular grids is that neighbouring cells are not similar to each other; the neighbours of an up triangle are all down triangles, and *vice versa*. Consequently, even for smooth data, the truncation error does not vary smoothly but has a grid-scale pattern and will tend to excite noise. Related to this, the most simple compact finite-difference approximation to the horizontal Laplacian on triangles is only first-order accurate, even for perfect equilateral triangles on a plane (analysis by Hui Wan, reported by Almut Gassmann, personal communication).

5. Rossby-wave propagation and geostrophic adjustment for discretisations of the shallow-water equations

The propagation of fast and slow waves in numerical models is strongly influenced by the grid structure. A poor representation of wave propagation, or the existence of computational modes, can lead to instability, inaccurate geostrophic adjustment, failure to maintain balance, or to a noisy solution. Analysis of the wave propagation characteristics of any proposed scheme can therefore give valuable information about its likely behaviour in practical application.

5.1. *Normal modes*

Normal modes are a useful tool, not only for understanding the fundamental properties of the continuous geophysical fluid dynamics equations, but also for analysing the stability and accuracy properties of their discrete analogues. The general procedure for obtaining normal modes is to:

- define a stationary basic-state atmosphere (usually at rest or in uniform translation or solid-body rotation);
- expand variables as small-amplitude perturbations about this state;
- insert these variables into the equations and linearise by neglecting the products of perturbations; and

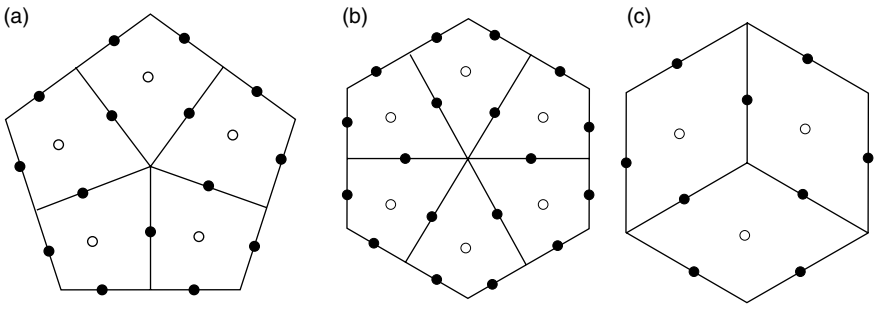

Figure 11. Some divisions of pentagons and hexagons into quadrilaterals: (a) a pentagon divided into five quadrilaterals, with quadrilateral centroids denoted by open circles (○), and quadrilateral midpoint edges by filled circles (●); (b) is as (a), but for a hexagon divided into six quadrilaterals; and (c) is as (b), but with division into three quadrilaterals.

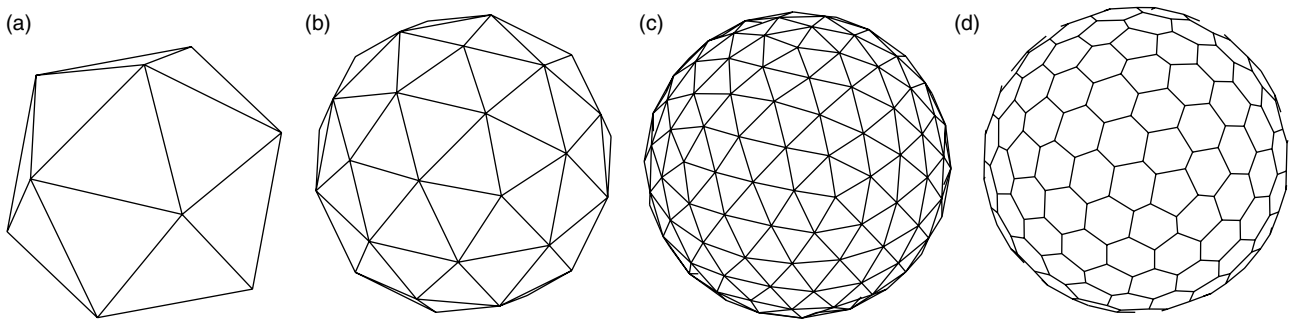

Figure 12. Icosahedral grid construction: (a) regular icosahedron, (b) after division of each triangular face into four subtriangles, (c) after decomposition into 4^2 subtriangles, and (d) the dual pentagonal-hexagonal grid of (c).

- solve the resulting eigenproblem to determine:
  - the number and type of modes;
  - their frequency of oscillation; and
  - their associated spatial structure.

The eigenproblem takes the form of a coupled set of linear differential equations for the continuous case, and a set of coupled linear algebraic equations for the discrete case.

The continuous normal modes of the shallow-water equations have been known since the time of Laplace. Their derivation and associated properties may be found in standard textbooks (e.g. Daley, 1991) for plane, spherical, and other geometries. A common feature of the continuous normal modes of all geometries is that there is a set of slowly propagating Rossby modes, and two sets of gravity modes (eastward- and westward-propagating) whose propagation speed depends on the depth of the fluid–the larger the depth, the faster the propagation. For large- and synoptic-scale flow, the Rossby waves are dynamically dominated by rotation (vorticity) and carry most of the energy, whereas the gravity modes are dynamically dominated by horizontal divergence.

5.2. *Rossby wave propagation and geostrophic adjustment*

The problems of Rossby wave propagation and geostrophic adjustment provide valuable information regarding the dispersion properties of numerical discretisations of the shallow-water equations and, by implication, how well complete models can be expected to perform in practice. Arakawa and colleagues, in some difficult-to-obtain references (e.g. Winninghoff, 1968; Mesinger and Arakawa, 1976; Arakawa and Lamb, 1977), pioneered the early work for numerical dispersion on staggered grids in Cartesian geometry. Attention herein is focused on more-accessible later studies that subsume this earlier pioneering work. An important criterion for judging how well a discretisation of the shallow-water equations performs is whether there is a good correspondence between the properties of the discrete normal modes and their continuous counterparts.

5.3. *Rectangular grids*

5.3.1. *On an f-plane*

The basic nomenclature (*A*–*E*) for unstaggered and staggered grids is that originally defined in Winninghoff (1968), Mesinger and Arakawa (1976) and Arakawa and Lamb (1977), and shown schematically in Figure 13.

A good introduction to geostrophic adjustment in discrete models on rectangular grids may be found in section 7.3 of Haltiner and Williams (1980). This is mainly based on the work of Schoenstadt (1980) for the 1D (linearised) shallow-water equations on an f-plane, using transfer function analysis in an unbounded domain. Essentially the same results can however be easily obtained in periodic domains by substituting discrete Fourier transforms for integral ones.

Schoenstadt (1980) examined the phase and group velocities of centred second- and fourth-order finite-difference and finite-element approximations to the (linearised) shallow-water equations written in primitive form (i.e. in terms of u and v, rather than vorticity and divergence). A good representation of group velocity is particularly important since it controls the propagation of energy, with its sign determining the direction of propagation. Consistent with other studies, both before and since, he showed that the representation of phase and group velocities is:

- very poor for the unstaggered *A*-grid;
- good for *B* and *C* staggerings;

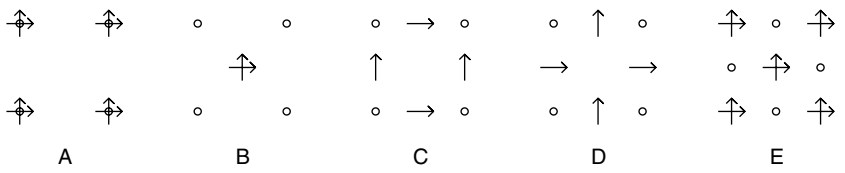

Figure 13. The five grids defined and analysed in Winninghoff (1968), Mesinger and Arakawa (1976) and Arakawa and Lamb (1977). Placement of variables: Φ at open circle points ($\circ$), u indicated by rightward-pointing arrows, and v by upward-pointing arrows.

- poor for D staggering;
- better for fourth-order discretisations than second-order ones;
- somewhat better for finite-element discretisations than finite-difference ones; and
- best overall for the C staggering (of the A–D staggerings).

Schoenstadt (1980)'s work was then extended by Williams (1981) to include vorticity-divergence discretisations on an unstaggered 1D grid, such as those proposed by Staniforth and Mitchell (1977, 1978), Staniforth and Daley (1979), Cullen and Hall (1979) and, more recently, by Randall (1994). Williams (1981) found that

- vorticity-divergence discretisations behave very similarly to, but somewhat better than, C-staggered discretisations of the primitive form of the equations.

The unstaggered vorticity-divergence formulation is now commonly referred to as the Z-grid. The Schoenstadt (1980) and Williams (1981) conclusions, for both primitive and vorticity-divergence discretisations of the 1D shallow-water equations on an f-plane, have been corroborated by the Wajsowicz (1986), Fox-Rabinovitz (1991) and Randall (1994) studies of analogous discretisations in 2D on an f-plane. Wajsowicz (1986) additionally considered 2D discretisations on a β-plane, as discussed below. Fox-Rabinovitz (1991) additionally considered time-staggered schemes.

5.3.2. *On a β-plane*

Whilst the f-plane assumption simplifies the analysis for inertia-gravity wave modes, it unfortunately leads to degenerate, stationary, Rossby modes. That they are stationary then makes it impossible to distinguish, on the basis of frequency, between legitimate physical, but degenerate, Rossby modes, and spurious, stationary, computational ones. Degeneracy means that any linear combination of zero-frequency eigenmodes is also a zero-frequency eigenmode, and this makes it difficult to deduce anything useful about the mode structures. Finally, the degenerate Rossby modes also lose the property of being dispersive. These limitations motivate the use of a β-plane to analyse Rossby-mode dispersion.

An alternative, halfway-house, approach on an f-plane is to introduce a uniform advection U in the x direction into the basic state, with a corresponding balancing term in the y direction, either from a transverse pressure gradient, or from an orographic gradient, or both. The Rossby modes then recover their individual, non-zero, frequencies kU for all wavenumbers k, instead of degenerate zero frequency. This permits an important distinction to be made between any stationary computational modes (if there are any), and individual Rossby modes. What is still missing from such an analysis, however, is the important retrogression, and dispersion, of Rossby waves due to the β effect: to a very good approximation, the phase speed on a β plane is

$$U - \frac{\beta}{\left(k^2 + l^2 + 1/\lambda^2\right)},$$

where k and l are x and y wavenumbers, respectively, and $\lambda = \sqrt{\Phi_0}/f_0$ is the Rossby radius, but this reduces to just U on an f-plane, with no dependence on β, k or l.

In Wajsowicz (1986), the inertia-gravity modes are first isolated by using a 2D f-plane. Geostrophic adjustment is then examined for finite-difference discretisations with the primitive form of the equations, leading to similar conclusions to those of 1D studies of geostrophic adjustment on an f-plane. Using a 2D β-plane, Wajsowicz (1986) next isolated the Rossby modes via a zero-order quasi-geostrophic balance assumption, and then examined their dispersion properties for various finite-difference discretisations on B and C-grids. Provided the horizontal gridlength is less than the Rossby radius, the C-grid is again the preferred grid. (Because of the dependence of the Rossby radius on $\sqrt{\Phi_0}$, i.e. on the equivalent depth in the context of baroclinic models, the C-grid does very well for the external and low-order internal modes, but struggles with the high-order internal modes if the horizontal resolution is too low. It is therefore important, when using a C-grid, to ensure that the horizontal resolution is high enough for given vertical resolution.)

Using the same approach as Wajsowicz (1986) for the dispersion of Rossby modes on a β-plane, Neta and Williams (1989) show that discretisations of the vorticity-divergence form of the equations leads to better Rossby-mode dispersion than discretisations on a C-grid. They attribute this to the fact that '*Rossby wave dynamics are partially or totally controlled by the vorticity equation*'.

For C-grid discretisations, the Coriolis terms must be spatially averaged and, as mentioned in Wajsowicz (1986), there are different ways of doing this. For a β-plane, Thuburn (2007) has shown that how, in detail, this is done has an important impact on the fidelity of the numerical representation of Rossby-wave dispersion. In particular, it was found that approximating $\left(fv, -fu\right)$ as $\left(\overline{f\overline{v}^{y}}^{x}, -\overline{f\overline{u}^{x}}^{y}\right)$ leads to good Rossby-wave dispersion, but approximating by $\left(f\overline{v}^{xy}, -f\overline{u}^{xy}\right)$, or by $\left(\overline{fv}^{xy}, -f\overline{u}^{xy}\right)$, does not (where an overbar indicates a horizontal average).

5.3.3. *On the sphere or in a spherical channel*

Geostrophic adjustment on the sphere is more complicated than on an f-plane or on a β-plane (Wiin-Nielsen, 1976, gives the analytic formulation for the sphere under the

assumption of no variation of the dependent variables in λ) but the (continuous) adjustment process on the sphere is, in many ways, very similar to that on a β-plane. However, the normal mode structures are now Hough functions (Hough, 1897), and their frequencies and structures must be evaluated numerically. Simpler approximate expressions may be found in the limit of large rotational Froude number $\Phi_0/(\Omega a)^2$, where Ω is the earth's rotation rate and a is its radius (Hough, 1897), and were used by Thuburn and Staniforth (2004). Also, exact solutions may be found if f is set to a constant (the 'f sphere') rather than given its usual dependence proportional to sine of latitude (Thuburn *et al.*, 2009).

Surprisingly little work appears to have been done on examining geostrophic adjustment for numerical discretisations of the shallow-water equations on lat–lon grids. The only such study of which we are aware is that of Thuburn and Staniforth (2004) for a lat–lon C-grid, which led to the more-detailed one for a β-plane (Thuburn, 2007). Thuburn and Staniforth (2004) treated variations of the dependent variables in λ analytically and, as later corroborated in Thuburn (2007) for a β-plane, found that the numerical Rossby wave dispersion is very sensitive to the details of how the Coriolis terms are approximated.

5.4. *Triangular and pentagonal-hexagonal geodesic grids*

The good dispersion properties of the staggered C-grid placement of variables, when using grids based on rectangles in biperiodic plane geometry, suggests that appropriate analogues of this placement might also be beneficial when using grids based on other polygons such as triangles, pentagons or hexagons. A key feature of the staggered C-grid placement of variables (Figure 13) is that the scalar variable ($\Phi \equiv gh$) is placed at the centre of a grid box, and the normal component of the velocity field is placed at the four midpoints of the edges of the box (i.e. u is placed on the left and right edges, and v is placed on the top and bottom edges). The natural generalisation of the C-grid to other polygons is thus to place the scalar variable ($\Phi \equiv gh$) at the centroid of the polygon, and the normal component of the velocity at the midpoints (or thereabouts) of the edges of the polygon.

5.4.1. *In the absence of Coriolis terms in plane geometry*

One of the first studies to examine numerical dispersion on triangular grids was the one by Walters and Carey (1983). They found that all three of the examined finite-element velocity-pressure pairs, defined on a grid composed of right-angled triangles, have spurious modes for the linear shallow-water equations. These particular element pairs correspond to either an A or a B placement of variables, with none of the examined pairs corresponding to a C placement. In their analysis, Walters and Carey (1983) showed that the spurious modes (in the examined context of no Coriolis terms) are associated with zero frequency. They noted that spurious modes in surface elevation can exist that have no effect on velocity and also, importantly, *vice versa*. Attention was focused almost exclusively on the spurious surface-elevation modes of zero velocity, since these were argued to be the most troublesome. For this family of spurious modes, *solution uniqueness is lost* since any multiple of a spurious mode can be added to any solution of the discrete equations and still satisfy them. It was also noted that this family of spurious modes is common to discretisations of both the shallow-water and the (compressible or incompressible) Navier–Stokes equations, and *the existence of spurious modes is a manifestation of loss of convergence*. Having made the important observation that a second set of spurious modes, for which $\Phi =$ constant, may exist, little further is said in Walters and Carey (1983) about them.

5.4.2. *On an f-plane*

Following up on the Walters and Carey (1983) analysis, Le Roux *et al.* (1998) observed that the existence of null spaces for $\mathbf{v}$ and Φ, and therefore the existence of computational modes, is implicitly linked to an imbalance in the ratio of the number of degrees of freedom for $\mathbf{v}$ and for Φ. They then compared nine finite-element pairs, defined on triangular meshes, and found that only one pair has the right ratio. Unfortunately this element pair is also unsatisfactory since it is analogous to an A-grid placement of variables with very poor dispersion properties (the group velocity has the wrong sign for many modes of propagation)! Since this study, there has been a number of other similar ones for geostrophic adjustment when using triangular grids on an f-plane with the primitive form of the shallow-water equations (e.g. Le Roux *et al.*, 1998, 2007; Le Roux, 2005; Hanert *et al.*, 2009). To date, no satisfactory placement of variables and choice of finite-element spaces appears to have been found for triangular grids.

Similar studies have been carried out for finite-difference schemes on hexagonal grids (Figure 1 of Ringler and Randall, 2002b, gives the definition of these). The best numerical dispersion is again obtained with the unstaggered vorticity-divergence form Z-grid (Ringler and Randall, 2002b; Ničković *et al.*, 2002). The Z-grid scheme can be mimicked whilst predicting the primitive variables $\mathbf{v}$ and Φ (the ZM-grid; Ringler and Randall, 2002b), but at the expense of introducing a branch of computational modes that need to be removed by filtering. Of the hexagonal A to E grids examined by Ničković *et al.* (2002), two have the right numbers of degrees of freedom to avoid branches of computational modes: the hexagonal A-grid, like the square A-grid, has poor inertio-gravity wave dispersion; the reduced C-grid, which omits every third velocity component of the full C-grid, sacrifices isotropy and accuracy. The full C-grid has the best inertio-gravity wave dispersion properties, similar to the square C-grid but somewhat more isotropic. However, for the particular C-grid discretization analysed by Ničković *et al.* (2002), the geostrophic modes were found to have frequencies of order f rather than zero. This inability to support stationary geostrophic modes implies a spontaneous numerical breakdown of balance, and was considered a serious limitation of the hexagonal C-grid. However, it was subsequently shown (Thuburn, 2008a) that, by an alternative discretisation of the Coriolis terms, the geostrophic modes could be made stationary. Nevertheless, the resulting dispersion relation has two branches of geostrophic modes, rather than the single branch supported by the continuous equations.

5.4.3. *On a β-plane*

Neta and Williams (1989) showed that good Rossby-mode dispersion is obtained on a 2D β-plane for the vorticity-divergence form of the equations when using linear elements

on isosceles triangles. In their analysis, they separated out the Rossby mode, as in Wajsowicz (1986), but then ignored the inertia-gravity modes, leaving unanswered the question of how well, or badly, these modes are represented. This could be examined by redoing the Williams (1981) analysis for geostrophic adjustment on an f-plane when using the vorticity-divergence form of the equations, but using linear elements on isosceles triangles instead of linear elements on rectangles. An important issue here, that does not appear to have been addressed, is whether degrees of freedom are lost in computing the velocity components at cell edges from the vorticity and divergence at cell centres; this does not happen on a rectangular grid but could happen on a triangular grid.

Le Roux *et al.* (2008), using a similar approach to Neta and Williams (1989) for the propagation of Rossby modes on rectangular grids, but using several finite-element pairs on triangular grids with the primitive form of the equations, found that some pairs lead to particularly significant errors for phase and group velocities.

For hexagonal C-grids, Thuburn (2008a) extended the f-plane analysis to examine Rossby-mode dispersion on a 2D β-plane. As noted by previous authors, the discrete solution not only has the three usual sets of modes (Rossby plus eastward and westward inertia-gravity), but also an additional set of Rossby modes. The question then is whether this additional set of Rossby modes are useful approximations to any continuous counterparts. Thuburn (2008a) showed that they are not. They can therefore be considered to be a set of unwelcome spurious modes. He also argued that a similar phenomenon occurs for discretisations on triangular grids: there are then *two* sets of spurious modes, but this time inertia-gravity modes, consistent with the work of Le Roux *et al.* (2008) and also of Danilov (2010). This means that discretisations on both triangular and hexagonal C-grids have a fundamental weakness inasmuch as both lead to branches of computational modes in the numerical dispersion relation. Measures will be needed to ensure that these computational modes are well controlled if such grids are used but, as discussed in Jablonowski and Williamson (2011), artificial control measures can have adverse and unexpected consequences. There is also an implied loss of efficiency in the use of available degrees of freedom, since higher resolution is likely to be needed to counterbalance the reduced accuracy caused by any applied remedial measures.

A further concern with triangular and hexagonal grids is that the associated discretisations may suffer from boundary-condition problems for limited-area applications, due to their computational modes. It is already very challenging to formulate sufficiently transparent open-boundary conditions for equation sets with multiple signal speeds when using a rectangular C-grid. It remains to be seen whether this can be done acceptably well for geodesic C-grids in the presence of a family (for pentagonal-hexagonal grids), or two families (for triangular grids), of computational modes.

5.4.4. *On the sphere*

Thuburn *et al.* (2009) examined how to construct C-type discretisations of the linearised shallow-water equations on arbitrary polygons (provided there exists a dual grid that is orthogonal to the original grid) that ensure not only that geostrophic modes are stationary, but also that they have good conservation properties. For their proposed scheme, they numerically calculated normal modes for a pentagonal-hexagonal icosahedral grid and a triangular icosahedral grid on the f sphere. For both grids, geostrophic modes are indeed stationary, as designed. However, consistent with the planar analysis (Thuburn, 2008a), the pentagonal-hexagonal grid has an extra set of geostrophic modes, whilst the triangular grid has two extra sets of inertio-gravity modes.

The discussion in section 6 of Thuburn *et al.* (2009) on computational modes makes three important points:

- C-grid discretisations on non-quadrilateral grids can support wave modes that propagate in an unphysical manner;
- the number and type of such modes, if they exist, depend on the underlying grid structure, and also upon the relative number of degrees of freedom for mass, vorticity and divergence (or, equivalently, for mass and momentum); and
- although wave propagation is strongly influenced by details of the discretisation, the existence of spurious modes depends only on the underlying grid structure and placement of variables.

5.5. *A unified approach*

Bernard *et al.* (2008, 2009a) presented an attractive unifying approach for analysis on a β-plane that permits *consistent* analysis of the propagation of inertia-gravity and Rossby modes. Instead of isolating the Rossby mode on a β-plane (via a WKB-type approximation that involves a zero-order quasi-geostrophic balance assumption), as is usually done, and then relying on a *separate* f-plane analysis to isolate the inertia-gravity modes, they instead handled all modes *simultaneously*. From the linearised shallow-water equations on a β-plane, they first eliminated u and Φ (without approximation) to obtain

$$\left[\frac{\partial}{\partial t}\left\{-\frac{\partial^2}{\partial y^2}+\frac{1}{\Phi_0}\left(\frac{\partial^2}{\partial t^2}+f^2\right)\right\}-\frac{\partial}{\partial x}\left(\beta+\frac{\partial^2}{\partial x\partial t}\right)\right]\left(\frac{\partial v}{\partial t}\right)=0, \tag{46}$$

where $f \equiv f_0 + \beta y$. Next they imposed channel boundary conditions and expanded v as

$$v(x,y,t) = Y(y)\exp\{\mathrm{i}(k_x x - \omega t)\}, \tag{47}$$

before substituting this into (46) to obtain

$$\frac{\mathrm{d}^2 Y}{\mathrm{d}y^2} - g(y)\,Y + k_y^2 Y = 0, \tag{48}$$

where

$$\left.\begin{aligned} g(y) &\equiv \frac{2f_0\beta}{\Phi_0}y + \frac{\beta^2}{\Phi_0}y^2, \\ k_y^2 &\equiv \frac{1}{\Phi_0}(\omega^2 - f_0^2) - \frac{\beta k_x}{\omega} - k_x^2. \end{aligned}\right\} \tag{49}$$

Equation (48), together with the wall boundary conditions $Y(-L/2) = Y(+L/2) = 0$, gives a Sturm–Liouville eigenproblem that is solved numerically to obtain 'exact' eigensolutions (exact to the extent that the numerical method to evaluate them is run at sufficiently high resolution).

The above is similar to the approach taken in Staniforth *et al.* (1993) for the normal modes in an f-plane channel with linear bottom slope, where the resulting eigenmodes turn out to be confluent hypergeometric functions. Following the procedure used in appendix A of Staniforth *et al.* (1993), it can be shown that the linearised equations of Bernard *et al.* (2008, 2009a), together with their wall boundary conditions, have both an energy principle and the crucial property that the exact solutions are neutrally stable. This is important, since any instability observed in an analysis of discretised equations must be attributable to the numerics, and not to any inherent instability in the underlying linearised equations.

For a certain choice of parameters, Bernard *et al.* (2008) showed that the eigenfrequencies agree qualitatively (but not particularly well quantitatively) with those of the usual mix of asymptotic solutions for Rossby modes on a β-plane, and inertia-gravity modes on an f-plane. For a discontinuous Galerkin discretisation, they then obtain analogous discrete eigenvalues and eigenmodes. These are then used to perform a convergence study for their discretisation.

5.6. *Forced response*

As well as the free normal modes and their representation by numerical schemes, the response to forcing on given space- and time-scales is also an important question. However, with the possible exception of spurious orographic resonance (see below), this question has received relatively little attention in the literature.

Flow over orography forces stationary gravity waves. Free gravity waves have their horizontal phase speed Doppler shifted by the background flow. If the background flow is strong enough to Doppler shift the horizontal phase speed to zero, then orographic forcing can produce a large-amplitude resonant response. But such a situation is very rare in the large-scale atmosphere because the horizontal phase speed of deep gravity waves is several hundred metres per second. However, when a semi-implicit semi-Lagrangian integration scheme is used with large time steps, the gravity-wave phase speed is artificially slowed, so that resonance is possible for much more modest wind speed. The problem may be mitigated to some extent by averaging the orography field or by off-centring in the time integration scheme (e.g. Tanguay *et al.*, 1992; Rivest *et al.*, 1994; Ritchie and Tanguay, 1996).

Another example of a spurious numerical resonant response to forcing was pointed out by Schneider (1987). He showed that vertical discretisation on a Lorenz grid is unable, in general, to provide a steady response to a steady diabatic forcing. The reason is that the steady forcing projects onto the zero-frequency computational mode supported by the Lorenz grid, producing a resonant growth that is linear in time.

A further example is provided by the linearised, 1D non-rotating shallow-water equations. These support gravity waves in which frequency ω is proportional to wavenumber k. However, when discretised using centred differences on an unstaggered grid (a 1D A-grid), $\omega(k)$ grows to a maximum for some k, before falling to zero for the maximum resolvable k. Thus, for any given ω, there are two values of k, rather than one, that satisfy the dispersion relation. Point forcing at some frequency ω will therefore produce a long-wave response that is an approximation to the correct response, but also a spurious short-wave response that has no physical counterpart.

Several authors have proposed shallow-water test cases in which forcing terms are chosen so as to achieve a specified exact solution (e.g. Marchesin, 1984; Dee and Da Silva, 1986; test case 4 of Williamson *et al.*, 1992). However, such test cases have not proved very popular with model developers. Nevertheless, effects that act as forcing on the dynamics (or coupling with the dynamics) are ubiquitous, such as orography, radiation, condensation, and subgrid models for convection, boundary-layer turbulence and unresolved gravity waves. The response of numerical schemes to such forcing or coupling should be a design priority for model developers, and we recommend further investigation of this topic. Some progress has recently been made on temporal discretisation aspects of physics–dynamics coupling (e.g. Staniforth and Wood, 2008, and references therein) but further work also needs to be undertaken on spatial discretisation aspects.

5.7. *Balancing degrees of freedom*

For good representations of geostrophic adjustment and Rossby-mode propagation, the above-cited studies indicate the need to balance the number of degrees of freedom between momentum and 'pressure' (as a surrogate of density), and that failure to do so leads to spurious, or very poorly represented, discrete modes. In this regard, the geometric properties of polyhedra can provide valuable insight. Assuming a C-grid placement of variables, with pressure at the centres of the faces and the normal component of momentum at the midpoints of the edges, the required balance is obtained when there are twice as many edges as faces.

The analysis of section 2 shows that the cube is the only regular polyhedron that has the required balance, and, in fact, *all* quadrilateral-based grids on the sphere have exactly twice the number of edges as faces: each face has exactly four sides, and each of these is exactly shared between two adjacent quadrilaterals (19).

For the triangular regular polyhedra, i.e. the tetrahedron, octahedron and the icosahedron, as well as other grids built entirely from triangles, the ratio is 3:2 (13) and there are too many pressure nodes. For the dodecahedron, the ratio is 5:2 (25) and there are too many momentum nodes. For the pentagonal-hexagonal geodesic grids, the ratio approaches 3:1 (42) for $F^{\mathrm{H}} \gg F^{\mathrm{P}}$, i.e. when there are many more hexagonal faces than pentagonal ones, and there are again too many momentum nodes.

6. Grid imprinting

As noted in section 2.3, only a small number of truly uniform tilings of the sphere exist, and these are all far too coarse for practical use. All other tilings of the sphere have some large-scale inhomogeneity, and there may be special points or special lines, such as those corresponding to the vertices and edges of the original cube on a cubed sphere, where the local grid structure is different from elsewhere. The dependence of the truncation error on the local grid structure is likely to leave some signal of the grid structure in the numerical solution. It is desirable to choose the grid and discretisation to minimise such grid imprinting.

Grid imprinting is most clearly visible when simulating a simple symmetrical flow. For example, using a linearised version of the shallow-water model of Tomita *et al.* (2001) on a subdivided icosahedral grid, with all variables placed at triangle vertices (i.e. a pentagonal-hexagonal *A*-grid), Tomita *et al.* (2002) initialised it with an analytic, balanced, solid-body, rotational flow, sampled at gridpoints. Because of the model's numerics, the exact analytic balance of the initial state is not preserved. Instead, the initial imbalance leads to a process of geostrophic adjustment. Tomita *et al.* (2002) then computed the (time-dependent) error fields and decomposed them in terms of Hough functions, which are the eigenfunctions of the (linear) shallow-water equations. They found that the resulting spectra for the Rossby and inertia-gravity modes have a spike for zonal wavenumber 5, and associated this with the underlying icosahedral grid structure, i.e. a manifestation of grid imprinting.

Thuburn and Li (2000) found that truncation errors in a hexagonal-icosahedral *Z*-grid shallow-water model broke the symmetry of a zonal wavenumber 4 Rossby–Haurwitz wave, artificially accelerating the instability of the Rossby–Haurwitz wave. The spurious triggering of instability by grid imprinting has been investigated more systematically by Lauritzen *et al.* (2010a). They compared the ability of eight different hydrostatic primitive equation models, and model variants, to maintain a steady, but baroclinically unstable, zonal flow. The model grids included lat–lon, cubed sphere, and geodesic grids, and different grid orientations relative to the flow were tested. Only the unrotated lat–lon grid models were able to maintain the zonal flow for 30 days at 2^o resolution. All of the models showed reduced grid imprinting with finer resolution. The grid imprinting was strongly sensitive to the details of the numerical scheme, as well as the grid structure itself.

For 3D models, the decrease of density with altitude will greatly amplify any signal forced by grid imprinting in the troposphere that subsequently propagates upwards. Thus, grid imprinting could be especially damaging for models that include the stratosphere and mesosphere.

For more complex, realistic flows, grid imprinting could manifest itself in more subtle and difficult-to-detect ways. For example, it could lead to a bias in preferred genesis regions for baroclinic instability. Grid imprinting could also lead to a bias in the wavenumber spectrum of planetary waves, since the most unstable baroclinic waves have zonal wavenumbers in the range of 4 to 10, and this correlates well with the symmetries of most, if not all, quasi-uniform grids. Coupling of dynamics with parametrised physical processes could potentially amplify any grid imprinting. Careful checking will be needed to ensure that model climate statistics are not contaminated by grid imprinting, for example by checking whether they are sensitive to re-orientation of the grid.

Another manifestation of grid imprinting may occur on grids for which the basic repeating unit is not a single grid cell but a small group of grid cells, for example a triangular grid, where the basic repeating unit is a pair of triangles, one up, one down, or a grid of triangles subdivided into quadrilaterals (section 3.5.4), where the basic repeating unit comprises a set of six differently oriented kite-shaped quadrilaterals (three for an up triangle plus three for a down one). For smooth data, the truncation error will have a grid-scale pattern reflecting the small-scale structure of the grid, resulting in a tendency to force grid-scale noise. Moreover, such grid-scale inhomogeneity can significantly impact the accuracy of the numerics. For example, combining a first-order accurate approximation of the divergence with a first-order accurate approximation of the gradient does not necessarily lead to a first-order accurate approximation of the Laplacian; this would be dependent on some cancellation of leading truncation errors, which in turn is likely to depend on certain symmetries of the local grid structure. For example, Hui Wan (reported by A. Gassmann, personal communication) has found that the natural centred-difference Laplacian operator on planar equilateral triangles is only first-order accurate, even though the divergence and gradient operators are both second-order.

7. Conclusion

A great many different ways of gridding the sphere are possible, and the choice of horizontal grid is a leading-order question in the design of any dynamical core for numerical modelling of the global atmosphere. The choice of grid will impact efficiency and scalability on the massively parallel computers needed to perform high-resolution simulations and forecasts. It will also affect the ability to obtain key properties that are considered essential or desirable for accuracy, such as conservation, avoiding computational modes, and minimising grid imprinting. More generally, it will also affect the choice of numerical methods for physics–dynamics coupling.

For scalability, a quasi-uniform grid is thought to be essential. If we assume a *C*-grid placement of variables, then the existence of a dual grid orthogonal to the primal grid is desirable, to facilitate the mimetic properties listed in the introduction, and the use of quadrilateral grid cells is desirable as a prerequisite for avoiding computational modes. So can we obtain a quasi-uniform orthogonal quadrilateral grid?

Orthogonal quadrilateral grids can be generated with the aid of conformal maps (section 3.2), and a variety of such grids are possible. However, they all have a number of singularities with associated resolution clustering (or resolution dilation), so the ratio of maximum to minimum grid length increases as the grid is refined. These grids are therefore not quasi-uniform.

One way to obtain quasi-uniformity on a quadrilateral grid is to give up orthogonality (section 3.4). Again a variety of such grids are possible. The issue now is whether all of the desirable properties, including the mimetic properties listed in the introduction, can be obtained on a quadrilateral *C*-grid without the existence of an orthogonal dual. This is currently an open research question.

An alternative way to obtain quasi-uniformity on a quadrilateral grid, without giving up orthogonality, is to use an overset grid (section 3.3), albeit at the cost of duplicated computation in the overset regions; once again a variety of such grids are possible. Recent progress has been made on the challenge of solving global elliptic problems on such grids for semi-implicit time integration schemes, which can significantly enhance computational efficiency. The outstanding issues with these grids, which merit further investigation, are: (i) whether a mass-conserving coupling at the overlaps can be achieved with sufficient accuracy and efficiency; (ii) whether the coupling at the overlaps compromises wave propagation, the adjustment process,

balance, forced response, or other aspects of accuracy; and (iii) the parallel efficiency of communication between the component grids.

More exotic grids might be able to achieve quasi-uniform quadrilaterals with an orthogonal dual. The kites grid (section 3.5.4) is one such example. However, although such grids have the right ratio of velocity degrees of freedom to pressure ones to avoid computational modes, their connectivity is not that of a logically rectangular grid, and they are inhomogeneous at the grid scale. Important questions requiring further investigation are (i) whether the grid-scale inhomogeneity leads to noisy solutions (possibly depending on the numerical method used), and (ii) whether dispersion analysis (section 5) confirms the hypothesis that computational modes are avoided.

An alternative approach is to retain quasi-uniformity and an orthogonal dual, but give up quadrilateral grid cells, as exemplified by triangular and pentagonal-hexagonal icosahedral grids (section 4). It is known that such grids support branches of computational modes in their numerical dispersion relations: two branches of inertio-gravity modes in the triangular case, an extra branch of Rossby modes in the pentagonal-hexagonal case. The outstanding issue here is whether these computational modes can be controlled well enough, in realistic applications, without unduly damaging the fidelity of the numerical solution, and its rate of convergence as resolution is increased. Even if they can, there is an associated loss of computational and communication efficiency. In the triangular case, grid-scale inhomogeneity may also be an issue.

Despite the appeal of the *C*-grid in terms of its accurate handling of fast waves and the adjustment process, the above discussion reveals that no currently available *C*-grid and scheme is known to satisfy all of the desired properties. It is appropriate, therefore, to consider briefly some of the alternatives.

A *Z*-grid placement of variables (section 5) is appealing because it retains the accurate wave dispersion properties of *C*-grids, whilst the number of mass plus divergence degrees of freedom always equals twice the number of vorticity degrees of freedom, making it likely that computational modes can be avoided without the restriction to quadrilateral grids. The principal drawback is that horizontal Poisson equations must be solved at each model level and time step to recover the horizontal wind components. This is a less-local elliptic problem than the Helmholtz problem that arises in semi-implicit time integration, and there are potential issues regarding its cost and scalability on massively parallel machines. However, recent progress suggests good scaling is possible up to at least eighty thousand processors (Ross Heikes and David Randall, personal communication). A further potential issue in the case of a triangular *Z*-grid is that degrees of freedom may be lost in computing velocity components from vorticity and divergence; it would be valuable to investigate this further.

Dynamical cores based on the spectral-element method on a cubed sphere have been shown to have excellent scaling on parallel computers (Dennis *et al.*, 2005) and a number of desirable mimetic properties (Taylor and Fournier, 2010). We are not aware of any dispersion analysis for the spectral-element method, but it seems likely that the use of a collocated grid for wind components and pressure will lead to retarded frequency and reversed group velocity for short waves, as on an *A*-grid. However, it may turn out that the high accuracy of the spectral-element method means that only the very shortest wavelengths are adversely affected. Another potential issue is that the quadrature points used in the spectral-element method are non-uniformly spaced, leading to the possibility of trapping of short waves in the highest-resolution regions. Some dispersion analysis for the spectral-element method to address these questions would be very valuable.

A largely overlooked aspect of the development of dynamical cores is the response of their numerical schemes to important forcings, such as orography, radiation, condensation, and subgrid models for convection, boundary-layer turbulence and unresolved gravity waves. The response of numerical schemes to such forcing or coupling should be a design priority for model developers, and we recommend further investigation of this topic.

In summary, developments in computing architecture are making it imperative for both operational prediction centres and developers of numerical models for research use to consider quasi-uniform grids on the sphere. However, all grids that have been proposed to date for global atmospheric modelling have known problems or issues that require further investigation. In this review we have attempted to summarise the strengths and weaknesses of the various grids, and to suggest some directions for further research that would be of value to the community.

Acknowledgements

Useful discussions on various aspects of this work with Tom Allen, Colin Cotter, Terry Davies, Almut Gassmann, Markus Gross, Ross Heikes, Tom Melvin, Abdessamad Qaddouri, David Randall, Chris Smith, David Thomson, Simon Vosper, Andy White, Nigel Wood and Mohamed Zerroukat are gratefully acknowledged. We also thank Nigel Wood, Jim Purser and an anonymous reviewer for their reviews of an earlier version of this paper which helped us improve it.

References

Adcroft A, Campin J-M, Hill C, Marshall J. 2004. Implementation of an atmosphere-ocean general circulation model on the expanded spherical cube. *Mon. Weather Rev.* **132**: 2845–2863.

Arakawa A, Lamb VR. 1977. Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods in Comput. Phys.* **17**: 174–265.

Arakawa A, Moorthi S. 1988. Baroclinic instability in vertically discrete systems. *J. Atmos. Sci.* **45**: 1688–1707.

Augenbaum J, Peskin C. 1985. On the construction of the Voronoi mesh on the sphere. *J. Comput. Phys.* **59**: 177–192.

Baba Y, Takahashi K, Sugimura T, Goto K. 2010. Dynamical core of an atmospheric general circulation model on a Yin-Yang grid. *Mon. Weather Rev.* **138**: 3988–4005.

Bacon DP, Ahmad NN, Boybeyi Z, Dunn TJ, Hall MS, Lee PCS, Sarma RA, Turner MD, Waight KT, Young SH, Zack JW. 2000. A dynamically adapting weather and dispersion model: The Operational Multiscale Environment model with Grid Adaptivity (OMEGA). *Mon. Weather Rev.* **128**: 2044–2076.

Berger MJ, Calhoun DA, Helzel C, LeVeque RJ. 2009. Logically rectangular finite volume methods with adaptive refinement on the sphere. *Phil. Trans. R. Soc. A* **367**: 4483–4496.

Bernard P-E, Deleersnijder E, Legat V, Remacle J-F. 2008. Dispersion analysis of discontinuous Galerkin schemes applied to Poincaré, Kelvin and Rossby waves. *J. Sci. Comput.* **34**: 26–47.

Bernard P-E, Remacle J-F, Legat V. 2009a. Modal analysis on unstructured meshes of the dispersion properties of the PNC-P1 pair. *Ocean Modelling* **28**: 2–11.

Bernard P-E, Remacle J-F, Comblen R, Legat V, Hillewaert K. 2009b. High-order discontinuous Galerkin schemes on general 2D manifolds applied to the shallow-water equations. *J. Comput. Phys.* **228**: 6514–6535.

Bourke W. 1972. An efficient, one-level, primitive-equation spectral model. *Mon. Weather Rev.* **100**: 683–689.

Browning GL, Hack JJ, Swarztrauber PN. 1989. A comparison of three numerical methods for solving differential equations on the sphere. *Mon. Weather Rev.* **117**: 1058–1075.

Calhoun DA, Helzel C, LeVeque RJ. 2008. Logically rectangular grids and finite volume methods for PDEs in circular and spherical domains. *SIAM Rev.* **50**: 723–752.

Charney JG, Fjortoft R, von Neumann J. 1950. Numerical integration of the barotropic vorticity equation. *Tellus* **2**: 237–254.

Chen C, Xiao F. 2008. Shallow-water model on a cubed sphere by multi-moment finite volume method. *J. Comput. Phys.* **227**: 5019–5044.

Comblen R, Legrand S, Deleersnijder E, Legat V. 2009. A finite-element method for solving the shallow-water equations on the sphere. *Ocean Modelling* **28**: 12–23.

Côté J, Gravel S, Méthot A, Patoine A, Roch M, Staniforth A. 1998. The operational CMC-MRB Global Environmental Muliscale (GEM) model. Part I: Design considerations and formulation. *Mon. Weather Rev.* **126**: 1373–1395.

Coxeter HSM. (ed) 1973. *Regular polytopes.* Dover: New York.

Cullen MJP. 1974. Integrations of the primitive equations on a sphere using the finite-element method. *Q. J. R. Meteorol. Soc.* **100**: 555–592.

Cullen MJP, Hall CD. 1979. Forecasting and general circulation results from finite-element models. *Q. J. R. Meteorol. Soc.* **105**: 571–592.

Daley R. 1991. *Atmospheric data analysis.* Cambridge University Press: Cambridge, UK.

Danilov S. 2010. On the utility of triangular C-grid type discretization for numerical modeling of large-scale ocean flows. *Ocean Dyn.* **60**: 1361–1369.

Davies T, Cullen MJP, Malcolm A, Mawson M, Staniforth A, White AA, Wood N. 2005. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q. J. R. Meteorol. Soc.* **131**: 1759–1782.

Dee DP, Da Silva AM. 1986. Using Hough harmonics to validate and assess nonlinear shallow-water models. *Mon. Weather Rev.* **114**: 2191–2196.

Dennis J, Fournier A, Spotz WF, St-Cyr A, Taylor MA, Thomas SJ, Tufo H. 2005. High resolution mesh convergence properties and parallel efficiency of a spectral element atmospheric dynamical core. *Int. J. High Perf. Comput. Appl.* **19**: 225–235.

Dudhia J, Bresch JF. 2002. A global version of the PSU-NCAR mesoscale model. *Mon. Weather Rev.* **130**: 2989–3007.

Ford R, Pain CC, Piggott MD, Goddard AJH, de Oliveira CRE, Umpleby AP. 2004a. A nonhydrostatic finite-element model for three-dimensional stratified oceanic flows. Part 1: Model formulation. *Mon. Weather Rev.* **132**: 2816–2831.

Ford R, Pain CC, Piggott MD, Goddard AJH, de Oliveira CRE, Umpleby AP. 2004b. A nonhydrostatic finite-element model for three-dimensional stratified oceanic flows. Part 2: Model validation. *Mon. Weather Rev.* **132**: 2832–2844.

Fournier A, Taylor MA, Tribbia JJ. 2004. The Spectral Element Atmosphere Model (SEAM): High-resolution parallel computation and localized resolution of regional dynamics. *Mon. Weather Rev.* **132**: 726–748.

Fox-Rabinovitz MS. 1991. Computational dispersion properties of horizontal staggered grids for atmospheric and ocean models. *Mon. Weather Rev.* **119**: 1624–1639.

Geleyn J-F, Caian M. 1997. Some limits to the variable-mesh solution and comparison with the nested-LAM solution. *Q. J. R. Meteorol. Soc.* **123**: 743–766.

Giraldo FX. 1997. Lagrange–Galerkin methods on spherical geodesic grids. *J. Comput. Phys.* **136**: 197–213.

Giraldo FX. 1998. The Lagrange–Galerkin spectral element method on unstructured quadrilateral grids. *J. Comput. Phys.* **147**: 114–146.

Giraldo FX. 2000. Lagrange–Galerkin methods on spherical geodesic grids: The shallow-water equations. *J. Comput. Phys.* **160**: 336–368.

Giraldo FX. 2001. A spectral element shallow-water model on spherical geodesic grids. *Int. J. Num. Fluids* **35**: 869–901.

Giraldo FX. 2005. Semi-implicit time-integrators for a scalable spectral element atmospheric model. *Q. J. R. Meteorol. Soc.* **131**: 2431–2454.

Giraldo FX. 2006. High-order triangle-based discontinuous Galerkin methods for hyperbolic equations on a rotating sphere. *J. Comput. Phys.* **214**: 447–465.

Giraldo FX, Rosmond TE. 2004. A scalable spectral element Eulerian atmospheric model (SEE-AM) for NWP: Dynamical core tests. *Mon. Weather Rev.* **132**: 133–153.

Giraldo FX, Warburton T. 2005. A nodal triangle-based spectral element method for the shallow-water equations on the sphere. *J. Comput. Phys.* **207**: 129–150.

Giraldo FX, Hesthaven JS, Warburton T. 2002. Nodal high-order discontinuous Galerkin methods for the spherical shallow-water equations. *J. Comput. Phys.* **181**: 499–525.

Giraldo FX, Hesthaven JS, Warburton T. 2003. A spectral element semi-Lagrangian (SESL) method for the spherical shallow-water equations. *J. Comput. Phys.* **190**: 623–650.

Giraldo FX, Lauter M, Handorf D, Dethloff K. 2008. A discontinuous Galerkin method for the shallow-water equations using spherical triangular coordinates. *J. Comput. Phys.* **2227**: 10226–10243.

Gopalakrishnan SG, Bacon DP, Ahmad NN, Boybeyi Z, Dunn TJ, Hall MS, Jin Y, Lee PCS, Mays DE, Madala RV, Sarma RA, Turner MD, Wait TR. 2002. An operational multiscale hurricane forecasting system. *Mon. Weather Rev.* **130**: 1830–1847.

Haltiner GJ, Williams RT. 1980. *Numerical Prediction and Dynamic Meteorology.* John Wiley & Sons: New York.

Hanert E, Walters RA, Le Roux DY, Pietrzak JD. 2009. A tale of two elements: P1NC-P1 and RT0. *Ocean Modelling* **28**: 24–33.

Harris LM, Lauritzen PH, Mittal R. 2011. A flux-form version of the conservative semi-Lagrangian multi-tracer transport scheme (CSLAM) on the cubed sphere grid. *J. Comput. Phys.* **230**: 1215–1237.

Heikes R, Randall DA. 1995a. Numerical integration of the shallow-water equations on a twisted icosahedral grid. Part I: Basic design and results of tests. *Mon. Weather Rev.* **123**: 1862–1880.

Heikes R, Randall DA. 1995b. Numerical integration of the shallow-water equations on a twisted icosahedral grid. Part II: A detailed description of the grid and an analysis of numerical accuracy. *Mon. Weather Rev.* **123**: 1881–1887.

Hoskins BJ, Simmons AJ. 1975. A multi-layer spectral model and the semi-implicit method. *Q. J. R. Meteorol. Soc.* **101**: 637–655.

Hough SS. 1897. On the application of harmonic analysis to the dynamical theory of the tides. Part II: On the general integration of Laplace's dynamical equations. *Phil. Trans. R. Soc. London* **191A**: 139–185.

Hyman JM, Shashkov M. 1997. Natural discretizations for the divergence, gradient, and curl on logically rectangular grids. *Computers Math. Applic.* **33**: 81–104.

Ii S, Xiao F. 2010. A global shallow-water model using a high-order multi-moment constrained finite-volume method and icosahedral grid. *J. Comput. Phys.* **229**: 1774–1796.

Itoh T, Shimada K. 2002. Automatic conversion of triangular meshes into quadrilateral meshes with directionality. *Int. J. CAD/CAM* **1**: 1–21.

Jablonowski C, Williamson DL. 2011. The pros and cons of diffusion, filters and fixers in atmospheric general circulation models. In *Numerical Techniques for Global Atmospheric Models*, *Lecture Notes in Computational Science and Engineering* 80. Lauritzen PH, Jablonowski C, Taylor MA, Nair RD. (eds) Springer: Berlin. 389–504.

Joe B. 1995. Quadrilateral mesh generation in polygonal regions. *Computer-Aided Design* **27**: 209–222.

Kageyama A. 2005. Dissection of a sphere and Yin-Yang grids. *J. Earth Simulator* **3**: 20–28.

Kageyama A, Sato T. 2004. The Yin-Yang grid: An overset grid in spherical geometry. *Geochem. Geophys. Geosyst.* **5**: Q09005.

Kurihara Y. 1965. Numerical integration of the primitive equations on a spherical grid. *Mon. Weather Rev.* **93**: 399–415.

Lanser D, Blom JG, Verwer JG. 2000. Spatial discretization of the shallow-water equations in spherical geometry using Osher's scheme. *J. Comput. Phys.* **165**: 542–565.

Lau TS, Lo SH, Lee CK. 1997. Generation of quadrilateral mesh over analytical curved surfaces. *Finite Elements Anal. Design* **27**: 251–272.

Lauritzen PH, Nair RD. 2008. Monotone and conservative cascade remapping between spherical grids (CaRS): Regular latitude–longitude and cubed-sphere grids. *Mon. Weather Rev.* **136**: 1416–1432.

Lauritzen PH, Jablonowski C, Taylor MA, Nair RD. 2010a. Rotated versions of the Jablonowski steady-state and baroclinic wave test cases: A dynamical core intercomparison. *J. Adv. Model. Earth Syst.* **2**: Art. 15, 34 pp.

Lauritzen PH, Nair RD, Ullrich PA. 2010b. A Conservative Semi-Lagrangian Multi-tracer transport scheme (CSLAM) on the cubed-sphere grid. *J. Comput. Phys.* **229**: 1401–1424.

Lauritzen PH, Jablonowski C, Taylor MA, Nair RD. (eds) 2011. *Numerical Techniques for Global Atmospheric Models. Lecture Notes in Computational Science and Engineering* 80. Springer: Berlin.

Läuter M, Handorf D, Rakowsky N, Behrens J, Frickenhaus S, Best M, Dethloff K, Hiller W. 2007. A parallel adaptive barotropic model of the atmosphere. *J. Comput. Phys.* **223**: 609–628.

Läuter M, Giraldo FX, Handorf D, Dethloff K. 2008. A discontinuous Galerkin method for the shallow-water equations in spherical triangular coordinates. *J. Comput. Phys.* **227**: 10226–10242.

Le Roux DY. 2005. Dispersion relation analysis of the PNC-P1 finite-element pair in shallow-water models. *SIAM J. Sci. Comput.* **27**: 394–414.

Le Roux DY, Staniforth A, Lin CA. 1998. Finite elements for shallow-water equation ocean models. *Mon. Weather Rev.* **126**: 1931–1951.

Le Roux DY, Rostand V, Pouliot B. 2007. Analysis of numerically induced oscillations in 2D finite-element shallow-water models Part I: Inertia-gravity waves. *SIAM J. Sci. Comput.* **29**: 331–360.

Le Roux DY, Rostand V, Pouliot B. 2008. Analysis of numerically induced oscillations in 2D finite-element shallow-water models Part II: Free planetary waves. *SIAM J. Sci. Comput.* **30**: 1970–1991.

Lee J-L, MacDonald AE. 2009. A finite-volume icosahedral shallow-water model on a local coordinate. *Mon. Weather Rev.* **137**: 1422–1437.

Lee LP. 1976. Conformal projections based on elliptic functions. Monograph No. 16, University of Toronto Press: Toronto, Canada. *Canad. Cartogr. Suppl.* **13**: 1–128.

Leopardi P. 2006. A partition of the unit sphere into regions of equal area and small diameter. *Electron. Trans. Numer. Anal.* **25**: 309–327.

Li X, Chen D, Peng X, Takahashi K, Xiao F. 2008. A multimoment finite-volume shallow-water model on the Yin-Yang overset spherical grid. *Mon. Weather Rev.* **136**: 3066–3086.

Loisel S, Côté J, Gander MJ, Laayouni L, Qaddouri A. 2010. Optimized domain decomposition methods for the spherical Laplacian. *SIAM J. Numer. Anal.* **48**: 524–551.

Majewski D, Liermann D, Prohl P, Ritter B, Buchhold M, Hanisch T, Paul G, Wergen W, Baumgardner J. 2002. The operational global icosahedral-hexagonal gridpoint model GME: Description and high-resolution tests. *Mon. Weather Rev.* **130**: 319–338.

Marchesin D. 1984. Using exact solutions to develop an implicit scheme for the baroclinic primitive equations. *Mon. Weather Rev.* **112**: 269–277.

Masuda Y, Ohnishi H. 1986. An integration scheme of the primitive equations model with an icosahedral-hexagonal grid system and its application to shallow-water equations. In *Short and Medium Range Numerical Weather Prediction.* Matsuno T. (ed.) Japan Meteorological Society: Tokyo. 237–240.

McGregor JL. 1996. Semi-Lagrangian advection on conformal-cubic grids. *Mon. Weather Rev.* **124**: 1311–1322.

McGregor JL. 1997. Semi-Lagrangian advection on a cubic gnomonic projection of the sphere. In *Numerical Methods in Atmospheric and Oceanic Modelling, the André J. Robert memorial volume.* Lin CA, Laprise R, Ritchie H. (eds) CMOS/NRC Press: Ottawa, Canada. 153–169.

McGregor JL. 2005. 'C-CAM: Geometric aspects and dynamical formulation'. Tech. report 70. CSIRO Atmospheric Research: Aspendale, Australia.

Mesinger F, Arakawa A. 1976. *Numerical methods used in atmospheric models. GARP Pub. Series* No. 17 1. World Meteorological Organisation: Geneva.

Miura H, Kimoto M. 2005. A comparison of grid quality of optimized spherical hexagonal-pentagonal geodesic grids. *Mon. Weather Rev.* **133**: 2817–2833.

Murray RJ. 1996. Explicit generation of orthogonal grids for ocean models. *J. Comput. Phys.* **126**: 251–273.

Nair RD, Thomas SJ, Loft RD. 2005a. A discontinuous Galerkin global shallow-water model. *Mon. Weather Rev.* **133**: 876–888.

Nair RD, Thomas SJ, Loft RD. 2005b. A discontinuous Galerkin transport scheme on the cubed sphere. *Mon. Weather Rev.* **133**: 814–828.

Neta B, Williams RT. 1989. Rossby wave frequencies and group velocities for finite-element and finite-difference approximations to the vorticity-divergence and primitive forms of the shallow-water equations. *Mon. Weather Rev.* **117**: 1439–1457.

Ničković S. 1994. On the use of hexagonal grids for simulation of atmospheric processes. *Contrib. Atmos. Phys.* **67**: 103–107.

Ničković S, Gavrilov M, Tošić I. 2002. Geostrophic adjustment on hexagonal grids. *Mon. Weather Rev.* **22**: 668–683.

Pain CC, Piggott MD, Goddard AJH, Fang F, Gorman GJ, Marshall DP, Eaton MD, Power PW, de Oliveira CRE. 2005. The numerical solution of parabolic and elliptic differential equations. *Ocean Modelling* **10**: 5–33.

Peng XD, Xiao F, Takahashi K. 2006. Conservative constraint for a quasi-uniform overset grid on the sphere. *Q. J. R. Meteorol. Soc.* **132**: 979–996.

Phillips NA. 1957. A map projection system suitable for large-scale numerical weather prediction. *J. Meteorol. Soc. Japan* **35**: (75th anniversary issue) 262–267.

Phillips NA. 1959. Numerical integration of the primitive equations on the hemisphere. *Mon. Weather Rev.* **87**: 333–345.

Phillips NA. 1962. 'Numerical integration of the hydrostatic system of equations with a modified version of the Eliassen finite-difference grid'. In *Proceedings of Internat. Symposium on NWP*, Tokyo. 109–120.

Piggott MD, Pain CC, Gorman GJ, Power PW, Goddard AJH. 2005. h, r, and hr adaptivity with applications in numerical ocean modelling. *Ocean Modelling* **10**: 95–113.

Purser RJ, Rančić M. 1997. Conformal octagon: An attractive framework for global models offering quasi-uniform regional enhancement of resolution. *Meteorol. Atmos. Phys.* **62**: 33–48.

Purser RJ, Rančić M. 1998. Smooth quasi-homogeneous gridding of the sphere. *Q. J. R. Meteorol. Soc.* **124**: 637–647.

Putman WM, Lin S-J. 2007. Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.* **227**: 55–78.

Qaddouri A. 2011. Nonlinear shallow-water equations on the Yin-Yang grid. *Q. J. R. Meteorol. Soc.* **137**: 810–818.

Qaddouri A, Lee V. 2011. The Canadian Global Environmental Multiscale model on the Yin-Yang grid system. *Q. J. R. Meteorol. Soc.* **137**: 1913–1926, DOI: 10.1002/qj.873.

Qaddouri A, Laayouni L, Loisel S, Côté J, Gander MJ. 2008. Optimized Schwarz methods with an overset grid for the shallow-water equations: Preliminary results. *Appl. Numer. Math.* **58**: 459–471.

Qaddouri A, Pudykiewicz J, Tanguay M, Girard C, Côté J. 2011. Experiments with different discretizations for the shallow-water equations on the sphere. *Q. J. R. Meteorol. Soc.* DOI: 10.1002/qj.976.

Rančić M, Purser RJ, Mesinger F. 1996. A global shallow-water model using an expanded spherical cube: Gnomonic versus conformal coordinates. *Q. J. R. Meteorol. Soc.* **122**: 959–982.

Randall DA. 1994. Geostrophic adjustment and the finite-difference shallow-water equations. *Mon. Weather Rev.* **122**: 1371–1377.

Randall DA, Ringler T, Heikes R, Jones P, Baumgardner J. 2002. Climate modeling with spherical geodesic grids. *Comput. Sic. Eng.* **4**: 32–41.

Ringler T, Randall DA. 2002a. A potential enstrophy and energy conserving numerical scheme for solution of the shallow-water equations on a geodesic grid. *Mon. Weather Rev.* **130**: 1397–1410.

Ringler T, Randall DA. 2002b. The ZM grid: An alternative to the Z grid. *Mon. Weather Rev.* **130**: 1411–1422.

Ringler TD, Heikes R, Randall DA. 2000. Modelling the atmospheric general circulation using a spherical geodesic grid: a new class of dynamical cores. *Mon. Weather Rev.* **128**: 2471–2490.

Ringler TD, Thuburn J, Klemp JB, Skamarock WC. 2010. A unified approach to energy and potential vorticity dynamics for arbitrarily structured C-grids. *J. Comput. Phys.* **229**: 3065–3090.

Ringler TD, Jacobsen D, Gunzberger M, Ju L, Duda M, Skamarock WC. 2011. Exploring a multi-resolution modeling approach within the shallow-water equations. *Mon. Weather Rev.* **139**: in press.

Ritchie H, Tanguay M. 1996. A comparison of spatially averaged Eulerian and semi-Lagrangian treatments of mountains. *Mon. Weather Rev.* **124**: 167–181.

Rivest C, Staniforth A, Robert A. 1994. Spurious resonant response of semi-Lagrangian discretizations to orographic forcing: Diagnosis and solution. *Mon. Weather Rev.* **122**: 366–376.

Ronchi C, Iacono R, Paolucci PS. 1996. The cubed sphere: A new method for the solution of partial differential equations in spherical geometry. *J. Comput. Phys.* **124**: 93–114.

Rossmanith JA. 2006. A wave propagation method for hyperbolic systems on the sphere. *J. Comput. Phys.* **213**: 629–658.

Sadourny R. 1972. Conservative finite-differencing approximations of the primitive equations on quasi-uniform spherical grids. *Mon. Weather Rev.* **100**: 136–144.

Sadourny R, Arakawa A, Mintz Y. 1968. Integration of the nondivergent barotropic vorticity equation with an icosahedral-hexagonal grid for the sphere. *Mon. Weather Rev.* **96**: 351–356.

Sasaki YK. 1976. Variational design of finite-difference schemes for initial value problems with an integral constraint. *J. Comput. Phys.* **21**: 270–278.

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.

Schmidt F. 1977. Variable fine mesh in a spectral global model. *Beitr. Phys. Atmos.* **50**: 211–217.
Schneider EK. 1987. An inconsistency in vertical discretization in some atmospheric models. *Mon. Weather Rev.* **115**: 2166–2169.
Schoenstadt AL. 1980. A transfer function analysis of numerical schemes used to simulate geostrophic adjustment. *Mon. Weather Rev.* **108**: 1248–1259.
Slingo J, Bates K, Nikiforakis N, Piggott M, Roberts M, Shaffrey L, Stevens I, Vidale PL, Weller H. 2009. Developing the next-generation climate system models: Challenges and achievements. *Phil. Trans. R. Soc. A* **367**: 815–831.
St-Cyr A, Jablonowski C, Dennis JM, Tufo HM, Thomas SJ. 2008. A comparison of two shallow-water models with non-conforming adaptive grids. *Mon. Weather Rev.* **136**: 1898–1922.
Staniforth A. 1997. André Robert (1929–1993): His pioneering contributions to numerical modelling. In *Numerical Methods in Atmospheric and Oceanic Modelling*, the André J. Robert memorial volume. Lin CA, Laprise R, Ritchie H. (eds) CMOS/NRC Press: Ottawa, Canada. 25–54.
Staniforth A, Côté J. 1991. Semi-Lagrangian integration schemes for atmospheric models – A review. *Mon. Weather Rev.* **119**: 2206–2223.
Staniforth A, Daley RW. 1979. A baroclinic finite element model for regional forecasting with the primitive equations. *Mon. Weather Rev.* **107**: 107–121.
Staniforth A, Mitchell HL. 1977. A semi-implicit finite-element barotropic model. *Mon. Weather Rev.* **105**: 154–169.
Staniforth A, Mitchell HL. 1978. A variable resolution finite-element technique for regional forecasting with the primitive equations. *Mon. Weather Rev.* **106**: 439–447.
Staniforth A, Wood N. 2008. Aspects of the dynamical core of a non-hydrostatic, deep-atmosphere, unified weather and climate prediction model. *J. Comput. Phys.* **227**: 3445–3464.
Staniforth AN, Williams RT, Neta B. 1993. Influence of linear depth variation on Poincaré, Kelvin, and Rossby waves. *J. Atmos. Sci.* **50**: 929–940.
Starius G. 1980. On composite mesh difference methods for hyperbolic differential equations. *Numer. Math.* **35**: 241–255.
Stuhne GR, Peltier WR. 1999. New icosahedral grid-point discretizations of the shallow-water equations on the sphere. *J. Comput. Phys.* **148**: 23–58.
Swinbank R, Purser RJ. 2006. Fibonacci grids: A novel approach to global modelling. *Q. J. R. Meteorol. Soc.* **132**: 1769–1793.
Tanguay M, Yakimiw E, Ritchie H, Robert A. 1992. Advantages of spatial averaging in semi-implicit semi-Lagrangian schemes. *Mon. Weather Rev.* **120**: 113–123.
Taylor MA, Fournier A. 2010. A compatible and conservative spectral element method on unstructured grids. *J. Comput. Phys.* **229**: 5879–5895.
Taylor MA, Tribbia J, Iskandarani M. 1997. The spectral element method for the shallow-water equations on the sphere. *J. Comput. Phys.* **130**: 92–108.
Temperton C, Hortal M, Simmons A. 2001. A two-time-level semi-Lagrangian global spectral model. *Q. J. R. Meteorol. Soc.* **127**: 111–128.
Thomas SJ, Loft RD. 2000. Parallel semi-implicit spectral element methods for atmospheric general circulation models. *J. Sci. Comput.* **15**: 499–518.
Thomas SJ, Loft RD. 2002. Semi-implicit spectral element method for the shallow-water equations on the sphere. *J. Sci. Comput.* **17**: 339–350.
Thomas SJ, Loft RD. 2005. The NCAR spectral element climate dynamical core: Semi-implicit Eulerian formulation. *J. Sci. Comput.* **25**: 307–322.
Thuburn J. 1997. A PV-based shallow-water model on a hexagonal-icosahedral grid. *Mon. Weather Rev.* **125**: 2328–2347.
Thuburn J. 2007. Rossby wave dispersion on the C-grid. *Atmos. Sci. Lett.* **8**: 37–42.
Thuburn J. 2008a. Numerical wave propagation on the hexagonal C-grid. *J. Comput. Phys.* **227**: 5836–5858.
Thuburn J. 2008b. Some conservation issues for the dynamical cores of NWP and climate models. *J. Comput. Phys.* **227**: 3715–3730.
Thuburn J, Li Y. 2000. Numerical simulations of Rossby–Haurwitz waves. *Tellus* **52A**: 181–189.
Thuburn J, Staniforth A. 2004. Conservation and linear Rossby-mode dispersion on the spherical C-grid. *Mon. Weather Rev.* **132**: 641–653.
Thuburn J, Ringler TD, Skamarock WC, Klemp JB. 2009. Numerical representation of geostrophic modes on arbitrarily structured C-grids. *J. Comput. Phys.* **228**: 8321–8335.
Tolstykh MA. 2003. Variable resolution global semi-Lagrangian atmospheric model. *Russian J. Numer. Anal. Math. Modelling* **18**: 347–361.
Tomita H, Satoh M. 2004. A new dynamical framework of nonhydrostatic global model using icosahedral grid. *Fluid Dyn. Res.* **34**: 357–400.
Tomita H, Tsugawa M, Satoh M, Goto K. 2001. Shallow-water model on a modified icosahedral geodesic grid by using spring dynamics. *J. Comput. Phys.* **174**: 539–613.
Tomita H, Satoh M, Goto K. 2002. An optimization of the icosahedral grid modified by spring dynamics. *J. Comput. Phys.* **183**: 307–331.
Ullrich PA, Lauritzen PH, Jablonowski C. 2009. Geometrically Exact Conservative Remapping (GECoRe): Regular latitude–longitude and cubed-sphere grids. *Mon. Weather Rev.* **137**: 1721–1741.
Ullrich PA, Jablonowski C, van Leer B. 2010. High-order finite-volume methods for the shallow-water equations on the sphere. *J. Comput. Phys.* **229**: 6104–6134.
Wajsowicz RC. 1986. Free planetary waves in finite-difference numerical models. *J. Phys. Ocean.* **16**: 773–789.
Walko RL, Avissar R. 2008a. The Ocean–Land–Atmosphere Model (OLAM). Part I: Shallow-water tests. *Mon. Weather Rev.* **136**: 4033–4044.
Walko RL, Avissar R. 2008b. The Ocean–Land–Atmosphere Model (OLAM). Part II: Formulation and tests of the nonhydrostatic dynamic core. *Mon. Weather Rev.* **136**: 4045–4062.
Walters RA, Carey GF. 1983. Analysis of spurious oscillation modes for the shallow-water and Navier-Stokes equations, *Computers and Fluids* **11**: 51–68.
Wang H, Tribbia JJ, Baer F, Fournier A, Taylor MA. 2007. A spectral element version of CAM2. *Mon. Weather Rev.* **135**: 3825–3840.
Weller H, Weller HG. 2008. A high-order arbitrarily unstructured finite-volume model of the global atmosphere: Tests solving the shallow-water equations. *Int. J. Num. Meth. Fluids* **56**: 1589–1596.
Weller H, Weller HG, Fournier A. 2009. Voronoi, Delaunay and block structured mesh refinement for solution of the shallow-water equations on the sphere. *Mon. Weather Rev.* **137**: 4208–4224.
Weller H, Ringler T, Piggott M, Wood N. 2010. Challenges facing adaptive mesh modeling of the atmosphere and ocean. *Bull. Amer. Meteorol. Soc.* **91**: 105–108.
White AA. 1997. Plato, polyhedra and weather forecasting. In *Mathematics masterclasses. Stretching the imagination.* Sewell M. (ed.) Oxford University Press: Oxford, UK. 194–215.
Wiin-Nielsen A. 1976. On geostrophic adjustment on the sphere. *Beitr. Phys. Atmos.* **49**: 254–271.
Williams RT. 1981. On the formulation of finite-element prediction models. *Mon. Weather Rev.* **109**: 463–466.
Williamson DL. 1968. Integration of the barotropic vorticity equation on a spherical geodesic grid. *Tellus* **20**: 642–653.
Williamson DL. 1969. Numerical integration of fluid flow over triangular grids. *Mon. Weather Rev.* **97**: 885–895.
Williamson DL. 1970. Integration of the primitive barotropic model over a spherical geodesic grid. *Mon. Weather Rev.* **98**: 512–520.
Williamson DL. 1971. A comparison of first- and second-order difference approximations over a spherical geodesic grid. *J. Comput. Phys.* **7**: 301–309.
Williamson DL. 1997. Climate simulations with a spectral, semi-Lagrangian model with linear grids. In *Numerical Methods in Atmospheric and Oceanic Modelling*, The André J. Robert memorial volume. Lin CA, Laprise R, Ritchie H. (eds). CMOS/NRC Press: Ottawa, Canada. 279–292.
Williamson DL. 2007. The evolution of dynamical cores for global atmospheric models. *J. Meteorol. Soc. Japan* **85B**: 241–269.
Williamson DL, Drake JB, Hack JJ, Jakob R, Swarztrauber PN. 1992. A standard test set for numerical approximations to the shallow-water equations in spherical geometry. *J. Comput. Phys.* **102**: 211–224.
Winninghoff FJ. 1968. 'On the adjustment toward a geostrophic balance in a simple promitive-equation model with application to the problem of initialization and objective analysis'. PhD thesis. Department of Meteorology: University of California, Los Angeles.
Yang C, Cao JW, Cai XC. 2010. A fully implicit domain decomposition algorithm for shallow-water equations on the cubed-sphere. *SIAM J. Sci. Comput.* **32**: 418–438.
Zerroukat M, Wood N, Staniforth A. 2004. SLICE-S: A Semi-Lagrangian Inherently Conserving and Efficient scheme for transport problems on the Sphere. *Q. J. R. Meteorol. Soc.* **130**: 2649–2664.
Zhang R-H, Shen X-S. 2008. On the development of the GRAPES – A new generation of the national operational NWP system in China. *Chinese Sci. Bull.* **53**: 3429–3432.