



# Runge–Kutta IMEX schemes for the Horizontally Explicit/Vertically Implicit (HEVI) solution of wave equations

Hilary Weller<sup>a,\*</sup>, Sarah-Jane Lock<sup>b,1</sup>, Nigel Wood<sup>c</sup>

<sup>a</sup> Department of Meteorology, University of Reading, UK

<sup>b</sup> ECMWF, Reading, UK

<sup>c</sup> Met Office, UK



## ARTICLE INFO

### Article history:

Received 20 November 2012

Received in revised form 7 June 2013

Accepted 17 June 2013

Available online 1 July 2013

### Keywords:

Implicit

Explicit

Runge–Kutta

Compressible

Atmosphere

Wave equations

## ABSTRACT

Many operational weather forecasting centres use semi-implicit time-stepping schemes because of their good efficiency. However, as computers become ever more parallel, horizontally explicit solutions of the equations of atmospheric motion might become an attractive alternative due to the additional inter-processor communication of implicit methods. Implicit and explicit (IMEX) time-stepping schemes have long been combined in models of the atmosphere using semi-implicit, split-explicit or HEVI splitting. However, most studies of the accuracy and stability of IMEX schemes have been limited to the parabolic case of advection–diffusion equations. We demonstrate how a number of Runge–Kutta IMEX schemes can be used to solve hyperbolic wave equations either semi-implicitly or HEVI. A new form of HEVI splitting is proposed, UfPreb, which dramatically improves accuracy and stability of simulations of gravity waves in stratified flow. As a consequence it is found that there are HEVI schemes that do not lose accuracy in comparison to semi-implicit ones.

The stability limits of a number of variations of trapezoidal implicit and some Runge–Kutta IMEX schemes are found and the schemes are tested on two vertical slice cases using the compressible Boussinesq equations split into various combinations of implicit and explicit terms. Some of the Runge–Kutta schemes are found to be beneficial over trapezoidal, especially since they damp high frequencies without dropping to first-order accuracy. We test schemes that are not formally accurate for stiff systems but in stiff limits (nearly incompressible) and find that they can perform well. The scheme ARK2(2,3,2) performs the best in the tests.

© 2013 The Authors. Published by Elsevier Inc. Open access under CC BY license.

## 1. Introduction

Many operational weather forecasting centres use semi-implicit time-stepping schemes because of their good efficiency. However, as computers become ever more parallel, horizontally explicit solutions of the equations of atmospheric motion might become an attractive alternative since implicit time-stepping introduces additional communication between processors. Implicit and explicit time-stepping schemes have long been combined to form semi-implicit models of the non-hydrostatic atmosphere [e.g. 22] which allow longer time-steps than fully explicit models. However the combination has not

\* Corresponding author.

E-mail address: [h.weller@reading.ac.uk](mailto:h.weller@reading.ac.uk) (H. Weller).

<sup>1</sup> Formerly at University of Leeds, UK.

always been made with consideration as to how the implicitly and explicitly represented terms interact and influence the stability or order of accuracy of the combined scheme. Recently, more attention has been paid to the way in which implicit and explicit time-stepping schemes are combined to form semi-implicit models of the global atmosphere [9,11,12,23].

Semi-implicit models treat acoustic and gravity waves implicitly and advection is usually semi-Lagrangian [e.g. 5] but sometimes explicit Eulerian [e.g. 21]. In either case, the time-step is not limited by the speed of sound, gravity wave speed or stratification [3]. With appropriate linearization, spatial discretisation and substitutions, this leads to a sparse, diagonally dominant matrix which can be solved with an iterative solver. With their ability to use long time-steps, semi-implicit methods can therefore be very efficient.

In HEVI models, the terms controlling the horizontal propagation of waves are treated explicitly whilst those terms controlling the vertical propagation are treated implicitly [e.g. 23] although Gassmann [10] treats gravity waves explicitly. HEVI solution techniques therefore overcome the worst stability constraints without the need for a global implicit solution with the associated global communication. The longest stable time-step is shorter but the computational cost per time-step is reduced, especially on massively parallel computers where communication slows down global matrix solutions.

Split-explicit methods involve taking shorter sub-steps for more rapidly varying terms of the governing equations. For example, in ocean models, the barotropic mode can be advanced on a short time-step while baroclinic modes are advanced on a large time-step [13] and in atmospheric models, split-explicit schemes often combine three different time-stepping schemes: an implicit scheme (such as trapezoidal or Crank–Nicolson) for the terms controlling vertical propagation of waves, forward–backward for the horizontal wave terms and a third-order Runge–Kutta scheme for Eulerian advection [e.g. 6,20]. Split-explicit solutions thus increase the time-step over HEVI models without the need for a global implicit solution.

In split-explicit methods, the long and short time-steps can be combined by splitting but Durran [8] showed that this leads to unacceptable errors. Alternatively they can be combined by using a lagged representation of the slow terms [20]. Durran [8] showed that this is much more accurate but usually unstable. The solution is to stabilise by applying time filtering [e.g. that of 2].

Implicit-explicit (IMEX) schemes have been developed for solving equations with fast and slow time-scales so that the fast terms are solved implicitly and the slow terms are solved explicitly paying particular attention to the stability and accuracy of the combined scheme. IMEX multistep methods (which use values from more than two time levels to achieve second order or higher accuracy) were analysed and tested by Giraldo [11] and Durran and Blossey [9] and were used to solve the hydrostatic primitive equations and the compressible Boussinesq equations. Multistep methods suffer from computational modes (although these are usually sufficiently damped so as not to spoil the solution [9,11]) and no linear (implicit) multistep method can be higher than second-order accurate and stable for all time-steps [4]. Lock et al. [16] calculate the theoretical stability and phase speed errors of a range of RK IMEX schemes with different HEVI splits for solving hyperbolic wave equations.

Ascher, Ruuth and Spiteri [ARS, 1] described how implicit and explicit Runge–Kutta (RK IMEX) schemes can be combined to create stable schemes where the combined scheme is second- or third-order accurate. These were tested on advection-diffusion equations (using the implicit part for the diffusion). Ascher et al. [1] paid attention to finding schemes that are stiffly accurate. That is, accurate in the limit that the fast term is infinitely fast. In the context of wave-equations and fluid-flow, this corresponds to finding accurate solutions for the slow modes while suppressing modes that are too fast to be resolved and for the atmosphere this means resolving Rossby waves and the slower gravity waves accurately while suppressing acoustic and fast gravity waves. The Ascher et al. [1] schemes are also “strong-stability preserving” (SSP) which means that a given norm of the solution does not increase. For example, new extrema are not produced.

Pareschi and Russo [19] proposed more RK IMEX schemes and tested them on advection-diffusion equations. Their schemes are SSP and asymptotically accurate (the order of accuracy is maintained in the stiff limit), which is a weaker constraint on the form of the scheme than the constraints that Ascher et al. [1] described for stiffly accurate. The schemes of Ascher et al. [1] and Pareschi and Russo [19] were both designed to damp high wave-numbers, unlike trapezoidal, which preserves the power in all wave-numbers. Kennedy and Carpenter [14] proposed some third- to fifth-order accurate Additive RK (ARK) IMEX schemes for convection–diffusion–reaction equations which were tested for semi-implicit and HEVI solutions of stratified flow by Giraldo et al. [12], who also proposed a new, second-order ARK scheme.

Three Runge–Kutta IMEX schemes were tested by Ullrich and Jablonowski [23] for the HEVI solution of the equations governing atmospheric motion. They tested the ARS(2,3,2) scheme of Ascher et al. [1] and also suggested the less computationally expensive but nearly as accurate Strang carryover scheme. This involves Strang splitting but the first implicit stage is cleverly re-used from the final implicit stage of the previous time-step and so there is only one implicit solution per time-step. Another novel approach taken by Ullrich and Jablonowski [23] is to use a Rosenbrock solution in order to treat all of the vertical terms implicitly rather than just the terms involved in wave propagation. A Rosenbrock solution is one iteration of a Newton solver. This circumvents the time-step restriction associated with vertical advection at the cost of slowing the vertical advection.

This paper considers the numerical solution of the equations governing compressible atmospheric waves. It investigates RK IMEX schemes applied both to a semi-implicit approach (treating acoustic and gravity waves implicitly and advection explicitly) and to a HEVI approach (Horizontally Explicit, Vertically Implicit) (Section 2.2). A range of Runge–Kutta IMEX schemes are analysed and tested for large-scale atmospheric modelling. We investigate the sensitivity to the choice of treating different terms implicitly and explicitly; in particular the standard semi-implicit with implicit gravity waves and two implementations of HEVI. A new type of HEVI splitting for IMEX methods is introduced in Section 2 which dramatically

improves accuracy and stability over a naive split. The conclusions from the numerical analysis (Section 3) are confirmed by the results of test cases (Section 5). Despite the apparent clear advantage of using a scheme that is accurate in the stiff limit, we test schemes that are not stiffly accurate and find advantages over other schemes, even for stiff system (nearly incompressible) in Section 5. Runge–Kutta IMEX schemes are defined in Section 2 and the splitting of terms of the compressible Boussinesq equations into explicit and implicit is described. A number of RK IMEX schemes are defined and analysed in Section 3 and results of all the schemes and all the splits for two test cases are given in Section 5. Final conclusions concerning the splitting, recommended schemes and utility for more complex equations are drawn in Section 6. The formulation of the Helmholtz equations by combining the momentum, temperature and continuity equations is described in Appendix A and the spatial discretisation is described in Appendix B.

## 2. Temporal discretisation

In this section, Runge–Kutta IMEX schemes are defined and three alternative ways of splitting the compressible Boussinesq equations into implicitly solved and explicitly solved terms are given.

### 2.1. Definition of Runge–Kutta IMEX

In order to define a Runge–Kutta IMEX scheme, we consider the solution of an ordinary differential equation of the form

$$\frac{dy}{dt} = s(y, t) + f(y, t) \quad (1)$$

where vectors  $y$ ,  $s$  and  $f$  can include a number of state variables and the time-scales associated with  $s(t, y)$  are relatively long whilst those associate with  $y$  can be very short so that  $f$  must be solved implicitly. A Runge–Kutta IMEX solution is defined by  $v$  sub-stages (indexed with  $j$ ) and one final stage in order to advance from time level  $n$  at time  $t^n$  to  $n+1$  at time  $t^{n+1} = t^n + \Delta t$ :

$$y^{(j)} = y^n + \Delta t \sum_{\ell=1}^{j-1} \tilde{a}_{j\ell} s(y^{(\ell)}, t^n + \tilde{c}_\ell \Delta t) + \Delta t \sum_{\ell=1}^j a_{j\ell} f(y^{(\ell)}, t^n + c_\ell \Delta t), \quad j = 1 \dots v, \quad (2)$$

$$y^{n+1} = y^n + \Delta t \sum_{j=1}^v \tilde{w}_j s(y^{(j)}, t^n + \tilde{c}_j \Delta t) + \Delta t \sum_{j=1}^v w_j f(y^{(j)}, t^n + c_j \Delta t). \quad (3)$$

Each Runge–Kutta IMEX scheme is defined by a double Butcher tableau:

$$\frac{\tilde{\mathbf{c}} | \tilde{\mathbf{A}}}{|\tilde{\mathbf{w}}^T|} = \begin{array}{c|cccccc} \tilde{c}_1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \tilde{c}_j & \tilde{a}_{j1} & \tilde{a}_{j2} & \cdots & \tilde{a}_{j\ell} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \tilde{c}_v & \tilde{a}_{v1} & \tilde{a}_{v2} & \cdots & \tilde{a}_{v\ell} & \cdots & 0 \\ \hline \tilde{w}_1 & \tilde{w}_2 & \cdots & \tilde{w}_\ell & \cdots & \tilde{w}_v \end{array}, \quad \frac{\mathbf{c} | \mathbf{A}}{|\mathbf{w}^T|} = \begin{array}{c|cccccc} c_1 & a_{11} & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ c_j & a_{j1} & a_{j2} & \cdots & a_{j\ell} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ c_v & a_{v1} & a_{v2} & \cdots & a_{v\ell} & \cdots & a_{vv} \\ \hline w_1 & w_2 & \cdots & w_\ell & \cdots & w_v \end{array}. \quad (4)$$

The explicit part must have  $\tilde{a}_{j\ell} = 0$  for  $j \geq \ell$  and a diagonally implicit Runge–Kutta scheme (DIRK) must have  $a_{j\ell} = 0$  for  $j > \ell$ . The  $a_{jj}$  terms require the implicit solution. A variety of Runge–Kutta IMEX schemes are defined and analysed in Section 3.

Pareschi and Russo [19] give the conditions on  $A$ ,  $\tilde{A}$ ,  $\mathbf{w}$ ,  $\tilde{\mathbf{w}}$ ,  $\mathbf{c}$  and  $\tilde{\mathbf{c}}$  for the overall scheme to be first-, second- and third-order accurate. If  $f$  is large in comparison to  $s$  then the system is stiff. Ascher et al. [1] describe a scheme that is accurate in the stiff limit as stiffly accurate. This can be achieved for if  $a_{v\ell} = w_\ell$ ,  $\ell = 1, \dots, v$ . For equations that are very stiff, [1] additionally suggest using  $\tilde{a}_{v\ell} = \tilde{w}_\ell$ ,  $\ell = 1, \dots, v$ . Pareschi and Russo [19] also give attention to stiffly accurate schemes which they describe as asymptotically accurate. However they derive asymptotically accurate schemes which do not need  $\tilde{a}_{v\ell} = \tilde{w}_\ell$ ,  $\ell = 1, \dots, v$ . We will pay attention to stability of wave equations in the stiff limit ( $f$  is arbitrarily large) and the explicit limit (when  $f$  is small).

### 2.2. Application of RK-IMEX to the compressible Boussinesq equations

The application of RK-IMEX schemes to the compressible Boussinesq equations in semi-implicit mode and two alternatives of HEVI mode is described here. All options handle the gravity wave terms implicitly. The formulation of the Helmholtz equation at each RK sub-stage is described in Appendix A.

### 2.2.1. Semi-implicit

We consider the solution of the two dimensional compressible Boussinesq equations [9]:

$$\frac{\partial \mathbf{u}}{\partial t} + \underbrace{\mathbf{u} \cdot \nabla \mathbf{u}}_{\text{slow}} + \underbrace{\nabla P}_{\text{fast}} = \underbrace{b \hat{\mathbf{g}}}_{\text{fast}} + \nabla \times \psi, \quad (5)$$

$$\frac{\partial P}{\partial t} + \underbrace{\mathbf{u} \cdot \nabla P}_{\text{slow}} + \underbrace{c_s^2 \nabla \cdot \mathbf{u}}_{\text{fast}} = 0, \quad (6)$$

$$\frac{\partial b}{\partial t} + \underbrace{\mathbf{u} \cdot \nabla b}_{\text{slow}} + \underbrace{w N^2}_{\text{fast}} = 0, \quad (7)$$

where  $\mathbf{u}$  is the velocity;  $w$  is the vertical component of  $\mathbf{u}$  (in the  $z$  direction);  $P \equiv \frac{p - \bar{p}(z)}{\rho_0}$  is the departure of pressure from a reference profile,  $\bar{p}(z)$ , normalised by a reference density,  $\rho_0$ ;  $b \equiv g \frac{\rho - \bar{\rho}(z)}{\rho_0}$  is the buoyancy with  $\rho$  the density and  $\bar{\rho}(z)$  a reference density profile;  $g$  and  $\hat{\mathbf{g}}$  are the magnitude and direction of gravity respectively and  $N^2 \equiv -\frac{g}{\rho_0} \frac{d\bar{\rho}}{dz}$  where  $N$  is the buoyancy frequency. The terms marked “fast” control acoustic and gravity wave propagation.  $\psi$  is an explicitly defined streamfunction used in the Durran and Blossey [9] test cases to force the flow.

The solution method for applying an RK-IMEX method to the compressible Boussinesq equations with this (semi-implicit) splitting, including the formulation of the Helmholtz equation at each RK sub-stage is described in [Appendix A](#).

### 2.2.2. HEVI UfPref and UfPreb

To solve the compressible Boussinesq equations with horizontal terms treated explicitly and vertical wave terms treated implicitly, the velocity vector,  $\mathbf{u}$ , is split into a horizontal component,  $u$  and a vertical component,  $w$ . The partitioning of terms given in Eqs. (5)–(7) is changed to:

$$\frac{\partial u}{\partial t} + \underbrace{\mathbf{u} \cdot \nabla u}_{\text{slow}} + \underbrace{\frac{\partial P}{\partial x}}_{\text{slow}} = -\frac{\partial \psi}{\partial z}, \quad (8)$$

$$\frac{\partial w}{\partial t} + \underbrace{\mathbf{u} \cdot \nabla w}_{\text{slow}} + \underbrace{\frac{\partial P}{\partial z}}_{\text{fast}} = \underbrace{b}_{\text{fast}} + \frac{\partial \psi}{\partial x}, \quad (9)$$

$$\frac{\partial P}{\partial t} + \underbrace{\mathbf{u} \cdot \nabla P}_{\text{slow}} + c_s^2 \left( \underbrace{\frac{\partial u}{\partial x}}_{\text{slow/fast}} + \underbrace{\frac{\partial w}{\partial z}}_{\text{fast}} \right) = 0, \quad (10)$$

$$\frac{\partial b}{\partial t} + \underbrace{\mathbf{u} \cdot \nabla b}_{\text{slow}} + \underbrace{w N^2}_{\text{fast}} = 0. \quad (11)$$

A simple partitioning is to treat  $\frac{\partial P}{\partial x}$  in (8) and  $\frac{\partial u}{\partial x}$  in (10) explicitly ( $\frac{\partial u}{\partial x}$  is assumed slow in (10)). This is referred to hereafter as UfPref – “ $\mathbf{u}$  forward, pressure forward” since both the  $u$ -equation (8) and the pressure equation (10) use forward time differencing for horizontal gradients. This type of splitting used by Giraldo et al. [12]. However most atmospheric models that use explicit methods in the horizontal [e.g. 20], employ a forward-backward method based on Mesinger [17] for the terms governing horizontal wave propagation. The equivalent partitioning for an IMEX scheme is to treat  $\frac{\partial P}{\partial x}$  in (8) explicitly and  $\frac{\partial u}{\partial x}$  in (10) implicitly ( $\frac{\partial P}{\partial x}$  slow and  $\frac{\partial u}{\partial x}$  fast). This is referred to hereafter as UfPreb – “ $\mathbf{u}$  forward, pressure backward” since the pressure equation (10) now uses backward time differencing. The UfPreb formulation does not respect the symmetry of the  $\frac{\partial P}{\partial x}$  and  $\frac{\partial u}{\partial x}$  terms in acoustic wave propagation but instead treats the components of the divergence in the same way in (10). UfPreb does not necessitate any additional implicit solutions. It just means using the most up to date value of  $u$  in Eq. (10) when calculating  $P$ . The two approaches will be compared using numerical analysis (Section 3) and numerical solutions (Section 5).

The formulation of the Helmholtz equations for each sub-stage, as described in [Appendix A](#), now leads to a Helmholtz equation in each column. These are cheap to solve since the size is the number of layers in the model and columns are not usually decomposed across processors so no parallel communications are needed for the Helmholtz solutions.

## 3. Runge–Kutta IMEX schemes and their linear stability

We will start by describing the stability analysis for Runge–Kutta IMEX schemes applied to wave equations. We will then define the familiar trapezoidal schemes with non-linear terms lagged as Runge–Kutta IMEX schemes and present the

stability analysis results. A number of other schemes will then be presented, such as one using Strang splitting and a number of strong stability preserving (SSP) and other schemes from the literature.

Following Pareschi and Russo [19] we use the notation [name] $k(s, \sigma, p)$  where  $k$  is the order of the explicit scheme,  $s$  is the number of implicit stages,  $\sigma$  is the number of explicit stages and  $p$  is the order of the whole scheme. For a scheme with  $v \times v$  matrices  $\tilde{A}$  and  $A$  in the Butcher tableau (Eq. (2)), we must have  $s \leq v$  and  $\sigma \leq v + 1$  with equality if and only if there are no rows of zeros, no repeated rows and  $\tilde{w}$  is not equal to the last row of  $\tilde{A}$  in the explicit tableau. (The number of explicit stages can include the final stage.)

### 3.1. Von Neumann stability analysis for wave equations

To investigate the stability properties of RK IMEX schemes, we perform a von Neumann stability analysis of two systems of wave equations, both of relevance to the atmosphere. We investigate first a generic wave equation consisting of an ODE with fast and slow contributions. From (1), we define  $\underline{\mathbf{y}} \equiv \mathbf{y}(t)$  and  $\underline{\mathbf{s}}(t, \underline{\mathbf{y}}) \equiv -is\underline{\mathbf{y}}$ ,  $\mathbf{f}(t, \underline{\mathbf{y}}) \equiv -if\underline{\mathbf{y}}$  to obtain:

$$\frac{d\underline{\mathbf{y}}}{dt} + is\underline{\mathbf{y}} + if\underline{\mathbf{y}} = 0. \quad (12)$$

Approximating the solution of (12) by use of an RK IMEX scheme, as described by (2)–(3), we can define the complex amplification factor  $A$  to satisfy  $\mathbf{y}^{n+1} = A\mathbf{y}^n$ . We can likewise define amplification factors  $A^{(j)}$  for the intermediate RK stages such that  $\mathbf{y}^{(j)} = A^{(j)}\mathbf{y}^n$ ; and therefore, for a general RK IMEX scheme,

$$A^{(j)} = 1 - is\Delta t \sum_{l=1}^{j-1} \tilde{a}_{jl} A^{(l)} - if\Delta t \sum_{l=1}^j a_{jl} A^{(l)}, \quad (13)$$

$$A = 1 - is\Delta t \sum_{j=1}^v \tilde{w}_j A^{(j)} - if\Delta t \sum_{j=1}^v w_j A^{(j)}. \quad (14)$$

Considered in the context of a HEVI simulation,  $s\Delta t$  and  $f\Delta t$  correspond to the Courant numbers associated with horizontally and vertically propagating waves respectively. For a stable scheme, we require  $|A| \leq 1$ , where  $|A| < 1$  indicates a scheme that is damping.

We also consider stability analysis of the 2D coupled system describing acoustic waves:

$$\frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} = 0, \quad (15)$$

$$\frac{\partial w}{\partial t} + \frac{\partial p}{\partial z} = 0, \quad (16)$$

$$\frac{\partial p}{\partial t} + c_s^2 \left( \frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} \right) = 0, \quad (17)$$

which can be rewritten as

$$\underline{\mathbf{y}}_t + S\underline{\mathbf{y}}_x + F\underline{\mathbf{y}}_x + G\underline{\mathbf{y}}_z = 0, \quad (18)$$

where  $\underline{\mathbf{y}} = [u, w, p]^T$  and

$$S = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ c_s^2 & 0 & 0 \end{pmatrix}, \quad G = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & c_s^2 & 0 \end{pmatrix}.$$

For an UfPreb HEVI solution of (18),  $S$  is treated explicitly and  $F$  and  $G$  are treated implicitly; whereas, for UfPref solution,  $(S + F)$  is treated explicitly and only  $G$  is treated implicitly.

To describe the stability analysis, we consider (first) the UfPreb solution of (18) with an RK IMEX scheme described by (2)–(3) such that

$$\underline{\mathbf{y}}^{(j)} = \underline{\mathbf{y}}^n - \Delta t \sum_{l=1}^{j-1} \tilde{a}_{jl} S\underline{\mathbf{y}}_x^{(l)} - \Delta t \sum_{l=1}^j a_{jl} (F\underline{\mathbf{y}}_x^{(l)} + G\underline{\mathbf{y}}_z^{(l)}), \quad (19)$$

$$\underline{\mathbf{y}}^{n+1} = \underline{\mathbf{y}}^n - \Delta t \sum_{j=1}^v \tilde{w}_j S\underline{\mathbf{y}}_x^{(j)} - \Delta t \sum_{j=1}^v w_j (F\underline{\mathbf{y}}_x^{(j)} + G\underline{\mathbf{y}}_z^{(j)}). \quad (20)$$

We focus the analysis on the stability of the time discretisation and therefore assume continuous spatial derivatives. From (18), solutions take the form

$$\underline{\mathbf{y}} = \underline{\mathbf{y}}_0 e^{i(k_x x + k_z z - \omega t)}$$

and therefore,  $\underline{\mathbf{y}}_x = ik_x \underline{\mathbf{y}}$  and  $\underline{\mathbf{y}}_z = ik_z \underline{\mathbf{y}}$  for substitution in (19)–(20). Furthermore, we define the complex amplification factor  $A$  to satisfy  $\underline{\mathbf{y}}^{n+1} = A \underline{\mathbf{y}}^n$ , such that  $A$  is a  $3 \times 3$  matrix; and similarly, the complex matrix  $A^{(j)}$  satisfies  $\underline{\mathbf{y}}^{(j)} = A^{(j)} \underline{\mathbf{y}}^n$ , which on substitution in (19)–(20) yields

$$A^{(j)} = I - i\Delta t \sum_{l=1}^{j-1} \tilde{a}_{jl} k_x S A^{(l)} - i\Delta t \sum_{l=1}^j a_{jl} (k_x F + k_z G) A^{(l)}, \quad (21)$$

$$A = I - i\Delta t \sum_{j=1}^v \tilde{w}_j k_x S A^{(j)} - i\Delta t \sum_{j=1}^v w_j (k_x F + k_z G) A^{(j)} \quad (22)$$

for the UfPreb formulation where  $I$  is the identity matrix.

Under an UfPref formulation, the amplification matrices become

$$A^{(j)} = I - i\Delta t \sum_{l=1}^{j-1} \tilde{a}_{jl} k_x (S + F) A^{(l)} - i\Delta t \sum_{l=1}^j a_{jl} k_z G A^{(l)}, \quad (23)$$

$$A = I - i\Delta t \sum_{j=1}^v \tilde{w}_j k_x (S + F) A^{(j)} - i\Delta t \sum_{j=1}^v w_j k_z G A^{(j)}, \quad (24)$$

the difference with (21)–(22) being the coefficients multiplying  $k_x F$ .

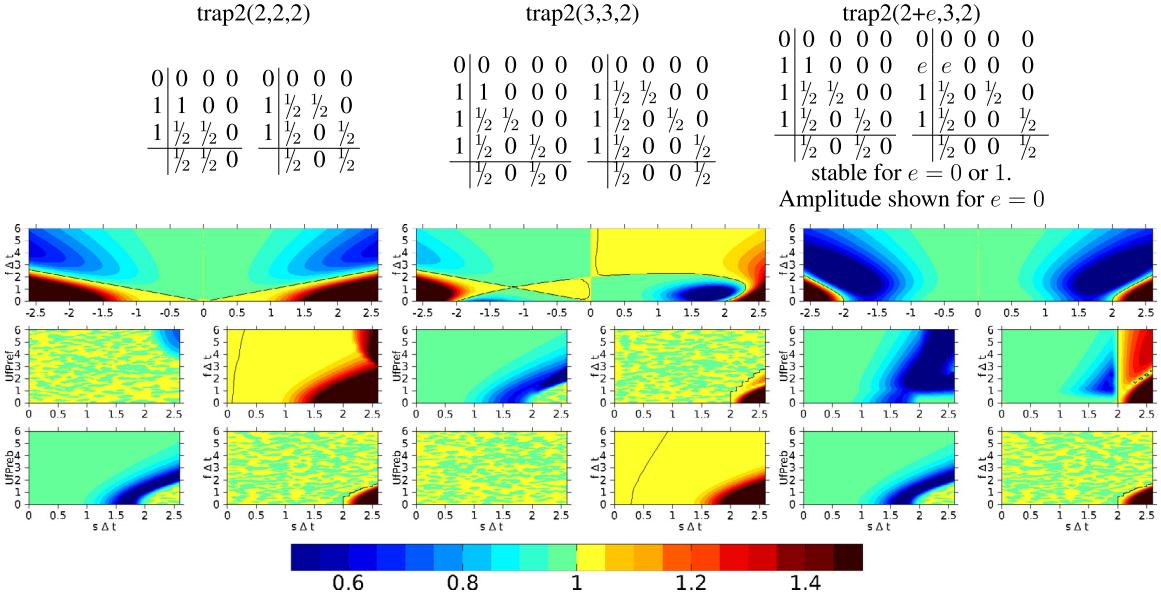
The amplitudes are found by rearranging Eqs. (21) and (23) for UfPreb and UfPref respectively, to give each  $A^{(j)}$  as a function of  $k_x \Delta t$  and  $k_z \Delta t$  and then substituting into Eqs. (22) and (24) respectively. For consistency with analysing the ODE, the Courant numbers  $c_s k_x \Delta t$  and  $c_s k_z \Delta t$  associated with the horizontally and vertically propagating acoustic waves, are referred to as  $s \Delta t$  and  $f \Delta t$  respectively. The amplitudes are calculated for ranges of  $s \Delta t$  and  $f \Delta t$  using GNU Octave. For stability in the 2D coupled system, we require absolute stability: that all three eigenvalues,  $\lambda$ , of  $A$  satisfy  $|\lambda| \leq 1$ . (We restrict consideration to absolute stability because our ultimate interest is in application of the schemes to more complex coupled systems of partial differential equations for which the discussion of Durran [7, Section 2.2.4] is relevant.) In the following analyses, we therefore consider the maximum magnitude of the three eigenvalues to indicate the stability of a given RK IMEX scheme and the minimum magnitude of the three eigenvalues to indicate the greatest damping resulting from the scheme.

### 3.2. Trapezoidal as a Runge–Kutta IMEX scheme

A trapezoidal scheme with deferred correction of explicit terms and a number of outer iterations can be written as a Runge–Kutta IMEX scheme with a number of stages. By iterating to update the explicit terms, there are many possibilities for different schemes. We restrict our attention to schemes that are stiffly accurate in the implicit and explicit Butcher tableau, implying that the final stage must be implicit (i.e.  $\tilde{w}$  is equal to the last row of  $\tilde{A}$  and  $w$  is equal to the last row of  $A$ ). Trapezoidal with 2 and 3 outer iterations can be defined by the Butcher tableau in Fig. 1 and are called trap2(2,2,2) and trap2(3,3,2) respectively (following Pareschi and Russo [19] notation). The amplification factor magnitudes for trap2(2,2,2) and trap2(3,3,2) are also shown in Fig. 1. For each scheme, the amplitude of the amplification factor for the ODE (12) is shown in the top row of the sub-figures in Fig. 1 for positive  $f \Delta t$  and positive and negative  $s \Delta t$ . For the ODE,  $|A(s \Delta t, f \Delta t)| = |A(-s \Delta t, -f \Delta t)|$  so the values for negative  $f \Delta t$  are not shown. Trap2(2,2,2) and trap2(3,3,2) both have regions of instability near the origin for the ODE (meaning that the scheme is unstable for small  $s \Delta t$  and  $f \Delta t$ ). This situation does not improve by using more stages (not shown). However for the acoustic wave equations (15)–(17), trap2(2,2,2) stability is dramatically improved in UfPreb mode and trap2(3,3,2) is improved in UfPref mode. For UfPref and UfPreb,  $|A(s \Delta t)| = |A(-s \Delta t)|$  and  $|A(f \Delta t)| = |A(-f \Delta t)|$  so only one quadrant is shown. For both the UfPref and UfPreb analysis, the largest and smallest magnitudes of the three eigenvalues are shown. Trap2(2,2,2) is unconditionally unstable for  $f \Delta t = 0$  for the ODE and in UfPref mode since the explicit scheme is the unstable Heun scheme but it is stable for  $s \Delta t \leq 2$  in UfPreb mode. With 3 iterations, trap2(3,3,2) is still unstable near the origin for the ODE but trap2(3,3,2) is stable up to  $s \Delta t = 2$  in UfPref mode.

For the ODE, trap2(3,3,2) is unstable in a region where  $s$  and  $f$  have different signs and therefore should partially cancel. Such cancellation is unlikely to occur for more general PDEs so it is not clear that the analysis of the ODE in this quadrant is relevant. However analysis of the ODE may be indicative of the behaviour of the schemes when used for other terms such as Coriolis or advection. It is therefore still worth considering.

Stability in the explicit limit (for  $f \Delta t = 0$ ) for the ODE can be improved by adding a final explicit stage. But this violates the stiff accuracy constraint. So, alternatively, stability can be improved by adding an entirely explicit stage before the other stages (trap2(2+e,3,2) as defined in Fig. 1). Using  $e = 0$  implies incrementing only the explicit terms before any implicit stages while using  $e = 1$  implies treating the first implicit stage explicitly. Both of these dramatically improve stability in the explicit limit. The analysis for  $e = 0$  is shown in Fig. 1.



**Fig. 1.** Double Butcher tableau for various forms of trapezoidal schemes and the resulting amplification factor magnitudes for ranges of values of  $s\Delta t$  and  $f\Delta t$ . The black contour is  $1 + 10^{-5}$  so larger values are unstable. The first row of each sub-plot shows the amplification factor magnitude for the generic wave equation (12) for positive and negative  $s\Delta t$  ( $|A(s\Delta t, f\Delta t)| = |A(-s\Delta t, -f\Delta t)|$ ). The second row shows the minimum (left) and maximum (right) amplification factor magnitudes for the 2D acoustic wave equations solved HEVI UfPref and the third row shows the magnitudes for the 2D acoustic wave equations solved HEVI UfPreb. The UfPref and UfPreb solutions are equal for  $\pm s\Delta t$  and  $\pm f\Delta t$  so only one quadrant is shown.

### 3.3. Runge–Kutta IMEX schemes from the literature

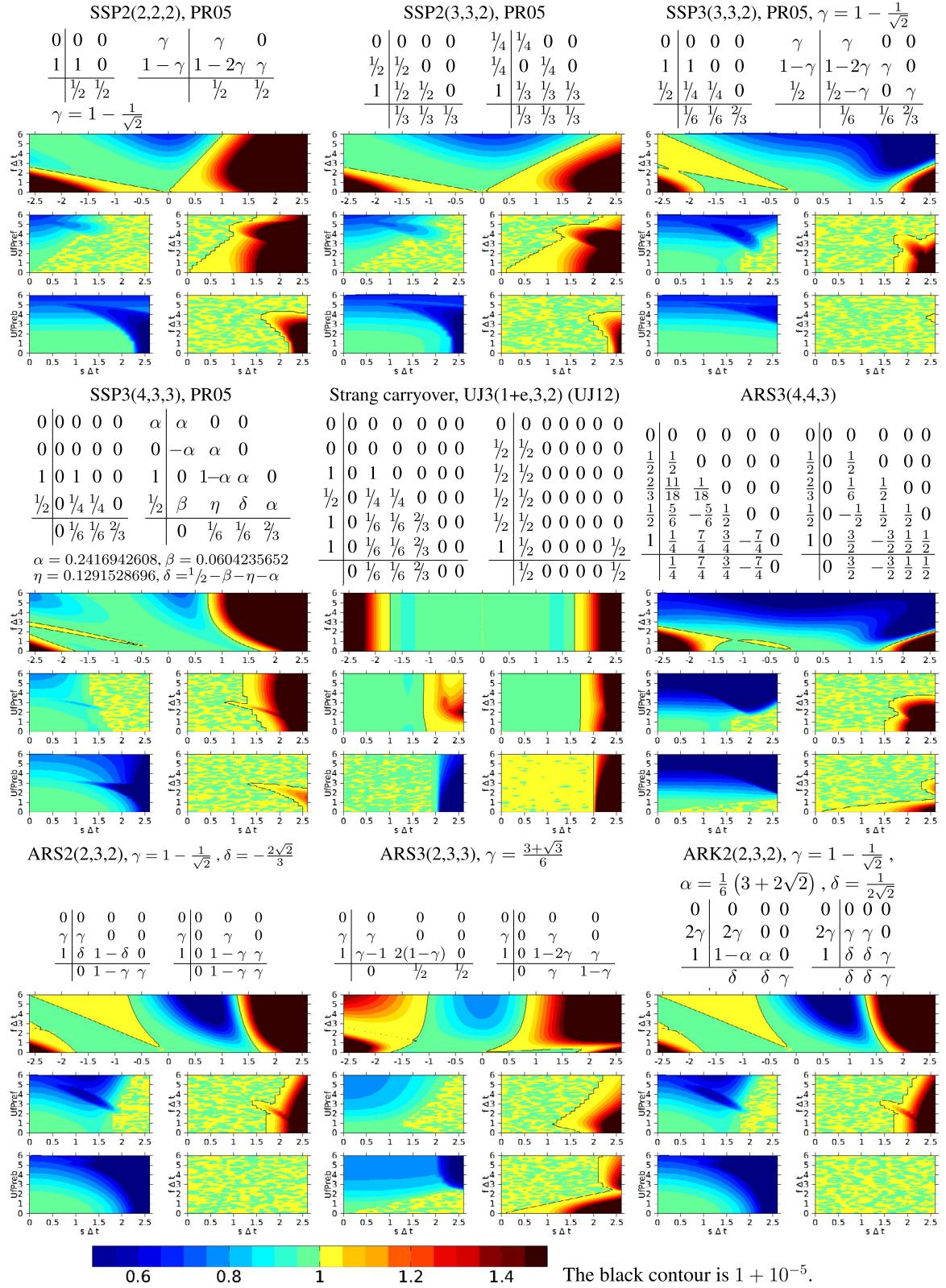
Schemes with no splitting increment the fast and slow terms by the same amount at each sub-stage (e.g. trap2(2,2,2) and trap2(3,3,2)) whereas schemes with complete splitting never increment the explicit and implicit terms in the same sub-stage (e.g. Strang splitting). Many Runge–Kutta IMEX schemes from the literature use partial splitting in order to achieve the maximum accuracy and stability for fewest sub-stages [e.g. 19]. These schemes have previously been analysed and used for advection-diffusion equations where the diffusion terms are handled implicitly. It is therefore necessary also to test them on wave equations with fast and slow terms. These schemes may be advantageous over the trapezoidal schemes because they damp high frequencies while retaining second-order accuracy, whereas trapezoidal requires off-centering to damp high frequencies which reduces the order of accuracy.

The Butcher tableau of a number of Runge–Kutta IMEX schemes from the literature are given in Fig. 2 and the resulting magnitudes of the amplification factors for wave equations are also shown. Kupka et al. [15] also defined some Runge–Kutta IMEX schemes. They were aiming to find SSP schemes which maximise accuracy for cost. We have tested their schemes but have not found any improvement over the Pareschi and Russo [19] schemes and so these schemes are not included.

Only the fully split, Strang carryover scheme (UJ3(1+e,3,2)) in Fig. 2 does not have regions of instability close to the origin for the ODE. For UJ3(1+e,3,2), the amplification factor is the product of the amplification factor (a function of  $s\Delta t$ ) arising from the explicit Butcher tableau with the amplification factor (a function of  $f\Delta t$ ) arising from the implicit Butcher tableau. This is a consequence of the complete (Strang) splitting used in UJ3(1+e,3,2) for solving an equation such as the ODE (12) in which the slow terms,  $s$ , and the fast terms,  $f$ , commute. For the other schemes which have no splitting or partial splitting, regions of instability close to the origin can be introduced.

Second-order schemes with only two explicit stages (such as SSP2(2,2,2)) are unconditionally unstable in the explicit limit for the generic wave equation. (It is well known that there are no 2-stage, second order, stable explicit schemes, e.g. Durran [6] when applied to the 2D acoustic wave equation.) SSP2(3,3,2) is also unconditionally unstable in the explicit limit for the ODE. The UfPref analysis is the same as the ODE in the explicit limit. However these schemes have good stability regions in UfPreb mode.

The other schemes, which have 3 or more explicit stages (SSP3(3,3,2), SSP3(4,3,3), ARS2(2,3,2), ARS(2,3,3) ARS3(4,4,3) and ARK2(2,3,2)), are stable in the explicit limit for the ODE up to some finite (non-zero) limit in  $s\Delta t$  but unstable for small, non-zero  $f\Delta t$  and  $s\Delta t$ . This instability was confirmed by solving the ODE with  $s\Delta t$  and  $f\Delta t$  in the marginally unstable region. However this instability is usually centred around  $s\Delta t = -f\Delta t$  where the ODE should give  $dy/dt = 0$ . It is not surprising that few schemes are unstable around this region since unless they are both discretized in the same way, the fast and slow terms do not always cancel exactly numerically. The stability in these regions is improved in UfPref mode. Mostly, UfPreb improves the stability region, but it does lead to unconditional instability for ARS3(2,3,3) and ARS3(4,4,3) in the explicit limit. Despite the lack of splitting in these two schemes, the otherwise different treatment of  $\frac{\partial P}{\partial x}$  and  $\frac{\partial u}{\partial x}$



**Fig. 2.** Double Butcher tableau of a Runge-Kutta IMEX schemes from Pareschi and Russo [19], Ullrich and Jablonowski [23], Ascher et al. [1] and Giraldo et al. [12] and the resulting amplification factor magnitudes. First row for generic wave equation (12), second row shows the minimum and maximum for HEVI UfPref and third row for HEVI UfPreb.

the horizontal acoustic wave propagation causes problems. SSP3(3,3,2), SSP3(4,3,3) ARS2(2,3,2) and ARK2(2,3,2) have good stability regions in both UfPref and UfPreb mode.

#### 4. Spatial discretisation

The spatial discretisation is defined for arbitrarily structured 3D grids in [Appendix B](#) and is implemented in OpenFOAM [18]. In the special case of a Cartesian grid with uniform resolution in the  $x$  and  $z$  directions, it is the same as the finite difference discretisation defined by Durran and Blossey [9], apart from the discretisation of the non-linear advection and the diffusion which use larger computational molecules than Durran and Blossey [9]. The discretisation presented will therefore be subject to spurious computational modes (which are not seen in the test cases presented). Pressure and velocity use C-grid staggering. A novel change of variables is described in order to implement Charney–Phillips staggering of pressure and buoyancy on arbitrarily structured grids. This involves treating the buoyancy as a vector and representing its normal component at each face between grid cells. However this implementation has not yet been tested on unstructured grids.

#### 5. Results

Section 5.1 presents results of the near-hydrostatic vertical slice test case of Durran and Blossey [9], solving the compressible Boussinesq equations. We are interested in the stiff limit so solving a near-hydrostatic problem with a non-hydrostatic model is relevant. In order to test the schemes with grid-scale forcing and larger time-steps (relative to the stratification) we also run this test case at coarser spatial resolution than was described by Durran and Blossey [9] in Section 5.2.

Only results from the schemes with at least three explicit stages are presented as these are generally the most stable and accurate.

##### 5.1. Durran and Blossey [9] near-hydrostatic vertical slice

The non-linear compressible Boussinesq equations (5)–(7) are solved on a 2D vertical slice for the near-hydrostatic test-case of Durran and Blossey [9] using the Runge–Kutta IMEX schemes defined in Section 3 and using semi-implicit, HEVI UfPref and HEVI UfPreb splitting of the compressible Boussinesq equations (Section 2.2). These are compared with a fully explicit, un-split simulation using the standard fourth-order Runge–Kutta scheme and a time-step of 0.5 s on the same grid. The diagnostics and errors are output after  $1.2 \times 10^5$  s (33.3 h). The initial conditions have zero vertical velocity, zero buoyancy perturbation, zero pressure perturbations and horizontal wind given by

$$u = 5 + z + 0.4(5 - z)(5 + z) \quad (\text{m s}^{-1}) \quad (25)$$

and the forcing streamfunction is:

$$\psi = \psi_0 \frac{\pi x}{L_x} \sin(\omega t) \exp\left[-\left(\frac{\pi x}{L_x}\right)^2 - \left(\frac{\pi z}{L_z}\right)^2\right] \quad (\text{m}^2 \text{s}^{-1}). \quad (26)$$

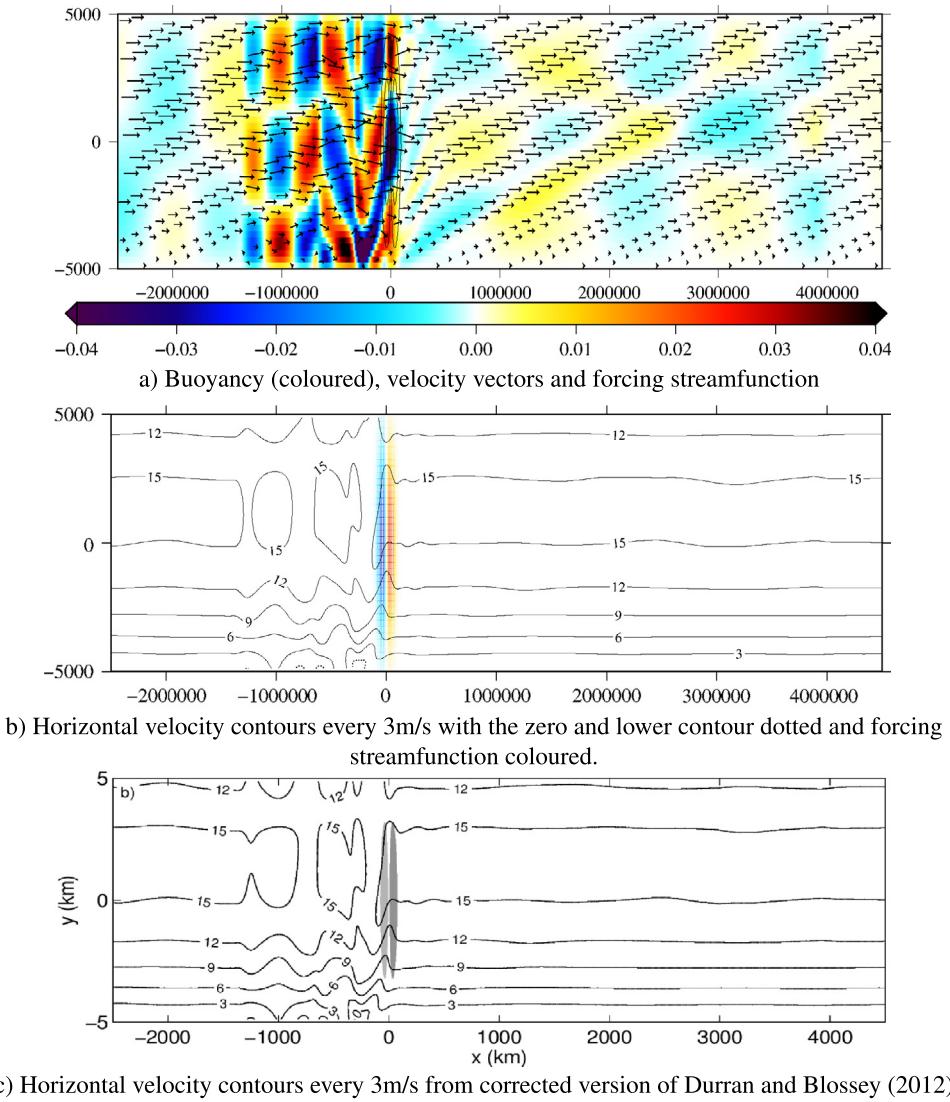
The parameters for the near-hydrostatic test-case of Durran and Blossey [9] are defined to be  $N = 0.02 \text{ s}^{-1}$ ,  $c_s = 350 \text{ m s}^{-1}$ ,  $\omega = 1.25 \times 10^{-4} \text{ s}^{-1}$ ,  $L_x = 160 \text{ km}$ ,  $L_z = 10 \text{ km}$ ,  $\psi_0 = 10 \text{ m}^2 \text{s}^{-1}$ ,  $z \in [-5, 5] \text{ km}$ ,  $\Delta x = 10 \text{ km}$ ,  $\Delta z = 250 \text{ m}$  and the hyper-diffusion coefficient,  $K = 1.17 \times 10^{-5} \text{ s}^{-1}$  (see [Appendix B](#)). The  $\pm x$  boundaries are periodic and the  $\pm z$  boundaries are free slip with fixed  $b = 0$  and zero vertical gradient of  $P$ .

The buoyancy, velocity vectors, horizontal velocity and forcing streamfunction at the diagnostic time are shown in [Fig. 3](#) in comparison to the horizontal velocity field in the corrected version of Durran and Blossey [9]. Our results are similar but not identical which is not surprising because the spatial discretisation is not identical. However this does not affect the conclusions of this work which are about the relative performance of the time-stepping schemes.

Following Durran and Blossey [9], the schemes are assessed by the normalised RMS buoyancy error ( $\ell_2$  error norm) in comparison to the fully explicit Runge–Kutta results. Buoyancy errors as a function of time-step for a variety of schemes and different splittings are shown in [Fig. 4](#). The schemes are tested at a variety of time-steps up to the maximum stable time-step.

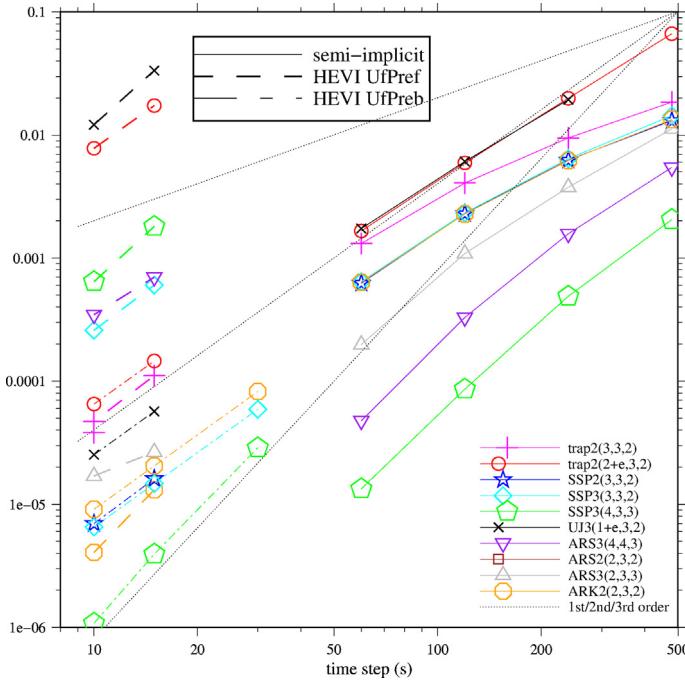
We can make a number of comments about the results in [Fig. 4](#):

- ARK2(2,3,2) is remarkable in that it is the only scheme that is accurate with both UfPreb and UfPref splitting. It is stable for larger time-steps with UfPreb splitting.
- UfPreb improves the stability and accuracy of most of the schemes when using HEVI splitting. The exceptions are trap2(3,3,2), ARS3(2,3,3) and ARS3(4,4,3) which are not stable with UfPreb splitting.
- The schemes that were analysed to be unconditionally unstable in the explicit limit in UfPref mode (trap2(2,2,2), SSP2(2,2,2), SSP2(3,3,2) and ARS2(2,2,2)) are indeed unstable in UfPref mode for this test case but stable in UfPreb mode. Of these, only SSP2(3,3,2) is shown in [Fig. 4](#).



**Fig. 3.** Results of the Durran and Blossey [9] near-hydrostatic vertical slice after 33.3 h using fully explicit fourth-order Runge–Kutta and a time-step of 0.5 s.

- ARS2(2,3,2) is not stable in any configuration. This is a surprise considering that the linear stability analysis gave identical results for ARS2(2,3,2) and ARK2(2,3,2). However ARK2(2,3,2) is stable with all the splittings for a wide range of time-steps.
- Some of the schemes dramatically lose accuracy in HEVI mode (SSP2(2,2,2), ARS2(2,2,2) and ARS3(4,4,3) but only ARS3(4,4,3) is shown). Durran and Blossey [9] also found that some schemes lose accuracy when gravity waves are treated implicitly and they attributed this to treating gravity waves with lower-order accuracy in the implicit part. However this would not explain the reduced accuracy of the above schemes since they have the same order accuracy in their implicit and explicit parts. However, splitting errors will inevitably cause problems when  $b$  and  $\frac{\partial P}{\partial x}$  are discretized differently, as happens in HEVI mode with implicit gravity waves.
- Using the trapezoidal based schemes, the accuracy as a function of time-step is maintained when running HEVI as opposed to semi-implicit as long as the right choice is made between using UfPref or UfPreb (UfPreb for trap2(2+e,3,2) but UfPref for trap2(3,3,2)). This is consistent with the lack of splitting for these schemes – the explicit and implicit parts use the same sub-stage time-steps (after any initial predictor stages) and they use the same weights. Therefore it does not make a big difference if terms are treated implicitly or explicitly, as long as the combination is stable.
- The SSP schemes and ARK2(2,3,2) do not lose accuracy in HEVI UfPreb mode and they are stable for a wide range of time-steps. It appears that the requirement for asymptotic accuracy is beneficial (rather than stiff accuracy, including a final implicit stage).



**Fig. 4.** Normalised RMS buoyancy errors of Durran and Blossey [9] near-hydrostatic test-case in comparison to explicit fourth-order Runge–Kutta with  $\Delta t = 0.5$  s. A time-step of 10 s corresponds to a horizontal acoustic Courant number of 0.35, a vertical acoustic Courant number of 14,  $\Delta t N = 0.2$  and an initial maximum advective Courant number of 0.0156.

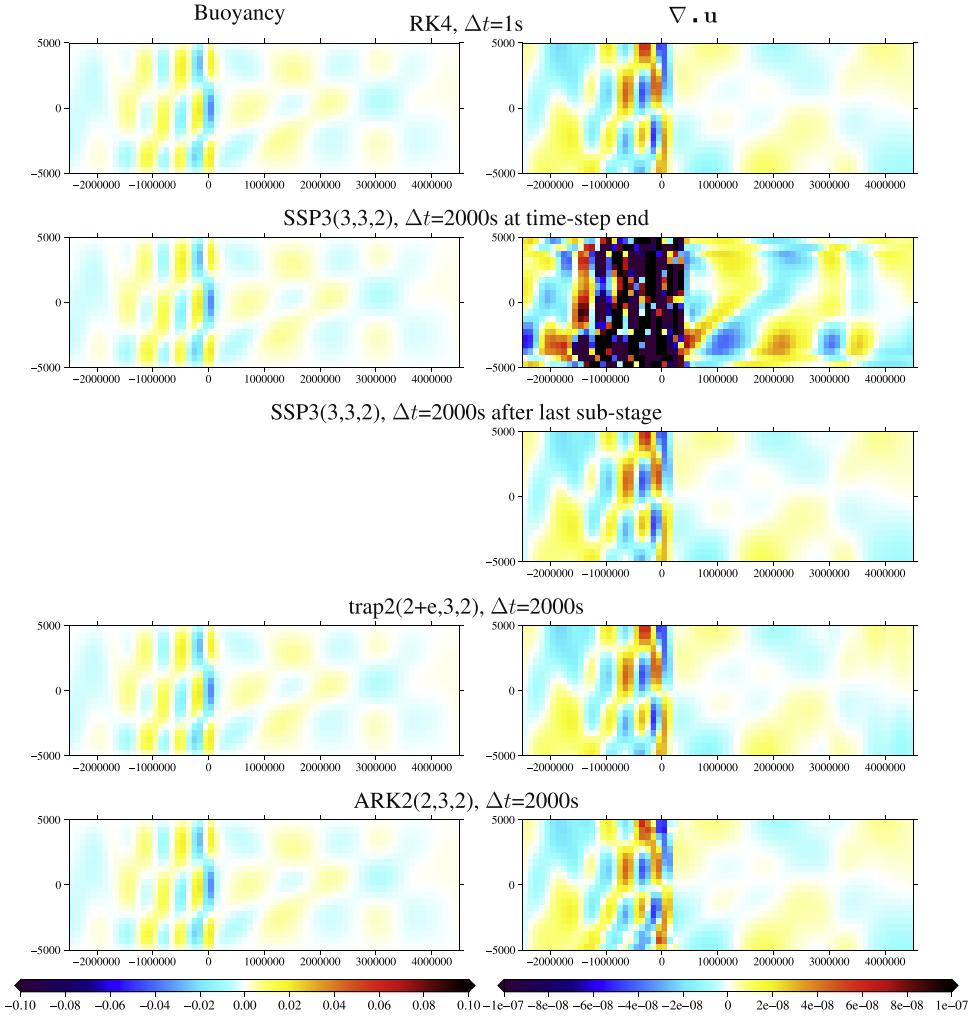
- Strang carryover (UIJ3(1+e,3,2)) performs well in HEVI UfPref mode, especially since it only uses one implicit stage. The Strang splitting introduces significant additional errors in UfPref mode but not in UfPreb mode.
- SSP3(3,3,2) and SSP3(4,3,3) are the most accurate and stable for the largest time-steps.

## 5.2. Vertical slice at coarser resolution

The SSP schemes that perform so well in Section 5.1 all have an explicit final stage which is a concern for simulating near incompressible flow for which we want  $\nabla \cdot \mathbf{u} = 0$  to be approximated very closely. When equations such as moisture advection are included, spurious moisture convergence could have serious consequences for precipitation predictions. The Durran and Blossey [9] near-hydrostatic test-case is therefore run at very coarse resolution ( $100 \text{ km} \times 500 \text{ m}$  instead of  $10 \text{ km} \times 250 \text{ m}$  in Section 5.1) to allow much longer time-steps and consequently not resolve the acoustic waves which generate divergent velocity.

The buoyancy and divergence for four of these coarse resolution simulations are shown in Fig. 5. The RK4 simulation with a time-step of 1 s is used as a reference. Results from SSP3(3,3,2) are presented because this scheme performed well in Section 5.1 despite not having a final implicit stage. Results from trap2(2+e,3,2) are also presented because this scheme is less accurate in Section 5.1 but does have a final implicit stage. Both SSP3(3,3,2) and trap2(2+e,3,2) predict the buoyancy well and trap2(2+e,3,2), which uses a final implicit stage, predicts divergence realistically. Conversely, SSP3(3,3,2) predicts erroneously strong divergence. This velocity field could lead to problems if it were used, for example, to advect moisture. A solution might be to use the velocity field from after the final implicit stage but before the final explicit stage (after the last sub-stage), which is shown in the third row of Fig. 5. This now recaptures the accuracy of the stiffly accurate scheme (trap2(2+e,3,2)) with the implicit final stage. The most important thing for an accurate representation of divergence at the end of the time-step is to have stiff accuracy in the implicit scheme. That is,  $w$  needs to be equal to the last row of  $A$ , while  $\tilde{w}$  need not be equal to the last row of  $\tilde{A}$ . For example, ARK2(2,3,2) has accurate divergence at the end of the time-step (final row of Fig. 5).

The normalised RMS errors in  $\nabla \cdot \mathbf{u}$  for all the schemes in semi-implicit, HEVI UfPref and UfPreb mode for a range of time-steps are shown in Fig. 6. For the schemes with an explicit final stage,  $\nabla \cdot \mathbf{u}$  is taken from after the last implicit stage rather than at the end of the time-step. Errors in  $\nabla \cdot \mathbf{u}$  from SSP3(4,3,3) converge only very slowly with decreasing time-step. This scheme is not stiffly accurate in the implicit or explicit Butcher tableau which reduces the accuracy. SSP2(3,3,2) and SSP3(3,3,2) give smaller and more convergent  $\nabla \cdot \mathbf{u}$  errors. The ARS schemes have small and convergent  $\nabla \cdot \mathbf{u}$  errors in semi-implicit mode but are either unstable or give large  $\nabla \cdot \mathbf{u}$  errors in HEVI mode. The trapezoidal schemes and Strang carryover (UIJ3(1+e,3,2)) give small and convergent errors in semi-implicit mode and also give small errors in HEVI mode for at least one of UfPref or UfPreb. This test appears to reveal a severe limitation with SSP3(4,3,3). At coarse resolution,



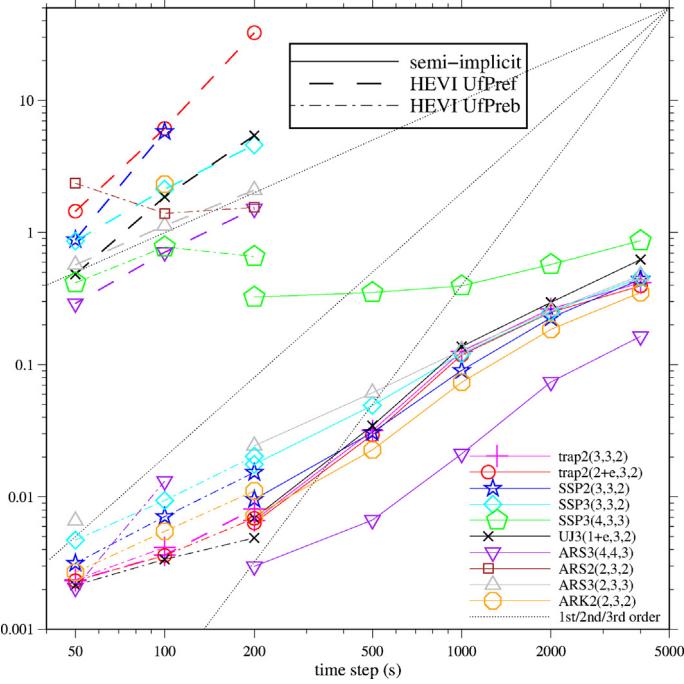
**Fig. 5.** Buoyancy and  $\nabla \cdot \mathbf{u}$  of the Durran and Blossey [9] near-hydrostatic vertical slice after 33.3 h at coarse resolution (100 km  $\times$  500 m).

looking at divergence errors, ARK2(2,3,2) now doesn't perform well in UfPref mode, however it is still very good in UfPreb mode.

## 6. Summary, conclusions and discussion

Linear stability analysis of a number of RK IMEX schemes is presented for solving a generic wave equation described by a complex valued ODE (12) with fast and slow frequencies. The fast frequency represents vertically propagating waves to be solved implicitly and the slow frequency represents horizontally propagating waves to be solved explicitly. Two methods of splitting the compressible Boussinesq equations in HEVI mode are then identified – UfPref (“ $\underline{\mathbf{u}}$  forward, pressure forward”) and UfPreb (“ $\underline{\mathbf{u}}$  forward, pressure backward”). (Forward implies using the explicit Butcher tableau and backward implies using the implicit tableau in the evolution of horizontally propagating acoustic waves.) Linear stability analysis of the PDEs describing acoustic wave propagation is then presented for both types of splitting. A range of Runge–Kutta IMEX schemes are analysed, including variations of trapezoidal implicit. Numerical experiments are undertaken for the range of time-stepping schemes using semi-implicit, HEVI UfPreb and HEVI UfPref splitting of the compressible Boussinesq equations. The experiments are the near-hydrostatic vertical slice test case of Durran and Blossey [9] and a version with coarser spatial resolution which has a larger time-step relative to the stratification and grid-scale forcing.

The analysis shows that many of the schemes are unconditionally unstable in the explicit limit (zero fast terms) when solving the ODE or the PDE using UfPref splitting. These schemes were also unstable in the numerical experiments. Other schemes have regions of instability at very small time-steps for solving the ODE where the slow and fast terms nearly cancel ( $s \approx -f$ ). However the stability analysis of the ODE in regions where the fast and slow terms have different signs was not found to be relevant to the stability analysis of the PDE describing acoustic wave propagation. The analysis of the PDE showed larger regions of stability, in line with the numerical experiments.



**Fig. 6.** Normalised RMS errors in  $\nabla \cdot \mathbf{u}$  of Durran and Blossey [9] near-hydrostatic test-case at coarse resolution ( $100 \text{ km} \times 500 \text{ m}$ ) in comparison to explicit fourth-order Runge–Kutta with  $\Delta t = 1 \text{ s}$ . A time-step of  $200 \text{ s}$  corresponds to a horizontal acoustic Courant number of 0.35, a vertical acoustic Courant number of 14,  $\Delta t N = 4$  and an initial maximum advective Courant number of 0.03125. For the schemes with an explicit final stage,  $\nabla \cdot \mathbf{u}$  is taken from after the last implicit stage rather than at the end of the time-step.

Some Runge–Kutta IMEX schemes are stiffly accurate, meaning that the final stage is identical to the final sub-stage. This implies that the final stage is implicit and the scheme is accurate for stiff systems. This seems essential since the atmosphere is close to divergence free. We have found that schemes that are not stiffly accurate in both the implicit and explicit Butcher tableau can give accurate approximations of buoyancy for stiff systems but inaccurate representations of  $\nabla \cdot \mathbf{u}$ . A possible way ahead is suggested. For schemes that are not stiffly accurate,  $\nabla \cdot \mathbf{u}$  is more accurately approximated at the end of the final implicit stage rather than at the end of the time-step. It would therefore be worth investigating using velocity and other fields from after the final implicit stage for input into other equations coupled to the core dynamical equations, such as moisture advection. Schemes that are not stiffly accurate may therefore still be useful for global atmospheric modelling.

All of the stable Runge–Kutta IMEX schemes considered perform well in semi-implicit mode (with implicit gravity waves) but there is more variation between the schemes in HEVI mode. Most of the schemes lose accuracy per time-step in HEVI mode, but the time-steps used are smaller and so accuracy is not lost over-all. Some recommendations about individual schemes are given:

- Efficiency comparisons have not been made, but Strang carryover (UJ3(1+e,3,2)) is the most cost effective in terms of having only one implicit stage per time-step. It performs well in semi-implicit mode and in HEVI UfPreb mode. The tests presented have not shown any evidence of problems due to its splitting.
- None of the ARS schemes are accurate for both test cases in HEVI mode and are therefore not recommended for the HEVI solution of wave equations.
- The trapezoidal scheme trap2(2+e,3,2) does not have any deficiencies for the analysis and tests undertaken, although it may be necessary to introduce damping of high frequencies for solving equations with noisy source terms.
- Some of the SSP schemes of Pareschi and Russo [19] seem promising. They lose very little accuracy in HEVI UfPreb mode in comparison to semi-implicit, some of them have long time-step constraints and some can accurately predict  $\nabla \cdot \mathbf{u}$  at coarse resolution. Overall, the SSP scheme that is performing best is SSP3(3,3,2).
- The ARK2(2,3,2) scheme of Giraldo et al. [12] appears to perform best overall. The analysis and test cases show it to be stable in UfPref and UfPreb mode and the experiments show that it is very accurate in HEVI UfPreb and semi-implicit mode.

### 6.1. Future work

The fast terms in the compressible Boussinesq equations are all linear and so they do not need to be linearised to be solved implicitly. To include fast, non-linear terms, linearization with outer iterations to update non-linearities can be used

[e.g. 24]. Another approach is to use a Newton solver [23]. If the linearization approach is used and if two implicit solutions were needed for every implicit Runge–Kutta sub-stage, the SSP schemes may not be so advantageous. Trapezoidal might be more efficient since sub-stages and outer-iterations to update non-linearities can be mixed since all sub-stage weights are the same. For other schemes, it may be beneficial not to update the explicit terms for each outer iteration of the implicit stage at each sub-stage. Or it may be sufficient to move the non-linearities to the explicit tableau. Further analysis, testing and comparison are necessary to find the most appropriate scheme for the Euler equations.

We have only looked at RK IMEX schemes with two Butcher tableaux. However it may be beneficial to have many more. For example split explicit could be defined and analysed using Butcher tableaux, or we could have different time-stepping schemes for horizontal and vertical waves, horizontal and vertical advection of dynamic and passive variables, sub-stepping for vertical advection and parameterisations. This would make the schemes much more difficult to analyse because the stability regions would be multi-dimensional. But more options should still be considered. In particular, when using a non-stiffly accurate scheme, it may be advantageous to use velocity fields from after a stiffly accurate stage for advecting quantities that rely on closely non-divergent winds. This has implications for the Butcher tableau for those terms.

## Acknowledgements

The careful review provided by two anonymous reviewers led to significant improvements. H.W. acknowledges support from NCAS Climate and NERC Grants NE/NE/H015698/1 and NE/I022086/1. S.J.L. acknowledges support from NERC Grant NE/I022094/1n. We would also like to thank Thomas Allen (Met Office) for useful input to this work.

## Appendix A. Formulation of the Helmholtz equations

In order to solve the compressible Boussinesq equations semi-implicitly, they are cast as a Helmholtz equation in one variable. The implicit solution of a Helmholtz equation leads to a sparse, diagonally dominant matrix which can be solved very efficiently. The Helmholtz problem is formulated using temporally discretized equations by taking the Schur complement in order to eliminate dependencies of pressure on velocity and of velocity on buoyancy. This involves temporally discretizing the buoyancy equation (7) and substituting  $b$  into the momentum equation (5) and then temporally discretizing the momentum equation (5) and substituting the resulting value of  $\mathbf{u}^j$  into the fast term of the pressure equation (6).

Each RK sub-stage for the buoyancy equation is

$$b^j = b^* - \Delta t a_{jj} w^j N^2 \quad (\text{A.1})$$

with the explicit part:

$$b^* = b^n - \Delta t \sum_{\ell=1}^{j-1} (\tilde{a}_{j\ell} \mathbf{u}^\ell \cdot \nabla b^\ell + a_{j\ell} w^{(\ell)} N^2) \quad (\text{A.2})$$

and each RK sub-stage for the momentum equation is

$$\mathbf{u}^j = \mathbf{u}^n - \Delta t \sum_{\ell=1}^{j-1} \tilde{a}_{j\ell} \mathbf{u}^\ell \cdot \nabla \mathbf{u}^\ell + \Delta t \sum_{\ell=1}^j a_{j\ell} (b^\ell \hat{\mathbf{g}} - \nabla P^\ell). \quad (\text{A.3})$$

Substituting (A.1) into (A.3) and re-arranging gives different equations for the horizontal component of  $\mathbf{u}^j$ ,  $\mathbf{u}_h^j$  and the vertical component,  $w^j = \mathbf{u}^j \cdot \hat{\mathbf{g}}$ :

$$\mathbf{u}_h^j = \mathbf{u}_h^* - \Delta t a_{jj} \nabla_h P^j, \quad (\text{A.4})$$

$$w^j = \frac{1}{1 + (\Delta t a_{jj} N)^2} \left( w^* - \Delta t a_{jj} \frac{\partial P^j}{\partial z} \right), \quad (\text{A.5})$$

where subscript  $h$  means the horizontal component,  $\mathbf{u}^*$  is given by

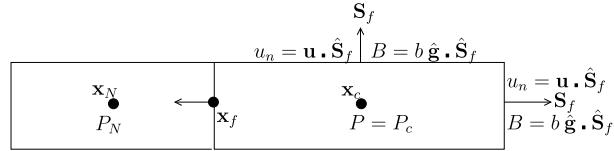
$$\mathbf{u}^* = \mathbf{u}^n + \Delta t \sum_{\ell=1}^{j-1} \tilde{a}_{j\ell} (-\mathbf{u}^\ell \cdot \nabla \mathbf{u}^\ell + b^\ell \hat{\mathbf{g}}) - \Delta t \sum_{\ell=1}^{j-1} a_{j\ell} \nabla P^\ell \quad (\text{A.6})$$

and  $w^*$  by

$$w^* = \mathbf{u}^* \cdot \hat{\mathbf{g}} + \Delta t a_{jj} b^*. \quad (\text{A.7})$$

The RK sub-stage for the pressure equation is

$$P^j = P^n - \Delta t \sum_{\ell=1}^{j-1} \tilde{a}_{j\ell} \mathbf{u}^\ell \cdot \nabla P^\ell - \Delta t c_s^2 \sum_{\ell=1}^j a_{j\ell} \nabla \cdot \mathbf{u}^\ell. \quad (\text{A.8})$$



**Fig. B.7.** Two cells of the grid with centres at  $x_c$  and the neighbour,  $x_N$  with the face between them centred at  $x_f$  and the placement of prognostic variables  $u_n$ ,  $B$  and  $P$  on the grid and the area vector,  $\hat{\mathbf{S}}_f$ .

Substituting  $\mathbf{u}^j = (\mathbf{u}_h^j, w^j)$  from Eqs. (A.4) and (A.5) into Eq. (A.8) gives a Helmholtz equation for every RK sub-stage:

$$\begin{aligned} P^j &= P^n - \Delta t \sum_{\ell=1}^{j-1} \tilde{a}_{j\ell} \mathbf{u}^\ell \cdot \nabla P^\ell - \Delta t c_s^2 \sum_{\ell=1}^{j-1} a_{j\ell} \nabla \cdot \mathbf{u}^\ell - \Delta t c_s^2 a_{jj} \nabla_h \cdot \mathbf{u}_h^* + (\Delta t c_s a_{jj})^2 \nabla_h^2 P^j \\ &\quad - \frac{\Delta t c_s^2 a_{jj}}{1 + (\Delta t a_{jj} N)^2} \frac{\partial w^*}{\partial z} + \frac{(\Delta t c_s a_{jj})^2}{1 + (\Delta t a_{jj} N)^2} \frac{\partial^2 P^j}{\partial z^2}. \end{aligned} \quad (\text{A.9})$$

Eq. (A.9) is solved implicitly at every sub-stage for  $P^j$  and  $\mathbf{u}^j$  is calculated by back-substitution from equations (A.4)–(A.7).

The final RK stage is simply:

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \Delta t \sum_{j=1}^v \tilde{w}_j \mathbf{u}^j \cdot \nabla \mathbf{u}^j + \Delta t \sum_{j=1}^v w_{j\ell} (b^j \hat{\mathbf{g}} - \nabla P^j), \quad (\text{A.10})$$

$$P^{n+1} = P^n - \Delta t \sum_{j=1}^v \tilde{w}_j \mathbf{u}^j \cdot \nabla P^j - \Delta t c_s^2 \sum_{j=1}^v w_j \nabla \cdot \mathbf{u}^j, \quad (\text{A.11})$$

$$b^{n+1} = b^n - \Delta t \sum_{j=1}^v \tilde{w}_j \mathbf{u}^j \cdot \nabla b^j - \Delta t \sum_{j=1}^v w_j N^2 w^j. \quad (\text{A.12})$$

## Appendix B. Spatial discretisation

The spatial discretisation is defined for arbitrarily structured, 3D grids. In the special case of a Cartesian grid with uniform resolution in  $x$  and  $z$  it is the same as the spatial discretisation defined by Durran and Blossey [9], apart from the discretisation of the non-linear advection and the diffusion. Pressure and velocity use C-grid staggering. A change of variables is described in order to implement Charney–Phillips staggering of pressure and buoyancy on arbitrarily structured grids. The spatial discretisation is therefore defined for arbitrarily structured grids.

The prognostic variables of the compressible Boussinesq equations (5)–(7) are  $u_n = \mathbf{u} \cdot \hat{\mathbf{S}}_f$  (on cell-faces),  $B = b \hat{\mathbf{g}} \cdot \hat{\mathbf{S}}_f$  (on cell-faces) and  $P$  (at cell-centres) where  $\hat{\mathbf{S}}_f$  is the outward pointing normal vector of each cell face with magnitude equal to the area of the face (area vector) (see Fig. B.7). Since  $\hat{\mathbf{g}} \cdot \hat{\mathbf{S}}_f = 0$  on faces whose area vector points horizontally,  $B$  is only non-zero on the interfaces between vertical layers on a Cartesian grid, so this placement is equivalent to Charney–Phillips on a Cartesian grid but generalises to other grids. A spatial discretisation that generalises to arbitrarily structured grids is used since this is more straightforward when implemented in OpenFOAM [18].

The discretized form of (5)–(7) is written:

$$\frac{\partial u_n}{\partial t} + (\mathbf{u}_F \cdot (\nabla \mathbf{u}_F)_F) \cdot \hat{\mathbf{S}}_f + \nabla_f P = B + (\nabla \times \psi)_f \cdot \hat{\mathbf{S}}_f - K(\nabla^2(\delta^2 \nabla^2(\delta^2 \mathbf{u}_C)))_F \cdot \hat{\mathbf{S}}_f, \quad (\text{B.1})$$

$$\frac{\partial P}{\partial t} + \mathbf{u}_C \cdot \nabla P + c_s^2 \nabla \cdot u_n = 0, \quad (\text{B.2})$$

$$\frac{\partial B}{\partial t} + (\mathbf{u}_F \cdot \nabla_f b_C) \hat{\mathbf{g}} \cdot \hat{\mathbf{S}}_f + u_n N^2 \hat{\mathbf{g}} \cdot \hat{\mathbf{S}}_f = 0 - K(\nabla^2(\delta^2 \nabla^2(\delta^2 b_C)))_F \hat{\mathbf{g}} \cdot \hat{\mathbf{S}}_f, \quad (\text{B.3})$$

where  $K$  is the hyper-diffusion coefficient,  $\delta$  is the spatial resolution in either the  $x$  and  $z$  direction, subscript  $f$  means that a variable lives on a face, subscript  $F$  means that a variable is linearly interpolated from the two neighbouring cell centres onto a face and subscript  $C$  means that a variable is reconstructed in a cell from face data.

Cell-centred vectors,  $\mathbf{u}_C$ , are reconstructed from face normal values,  $u_n$ , using the reconstruction:

$$\mathbf{u}_C = \left( \sum_{\text{faces}} \hat{\mathbf{S}}_f \hat{\mathbf{S}}_f^T \right)^{-1} \sum_{\text{faces}} u_n \hat{\mathbf{S}}_f, \quad (\text{B.4})$$

where  $\sum_{\text{faces}} \hat{\mathbf{S}}_f \hat{\mathbf{S}}_f^T$  is a tensor. It can easily be seen that this form preserves uniform velocity fields on 3D arbitrarily-structured grids. On uniform, structured grids, it reproduces the Cartesian form  $(\bar{u}^x, \bar{w}^z)$ . The full velocity on the face,  $\mathbf{u}_F$ , is then calculated by linearly interpolating neighbouring values of  $\mathbf{u}_C$ . In order to complete the discretisation of the  $\mathbf{u}_F \cdot (\nabla \mathbf{u}_F)_F$  term in Eq. (B.1), we define that  $\nabla \mathbf{u}_F$  is calculated at the cell-centre using Gauss's theorem (which requires no additional approximations):

$$\nabla \mathbf{u}_F = \frac{1}{V} \sum_{\text{faces}} \mathbf{u}_F \mathbf{S}_f, \quad (\text{B.5})$$

where  $V$  is the cell volume. This discretisation of  $(\nabla \mathbf{u}_F)_F$  has two levels of interpolation unlike a pure C-grid which only has one. More computational modes will therefore be permitted. However none are not forced in the test-cases presented which all have very small non-linear terms.

The gradient at a face (in the direction of the normal) of a cell-centred variable is the simple two-point difference:

$$\nabla_f P = \frac{P_N - P_c}{(\mathbf{x}_N - \mathbf{x}_c) \cdot \hat{\mathbf{g}}}, \quad (\text{B.6})$$

where  $P_N$  and  $P_c$  are the pressures in the cells either side of face  $f$  (see Fig. B.7) and  $\mathbf{x}_N$  and  $\mathbf{x}_c$  are the locations of the cell centres. The cell-centred gradient is calculated using linear interpolation and Gauss's theorem:

$$\nabla P = \frac{1}{V} \sum_{\text{faces}} P_F \mathbf{S}_f. \quad (\text{B.7})$$

The cell-centred  $b_C$  is given by  $b_C = B_C \cdot \hat{\mathbf{g}}$ . Finally,  $(\mathbf{u} \cdot \hat{\mathbf{g}})(\hat{\mathbf{g}} \cdot \hat{\mathbf{S}}_f)$  has been replaced by  $u_n(\hat{\mathbf{g}} \cdot \hat{\mathbf{S}}_f)$  since on grids aligned with  $\hat{\mathbf{g}}$ , these are equal.

The hyper-diffusion terms in (B.1) and (B.3) are solved explicitly and, sub-optimally, have an additional and unnecessary level of averaging. The cell-centred, diagnostic variables,  $\mathbf{u}_C$  and  $b_C$  are used and the cell-centred laplacian of a cell-centred variable,  $\Psi$ , is discretized as:

$$\nabla^2(\delta^2 \Psi) = \frac{1}{V} \sum_{\text{faces}} |\mathbf{S}_f| \delta_f^2 \frac{\Psi_N - \Psi_c}{\delta_f} \quad (\text{B.8})$$

where  $\delta_f = (\mathbf{x}_N - \mathbf{x}_c) \cdot \hat{\mathbf{S}}_f$  as in Eq. (B.6) and  $\Psi_N$  and  $\Psi_c$  are the values in the neighbour cell and the cell,  $c$  respectively. After the laplacian is taken twice to calculate the hyper-diffusion, the result is interpolated onto the faces. The two levels of interpolation (first using  $\mathbf{u}_C$  and  $b_C$  rather than the prognostic variables which are on the faces) and then interpolating the hyper-diffusion from cell centres onto faces, will allow grid-scale oscillations and are less accurate than if the laplacian of the prognostic variables had been calculated directly. However this was easier to code and grid-scale oscillations are not generated in the simulations described below. This sub-optimal discretisation therefore does not affect the conclusions of this paper.

## References

- [1] U. Ascher, S. Ruuth, R. Spiteri, Implicit-explicit Runge–Kutta methods for time-dependent partial differential equations, in: Workshop on Innovative Time Integrators, Ctr. Math. & Comp. Sci., Amsterdam, Netherlands, Oct. 30–Nov 01, 1996, Appl. Numer. Math. 25 (2–3) (1997) 151–167, [http://dx.doi.org/10.1016/S0168-9274\(97\)00056-1](http://dx.doi.org/10.1016/S0168-9274(97)00056-1).
- [2] R. Asselin, Frequency filter for time integrations, Mon. Weather Rev. 100 (1972) 487–490, [http://dx.doi.org/10.1175/1520-0493\(1972\)100<0487:FFTI>2.3.CO;2](http://dx.doi.org/10.1175/1520-0493(1972)100<0487:FFTI>2.3.CO;2).
- [3] M. Cullen, A test of a semi-implicit integration technique for a fully compressible nonhydrostatic model, Quart. J. Roy. Meteor. Soc. 116 (495) (1990) 1253–1258, <http://dx.doi.org/10.1002/qj.49711649513>.
- [4] G. Dahlquist, A special stability problem for linear multistep methods, BIT Numer. Math. 3 (1) (1963) 27–43, <http://dx.doi.org/10.1007/BF01963532>.
- [5] 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, Quart. J. Roy. Meteor. Soc. 131 (608) (2005) 1759–1782, <http://dx.doi.org/10.1256/qj.04.101>.
- [6] D. Durran, *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*, Springer, ISBN 0387983767, 1999.
- [7] D. Durran, *Numerical Methods for Fluid Dynamics with Applications to Geophysics*, 2nd edition, Texts Appl. Math., vol. 32, Springer, ISBN 1441964118, 2010.
- [8] D. Durran, Stabilizing fast waves, Chapter 6, in: *Numerical Techniques for Global Atmospheric Models*, in: Lect. Notes Comput. Sci. Eng., vol. 80, Springer, ISBN 978-3-642-11640-7, 2011.
- [9] D. Durran, P. Blossey, Implicit-explicit multistep methods for fast-wave-slow-wave problems, Mon. Weather Rev. 140 (4) (2012) 1307–1325, <http://dx.doi.org/10.1175/MWR-D-11-00088.1>.
- [10] A. Gassmann, A global hexagonal C-grid non-hydrostatic dynamical core (ICON-IAP) designed for energetic consistency, Quart. J. Roy. Meteor. Soc. 139 (670) (Jan. 2013) 152–175, <http://dx.doi.org/10.1002/qj.1960>.
- [11] F. Giraldo, Semi-implicit time-integrators for a scalable spectral element atmospheric model, Quart. J. Roy. Meteor. Soc. 131 (610) (2005) 2431–2454, <http://dx.doi.org/10.1256/qj.03.218>.
- [12] F. Giraldo, J. Kelly, E. Constantinescu, *Implicit-explicit formulations of a three-dimensional nonhydrostatic unified model of the atmosphere (NUMA)*, SIAM J. Sci. Comput. (2013), in press.
- [13] R. Higdon, A two-level time-stepping method for layered ocean circulation models: Further development and testing, J. Comput. Phys. 206 (2) (2005) 463–504, <http://dx.doi.org/10.1016/j.jcp.2004.12.011>.

- [14] C. Kennedy, M. Carpenter, Additive Runge–Kutta schemes for convection–diffusion–reaction equations, *Appl. Numer. Math.* 44 (2003) 139–181, [http://dx.doi.org/10.1016/S0168-9274\(02\)00138-1](http://dx.doi.org/10.1016/S0168-9274(02)00138-1).
- [15] F. Kupka, N. Happenhofer, I. Higueral, O. Kock, Total-variation-diminishing implicit–explicit Runge–Kutta methods for the simulation of double-diffusive convection in astrophysics, *J. Comput. Phys.* 231 (2012) 3561–3586, <http://dx.doi.org/10.1016/j.jcp.2011.12.031>.
- [16] S.-J. Lock, N. Wood, H. Weller, Numerical analyses of Runge–Kutta implicit–explicit schemes for horizontally-explicit vertically-implicit solutions of atmospheric models, *Quart. J. Roy. Meteor. Soc.*, in press.
- [17] F. Mesinger, Forward–backward scheme, and its use in a limited area model, *Contrib. Atmos. Phys.* 50 (1977) 200–210.
- [18] OpenFOAM. The opensource CFD toolbox. [Available online at <http://www.openfoam.org>], cited 2012. The OpenCFD Foundation.
- [19] L. Pareschi, G. Russo, Implicit–explicit Runge–Kutta schemes and application to hyperbolic systems with relaxation, *J. Sci. Comput.* 25 (1/2) (2005) 129–155, <http://dx.doi.org/10.1007/s10915-004-4636-4>.
- [20] W. Skamarock, J. Klemp, The stability of time-split numerical methods for the hydrostatic and the nonhydrostatic elastic equations, *Mon. Weather Rev.* 120 (1992) 2109–2127, [http://dx.doi.org/10.1175/1520-0493\(1992\)120<2109:TSOTSN>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1992)120<2109:TSOTSN>2.0.CO;2).
- [21] J. Szmelter, P. Smolarkiewicz, An edge-based unstructured mesh discretisation in geophysical framework, *J. Comput. Phys.* 229 (2010) 4980–4995, <http://dx.doi.org/10.1016/j.jcp.2010.03.017>.
- [22] M. Tapp, P. White, A non-hydrostatic mesoscale model, *Quart. J. Roy. Meteor. Soc.* 102 (432) (1976) 277–296, <http://dx.doi.org/10.1002/qj.49710243202>.
- [23] P. Ullrich, C. Jablonowski, Operator-split Runge–Kutta–Rosenbrock methods for nonhydrostatic atmospheric models, *Mon. Weather Rev.* 140 (4) (2012) 1257–1284, <http://dx.doi.org/10.1175/MWR-D-10-05073.1>.
- [24] K.-S. Yeh, J. Côté, S. Gravel, A. Méthot, A. Patoine, M. Roch, A. Staniforth, The CMC-MRB global environmental multiscale (GEM) model. part III: Nonhydrostatic formulation, *Mon. Weather Rev.* 130 (2) (2002) 339–356, [http://dx.doi.org/10.1175/1520-0493\(2002\)130<0339:TCMGEM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2002)130<0339:TCMGEM>2.0.CO;2).