Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 227 (2008) 3445–3464
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Aspects of the dynamical core of a nonhydrostatic, deep-atmosphere, unified weather and climate-prediction model ☆

Andrew Staniforth *, Nigel Wood

Met Office, FitzRoy Road, Exeter, Devon EX1 3PB, UK

Received 18 July 2006; received in revised form 1 November 2006; accepted 3 November 2006
Available online 28 December 2006

Abstract

The dynamical core, which governs the evolution of resolved fluid-dynamical processes, is a critical element of any atmospheric model. Its governing equations must include all relevant dynamical terms, and the numerical formulae used to approximate them must be accurate, stable and efficient. This is particularly so in a unified modeling environment in which the same dynamical core is used for both operational weather prediction and long term climate simulations.

Recent research at the Met Office on unified dynamical core issues is reviewed. Aspects covered include: properties of various equation sets; vertical coordinates; semi-Lagrangian advection and conservation; trajectory computation and dynamical equivalence; horizontal and vertical discretization; and coupling of physical parameterizations to a dynamical core.

Keywords: Normal modes; NWP; Physics-dynamics coupling; Semi-implicit; Semi-Lagrangian; Time-staggered

1. Introduction

Numerical meteorological models are run at large scale (either global or regional) primarily for two purposes: (a) operational weather forecasting (numerical weather prediction (NWP)); and (b) long range climate simulation. Historically, different types of models have been used for each of these applications even within the same institution. Since the start of numerical meteorology (the first operational forecasts were produced just over 50 years ago, see e.g. [1] for a review) the subject has matured significantly and the available computing power has increased dramatically. Additionally, the institutions producing operational forecasts and climate predictions have come under increasing pressure to make economies of scale and to operate as efficiently as possible. This has led to the development of two strategies: either community models or, particularly for

* Corresponding author. Tel.: +44 1392 884033; fax: +44 1392 885681.
E-mail address: Andrew.Staniforth@metoffice.gov.uk (A. Staniforth).

doi:10.1016/j.jcp.2006.11.009

operational centers, unified models. In a unified modeling environment the same model is used for both operational weather prediction (both global and regional) and long term climate modeling.

The Met Office has had a unified approach to NWP and climate modeling, over a broad range of spatial and temporal scales, since the early 1990's [2]. This approach, though economic in terms of maintenance and transferability of expertise, imposes severe constraints: while small errors in conservation of, for example, the mass of dry air, are typically insignificant for regional forecasts of a day or two, the cumulative effect in a climate simulation of several hundred years can be devastating! Additionally, approximations (such as the hydrostatic approximation – see Section 2.1) that perform well at global and synoptic scales are no longer applicable at scales of a few kilometers. It is this kilometer scale that many operational centers are increasingly aspiring to resolve since this is seen as a primary means to obtain accurate prediction of extreme events. Also the numerics of the model must perform robustly and appropriately accurately with timesteps and spatial scales, which vary by two orders of magnitude. This is a significant constraint in an operational setting.

The dynamical core, which governs the evolution of resolved fluid-dynamical processes, is a critical element of any NWP or climate-simulation model. Essential to its performance is the form of the continuous governing equations and the numerical formulae used to approximate them.

The continuous equations therefore need to be written in a form that allows the numerical approximation to be accurate but also highly efficient in order to meet operational computational schedules. In addition, the numerical formulae must be numerically stable for long time integrations. Improving the efficiency of the dynamical core while maintaining, or even improving, accuracy allows optimization of the model's resolution for given computing time. This aspect will be critical to the use of higher and higher resolution models (the kilometer scale issue discussed above). In addition, the underlying continuous equations possess certain key conservation properties. If the numerical analogue of these equations is to be accurate it should, as far as possible, preserve such properties.

Research at the Met Office has therefore been undertaken into the most appropriate form of the underlying continuous equations, and into the development and exploitation of numerical schemes with improved accuracy, robustness and conservation properties. The approach being taken is to continue to improve, where feasible, the existing dynamical core [3,4] while providing a development path for its future replacement.

Other key elements of a unified modeling system are data assimilation and the parameterization of non-fluid dynamical processes and fluid-dynamical processes that are not resolved by the dynamical core (the so-called physics parameterizations). Both of these topics are discussed elsewhere in this issue. However, it is increasingly recognized that, while improvement of both the dynamical core and the physics parameterizations is necessary, this will only have limited impact on the overall performance of the model unless significant attention is also paid to how those two elements of the model are coupled together numerically. A focus of effort within the Met Office has therefore recently been to attempt to understand what the numerical issues of that coupling are.

This paper is an updated and extended version of one that appeared in the proceedings of an ECMWF workshop [5]. As in that report, and in the interest of limiting the length of the present paper, references to the literature are mostly limited to published papers of recent work carried out at the Met Office. Relevant linkages to the broader literature are however available in the papers cited herein.

The paper is organized as follows. The governing dynamical equations are presented and discussed in Section 2, together with issues regarding the vertical coordinate. Various aspects associated with the temporal discretization of the equations are discussed in Section 3 while those aspects associated with spatial discretization are discussed in Section 4. The problem of coupling the dynamical core with the physics is aired in Section 5 before conclusions are drawn in Section 6.

2. Governing dynamical equations and vertical coordinates

2.1. Equation sets and their properties

When designing a dynamical core, it is necessary to identify an appropriate set of governing dynamical equations for subsequent discretization. The very limited computing capability available in the early days of NWP and climate modeling, together with inefficient explicit time discretization schemes, led to the

Table 1
Equation sets for dynamical cores using the nomenclature of [7]

	Deep	Shallow
Nonhydrostatic	Nonhydrostatic deep equations	Nonhydrostatic shallow
Hydrostatic	Quasi-hydrostatic	Hydrostatic primitive

adoption of various approximations to the fully compressible fluid-dynamical equations. Today, all models still make some kind of approximation to these equations, albeit far fewer than in the past. For example, most current global models are based on the hydrostatic primitive equations. The advantage of this is that vertically-propagating acoustic oscillations are absent via the hydrostatic assumption. This avoids the very restrictive timestep of an explicit time discretization of the acoustic modes. However, it invalidates the equations for many mesoscale flows. Additionally the shallow-atmosphere assumption is made, which is based on simple scaling arguments and implies omitting various Coriolis and metric terms for dynamical consistency. So what equation set should one use for a dynamical core that is applicable at all scales?

It is highly desirable, if not essential, that the equations be dynamically consistent, i.e. that they possess conservation principles for energy, angular momentum and potential vorticity, and have a Lagrangian form of the momentum equation. Four such models are identified in [6,7], which correspond to whether approximations of hydrostatic and shallow-atmosphere type are, or are not, individually made (see Table 1). The case for using the deep-atmosphere equations, i.e. not making the shallow-atmosphere approximation, primarily amounts to the desirability of retaining a complete representation of the Coriolis force, including the $2\Omega\cos\phi$ terms, where Ω is Earth's rotation rate and ϕ is latitude. In [7], it is also shown that: the spherical-geopotential approximation (which is made in all terrestrial atmospheric models of which we are aware) requires neglect of latitudinal variation of apparent gravity; making the shallow-atmosphere approximation implies that apparent gravity should not vary as a function of r, distance from the Earth's center; and, in contradistinction, the deep-atmosphere equations require apparent gravity to vary as $1/r^2$. Additionally, more recent work (A.A. White 2006 private communication) has shown that the non-Euclidean nature of the shallow-atmosphere approximation significantly complicates, at least at a conceptual level, the implementation of the semi-Lagrangian scheme (see Section 3 herein for discussion of some other aspects of the semi-Lagrangian scheme).

The most complete equation set of the quartet considered in [6,7] is the deep nonhydrostatic one. To facilitate later discussion, it is convenient to state this equation set here in the (λ,ϕ,s) coordinate system, where λ and ϕ are longitude and latitude, respectively, and s is a generalized vertical coordinate that is any monotonic function of r. Setting $s\equiv r$ then gives the equations in standard spherical polar coordinates (λ,ϕ,r). Following the derivation given in [8], the resulting dry equations for a rotating spherical deep atmosphere are, in standard notation,

$$\frac{\mathrm{D}u}{\mathrm{D}t}-\frac{uv\tan\phi}{r}+\frac{uw}{r}-2\Omega v\sin\phi+2\Omega w\cos\phi+\frac{1}{\rho r\cos\phi}\left(\frac{\partial p}{\partial\lambda}-\frac{\partial r}{\partial\lambda}\frac{\partial s}{\partial r}\frac{\partial p}{\partial s}\right)=F^u, \tag{1}$$

$$\frac{\mathrm{D}v}{\mathrm{D}t}+\frac{u^2\tan\phi}{r}+\frac{vw}{r}+2\Omega u\sin\phi+\frac{1}{\rho r}\left(\frac{\partial p}{\partial\phi}-\frac{\partial r}{\partial\phi}\frac{\partial s}{\partial r}\frac{\partial p}{\partial s}\right)=F^v, \tag{2}$$

$$\frac{\mathrm{D}w}{\mathrm{D}t}-\frac{(u^2+v^2)}{r}-2\Omega u\cos\phi+g+\frac{1}{\rho}\frac{\partial s}{\partial r}\frac{\partial p}{\partial s}=F^w, \tag{3}$$

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\rho r^2\cos\phi\frac{\partial r}{\partial s}\right)+\left(\rho r^2\cos\phi\frac{\partial r}{\partial s}\right)\mathfrak{D}=0, \tag{4}$$

$$\frac{\mathrm{D}\theta}{\mathrm{D}t}=F^\theta, \tag{5}$$

where

$$p=\rho RT, \tag{6}$$

$$\theta = T\left(\frac{p}{p_0}\right)^{\frac{R}{c_p}}, \tag{7}$$

$$(u, v, w) \equiv \left(r\cos\phi\frac{D\lambda}{Dt}, r\frac{D\phi}{Dt}, \frac{Dr}{Dt}\right), \tag{8}$$

$$\dot{s} \equiv \frac{Ds}{Dt} = \frac{\partial s}{\partial r}\left(w - \frac{\partial r}{\partial t} - \frac{u}{r\cos\phi}\frac{\partial r}{\partial \lambda} - \frac{v}{r}\frac{\partial r}{\partial \phi}\right), \tag{9}$$

$$\mathfrak{D} = \frac{\partial}{\partial \lambda}\left(\frac{u}{r\cos\phi}\right) + \frac{\partial}{\partial \phi}\left(\frac{v}{r}\right) + \frac{\partial \dot{s}}{\partial s}, \tag{10}$$

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \frac{u}{r\cos\phi}\frac{\partial}{\partial \lambda} + \frac{v}{r}\frac{\partial}{\partial \phi} + \dot{s}\frac{\partial}{\partial s}, \tag{11}$$

(F^u, F^v, F^w) and F^θ are any parameterized source/sink terms, $g \equiv d\Phi/dr$ is the gravitational acceleration, $\Phi = \Phi(r)$ being the geopotential, R is the gas constant, c_p is the specific heat at constant pressure, $\mathfrak{D}$ is a pseudo-divergence, and partial derivatives with respect to λ, ϕ and t are evaluated holding s constant. Eqs. (1)–(8) are respectively: the three components of the momentum equation; the continuity equation for density ρ; the thermodynamic equation for potential temperature θ; the equation of state; the definition of potential temperature in terms of absolute temperature T and pressure p; and the definition of the wind components u, v, w in (λ, ϕ, r) coordinates.

For realistic atmospheric models, various modifications to the above equation set, commensurate with the envisaged application, are required to incorporate moisture and other physical and chemical species. However, the mixing ratios of these various species are small in the Earth's atmosphere. Therefore, their inclusion results in only relatively minor changes to terms involving thermodynamic variables, together with the addition of various transport equations of the form

$$\frac{DX}{Dt} = F^X, \tag{12}$$

where X represents any of several mixing ratios of moisture in various phases and of other included quantities such as aerosols or chemical species.

The strategy for the dynamical core described in [3] is to avoid unnecessary approximations. Accordingly, the above fully compressible, nonhydrostatic equations have been adopted. The vertical acceleration term is retained in (3) (it is dropped in the quasi-hydrostatic equation set, and in the hydrostatic primitive equation set – see Table 1). Furthermore, the shallow atmosphere approximation (a feature of both the nonhydrostatic shallow and the hydrostatic primitive equation sets – again see Table 1) is not made, thereby retaining the $2\Omega\cos\phi$ Coriolis terms together with all metric terms in (1)–(3), while avoiding setting r to the Earth's radius where it appears as a metric factor in various remaining terms. The Met Office currently uses the dynamical core described in [3] as the cornerstone of its Unified Model, which is used for NWP and climate applications over a broad range of spatial scales, from global to mesoscale. It is believed to be the only deep-atmosphere nonhydrostatic dynamical core used to date for operational modeling of the Earth's atmosphere.

Note that although the shallow-atmosphere approximation is not made in this dynamical core, the spherical geopotential approximation, which in particular neglects the explicit representation of the Earth's equatorial bulge, nevertheless remains. Because the Earth's oblateness is of order 0.3%, its neglect is generally believed to be unimportant. A weakness in this argument is that although the Earth's oblateness is small, its impact could conceivably be much larger due to systematic cumulative effects, particularly for climate simulations. The issue thus arises as to the feasibility of putting the well-foundedness of this approximation to the test by relaxing the spherical geopotential approximation in a future Met Office dynamical core. One way of achieving this is to represent geopotential surfaces as oblate spheroids. Recent work by A.A. White (personal communication) indicates that this could be achieved within the Met Office's dynamical core without incurring prohibitive computational cost.

Normal-mode analysis provides further insight into the properties and validity of governing equation sets, e.g. by identifying which terms should, for computational stability, be treated implicitly in time. Until recently, such studies made the shallow-atmosphere approximation and neglected the vertical variation of gravity. However these two constraints are relaxed in [9,10]. For terrestrial parameters, neither constraint has significant impact on the spatial form of the energetically significant components of most normal modes, with only slight changes (less than 1%) in frequency [9]. However, relaxing the shallow-atmosphere approximation does lead to significant changes in the tropical structure of long-zonal-wavelength internal acoustic modes, primarily due to the presence of the $2\Omega\cos\phi$ Coriolis terms – this could be important in the presence of forcing, e.g. due to tropical convection. Relaxing the shallow-atmosphere approximation also leads to nonzero vertical velocity and potential temperature fields for external acoustic and Rossby modes; in contrast, these fields are identically zero when the shallow-atmosphere approximation is made. Inclusion of realistic vertical variation in the gravitational acceleration leads to a small but systematic decrease in the magnitude of normal mode frequencies, with the largest differences found being less than 1.5%.

Further insight into the role of the $2\Omega\cos\phi$ terms, associated with the horizontal component of the Earth's rotation vector and a deep atmosphere, is provided in [10] by deriving normal modes for an f–F plane, where $f \equiv 2\Omega\sin\phi$ and $F \equiv 2\Omega\cos\phi$ are both set constant. A particularly surprising result is that the inclusion of the F terms gives rise, in planar geometry with rigid lower and upper boundaries, to an additional branch to the normal mode dispersion relation. The additional modes are inertial in character, have frequency very close to f, and have extremely strong vertical tilt. These modes, and the importance of the $2\Omega\cos\phi$ terms for dynamical cores, are further examined in [11–14].

For a finite-difference model to represent well the behaviour of the free atmosphere, it must capture accurately the structures of the analytic normal modes. Therefore, the structures of analytic normal modes can have implications for the choice of prognostic variables and grid staggering. In the absence of other considerations, it is concluded in [9] that density and temperature should be analytically eliminated in favour of pressure and potential temperature as the prognostic thermodynamic variables, since the structure of density and temperature for high vertical wavenumbers would not be accurately captured on either the horizontal velocity levels or the vertical velocity levels. Also, potential temperature and vertical velocity should be staggered in the vertical with respect to the other dynamic prognostic variables, the so-called Charney–Phillips grid.

Following up on this work, a broad range of vertical configurations with different choices of vertical grid staggerings, different choices of prognostic thermodynamic variables, and different coordinate systems, are examined in [15]. Performance is categorized according to a configuration's ability to discretely represent the structure and frequencies of the analytic normal modes of the linearized compressible Euler equations, the categories being: optimal; near optimal; existence of a single zero-frequency computational mode; existence of inertial frequency decoupled modes; and existence of other, serious, miscellaneous deficiencies. It is concluded that: (a) heuristic arguments, such as the amount of averaging and coarse differencing, and the arguments given in [9], and summarized above, are useful guides to whether a particular configuration is optimal or not; (b) the number of degrees of freedom in the discretization is an accurate guide to the existence of computational modes; (c) there is only minor sensitivity to whether equations for thermodynamic variables are discretized in advective or flux form; and (d) an accurate representation of acoustic modes is a prerequisite for accurate representation of inertia-gravity modes which, in turn, is a prerequisite for accurate representation of Rossby modes.

Among the height-coordinate configurations examined in [15], one in particular stands out as being particularly good, since it not only has significantly better wave dispersion properties than all the others, but also has no computational modes. This configuration is in fact that argued for in [9] – see above. In the notation of [15], it corresponds to $(w\theta,\ uvp)$ meaning that potential temperature θ and pressure p are the prognostic thermodynamic variables, and p and the horizontal wind components (u, v) are vertically staggered with respect to the vertical velocity component, w, and θ. The mass variable, density ρ, is not a prognostic variable for this configuration. As pointed out in [16], this makes it difficult to formulate an inherently mass-conserving numerical scheme, a highly desirable property particularly for models used for climate simulation. The alternative configuration $(w\theta, uv\rho)$ facilitates mass conservation, has no computational modes and well represents acoustic and inertio-gravity modes, but is sub-optimal since it has the undesirable property of significantly retarding higher internal Rossby modes. This leads to the apparent dilemma of either giving up inherent mass conservation, or accepting sub-optimal performance. What to do? In [16], it is shown that the performance of the sub-optimal $(w\theta, uv\rho)$ configuration can be made optimal by an appropriate modification of the discretization of the pressure gradient term, resulting in a mass-conserving optimal configuration that resolves this dilemma.

The principal tool used to develop approximate equation sets, and to assess their validity as a function of flow regime, has been scale analysis which has proven quite subtle to apply. In [17], it is shown that normal-mode analysis provides a useful complementary tool for assessing the validity of various anelastic, hydrostatic and pseudo-incompressible equation sets for both small- and large-scale flows, and leads to the following conclusions. While of key importance for small-scale theoretical studies and process modeling, the anelastic equations are not recommended for either operational NWP or climate simulation at any scale. The pseudo-incompressible set [18] appears to be viable for NWP, but only at short horizontal scales since, at large horizontal scales, the frequencies of deep gravity modes are distorted. For global nonhydrostatic modeling, only the fully compressible equations appear suitable. Advances in numerical techniques in the past decade or so (e.g. [19,3]), allow these to be integrated in a computationally efficient manner. Note however that stability with implicit and semi-implicit timestepping and long timesteps is achieved by spuriously retarding the fast-propagating modes responsible for the timestep limitations of explicit schemes. Should fast-propagating oscillations carry non-negligible energy for a given application, then the timestep would have to be shortened in order to properly represent the physics, and the timestep advantage over an explicit treatment might then be lost.

2.2. Vertical coordinates

Once an equation set for a dynamical core has been chosen, the next issue to be addressed is a suitable choice of vertical coordinate. The hydrostatic primitive equations were reviewed and analyzed in [20] using a generalized vertical coordinate, defined to be any variable which is a single-valued monotonic function of geometric height. This influential review has proven to be a valuable reference, much cited by atmospheric modellers, but is not directly applicable to nonhydrostatic equation sets nor to deep-atmosphere ones. Consequently, the analysis of [20] is extended in [8] by: relaxing the hydrostatic and shallow-atmosphere assumptions; no longer constraining the upper boundary to be a coordinate surface, to permit more general upper boundary conditions; and, in addition to examining the energetics, also examining axial angular momentum conservation to determine its sensitivity to the choice of upper boundary condition. This leads to a formulation, summarized in (1)–(11), of the deep-atmosphere nonhydrostatic Euler equations using a generalized vertical coordinate. It includes, as a special case, the formulation of the Met Office's new dynamical core [3] in a height-based terrain-following coordinate.

It is found for a generalized vertical coordinate that the implied principles of energy and axial angular momentum conservation (in the absence of zonal mechanical forcing and mountain torque) depend on the form of the upper boundary. In particular, for a modeled atmosphere of finite extent, global energy conservation is only obtained for a rigid lid, fixed in space and time. To additionally conserve global axial angular momentum, the height of the lid cannot vary with longitude. This result has been shown to be independent of whether the atmosphere is shallow or deep, and hydrostatic or nonhydrostatic. In particular, models that impose a material surface with constant (non-zero) pressure at the upper boundary, do not conserve total energy and axial angular momentum, although they may possess an energy-like invariant (see Section 2.3). This is consistent with, and generalizes, the analysis in [20] for a shallow hydrostatic atmosphere. There it was demonstrated that total energy is conserved for a rigid lid in height coordinates. However it is not generally conserved for an isobaric lid in pressure coordinates, but it was shown that a pseudo-energy invariant exists instead.

Today's atmospheric models are usually formulated in terms of terrain-following coordinates. [However, such coordinate systems are not without their problems particularly at high resolution near steep mountains.] For shallow-atmosphere hydrostatic models it is natural and convenient to use pressure as the vertical coordinate (i.e. an isobaric coordinate), which has the advantage of making the continuity equation a diagnostic relation. A family of ''terrain-following hydrostatic-pressure'' coordinates, also referred to as mass coordinates, is introduced in [21]. This approach is valid for nonhydrostatic shallow atmospheres, retains the acoustic modes, and leads to a diagnostic continuity equation with no need for any approximations other than those of a shallow atmosphere. A simple example is

$$s = \eta \equiv \frac{\pi - \pi_T}{\pi_S(\mathbf{x}) - \pi_T}, \tag{13}$$

where

$$\pi(\mathbf{x}, z, t) \equiv \pi_T + \int_z^{z_T} g\rho(\mathbf{x}, z', t)\, dz' \left(\Rightarrow \frac{\partial \pi}{\partial z} = -\rho g \right) \tag{14}$$

is "hydrostatic pressure", z is geometric height, $\mathbf{x}$ is the horizontal position vector, and π_T is constant. For a hydrostatic shallow atmosphere, the η coordinate (13) reduces to the traditional sigma-coordinate $\sigma = (p - p_S)/(p_S - p_T)$, since π reduces to total pressure p in the hydrostatic limit.

Using the analysis of [8], it is shown in [22] that the terrain-following coordinate η of [21], based on hydrostatic pressure π for the shallow-atmosphere Euler equations, can be generalized to the deep-atmosphere Euler equations. For a shallow atmosphere, the terrain-following coordinate of [21] can be interpreted equivalently as being either based on "hydrostatic pressure" or on mass. However the generalization to deep atmospheres is such that the analogous quantity is based on mass and not on pressure. This is because, for a deep atmosphere, the cross-sectional areal element increases with height, whereas it is constant for a shallow atmosphere, resulting in different volume elements. A consequent benefit of the mass-based generalization given in [22] of the coordinate of [21] is that an existing (shallow-atmosphere) hydrostatic primitive-equations model which uses a pressure-based terrain-following vertical coordinate, could be modified for nonhydrostatic deep-atmosphere applications, without the need to substantially change the scientific and computing infrastructure in which it is embedded.

2.3. Energetics

As mentioned above, in the absence of external momentum and thermal forcing and net surface torque, it is shown in [8] that deep (and shallow) atmospheres of finite extent are only guaranteed to globally conserve energy and axial angular momentum if a rigid lid upper boundary condition is applied. Otherwise energy is not conserved. Elastic isobaric upper lids (where the lid is a specified isobaric, i.e. constant pressure, surface) are however popular and have merit. The presence of global invariants provides a constraint on the system which can be useful when designing effective numerical schemes. An energy-like invariant is known to exist for *shallow hydrostatic* atmospheres with an elastic isobaric lid. In [23], a generalization of this invariant is derived. It is $(E + p_T/\rho)$, where $p_T = p_T(\lambda, \phi)$ is the pressure at the lid, and ρ is density. This energy-like invariant: (a) is valid independently of whether the atmosphere is assumed *deep* or *shallow*, and *hydrostatic* or *non-hydrostatic*; and (b) subsumes previous shallow-atmosphere energy-like invariants in the atmospheric literature. Note that while $\rho E + p_T$ *is* globally conserved (in the sense that its volume integral is an invariant of the system) with an elastic isobaric lid, the true total energy ρE of the system is *not* conserved. The difference between the two, viz. the contribution of $p_T(\lambda, \phi)$, represents the work done by the stationary pressure applied at the upper surface as the height of that surface changes.

An important practical issue, raised but left unanswered in [23], is whether it is better to impose a rigid or an elastic lid for an atmosphere of finite extent. When the atmosphere undergoes heating, is it better to consider that this is done at constant pressure or at constant volume? This remains a moot point.

3. Temporal discretization of the dynamical equations

When forecasting the state of the atmosphere, it is essential to use efficient discretization methods. The simplest time discretizations, e.g. the well-known leapfrog time scheme used in the early days of atmospheric modeling, are explicit. Their timestep is restricted, for reasons of computational stability, by the speed of the fastest-propagating modes. Normal mode analysis – see discussion in Section 2.1 – reveals that the fastest modes of propagation are acoustic waves and horizontally propagating external gravity waves, with propagation speeds of the order of 350 m s^{-1}. Although these modes carry little energy, they are the ones that most seriously restrict the timestep length, and hence computational efficiency. The restrictions are particularly severe for global finite-difference or finite-element methods that use a latitude–longitude grid, since the convergence of the meridians at the poles results in very small meshlengths, and consequently in very small timesteps.

The essence of the semi-implicit (SI) method – see [24] for a historical review of its development – is to treat in a time-implicit manner the linear part of the terms responsible for the fastest-propagating modes (as identified by normal-mode analysis), and to treat the other terms (including nonlinear perturbations) in a time-explicit manner. This results in a retardation of the fastest-propagating modes, and also in the overhead of solving an elliptic-boundary-value problem. In practice it is found that this overhead is small compared with the efficiency advantage gained by being able to stably integrate with long timesteps, with negligible loss of accuracy (since the fastest-propagating modes generally carry little energy). This has resulted in the widespread adoption of the SI scheme in atmospheric (and other) models, it being a particularly attractive approach for stiff systems. Recently – see Section 3.2 – an alternative to the SI scheme, the regularized time-staggered scheme, has been proposed and analyzed.

Stability analysis shows – see reviews of [25,24] – that applying an implicit or SI treatment of fast-propagating modes addresses the timestep constraint of these modes. However, if it is coupled with an Eulerian treatment of advection, then the local Courant number ($U\Delta t/\Delta x$) is constrained to be somewhat less than unity. [For a latitude–longitude mesh, this is again particularly restrictive in polar regions due to the convergence of the meridians.] However, analysis also shows that the temporal truncation error is generally still much smaller than the spatial truncation error [26], at least for large- and synoptic-scale flows. The timestep can therefore be increased with negligible loss of accuracy provided that a stable and accurate method to handle advection with long timesteps can be found. This motivates – see [26,27] and the review of [25] – the use of a semi-Lagrangian treatment of advection, which is stable for Courant numbers much greater than unity. Coupling a semi-Lagrangian treatment of advection with a SI treatment of the fastest-propagating modes then results in a scheme where the timestep can be chosen on the basis of accuracy rather than being constrained by stability. Semi-implicit semi-Lagrangian methods are now used in many atmospheric models, e.g. [3,19,28,29].

3.1. Semi-Lagrangian discretization

3.1.1. Semi-Lagrangian advection and conservation

Semi-Lagrangian (SL) schemes are widely used for the advection component of many modern operational atmospheric models due to their increased efficiency and stability compared to Eulerian schemes. However, a common disadvantage of interpolating SL schemes is the lack of mass and tracer conservation. Though mass conservation may not be critical for short period NWP simulations, it is very important for long period simulations such as those of climate studies. Over a long simulation period, the total mass can drift significantly if no correction is applied. Hence, SL schemes which are inherently mass conserving are desirable. The challenge is to not only achieve inherent conservation, but to do so while remaining computationally efficient compared with a traditional interpolating SL scheme. This motivated the development of the semi-Lagrangian inherently-conserving and efficient (SLICE) algorithm [30].

There are two ingredients. The first is to rewrite the Eulerian flux form

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{u}) = 0, \tag{15}$$

where ρ is a scalar field transported by velocity $\mathbf{u}$, in a finite-volume Lagrangian form

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_{\partial V}\rho\,\mathrm{d}V = 0, \tag{16}$$

where ∂V is a fluid parcel or Lagrangian control volume. This equation is then integrated in time to obtain

$$M_{\mathrm{a}}^{n+1} = M_{\mathrm{d}}^{n}, \tag{17}$$

where M_{a}^{n+1} is its mass at time $(n+1)\Delta t$ centered on the arrival location $\mathbf{x}_{\mathrm{a}}$, and M_{d}^{n} its mass at time $n\Delta t$ centered on the departure location $\mathbf{x}_{\mathrm{d}}$. The second, inspired by cascade interpolation [31], is the use of a cascade remapping strategy to very efficiently decompose a multi-dimensional remapping problem (from Eulerian control volumes to Lagrangian ones, or vice-versa) into a number of much-simpler one-dimensional (1D) remapping problems – see [30] for details. An important property of cascade remapping is that it preserves characteristics of the flow, thus minimizing splitting errors. Overall, it is found that in addition to exactly conserving mass, the SLICE algorithm is also competitive with standard non-conserving semi-Lagrangian schemes from the viewpoints of both computational efficiency and accuracy.

This algorithm [30], in planar geometry, is extended to spherical geometry in [32]. It has no restriction on Courant numbers and again achieves comparable, or better accuracy, as standard non-conserving and other published conserving SL schemes over the sphere. The algorithmic complexity has been a major design constraint with the aim of achieving this extension in a straightforward and flexible way without a major computational overhead.

The extension of the SLICE algorithm to allow monotonicity (and positive-definiteness) to be efficiently imposed in both planar and spherical geometry is described in [33]. Monotonicity and positive-definiteness are essential for algorithms that transport physical and chemical atmospheric species. This extension operates by first identifying where monotonicity is violated (the detection stage), and by then locally reducing the order of the piecewise polynomial used in the remapping algorithm until monotonicity is regained (the reduction stage). A global minimum and/or a global maximum can similarly be imposed and positive-definiteness is achieved by setting the global minimum to be zero. The resulting monotonicity scheme is more selective and less damping in the smooth part of the solution than other filters.

For the SLICE schemes discussed above, the 1D remapping is accomplished using a novel piecewise cubic method (PCM) [30]. A somewhat simpler and more efficient 1D remapping algorithm, termed the parabolic spline method (PSM), is described in [34]. Of all piecewise parabolic functions that satisfy a given mass distribution, it is shown that PSM yields an optimal reconstruction since it possesses the minimum norm (or curvature) and the best approximation properties. A truncation error analysis shows that although it has a similar truncation error in the converged limit as that of the widely used piecewise parabolic method (PPM – see [35]) for infinitely differentiable functions, PSM is more accurate than PPM for problems with slow spectral decay, such as those encountered in typical atmospheric modeling applications. Additionally, an operation count shows PSM to be 60% more efficient than PPM.

Illustrative comparative results for the challenging, non-smooth, deformational problem on the sphere – see [33] for its definition and parameters – are displayed in Fig. 1, using SLICE with PSM remapping, without (Fig. 1a) and with (Fig. 1c) monotonicity, and using a conventional bicubic non-conserving SL algorithm (Fig. 1b). The exact result is displayed in Fig. 1d for comparison.

For further examples of applications of SLICE in multiple dimensions see [36,37].

3.1.2. *Trajectories and dynamical equivalence*

A crucial component of any SL scheme is the computation of the trajectories (or displacement vector) via numerical approximation of $\mathrm{D}\mathbf{x}/\mathrm{D}t = \mathbf{u}(\mathbf{x}, t)$, where $\mathbf{x}$ is now the three-dimensional (3D) position vector. This computation has an important influence on the stability and accuracy of the discretization of the governing equations. It is shown by [38] that the departure-point and momentum equations, given respectively by

$$\frac{\mathrm{D}\mathbf{x}}{\mathrm{D}t} = \mathbf{u}(\mathbf{x}, t), \quad \frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} = \mathbf{F}, \tag{18}$$

where $\mathbf{F}$ is the local force per unit mass, analytically convey the same information as the departure-point, angular momentum and scalar-product equations

$$\frac{\mathrm{D}\mathbf{x}}{\mathrm{D}t} = \mathbf{u}(\mathbf{x}, t), \quad \frac{\mathrm{D}(\mathbf{x} \times \mathbf{u})}{\mathrm{D}t} = \mathbf{x} \times \mathbf{F}, \quad \frac{\mathrm{D}(\mathbf{x} \cdot \mathbf{u})}{\mathrm{D}t} = \mathbf{u}^2 + \mathbf{x} \cdot \mathbf{F}. \tag{19}$$

White [38] refers to this property as dynamical equivalence.

The conditions under which this also holds for discrete forms of these equation sets are analyzed in [38,39]. For time-centered discretizations, once a rule has been chosen for the approximation of averages along the trajectories of vector and scalar products, dynamical equivalence implies a particular discrete form of the departure-point equation [38]. For time-decentered discretizations, two types of discretization of the departure-point equation are identified in [39] which preserve dynamical equivalence: some existing discretizations are then found to be approximations to them. It is also shown how to incorporate physical forcings and/or

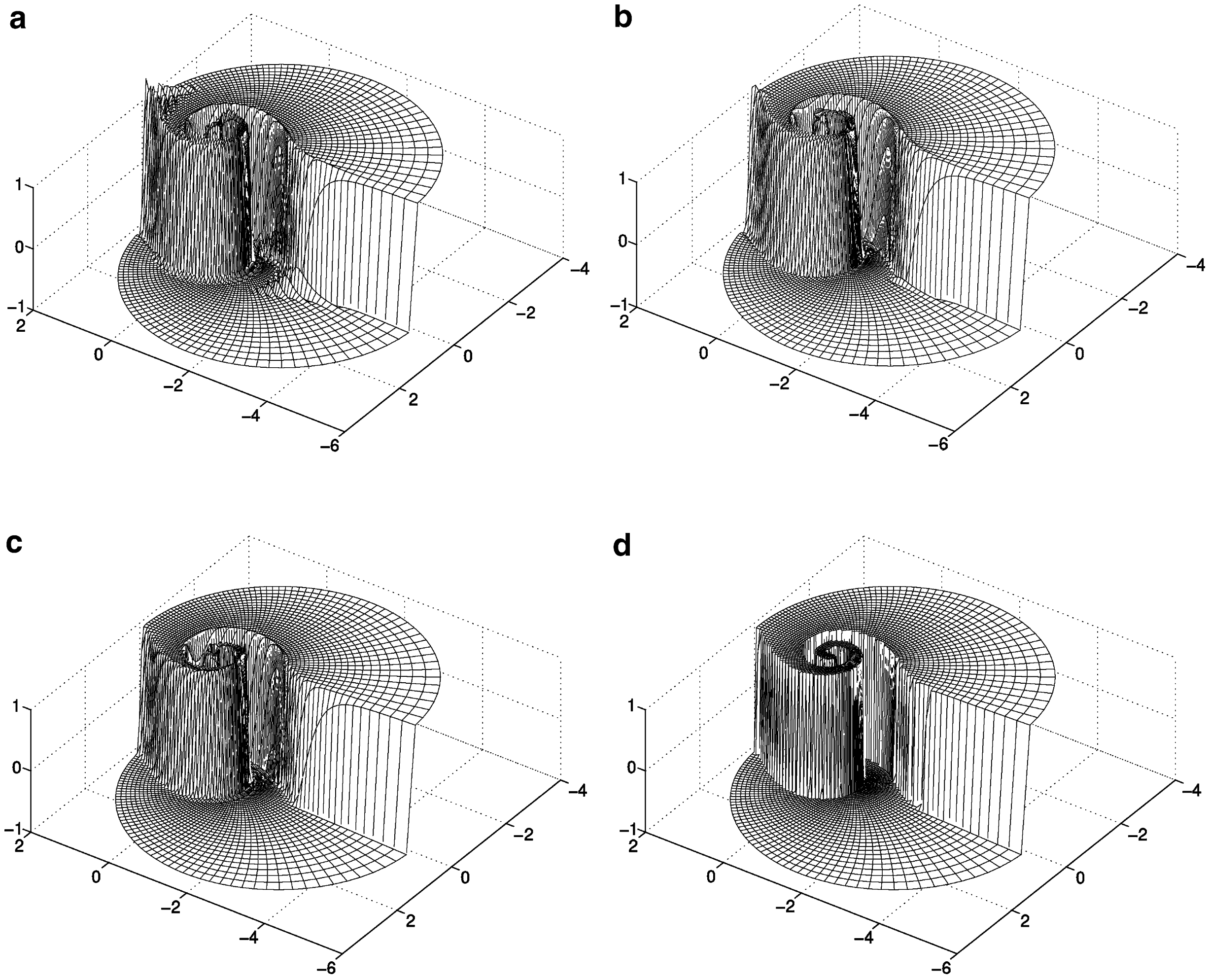

Fig. 1. Solutions, projected on a tangent plane, after 64 timesteps for non-smooth deformational flow on a sphere – see [33] for definition of problem and parameters: (a) SLICE (PSM), *without* monotonicity; (b) bicubic interpolating SL, *without* monotonicity; (c) SLICE (PSM), *with* monotonicity; (d) analytic.

predictor–corrector dynamics into the formulation. Furthermore, two simple model problems (for solid-body rotation and wave motion) are used to provide further insight into the accuracy and stability properties of various departure-point schemes, including: dynamically-equivalent schemes; approximations to these; and several existing schemes. Cordero et al. [40] further analyzed the impact of the trajectory computation on numerical stability – see Section 4.2.2 herein for a summary of this work.

3.2. Regularized time-staggered scheme

A recently proposed alternative to the SI scheme, which preserves unconditional stability of propagating gravity modes, is a time-staggered discretization combined with a regularization of the continuous governing shallow-water equations [41]. In this approach, a "regularized" pressure is determined from the pressure field by solving an elliptic-boundary-value problem, and the regularized pressure is used in place of the pressure in the momentum equation. By time staggering the momentum variables with respect to the pressure variable within a Lagrangian framework, the momentum and continuity equations can then be time discretized in an explicit leapfrog manner. Just as for the SI scheme, the need to solve an elliptic-boundary-value problem is the overhead incurred in order to stably integrate with much larger timesteps than those permitted by explicit time schemes.

To reduce to its essence the formulation of a time-staggered discretization to the continuous regularized equations, consider the linear 1D gravity wave equations

$$u_t = -gh_x, \tag{20}$$

$$h_t = -Hu_x, \tag{21}$$

where g is the constant acceleration due to gravity, and $h(x,t)$ is the perturbation fluid depth about the constant mean fluid surface height H of a resting basic state. The first step, before discretization, is to replace the continuous momentum equation (20) by the regularized momentum equation

$$u_t = -g\tilde{h}_x, \tag{22}$$

where $\tilde{h}$ plays the role of regularized pressure. It satisfies the Helmholtz equation

$$\left(1+\gamma^2-\alpha^2\frac{\partial^2}{\partial x^2}\right)\tilde{h} = h, \tag{23}$$

where α^2 is an arbitrary smoothing length scale, and γ^2 is an additional smoothing parameter. The second step, after regularization, is to discretize (21)–(23) in a time-staggered manner as

$$\frac{h^{n+\frac{1}{2}}-h^{n-\frac{1}{2}}}{\Delta t} = -Hu_x^n, \tag{24}$$

$$\left(1+\gamma^2-\alpha^2\frac{\partial^2}{\partial x^2}\right)\tilde{h}^{n+\frac{1}{2}} = h^{n+\frac{1}{2}}, \tag{25}$$

$$\frac{u^{n+1}-u^n}{\Delta t} = -g\tilde{h}_x^{n+\frac{1}{2}}, \tag{26}$$

where u is defined at integer time levels, h and $\tilde{h}$ at half-integer time levels, and derivatives are evaluated using, for example, centered finite differences or the spectral method. Thus, given the state vector $(h^{n-1/2}, u^n)$, the solution is advanced in time by successively applying (24)–(26) to obtain the new state vector $(h^{n+1/2}, u^{n+1})$.

The effect of the regularization on the *time-continuous* equations is governed by its two arbitrary parameters α and γ. Returning now to the more general case of the rotating 2D shallow-water equations with orography, by examining linear perturbations of the time-continuous equations, the forced response of the time-continuous regularized equations can be shown to be close to that of the unregularized equations provided $\gamma \ll 1$ and $\alpha \ll L_R$, where $L_R \equiv \sqrt{gH}/f$ is the Rossby radius of deformation [41]. Furthermore, and as expected, the free response (the inertia-gravity waves) of the time-continuous regularized equations approaches that of the unregularized ones as $\alpha \to 0$ and $\gamma \to 0$. For non-zero values of α, the inertia-gravity waves are increasingly retarded as their wavenumber increases (reminiscent of the effect on discrete inertia-gravity waves of the SI scheme). Increasing γ away from zero also retards the inertia-gravity waves but the effect, in isolation from α, is independent of the wavenumber.

To determine appropriate values for α and γ, the linear free and forced responses of the time-staggered scheme, applied to the rotating 2D regularized shallow-water equations with orography, are determined in [41]. It is found that numerical stability of the free solution of this scheme requires

$$\alpha \geqslant \frac{\sqrt{gH}\Delta t}{2}. \tag{27}$$

For accuracy, α should be as small as possible (recall that $\alpha \equiv 0$ for the original unregularized equations), thus the optimal choice for the smoothing length is $\alpha = \sqrt{gH}\Delta t/2$ since increasing α beyond this value unnecessarily reduces accuracy. Furthermore, comparing with the corresponding result for the SI scheme, it is found that if α assumes its optimal value and the choice $\gamma = f\Delta t/2$ is made then, surprisingly, the two schemes give exactly the same linear numerical dispersion relation for the *free* response, i.e. for the inertia-gravity waves. Additionally, the regularized time-staggered discretization yields a similar result to the analytic *forced* response (which is exactly captured by the SI discretization) provided $(\alpha/L_R)^2 \ll 1$ and $\gamma^2 \ll 1$. With $\alpha = \sqrt{gH}\Delta t/2$ and $\gamma = f\Delta t/2$, then $\alpha/L_R = \gamma$. The two conditions for validity then reduce to the same requirement, namely that the timestep should be chosen such that $|f\Delta t/2| \ll 1$.

Although this latter condition is usually respected for models of the Earth's atmosphere, it is nevertheless a disadvantage of the time-staggered discretization of the regularized equations with respect to the SI discretization of the unregularized equations. A further, serious, disadvantage is that an initially balanced state of the atmosphere (i.e. a state for which $\nabla \cdot (D\mathbf{u}/Dt)$ vanishes) is not retained, something that is particularly important for data assimilation. These two deficiencies are however addressed in [42], where a revised regularization procedure is introduced that takes into account the forcing terms in the equations, and only impacts the *unbalanced* components of the flow while leaving balanced components untouched. Linear analysis of the discrete equations shows that the solutions are neutrally stable provided the regularization parameter, α, again satisfies (27). The optimal choice, in terms of accuracy and computational efficiency for the discrete equations, is to again choose equality in (27) so that α takes its smallest value permitted for unconditional stability. With this choice for α, used now with $\gamma \equiv 0$, it is found that not only is the free linear response of the time-staggered leapfrog discretization of the regularized equations identical to that of the SI discretization of the unregularized equations, as it is for the original regularized formulation, but so also is the forced linear response.

This linear equivalence, which holds in the absence of advection, suggests that perhaps the time-staggered leapfrog discretization of the equations, with the new regularization, could be used as the basis for a viable alternative to the SI scheme by coupling the time-staggered leapfrog discretization of the regularized equations with a semi-Lagrangian treatment of advection. If so, a potential side benefit would be the potential to simplify and improve the physics/dynamics coupling (see Section 5 for a discussion of related issues). Such a time-staggered semi-Lagrangian (TSSL) discretization of the rotating regularized shallow-water equations, spatially discretized on a staggered Arakawa *C* grid, is therefore proposed and analyzed in [43] (a similar but alternative discretization is given in [44]). The discretization of [43] is second-order accurate in both time and space. A further aspect, crucial for stability reasons, is that the discretizations of the kinematic and momentum equations are tightly, and *implicitly*, coupled when advancing momentum from one timestep to the next. Linear analysis, that now includes uniform advection in the basic state, shows that unconditional stability is achieved provided the regularization parameter α is chosen to have the same value as in the non-advective analysis of [42]. An example application, in the absence of orographic forcing, of the scheme to a fully nonlinear case of two interacting vortices indicates the practical potential of this spatio-temporal discretization [43]. Figs. 2 and 3 (S. Reich, personal communication) show the time evolution, in an *f*-plane shallow-water model, of the potential vorticity due to a pair of interacting vortices. Details of the simulation are given in [43]. Fig. 2 shows the result using the regularized time-staggered discretization with a timestep of 20 min whereas Fig. 3 shows the same results but obtained, in the absence of regularization, using a timestep of only 1 min. As can be seen from the two figures, the two simulations produce very similar results for the PV evolution. Comparison of the regularized, 20 min timestep simulation with a comparable SISL simulation is presented in [43].

An important lesson to learn from the history of the development of semi-implicit semi-Lagrangian (SISL) schemes is that it is important to thoroughly understand not only the free response but also the forced response. The 2D analysis of [45] explained, after the fact, the unexpected noise problems observed in full 3D models in the presence of orography, and they also proposed what has become a widely used solution, namely temporal off-centering. With the benefit of hindsight, this problem, and its solution, could have been anticipated had an appropriate analysis of the forced response of the shallow-water equations been performed *before* the development of full 3D SISL models. It is therefore a natural and necessary next step in the development of TSSL schemes to analyze whether spurious orographic resonance is possible, and this is done in [46].

It is found that the TSSL scheme also shares with the SISL discretization the possibility of spurious orographic resonance. Additionally, such resonance occurs in the TSSL case at a Courant number half as small as in the SISL case. However, a procedure, akin to the off-centering usually employed in the SISL scheme, has been proposed for the TSSL scheme. This consists of adding a term, proportional to the discrete time tendency of the unregularized geopotential field, to the regularization equation. The constant of proportionality, ε, determines how much off-centering is applied. Analysis shows that the off-centered TSSL scheme then remains

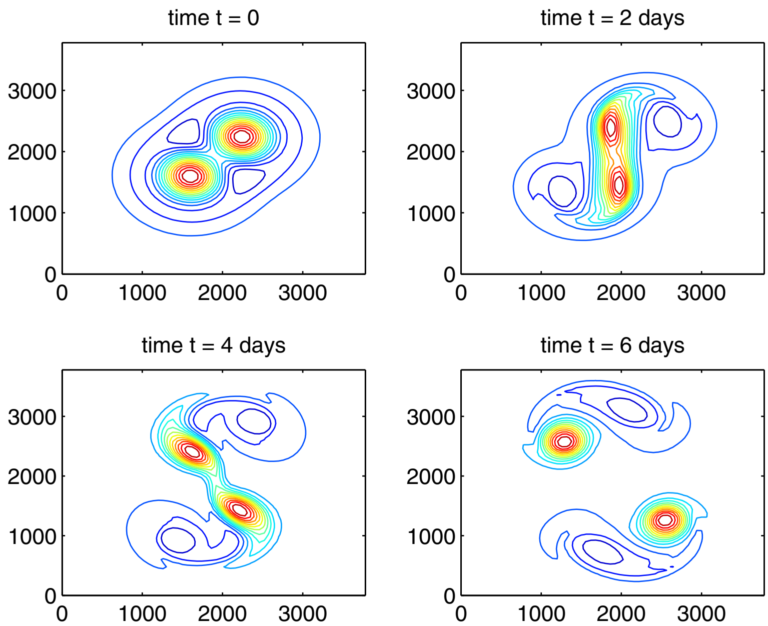

Fig. 2. Time-staggered semi-Lagrangian computation of time evolution of PV field *with* regularization and a timestep of 20 min – see [43] for definition of problem and parameters.

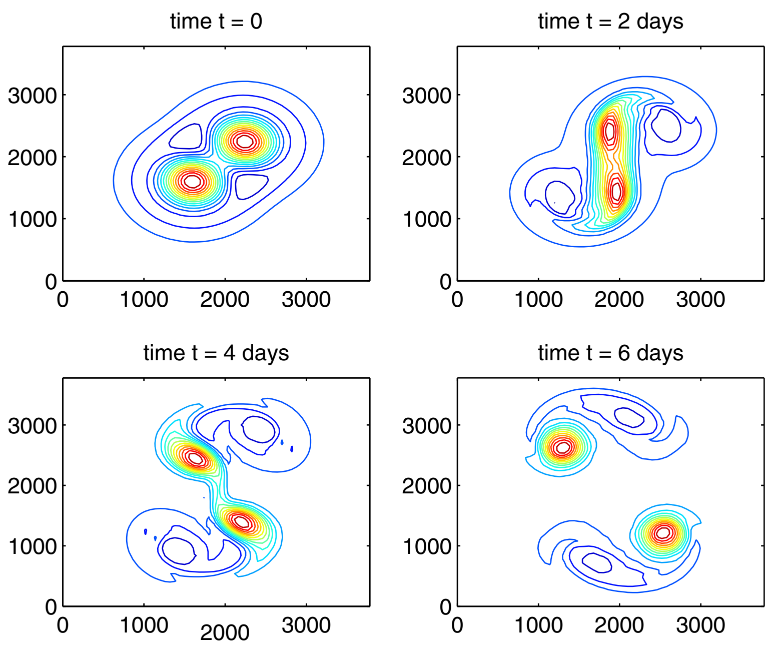

Fig. 3. Same as Fig. 2 except *without* regularization and the timestep is 1 min.

unconditionally stable (with the same appropriate choice for the regularization parameter) provided $\varepsilon \geqslant 0$. Additionally, and crucially, resonance cannot occur provided ε is chosen to be strictly greater than zero. Also, there is no computational mode.

How to extend the TSSL scheme to the fully-compressible equations to not only handle the propagation of gravity modes, as in [43], but also the propagation of acoustic modes, is outlined in [47] and analyzed using normal modes.

4. Spatial discretization of the dynamical equations

4.1. Horizontal discretization

There are many numerical methods for spatial discretization and most have been, or are used, in operational NWP and climate models. These include the spectral method, finite-element methods, finite-volume methods and the closely related finite-difference method [48]. Within the Met Office, finite-difference methods have always been the preferred approach, albeit with some influence from the finite volume approach. Even within that limitation, though, there are many ways of gridding the sphere, including the use of triangles and more quasi-uniform grids such as the conformal cubic grid. However, certainly the simplest approach, yet (with the advent of the semi-Lagrangian scheme) still a competitive one, is a quadrilateral mapping of the sphere based on a latitude–longitude grid (this degenerates to a triangular mapping at the poles). This is the approach currently used in the unified model of the Met Office (and likely to remain so for the immediate future).

4.1.1. Variable horizontal resolution

Like many meteorological organizations, regional forecasting and climate applications at the Met Office currently use a non-interacting or one-way nesting strategy, whereby a lower-resolution configuration of its model provides the lateral boundary conditions for a higher-resolution configuration run over a region of interest. It is well known – e.g. [49] – that there are a number of theoretical and practical problems associated with this approach. It is argued in [28] that it is preferable to use a fully-interacting variable-resolution (stretched grid) strategy, whereby a single global model is integrated with resolution focused on a uniform-resolution sub-domain of a rotated latitude–longitude mesh. Following the successful implementation of this strategy for operational NWP at the Canadian Meteorological Centre [28,50], and its use for regional climate simulation [51], the formulation of the Met Office's dynamical core has been generalized to variable resolution – see [4] for details. Implementation of this generalization is currently underway at the Joint Centre for Mesoscale Meteorology at Reading University.

4.1.2. Conservation and Rossby-mode propagation on the sphere

Spatial discretizations of the linearized shallow-water equations on a spherical C-grid are analyzed in [52]. Constraints are derived therein that ensure conservation of mass, angular momentum and energy, and generalize published results (e.g. [53]) to the case of non-uniform and rotated grids (but restricted to the linearized equations). Grids with either meridional velocity v, or azimuthal velocity u and free-surface height h, stored at the poles are considered. Energy conservation is shown to be problematic for grids with u and h stored at the poles. It is also found that an inappropriate averaging of the Coriolis terms leads to a misrepresentation of the Rossby modes with shortest meridional scale. The appropriate averaging is shown to not only address this problem but to be compatible with the constraints required for conservation.

4.2. Vertical discretization

4.2.1. Analysis of a new finite-element vertical discretization

The choice of vertical discretization method is an important aspect of designing a dynamical core, the most popular being the use of *low*-order *finite differences* with a variety of vertical staggerings of dependent variables. Recently, a new *high*-order *finite-element* (FE) vertical discretization scheme has been proposed for a hydrostatic primitive equation model using a terrain-following pressure-based vertical coordinate and an unstaggered grid [54]. This motivated its mathematical analysis in [55].

The essence of the scheme is an accurate FE algorithm for evaluating vertical integrals using either linear or cubic splines, respectively denoted "linear FE" and "cubic FE" in [54]. They also describe and evaluate two further methods: a finite-difference based scheme; and a "cubic collocation" scheme. The cubic collocation scheme first constructs a cubic-spline interpolant of the integrand and then analytically integrates it. It is observed in [54] that: (a) the empirically estimated truncation errors of the linear FE and cubic collocation schemes are both of fourth order; (b) the measured errors of these two schemes are not only of the same order but identical (to the two significant figures given); and (c) the empirically estimated truncation error of the cubic FE scheme is of eighth order. In addition to explaining these results, the analysis of [55] shows that: (d) the truncation errors of the linear FE and cubic FE schemes applied to the integral form of the equations are respectively four and eight times smaller than those obtained by applying FE's to the differential form; and (e) the cubic FE scheme is formally equivalent for uniform resolution to a new ‘‘heptic collocation’’ scheme, in which a seventh-order spline is analytically integrated.

4.2.2. Discrete normal-mode analysis

Normal modes are fundamental solutions of linearizations of equation sets and are useful in a number of contexts. As mentioned above, they can be used to assess the validity of various equation sets as a function of scale [17], and also to guide the choice of prognostic variables and vertical grid staggering [9]. In [4], spatially continuous normal modes are used to examine the stability properties of the dynamical core of [3] by solving a derived polynomial dispersion relation. Such an approach focuses on the stability characteristics of the time discretization and has the virtue of simplicity, but neglects the impact of the spatial discretization, including non-uniform resolution, and the application of boundary conditions – see [56] for a discussion of possible problems that can be encountered. These neglected aspects may however be included using the framework of matrix stability analysis.

For a given temporal and spatial discretization, the matrix stability analysis proceeds by first linearizing the discretizations about a basic state, and then expressing the resulting set of linear difference equations as $\mathbf{A}\mathbf{x}^{n+1} = \mathbf{B}\mathbf{x}^{n}$. Here the matrices $\mathbf{A}$ and $\mathbf{B}$ together define the vertical discretization; $\mathbf{x}^n \equiv [u, v, w, \theta, \rho, \pi]^{\mathrm{T}}$ is the transpose of the model's discrete state vector at time level n; and $u \equiv [u_1, u_2, \ldots]^{\mathrm{T}}$ is the vector of values of u at the set of discrete vertical levels $z = z_1, z_2, \ldots$, with similar definitions for the other model variables v, w, θ, ρ and π. The generalized discrete eigenproblem $\mathbf{B}\mathbf{x}^n = \lambda\mathbf{A}\mathbf{x}^n$ is obtained by setting $\mathbf{x}^{n+1} \equiv \lambda\mathbf{x}^n$, and the discretization is stable provided $|\lambda| \leqslant 1$.

In [40], this framework is used to assess the impact, in a 1D (column model) version of the dynamical core of [3], of (one- and two-term) extrapolated trajectory schemes on the stability properties of centered SISL schemes. It is found that, in the absence of any controlling mechanism, both extrapolated trajectory schemes are unstable. Additionally they can significantly distort the vertical structure of the acoustic modes. Though not studied there, the analogous distortion of Rossby and gravity waves could be expected to be deleterious to a forecast model. In contrast, an interpolated trajectory scheme is found to be stable and to accurately capture the vertical structures of the normal modes.

5. Physics-dynamics coupling

Physics parameterization packages, which model unresolved fluid-dynamical processes together with non-fluid-dynamical ones, are key elements in the success of numerical weather and climate prediction models. The accuracy and complexity of these schemes continues to increase apace. Similarly, the accuracy of dynamical cores has continued to steadily improve. However, a chain is only as strong as its weakest link, e.g. two second-order components coupled in a first-order manner imply a first-order model. The link that couples the physics package to the inviscid, adiabatic dynamical core has received little attention. It is therefore important for the continued improvement of models that the virtues and vices of the various strategies employed in such coupling are well understood, and that the vices are addressed.

In numerical models a distinction is usually made between fast and slow timescales because of differing stability considerations [57–59]. An explicit time discretization generally has the virtue of simplicity. For a slow timescale process, computational efficiency is usually not hindered by an associated stability-limited timestep and an $\mathrm{O}(\Delta t)$ accurate discretization therefore arguably suffices. However, for a fast timescale process, an explicit time-discretization generally unduly limits the timestep due to an overly-restrictive stability condition. Therefore a more costly implicit time-discretization is usually adopted. Even so, while this can address the stability issue, if the resulting discretization is only $\mathrm{O}(\Delta t)$ accurate, then the timestep may still be unduly limited due to time truncation error [60]. This motivates an $O(\Delta t^2)$-accurate implicit time discretization of fast processes.

However, in a model there are several distinct processes (e.g. the dynamical core and each component of the physics package) each with their own timescale(s). The use of an implicit scheme to solve simultaneously for the time tendency of the complete model, though appealing, is currently prohibitively expensive, at least in an operational setting, and is likely to remain so for the foreseeable future. This is because of the expense of solving a modified Helmholtz problem which consists of contributions from both the dynamics *and* the physics package. The solution is to apply some form of splitting in which the time tendency due to the different elements of a model are evaluated separately, and then combined in some way to generate the complete model tendency. All operational models employ some form of splitting. The problem is that splitting in general introduces errors additional to the truncation errors associated with each individual process. With large timesteps, of the size permitted by SISL schemes, such errors can dominate the model error. The question is therefore: "How to determine the optimal way of performing such splitting?"

A methodology for analyzing the numerical properties of such splitting schemes is developed in [61,62]. A canonical problem is introduced to idealize both the dynamics (with terms to represent both fast and slow propagating modes), and the parameterizations of fast and slow, oscillatory and damped, physical processes. It permits the examination of a broad set of physics–dynamics coupling issues, while keeping the analysis tractable. Any given coupling scheme can be assessed in terms of its numerical stability and of the accuracy of both its transient and steady-state responses. The methodology essentially examines the forced evolution equation for the amplitude of a normal mode – this again underlines the underpinning importance of normal modes to understanding the behaviour of numerical schemes.

For the reasons discussed above, fully implicit coupling is impracticable, as is fully explicit coupling due to timestep restrictions. A popular approach is "split-implicit" coupling in which a dynamics predictor is followed by a physics corrector. It addresses the stability issue of an explicit coupling while keeping the physics discretization distinct from the dynamics discretization. However, using the framework of [61,62], it is found that the steady-state solution is corrupted and the forced response can be spuriously amplified by an order-of-magnitude. This motivated the "symmetrized split-implicit" coupling in which two physics discretizations are arranged symmetrically around a dynamics sub-step. The analysis shows that this addresses the stability and accuracy deficiencies of an explicit coupling while still correctly representing the exact steady-state solution for constant forcing. It also keeps the physics discretization distinct from the dynamics one. It partially shares the disadvantage of the fully implicit model inasmuch as the second physics sub-step is an implicit discretization of the highly nonlinear physics. However the usual column-based physical parameterizations are such that the discrete set of nonlinear equations can be solved column-by-column, greatly reducing the computational cost.

This early work was done in the context of a physics package comprising only one component. In a typical model, however, there are at least four distinct components, each with different characteristics. The work of [63,64] therefore extends the above-described framework to examine the coupling of a mix of physical parameterizations of various damping and oscillatory processes associated with a range of timescales. Various coupling strategies were examined therein, but none were found to perform uniformly well leading to only general, rather than specific, conclusions. For example, two generic splitting schemes were examined: sequential-splitting, in which the model's tendency is updated sequentially using the tendency due to each physics component in turn; and parallel-splitting, in which the model's tendency is updated simply by summing, independently, the tendencies of each physics component. It was found that sequential splitting is more flexible in its ability to eliminate splitting errors than parallel splitting. A disadvantage is that the sequential approach is sensitive to the order in which the physics components are applied. In practice a mix of sequential schemes for the fast timescale physics, and parallel schemes for the slow timescale ones, appears to optimize the overall coupling strategy. It is then found that the slower processes, such as radiation, should appear near the center of the timestep, with the faster processes, such as boundary layer diffusion, coupled implicitly at the end of the timestep.

Following up on this work, idealizations of several specific coupling formulations, considered to be paradigms of actual approaches used in atmospheric models, were then investigated in [65] with attention focused on two important properties. These are: (a) the ability to exactly maintain a steady state, since this implies a correct balance between the individual physics and dynamics processes, and thereby reduces systematic errors and biases; and (b) the possibility of achieving second-order accuracy in time. Although none of the operationally-inspired idealizations fully meets both of these evaluation criteria, an idealization of an experimental multiple-sweep predictor–corrector scheme proposed in [66] does, albeit at additional cost, and has much to recommend it.

As analyzed in [67,68], incorporation of the parameterized boundary layer vertical diffusion into an atmospheric model can produce spurious two-timestep oscillations in boundary-layer tendencies, and even instability, due to the inherently very fast timescales of this physical process. In [69], a generalization of two of the implicit schemes proposed in [67] to mitigate this behaviour is presented, analyzed and demonstrated by numerical simulation. It is found that this implicit scheme provides an affordable, stable, and well-behaved alternative to the prohibitively-expensive strategy of greatly reducing the timestep to be commensurate with the process timescale. It is also well suited for use as a component of more general predictor–corrector couplings such as those described in [66,70].

The framework of [61,62] can also be used to analyze the problem of spurious computational resonance in a SISL model. Traditionally, this has arisen in the presence of stationary spatial forcing, specifically that due to orography [45]. In this case, spurious resonance is absent when a Courant number restriction on timestep is satisfied. It is shown in [61] that time-dependent forcing, such as that due to the physics package, can also give rise to spurious resonance. Importantly though, the Courant number limitation on the timestep is then twice as restrictive as that for stationary forcing, thereby exacerbating the problem of spurious computational resonance with long timesteps.

6. Conclusion

A unified approach to NWP and climate modeling for multiscale applications implies a strong constraint on the design of a dynamical core. Recent research at the Met Office on a number of such dynamical core issues has been reviewed herein. These include the following:

- The continuous system requires consideration of a suitable equation set and vertical coordinate. Inter alia, these, together with the boundary conditions, determine the energetics of the system that is to be simulated numerically.
- The temporal discretization determines the stability and efficiency of the scheme. The semi-implicit method, when coupled with a semi-Lagrangian scheme, is a proven approach. However, a regularized, time-staggered method, again when coupled to a semi-Lagrangian scheme, presents an interesting and plausible alternative.
- The semi-Lagrangian method itself presents challenges with regard to preserving accurately any of the conservation aspects of the underlying equations. There are also underlying issues of stability and accuracy associated with the departure-point computation – a key element of the semi-Lagrangian approach.
- As well as the temporal discretization, the spatial discretization is of course also fundamental to the performance of the model.
- Finally (at least within the terms of this review!), the coupling of the dynamical core with the physical parameterizations is becoming increasingly recognized as potentially a limiting factor, for given computational effort, to improving the accuracy of an NWP/climate model.

Acknowledgments

The work presented herein has benefited from discussions and collaboration with many people over many years. It is a pleasure to particularly acknowledge Drs. Terry Davies, Michail Diamantakis, Mark Dubal, Andy Malcolm, Andy White and Mohamed Zerroukat, of Dynamics Research within the Met Office, and also Prof. John Thuburn of Exeter University and Prof. Sebastian Reich of Potsdam University. The helpful comments of the Editors and two anonymous reviewers are also gratefully acknowledged.

References

[1] A. Spekat (Ed.), Proceedings of Symposium on 50th Anniversary of NWP, Potsdam, Germany, 9–10 March 2000, Deutsche Meteorologische Gesellschaft e.V., Berlin, 2001, pp. 255.

[2] M.J.P. Cullen, The unified forecast/climate model, Meteorol. Mag. 122 (1993) 81–94.

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

[4] A. Staniforth, A. White, N. Wood, J. Thuburn, M. Zerroukat, E. Cordero, T. Davies, The Joy of U.M. 6.3- model formulation, Unified Model Documentation Paper No. 15, available online at http://www.metoffice.com/research/nwp/publications/papers/unified_model/index.html (2006).

[5] A. Staniforth, N. Wood, Recent research for dynamical cores of nonhydrostatic, deep-atmosphere, unified models, in: ECMWF Seminar Proceedings: Recent Developments in Numerical Methods for Atmosphere and Ocean Modelling, ECMWF, Reading, 2004, pp. 15–26.

[6] A. Staniforth, Developing efficient unified nonhydrostatic models, in: A. Spekat (Ed.), Proceedings of Symposium on 50th Anniversary of NWP, Potsdam, Germany, 9–10 March 2000, Deutsche Meteorologische Gesellschaft e.V., Berlin, 2001, pp. 185–200, 255pp.

[7] A.A. White, B.J. Hoskins, I. Roulstone, A. Staniforth, Consistent approximate models of the global atmosphere: shallow, deep, hydrostatic, quasi-hydrostatic and non-hydrostatic, Q. J. R. Meteorol. Soc. 131 (2005) 2081–2107.

[8] A. Staniforth, N. Wood, The deep-atmosphere equations in a generalized vertical coordinate, Mon. Weather Rev. 131 (2003) 1931–1938.

[9] J. Thuburn, N. Wood, A. Staniforth, Normal modes of deep atmospheres. I: spherical geometry, Q. J. R. Meteorol. Soc. 128 (2002) 1771–1792.

[10] J. Thuburn, N. Wood, A. Staniforth, Normal modes of deep atmospheres. II: f–F-plane geometry, Q. J. R. Meteorol. Soc. 128 (2002) 1793–1806.

[11] A. Kasahara, On the nonhydrostatic atmospheric models with inclusion of the horizontal component of the Earth's angular velocity, J. Meteorol. Soc. Japan 81 (2003) 935–950.

[12] A. Kasahara, The roles of the horizontal component of the Earth's angular velocity in nonhydrostatic linear models, J. Atmos. Sci. 60 (2003) 1085–1095.

[13] D.R. Durran, C. Bretherton, Comments on "The roles of the horizontal component of the Earth's angular velocity in nonhydrostatic linear models", J. Atmos. Sci. 61 (2004) 1982–1986.

[14] A. Kasahara, Reply to 'Comments on "The roles of the horizontal component of the Earth's angular velocity in nonhydrostatic linear models"', J. Atmos. Sci. 61 (2004) 1987–1991.

[15] J. Thuburn, T.J. Woollings, Vertical discretizations for compressible Euler equation atmospheric models giving optimal representation of normal modes, J. Comput. Phys. 203 (2005) 386–404.

[16] J. Thuburn, Vertical discretizations giving optimal representation of normal modes: sensitivity to the form of the pressure gradient term, Q. J. R. Meteorol. Soc. (submitted).

[17] T. Davies, A. Staniforth, N. Wood, J. Thuburn, Validity of anelastic and other equation sets as inferred from normal-mode analysis, Q. J. R. Meteorol. Soc. 129 (2003) 2761–2775.

[18] D.R. Durran, Improving the anelastic approximation, J. Atmos. Sci. 46 (1989) 1453–1461.

[19] M. Tanguay, A. Robert, R. Laprise, A semi-implicit semi-Lagrangian fully compressible regional forecast model, Mon. Weather Rev. 118 (1990) 1970–1980.

[20] A. Kasahara, Various vertical coordinate systems used for numerical weather prediction, Mon. Weather Rev. 102 (1974) 509–522.

[21] R. Laprise, The Euler equations of motion with hydrostatic pressure as an independent variable, Mon. Weather Rev. 120 (1992) 197–207.

[22] N. Wood, A. Staniforth, The deep-atmosphere Euler equations with a mass-based vertical coordinate, Q. J. R. Meteorol. Soc. 129 (2003) 1289–1300.

[23] A. Staniforth, N. Wood, C. Girard, Energy and energy-like invariants for deep non-hydrostatic atmospheres, Q. J. R. Meteorol. Soc. 129 (2003) 3495–3499.

[24] A. Staniforth, André Robert (1929–1993): his pioneering contributions to Numerical Modelling, in: C.A. Lin, R. Laprise, H. Ritchie (Eds.), Numerical Methods in Atmospheric and Oceanic Modelling, The André J. Robert Memorial Volume, CMOS/ NRC Press, Ottawa, 1997, pp. 25–54.

[25] A. Staniforth, J. Côté, Semi-Lagrangian integration schemes for atmospheric models – a review, Mon. Weather Rev. 119 (1991) 2206–2223.

[26] A. Robert, A stable numerical integration scheme for the primitive meteorological equations, Atmos. Ocean 19 (1981) 35–46.

[27] A. Robert, A semi-Lagrangian and semi-implicit numerical integration scheme for the primitive meteorological equations, Japan Meteorol. Soc. 60 (1982) 319–325.

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

[29] C. Temperton, M. Hortal, A. Simmons, A two-time-level semi-Lagrangian global spectral model, Q. J. R. Meteorol. Soc. 127 (2001) 111–128.

[30] M. Zerroukat, N. Wood, A. Staniforth, SLICE: a semi-Lagrangian inherently conserving and efficient scheme for transport problems, Q. J. R. Meteorol. Soc. 128 (2002) 2801–2820.

[31] J. Purser, L.M. Leslie, An interpolation procedure for high-order three-dimensional semi-Lagrangian models, Mon. Weather Rev. 119 (1991) 2492–2498.
[32] M. Zerroukat, N. Wood, A. Staniforth, SLICE-S: a semi-Lagrangian inherently conserving and efficient scheme for transport problems on the sphere, Q. J. R. Meteorol. Soc. 130 (2004) 2649–2664.
[33] M. Zerroukat, N. Wood, A. Staniforth, A monotonic and positive-definite filter for a Semi-Lagrangian inherently conserving and efficient (SLICE) scheme, Q. J. R. Meteorol. Soc. 131 (2005) 2923–2936.
[34] M. Zerroukat, N. Wood, A. Staniforth, The Parabolic Spline Method (PSM) for conservative transport problems, Int. J. Numer. Meth. Fluid 11 (2006) 1297–1318.
[35] P. Colella, P. Woodward, The piecewise parabolic method (PPM) for gas-dynamical simulations, J. Comput. Phys. 54 (1984) 174–201.
[36] M. Zerroukat, N. Wood, A. Staniforth, Application of the Parabolic Spline Method (PSM) to a multi-dimensional conservative transport scheme (SLICE), J. Comput. Phys. (accepted to publication).
[37] M. Zerroukat, N. Wood, A. Staniforth, A three-dimensional conservative semi-Lagrangian scheme (SLICE-3D) for transport problems, in: J. Côté (Ed.), Research Activities in Atmospheric and Oceanic Modelling, CAS/ JSC Working Group on Numerical Experimentation Report No. 36, World Meteorological Organisation WMO/TD No. 1347, pp. 03.21–03.22.
[38] A.A. White, Dynamical equivalence and the departure point equation in semi-Lagrangian numerical models, Q. J. R. Meteorol. Soc. 129 (2003) 1317–1324.
[39] A. Staniforth, A. White, N. Wood, Analysis of semi-Lagrangian trajectory computations, Q. J. R. Meteorol. Soc. 129 (2003) 2065–2085.
[40] E. Cordero, N. Wood, A. Staniforth, Impact of semi-Lagrangian trajectories on the discrete normal modes of a non-hydrostatic vertical column model, Q. J. R. Meteorol. Soc. 131 (2005) 93–108.
[41] J. Frank, S. Reich, A. Staniforth, A. White, N. Wood, Analysis of a regularized, time-staggered discretization method and its link to the semi-implicit method, Atmos. Sci. Lett. 6 (2005) 97–104.
[42] N. Wood, A. Staniforth, S. Reich, An improved regularization for time-staggered discretization and its link to the semi-implicit method, Atmos. Sci. Lett. 7 (2006) 21–25.
[43] A. Staniforth, N. Wood, S. Reich, A time-staggered semi-Lagrangian discretization of the rotating shallow-water equations, Q. J. R. Meteorol. Soc. (submitted).
[44] S. Reich, Linearly implicit timestepping methods for numerical weather prediction, BIT 46 (2006) 607–616.
[45] C. Rivest, A. Staniforth, A. Robert, Spurious resonant response of semi-Lagrangian discretisations to orographic forcing: diagnosis and solution, Mon. Weather Rev. 122 (1994) 366–376.
[46] A. Staniforth, N. Wood, Analysis of the response to orographic forcing of a time-staggered semi-Lagrangian discretization of the rotating shallow-water equations, Q. J. R. Meteorol. Soc. (submitted).
[47] M. Dubal, A. Staniforth, N. Wood, S. Reich, Analysis of a regularized time-staggered scheme to a vertical slice model, Atmos. Sci. Lett. (submitted).
[48] ECMWF Seminar Proceedings: recent developments in numerical methods for atmospheric modelling, ECMWF, Reading, 1999, 453pp.
[49] A. Staniforth, Regional modelling: a theoretical discussion, Meteorol. Atmos. Phys. 63 (1997) 15–29.
[50] J. Côté, J.-G. Desmarais, S. Gravel, A. Méthot, A. Patoine, M. Roch, A. Staniforth, The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part II: mesoscale results, Mon. Weather Rev. 126 (1998) 1373–1395.
[51] M.S. Fox-Rabinovitz, J. Côté, B. Dugas, M. Déqué, J. McGregor, Variable-resolution GCMs: Stretched-Grid Model Intercomparison Project (SGMIP), J. Geophys. Res. 111 (2006) D16104, doi:10.1029/2005JD006520.
[52] J. Thuburn, A. Staniforth, Conservation and linear Rossby-mode dispersion on the spherical C grid, Mon. Weather Rev. 132 (2004) 641–653.
[53] A. Arakawa, V.R. Lamb, A potential enstrophy and energy conserving scheme for the shallow-water equations, Mon. Weather Rev. 109 (1981) 18–36.
[54] A. Untch, M. Hortal, A finite-element scheme for the vertical discretization in the semi-Lagrangian version of the ECMWF forecast model, Q. J. R. Meteorol. Soc. 130 (2004) 1505–1530.
[55] A. Staniforth, N. Wood, Comments on 'A finite-element scheme for the vertical discretization in the semi-Lagrangian version of the ECMWF forecast model' by A. Untch and M. Hortal, Q. J. R. Meteorol. Soc. 131 (2005) 765–772.
[56] A. Staniforth, N. Wood, An unsuspected boundary-induced temporal computational mode in a two-time-level discretization, Mon. Weather Rev. 133 (2005) 712–720.
[57] W.W. Grabowski, P.K. Smolarkiewicz, Two-time-level semi-Lagrangian modeling of precipitating clouds, Mon. Weather Rev. 124 (1996) 487–497.
[58] D.L. Williamson, Time-split versus process-split coupling of parameterizations and dynamical core, Mon. Weather Rev. 130 (2002) 2024–2041.
[59] A. Beljaars, P. Bechtold, M. Köhler, J.-J. Morcrette, A. Tompkins, P. Viterbo, N. Wedi, The numerics of physical parametrization, in: ECMWF Seminar Proceeding: Recent Developments in Numerical Methods for Atmosphere and Ocean Modelling, ECMWF, Reading, 2004, pp. 113–134.
[60] A. Caya, R. Laprise, P. Zwack, Consequences of using the splitting method for implementing physical forcings in a semi-implicit semi-Lagrangian model, Mon. Weather Rev. 126 (1998) 1707–1713.
[61] A. Staniforth, N. Wood, J. Côté, Analysis of the numerics of physics–dynamics coupling, Q. J. R. Meteorol. Soc. 128 (2002) 2779–2799.

[62] A. Staniforth, N. Wood, J. Côté, A simple comparison of four physics–dynamics coupling schemes, Mon. Weather Rev. 130 (2002) 3129–3135.
[63] M. Dubal, N. Wood, A. Staniforth, Analysis of parallel versus sequential splittings for time-stepping physical parameterizations, Mon. Weather Rev. 132 (2004) 121–132.
[64] M. Dubal, N. Wood, A. Staniforth, Mixed parallel-sequential split schemes for time-stepping multiple physical parameterisations, Mon. Weather Rev. 133 (2005) 989–1002.
[65] M. Dubal, N. Wood, A. Staniforth, Some numerical properties of approaches to physics–dynamics coupling for NWP, Q. J. R. Meteorol. Soc. 132 (2006) 27–42.
[66] M.J.P. Cullen, Alternative implementations of the semi-Lagrangian semi-implicit schemes in the ECMWF model, Q. J. R. Meteorol. Soc. 127 (2001) 2787–2802.
[67] E. Kalnay, M. Kanamitsu, Time schemes for strongly nonlinear damping equations, Mon. Weather Rev. 116 (1988) 1945–1958.
[68] C. Girard, Y. Delage, Stable schemes for nonlinear vertical diffusion in atmospheric circulation models, Mon. Weather Rev. 118 (1990) 737–745.
[69] M. Diamantakis, N. Wood, T. Davies, An improved implicit predictor–corrector scheme for boundary layer vertical diffusion, Q.J. R. Meteorol. Soc. 132 (2006) 959–978.
[70] M. J.P. Cullen, D.J. Salmond, On the use of a predictor–corrector scheme to couple the dynamics with the physical parametrizations in the ECMWF model, Q. J. R. Meteorol. Soc. 129 (2003) 1217–1236.