



Current and Emerging Time-Integration Strategies in Global Numerical Weather and Climate Prediction

Gianmarco Mengaldo¹ · Andrzej Wyszogrodzki² · Michail Diamantakis¹ · Sarah-Jane Lock¹ · Francis X. Giraldo³ · Nils P. Wedi¹

Received: 4 January 2018 / Accepted: 5 February 2018 / Published online: 23 February 2018
© The Author(s) 2018. This article is an open access publication

Abstract

The continuous partial differential equations governing a given physical phenomenon, such as the Navier–Stokes equations describing the fluid motion, must be numerically discretized in space and time in order to obtain a solution otherwise not readily available in closed (i.e., analytic) form. While the overall numerical discretization plays an essential role in the algorithmic efficiency and physically-faithful representation of the solution, the time-integration strategy commonly is one of the main drivers in terms of cost-to-solution (e.g., time- or energy-to-solution), accuracy and numerical stability, thus constituting one of the key building blocks of the computational model. This is especially true in time-critical applications, including numerical weather prediction (NWP), climate simulations and engineering. This review provides a comprehensive overview of the existing and emerging time-integration (also referred to as time-stepping) practices used in the operational *global* NWP and climate industry, where *global* refers to weather and climate simulations performed on the *entire globe*. While there are many flavors of time-integration strategies, in this review we focus on the most widely adopted in NWP and climate centers and we emphasize the reasons why such numerical solutions were adopted. This allows us to make some considerations on future trends in the field such as the need to balance accuracy in time with substantially enhanced time-to-solution and associated implications on energy consumption and running costs. In addition, the potential for the co-design of time-stepping algorithms and underlying high performance computing hardware, a keystone to accelerate the computational performance of future NWP and climate services, is also discussed in the context of the demanding operational requirements of the weather and climate industry.

1 Introduction

Over the past few decades, operational NWP and climate models have evolved tremendously thanks to the continuous improvement of computing technologies and of the underlying algorithms at the foundations of these models. Today, these algorithms are facing a major challenge, as NWP and climate models are transitioning to more sophisticated Earth System Models (ESM), that will incorporate more components, and towards more accurate representation of the flow physics. In addition, the

hardware used to perform the simulations is also undergoing a dramatic change, with many-core architectures and co-processors—e.g., graphical processing units (GPU) and Intel’s Many Integrated Core processor architecture (MIC)—becoming the prevailing technologies. Therefore, it is necessary to review the core algorithmic strategies—in particular, the numerical discretizations and the time-integration strategies—currently adopted in the industry, to understand their potential in the future landscape.

To begin the discussion, we first need to characterize in more detail the models adopted in weather and climate applications. A numerical weather or climate model is constituted by a set of prognostic partial differential equations (PDEs) governing the fluid motion in the atmosphere (i.e., the geophysical flow) and by all those physical processes acting at a subgrid scale, whose statistical effects on the mean flow are expressed as a function of resolved-scale quantities [64]. The former is known as the *dynamical core* and represents the scale-resolved part of the model,

✉ Gianmarco Mengaldo
gianmarco.mengaldo@ecmwf.int

¹ ECMWF, Shinfield Park, Reading RG2 9AX, UK

² IMGW-PIB, Podleśna 61, Warsaw, Poland

³ Department of Applied Mathematics, Naval Postgraduate School, Monterey, CA 93943, USA

whereas the latter are referred to as the *physical parameterizations* and include the under-resolved processes. The dynamical core is usually described by the laws of thermodynamics and Newton's law of motion for the fluid air (e.g., compressible Navier–Stokes or Euler equations). Typical examples of physical parameterizations are convective processes, cloud microphysics, solar radiation, boundary layer turbulence and drag processes, which are handled by vertically-columnar submodels. The overall model is then discretized in space and time via suitable algorithmic blocks to approximate the continuous system of equations, thereby providing a solution otherwise not readily achievable.

In the short description above, we did not distinguish between a weather and a climate model, but we treated them as though they were identical. This deserves an explanation. While there are certainly overlaps between them, generally the time-scale and scope of a weather model differ substantially from those of a climate model. For example, the typical forecast window of weather models spans a range up to several days ahead, whilst for climate models, the forecast range is several months to years ahead. However, the set of PDEs defining the dynamical core are similar, if not identical; in fact, weather models are also referred to as the “higher-resolution siblings of the climate models' atmospheric component” [85], since they typically have higher spatial and temporal resolution than climate models. Also, they include a smaller number of physical processes, although both weather and climate models are evolving towards Earth System Models ESMs, that will include dynamic oceans, cryospheres and biochemical cycles [85]. Yet, despite these differences, in both weather and climate the “numerical engine” used to solve the PDEs constituting the model must be effective under several evaluation metrics in order to provide a ‘fast’ and ‘quantitatively satisfactory’ forecast. Hereafter, we will make no distinctions between NWP and climate models, as the scope of this review targets a shared aspect of both, despite the usually different operational and scientific objectives of weather and climate applications. In addition, we will refer to NWP operational constraints omitting those of climate simulations, as they are commonly the most severe.

The “numerical engine” employed to discretize the equations governing weather and climate models is the key to achieve the desired and commonly demanding performance required in operational NWP and climate simulations. From this perspective, the globally scale-resolved numerical solution, also referred to as direct numerical simulation (DNS), of weather and climate is not feasible due to the continuous spectrum of spatial and temporal scales and their nonlinear interactions that would require computational resources not existent today (and that will

not be available in the near future). On the other hand, large-eddy simulations (LES) of weather and climate are an emerging area of atmospheric (and oceanic) research, for both *limited-area* models (i.e., models that work on a limited portion of the Earth) and some *global* models (i.e., models that work on the entire planet) [44]. However, the major *global* operational and research centers use the compressible Euler equations (often further approximated due to the dominant hydrostatic balance of the atmosphere), together with physical subgrid-scale processes. This system of prognostic equations constitutes the backbone of all modern operational weather and climate models; their numerical solution together with the physical parameterizations and boundary conditions provide the spatio-temporal evolution of the atmosphere in terms of wind, pressure, temperature, density, including moist variables such as specific humidity, rain, snow, cloud water and ice, precipitation, and other atmospheric constituents. Although the overall numerical discretization strategy affects the quality metrics of the forecast system, the time integration of the PDEs governing geophysical flows in global weather and climate applications—that is the focus of this review—constitutes one of the most important aspects in the design of the computational model. The time integration in fact drives several key aspects of a NWP or climate model:

- (1) Solution accuracy,
- (2) Effectiveness of uncertainty quantification,
- (3) Time-to-solution,
- (4) Energy- or money-to solution, and
- (5) Robustness (e.g., numerical stability).

The forecast accuracy and the quantification of its uncertainty bounds are the highest goals for a successful NWP or climate model, together with a strict requirement on the timeliness of delivery of the forecast. The latter is strongly limited by the time of arrival of the global observations that input into the model initialisation and the deadline to deliver the forecast. The data assimilation of observations to derive the initial conditions (the analysis) already makes heavy use of and strongly depends on the quality of the forecast model and its efficiency. For instance, in terms of time-to-solution, the time threshold required operationally to run the entire model is 8.5 min per simulated forecast day, as defined in 2014 in the dynamical core evaluation criteria by the Advanced Computing Evaluation Committee (AVEC) for the Next Generation Global Prediction Systems (NGGPS).¹ This time constraint casts huge demands in terms of algorithmic efficiency for both parallel scalability as well as single-node performance that can strongly influence the choice of the numerical

¹ http://www.weather.gov/sti/stimodeling_nggps.

discretization strategy (e.g., time-integration) that is to be adopted. For example, for NWP, it is deemed acceptable to make some compromises on the level of required machine accuracy for each physical process (given the underlying uncertainty of some of the modeled aspects) in order to achieve better time-to-solution. Most recently, many operational centers are also including the energy-to-solution as an essential requirement to achieve a sustainable and cost-effective path to the future, given that the cost for electricity to run a state-of-the-art global NWP assimilation and forecast system is extremely demanding and will grow as the models become more complex and refined [9, 103]. Finally, the numerical stability and associated reliability of forecast delivery allows a robust operational workflow that guarantees, on the one hand, that the operational model will not fail (e.g., crash), and, on the other hand, that it will provide reproducible results.

Today there are a number of highly successful strategies that have emerged as the method of choice for the temporal (and spatial) discretization of the set of PDEs underlying NWP and climate models. Yet, the evolution of high-performance computing architectures requires a careful review of these strategies. From this perspective, the development of novel mathematical algorithms and their combination with existing successful methods, as well as hardware–software co-design are becoming an essential activity undertaken by many practitioners in the weather industry. In this work, we provide a broad overview on the different numerical time-integration strategies used to solve the PDEs arising in *global* NWP and climate models, emphasizing the most prominent techniques adopted in the weather community operationally in the past few decades, and describing the emerging solutions that operational weather centers (and the European Centre for Medium Range Weather Forecasts (ECMWF) in particular) are considering for the future. This review also aims to clearly categorize the currently adopted and emerging time-integration strategies under a more structured nomenclature.

The rest of the review is organized as follows. In Sect. 2, we introduce the set of equations used in weather and climate models. In Sect. 3, we categorize the most prominent time-integration approaches used in the industry. In Sect. 4, we highlight the time-stepping strategies adopted by the main operational and research models. In Sect. 5, we introduce three time-integration schemes that are being considered as potentially competitive for the future. Finally, in Sect. 6, we discuss the possible evolution that the weather and climate industry might undergo in the near to long-term future.

2 Equations Modeling the Atmosphere

Operational global NWP and climate models employ the compressible Euler equations to describe the fluid motion in the atmosphere, where the missing viscous stresses, and the sensible and latent heat fluxes as a result of diabatic processes are modeled as part of the physical parameterizations, and represented on the right-hand side of the equations as forcing terms. These equations are usually written in spherical coordinates—or coordinate mappings to a tangential plane for limited-area studies—and include the effects of gravity and the Earth rotation (i.e., the Coriolis force).

The formulation of the Euler equations can be either conservative (also referred to as flux-type through including the action of the continuity equation) or non-conservative (also referred to as advective-type). The form chosen has implications on the formal accuracy of global integrals such as mass, momentum and energy, important for climate projections, as well as local conservation properties important for very high-resolution simulations of clouds and convection. The chosen (type-) formulation of the equations influences the numerical schemes that can be most efficiently employed.

The compressible Euler equations written in conservative form are given as

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (1a)$$

$$\frac{\partial (\rho \mathbf{u})}{\partial t} + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \nabla p = -\rho \mathbf{g} - 2\rho(\boldsymbol{\omega} \times \mathbf{u}) + \mathcal{P}, \quad (1b)$$

$$\frac{\partial (\rho \theta)}{\partial t} + \nabla \cdot (\rho \theta \mathbf{u}) = \mathcal{Q}, \quad (1c)$$

where ρ is the density, p is the pressure, \mathbf{u} is the velocity vector, \mathbf{g} is the gravitational force, $\boldsymbol{\omega}$ is the vector representing the Earth rotation, θ is the potential temperature, that is related to pressure, p , and temperature, T , via $\theta = T/\pi$, with $\pi = (p/p_0)^{R/c_p}$ being the Exner pressure and p_0 , R and c_p being a reference pressure, the gas constant and the heat capacity at constant pressure, respectively. Equation (1) needs to be complemented by an equation of state, that is:

$$p = p_0 \left(\frac{\rho R \theta}{p_0} \right)^{c_p/c_v}, \quad (2)$$

where c_v is the heat capacity at constant volume. In addition, the terms \mathcal{P} and \mathcal{Q} in Eq. (1) represent the physical parametrization for the momentum and energy equation, respectively. The system of Eq. (1) is just one of several possibilities to write the equations governing the

atmosphere (for alternatives see, for instance, [41, 42]). In addition, here we do not formally define the physical parametrization terms \mathcal{P} and \mathcal{Q} , as they are not relevant for the purpose of this review. The reader interested in these aspects can refer to [9] (Fig. 1).

The system of Euler equations complemented with the parametrized physical processes simulates a wide range of characteristic waves whose propagation speeds, in addition to the advective time-scale (of the wind), dictate the resolvable temporal and spatial scales of the solution. These span from slower planetary waves (Rossby waves), to faster propagating Kelvin waves (in equatorial regions), and ubiquitous inertia-gravity waves as well as acoustic waves (the latter only if not filtered by the hydrostatic simplifications), see Fig. (1). In NWP, large-scale Rossby and Kelvin waves as well as inertia-gravity waves are important features that are ideally resolved within the model, while the energetic impact of acoustic waves is small and their subsequent impact on the weather forecast are usually considered marginal. The latter is a key aspect in NWP, since acoustic waves impose a huge restriction on the maximum time-step that can be used when an explicit time-integration scheme is adopted. In particular, the vertically-propagating acoustic waves are the most restrictive in terms of time-step because the vertical spatial discretization is much finer in global simulations than the horizontal discretization, that is $z_h \ll s_h$. In the past few decades, most models adopted in global NWP have widely exploited the hydrostatic approximation, where the upward-directed pressure gradient force (i.e., the decrease of pressure with height) is balanced by the downward-directed gravitational pull of the Earth. This approximation results in simplifying the vertical momentum equation as follows

$$\frac{\partial p}{\partial z} = -g\rho, \quad (3)$$

where z is the vertical coordinate. The hydrostatic approximation is well satisfied in the global atmosphere

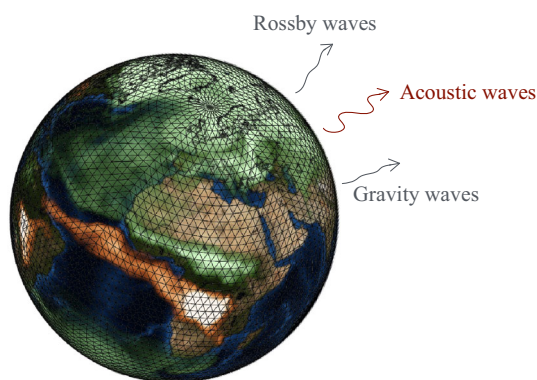


Fig. 1 Characteristic waves in the atmosphere

over a wide range of scales [53], and the resulting hydrostatic NWP model does not have any acoustic wave propagation in the vertical direction, thus removing the associated time-step restrictions. In addition, hydrostatic models substantially simplify the treatment of vertical boundary conditions, with important implications on the stability in complex terrain of the underlying numerical discretization. These features facilitate the adoption of time-stepping strategies that are accurate and stable for very large time-steps, $\Delta t \sim [120-1800]$ s for horizontal resolutions $\Delta s \sim [2-18]$ km. However, in global (sub-) km-scale simulations, as required to make fundamental advances in resolving existing model uncertainties with respect to cloud-radiation interactions, the hydrostatic approximation is questionable, although not proven to be inadequate for weather and climate simulations at least up to $\mathcal{O}(1)$ km resolutions, c.f. [53]. Many global operational NWP centers also operate local sub-km grid refinements, and driven by the desire to maintain a single code framework, will in the next decade transition from hydrostatic models to non-hydrostatic models. This transition has also important consequences on the numerical discretization strategies that can be adopted. In a non-hydrostatic model the vertically-propagating acoustic waves need to be appropriately treated to avoid severe restrictions on the time-step that can be used when explicit time-integration schemes are employed. There are different solutions proposed in the literature, such as *a priori* filtering of acoustic modes, i.e. sound-proof models, or by solving the compressible Euler equations and treating the acoustic waves via implicit-explicit or fully-implicit time-integrators [4, 10, 30, 90, 92]. While filtering the acoustic modes is attractive, the numerical solution procedure of the resulting filtered equations may be more difficult, hence there is still no consensus in the weather community towards one particular model [69], although the unfiltered compressible Euler equations with a numerical handling of acoustic modes is currently favored. The particular algorithm needs to consider physical and numerical constraints as for example the lower boundary treatment, conservative properties of the cloud resolving algorithms or the correctness of the treatment of acoustic and gravity waves (cf. [1]).

In this review, we consider time-integration strategies that target both hydrostatic and non-hydrostatic models and that are adopted by the major global operational centers. In the following we will denote, without lack of generality, the system of equations governing the NWP model as follows:

$$\frac{\partial \mathbf{y}}{\partial t} = \mathcal{R}(\mathbf{y}, t), \quad (4)$$

where $\mathcal{R}(\mathbf{y}, t)$ denotes the right-hand side of the system of equations, \mathbf{y} are the prognostic variables and the system is complemented by suitable initial and boundary conditions, which make it formally an initial boundary value problem (IBVP). In the following, we will manipulate Eq. (4) to highlight the characterizing aspects of each time-integration strategy that will be described.

3 Time-Integration Strategies in Global NWP and Climate Modeling

Broadly speaking, there are two distinct, high-level classes of time-integration schemes for IBVP (4) such as those arising in NWP and climate models:

- A. Eulerian-based time-integration (EBTI), where spatial and temporal discretizations are viewed as independent from each other; and
- B. Lagrangian or path-based time-integration (PBTI), where space and time are solved together, or where temporal derivatives may be expressed as spatial derivatives.

EBTI and PBTI have different characteristics in terms of stability properties, accuracy, suitability to emerging hardware, etc. and their application in the context of NWP has been rather different in the past few decades, with PBTI methods, especially semi-implicit semi-Lagrangian schemes, as the method of choice in operational global NWP. Both EBTI and PBTI methods can be explicit (current time-level is calculated using information coming from the previous time-steps only) or implicit (current time-level is obtained by solving a nonlinear problem that uses information from the current time-step). In the following, we describe in more detail each of the two classes and, for each class, we outline the prevailing operational time-integration practices adopted by the main weather and climate centers (cf. Table 1). In particular, we focus on two EBTI time-integration methods, namely (i) split-explicit (SE) schemes and (ii) horizontally-explicit vertically-implicit (HEVI) schemes, and on the PBTI-based family of methods referred to as (iii) semi-implicit semi-Lagrangian schemes.

3.1 Eulerian-Based Time-Integration (EBTI)

EBTI schemes [84], also referred to as Method Of Lines-based schemes (MOL), recast the IBVP constituted by the set of prognostic PDEs describing the physical model (4), into two sequential problems, a semi-discrete boundary value problem (BVP), where the equations are discretized in space, and an initial value problem (IVP), where the

spatially discretized equations are discretized in time via a suitable time-integration scheme, as depicted in Fig. 2.

In fact, the continuous system of PDEs (4), after being spatially-discretized, is reduced to a semi-discrete BVP that is formally a system of ordinary differential equations (ODEs)

$$\frac{d\mathbf{y}_h}{dt} = \mathcal{R}_h(\mathbf{y}_h, t) \quad (5)$$

where \mathbf{y}_h is the vector of spatially discretized prognostic variables and \mathcal{R}_h is the spatially-discretized right-hand side. Equation (5) can be solved to a desired time accuracy at all spatial locations of the model domain

$$\frac{\mathbf{y}_h^{n+1} - \mathbf{y}_h^*}{\alpha \Delta t} = \mathcal{R}_h(\tilde{\mathbf{y}}_h), \quad (6)$$

where n indicates time-level t^n , $\Delta t = (t^{n+1} - t^n)$ is the time-step and the factor α signifies the time-interval over which the temporal approximation is made. In addition, \mathbf{y}_h^* and $\tilde{\mathbf{y}}_h$ are combinations of model solutions. In particular, the first contains only known quantities (i.e., quantities from previous time-steps), while the second contains either quantities from previous time-steps only—explicit time-integration—or includes also future quantities—implicit time-integration. The right-hand side $\mathcal{R}_h(\mathbf{y}_h, t)$ of Eq. (5), includes the spatially-discretized (nonlinear) advection term $(\mathbf{u}_h \cdot \nabla_h) \mathbf{y}_h$ and terms describing wave propagation. The fastest of these terms cast the most severe restrictions regarding the time-step that can be adopted for explicit time-integration schemes. In particular, fast gravity and acoustic waves need to be handled appropriately to avoid time-steps that will be otherwise too small to be used in the context of operational global NWP and climate simulations.

EBTI schemes are conceptually simpler than PBTI and encapsulate two subcategories [67],

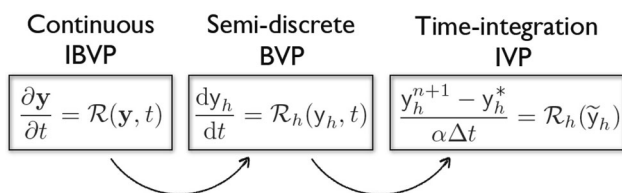
- i. multistage Runge–Kutta methods, that use multiple stages between two consecutive time-levels (or time-steps), discarding information from earlier time-steps; and
- ii. linear multistep methods, that use information from multiple earlier time-steps.

These two subcategories can be effectively represented within the General Linear (GL) method proposed by Butcher in 1987 [15], a unifying framework for which there exist many reviews in the literature—see for instance [16, 51]—as well as several implementation strategies—see for example [102]. Both subcategories, multistage Runge–Kutta and linear multistep methods, can be fully-implicit (the current time-level is obtained by solving a nonlinear problem that uses information from the current time-step) and fully-explicit (the current time-level is

Table 1 Time-stepping strategies adopted by the main operational global NWP and climate models as of 2017

Institute	Model	Type	Equation	Time integration	Class	Date	Country
ECMWF	IFS	NWP	H	SISL	PBTI	1991	Europe
MF	ARPEGE	NWP	H	SISL	PBTI	1991	France
UKMO	UM	NWP	NH	SISL	PBTI	2002	U.K.
JMA	GSM	NWP	H	SISL	PBTI	2005	Japan
CMA	GRAPES	NWP	H	SISL	PBTI	2010	China
NRL	NAVGEM	NWP	H	SISL	PBTI	2013	U.S.
EC	GEM	NWP	NH	SISL	PBTI	2014	Canada
NCEP	GFS	NWP	H	SISL	PBTI	2014	U.S.
DWD/MPI	ICON	NWP&Climate	NH	HEVI	EBTI	2015	Germany
MPI	ECHAM	Climate	H	SISL	PBTI	2003	Germany
HC	HADGEM	Climate	NH	SISL	PBTI	2009	U.K.
NOAA/GFDL	FV3	RD	H/NH	SISL	PBTI	2004	U.S.
CCSR/JAMSTEC	NICAM	RD	NH	SE-HEVI	EBTI	2008	Japan
NCAR	CAM-SEM	RD	H	SE-explicit	EBTI	2012	U.S.
NCAR	MPAS	RD	NH	SE-HEVI	EBTI	2013	U.S.
NCAR	CAM6	RD	H	HESL	PBTI	2015	U.S.
NPS/NRL	NUMA/ NEPTUNE	RD	NH	IMEX/HEVI	EBTI	2016	U.S.

RD models indicate those planned for future NWP and/or climate applications. Note that H stands for hydrostatic model, while NH stands for non-hydrostatic (where used in a global operational setup); SISL denotes the semi-implicit semi-Lagrangian method; HEVI indicates horizontally-explicit vertically-implicit schemes and SE denotes split-explicit schemes; IMEX indicates implicit/explicit schemes; and HESL denotes a horizontally-explicit (flux-form) semi-Lagrangian method

**Fig. 2** Conceptual schematics of EBTI schemes

calculated using information coming from the previous time-steps only). Representative classes of time-integration schemes embedded in the GL method consist of implicit multistep methods such as Adams–Moulton (AM) [22] and backward differentiation (BDF) methods [13, 20, 21], implicit multistage Runge–Kutta schemes such as diagonally (DIRK) and singly-diagonally (SDIRK) implicit Runge–Kutta schemes [3, 19, 59], explicit multistep methods, such as leapfrog and Adams–Bashforth methods [28, 43], explicit Runge–Kutta schemes, such as the fourth-order Runge–Kutta scheme [55] and partitioned methods, such as Implicit–Explicit (IMEX) schemes, whereby the operators are linearized in some fashion with—e.g., two Butcher tableaux, one explicit and one implicit [5, 40, 106].

While EBTI schemes are widely used in computational fluid dynamics, especially in the engineering sector [18, 52], their adoption in the weather and climate communities has been less widespread, with SE schemes

[54, 88, 107] and horizontally-explicit vertically-implicit schemes [8, 40, 63]—i.e., schemes where the horizontal direction is treated explicitly and the vertical is treated implicitly—becoming more prominent but still confined mainly to research and limited-area models (with very few exceptions—see Table 1). Within this context, Eq. (5) is further expressed as

$$\frac{\partial \mathbf{y}}{\partial t} = \mathcal{R}_f(\mathbf{y}, t) + \mathcal{R}_g(\mathbf{y}, t), \quad (7)$$

where $\|\mathcal{R}_f\| \gg \|\mathcal{R}_g\|$ (by some norm). The difference in magnitude of \mathcal{R}_f and \mathcal{R}_g comes from two aspects:

1. solutions to the continuous model (7) comprise fast and slow modes, i.e. \mathcal{R}_f (for the fast modes) and \mathcal{R}_g (for the slow modes) can describe processes that differ by orders of magnitude with respect to the time-scale of their propagation; and in addition,
2. in the discretized model, as already highlighted, the grid-spacings used to resolve the horizontal and vertical directions are highly anisotropic ($z_h \ll s_h$), reflecting the different scales that characterize the important processes in each direction. Since the terms on the right-hand side of (7) include spatial gradients, then, if \mathcal{R}_f represents contributions from the vertical direction and \mathcal{R}_g from the horizontal, the mesh-

anisotropy leads to a separation of scales between the two terms.

The separation of scales that arises between the vertical and horizontal directions is a key attribute for EBTI approaches in global NWP and climate simulations, as it motivates the use of different solution methods in the two directions. The “special-ness” of the vertical direction is further enhanced by the typical method of domain decomposition for parallelization of modern atmospheric models. The domain decomposition is limited to the horizontal dimension, with entire vertical columns preserved on each processor (based on the principle that important physical processes, such as radiative balance, act through the entire atmospheric depth, more or less in a pure vertical direction). EBTI approaches exploit the locality of the model data in the vertical direction in their choice of solution methods.

In the rest of this subsection, two EBTI approaches will be discussed—both are formally horizontally-explicit, vertically-implicit approaches, but the “split-explicit” approach, discussed in Sect. 3.1.1, adds an additional level of complexity in its use of sub-steps to handle the integration of fast processes. The “HEVI” approaches discussed in Sect. 3.1.2 highlight more recent developments, motivated by high-resolution global atmospheric models, which do not use sub-stepping.

3.1.1 Split-Explicit Schemes

Split-explicit schemes have been mainly used for high-resolution atmospheric simulations, where the horizontal mesh spacing is $\leq \mathcal{O}(10)$ km, thus being confined to local area models in the past few decades. One of the first SE (or “time-splitting”) approaches presented in the literature is in fact [57] (hereafter “KW78”). This was a limited-area cloud-resolving (grid-spacing of 1 km) atmospheric dynamical model, where advection, mixing and buoyancy were identified as the physically important processes in the system. The choice of compressible equations adopted however, meant that relatively fast acoustic waves were also present. Their system could be characterized by further extending (7) to

$$\frac{\partial \mathbf{y}}{\partial t} = \mathcal{R}_{f_s}(\mathbf{y}, t) + \mathcal{R}_{f_v}(\mathbf{y}, t) + \mathcal{R}_g(\mathbf{y}, t), \quad (8)$$

where \mathcal{R}_{f_s} and \mathcal{R}_{f_v} represent horizontally- and vertically-propagating fast waves respectively, and \mathcal{R}_g represents the slower modes.

Under the SE approach, a scheme with the required accuracy over an appropriate time-step, Δt , is chosen to solve the physically important terms in \mathcal{R}_g —KW78 used the 3-time-level leapfrog scheme. The fast terms are advanced on short sub-steps, $\Delta \tau$, such that $\Delta t = M \Delta \tau$

($M > 1$), using simpler and cheaper schemes that guarantee stability, but sacrifice accuracy. For terms in \mathcal{R}_{f_s} , the single-step 2nd-order forward-backward scheme [68] was used, where the horizontally-propagating acoustic wave terms of the continuity equation are integrated forward in time and those of the momentum equation are integrated backward in time. For \mathcal{R}_{f_v} , the implicit trapezoidal (or Crank–Nicholson) scheme was used. This yields a tridiagonal system of equations for each vertical model column, which is computationally simple to compute. A pictorial representation of a time-integration step with the leapfrog-based SE approach used in KW78 is presented in Fig. 3a.

The SE approach gains efficiency by computing the contributions from \mathcal{R}_g only on the longer time-step, Δt . The terms in \mathcal{R}_g include advection and mixing terms that require a relatively large stencil of data and are therefore more computationally expensive. Meanwhile, the contributions from \mathcal{R}_f (due to the fast waves) are computed every sub-step, $\Delta \tau$, but involve only immediate neighbour data-points to calculate local gradients. The latter aspect is particularly attractive for emerging computing technologies, given the reduced communication-to-flop ratio required, thus favoring co-processors, accelerators and many-core architectures.

Based on a stability analysis of the KW78 SE approach, [87] argued that greater efficiency could be gained by handling the buoyancy terms with the implicit scheme on the sub-step alongside the vertically-propagating acoustic waves. Under this approach, the longer time-step Δt is limited by the maximum speed of advection. Meanwhile, the sub-step $\Delta \tau$ continues to be limited by the horizontally-propagating acoustic waves. In addition, it was found that the SE approach needs some damping in its formulation to ensure an acceptable stability region. Skamarock and Klemp [87] proposed a “divergence damping” term to filter the acoustic modes in their analysis of the leapfrog-based KW78 approach. More recently, [35] has demonstrated the importance of an isotropic application of the divergence damping to the acoustic waves. Alternatively, they proposed using simple off-centering for the sub-step implicit solver. Baldauf [6] includes a comprehensive stability analysis of RK-based SE approaches, where the free parameters associated with the various components of the method are optimised, in terms of accuracy and stability. In particular, optimal values are proposed for the off-centering of the implicit solution of the acoustic and buoyancy terms, and the magnitude of the divergence damping applied to the acoustic waves.

The precise integration schemes used in a SE approach are open to choice: simple, efficient methods for the fast components; and a method with good accuracy and an acceptable window of stability (in terms of time-step

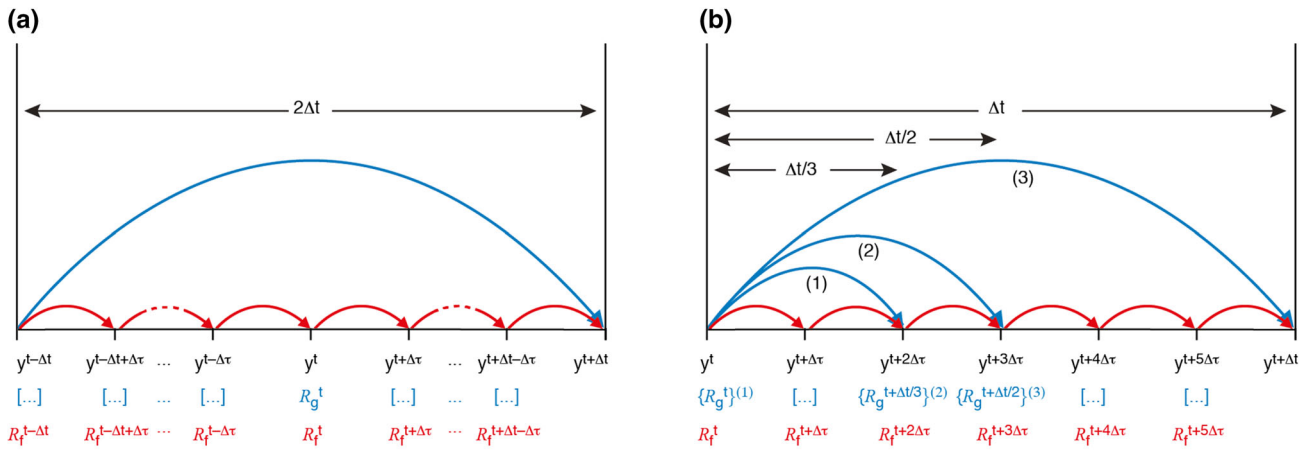


Fig. 3 Illustrations of two commonly used SE integration steps for solving (8): the terms in \mathcal{R}_g contribute to the model solution using a leapfrog to step from $t - \Delta t$ to $t + \Delta t$; and **b** a 3-stage 3rd-order

Runge–Kutta scheme to step from t to $t + \Delta t$ using contributions from predictor-stages at $t + \Delta t/3$ and $t + \Delta t/2$. In both cases, contributions from the fast terms \mathcal{R}_f and \mathcal{R}_f are updated at each sub-step $\Delta\tau$

length) for the slow components. A SE approach has been adopted in a number of active research and operational limited-area atmospheric models: the JMA’s non-hydrostatic Mesoscale Model (JMA-NHM) continues to use a leapfrog-based approach for the long time-step integrations, but includes some low-order advection components in the short time-step computations to improve the computational stability [79, 80]. Other groups have moved towards single-time-level, multistage explicit Runge–Kutta schemes to integrate the slow components—both the COSMO [6, 7] and WRF [56, 88] models use a 3-stage 3rd-order Runge–Kutta (RK) scheme for integrating the slow components, retaining the forward–backward and trapezoidal schemes (previously described) for the fast components (following the analyses of RK methods for time-splitting in [108, 109]). Figure 3b illustrates the 3-stage RK-based approach.

With the recent developments of *global* models with “very” high resolution (grid-spacings $\leq \mathcal{O}(10)$ km), SE methods are now being used also for global NWP: the MPAS model [89] has adopted the SE approach (as described in [56]) based directly on the successful experiences with the 3-stage 3rd-order RK-based SE approach in WRF. The high resolution global non-hydrostatic model, NICAM, also uses a RK-based SE approach (with options for 2nd- or 3rd-order RK schemes) [82, 83].

3.1.2 Horizontally-Explicit Vertically-Implicit (HEVI) Schemes

Similarly to SE approaches, HEVI schemes are becoming more and more attractive due to the latest advancements in computing that are driving the development of “very” high-resolution global NWP models. For global NWP, the stratosphere plays a significant role in the global circulation

[48, 50, 71]. Inclusion of a well-represented stratosphere has implications for the chosen time-integration methods, since the stratospheric polar jet (which contributes via the advection term) reaches speeds exceeding 100 m s^{-1} , i.e., the advective Courant number approaches the acoustic one. As highlighted in [34], in this context the efficiency gains from the SE approach become less clear: the horizontal splitting (which defines the sub-stepping) in SE schemes is only relevant when there is a scale-separation between fast insignificant and slow significant processes. With the acoustic and advective Courant numbers being similar, the sub-step $\Delta\tau$ and the model time-step Δt are constrained by similar stability limits and little efficiency can be gained from sub-stepping. In addition, as already noted, SE models require artificial damping to ensure stabilization, with atmospheric models typically employing divergence damping (see, e.g., [6]). HEVI-based alternatives can be efficiently used for global non-hydrostatic equation sets and do not have the drawbacks affecting SE schemes.

Consider again (7), describing the atmosphere as containing contributions from two scale-separated processes:

$$\frac{\partial \mathbf{y}}{\partial t} = \mathcal{R}_g(t, \mathbf{y}) + \mathcal{R}_f(t, \mathbf{y}), \quad \|\mathcal{R}_f\| \gg \|\mathcal{R}_g\|,$$

where, for global atmospheric models, the scale-separation occurs due to the order-of-magnitude difference in grid-spacings in the horizontal and vertical directions, that is $z_h \ll s_h$. In this context, \mathcal{R}_f contains vertically-propagating processes and \mathcal{R}_g contains horizontally-propagating terms. The problem naturally lends itself to an IMEX (Implicit–Explicit) approach, examples of which have been widely analyzed in the literature for use in very stiff diffusion-dominated (parabolic) systems. For the atmospheric (hyperbolic) system, IMEX schemes have only recently been analyzed in the context of HEVI solutions. The analyses

have tended to focus on (single time-step) Runge–Kutta (RK) based approaches [8, 23, 40, 63, 100, 106], which inherently avoid any problems associated with the computational modes that derive from multi-step methods.

A ν -stage IMEX-RK method can be expressed by a so-called “double Butcher tableau”

$$\begin{array}{c} : \\ \tilde{c} | \tilde{A} \\ \tilde{b} \end{array} = \frac{\begin{array}{c} \tilde{c}_1 \\ \vdots \\ \tilde{c}_\nu \end{array} \left| \begin{array}{ccc} \tilde{\alpha}_{11} & \cdots & \tilde{\alpha}_{1\nu} \\ \vdots & & \vdots \\ \tilde{\alpha}_{\nu 1} & \cdots & \tilde{\alpha}_{\nu\nu} \end{array} \right.}{\begin{array}{c} \tilde{b}_1 \\ \cdots \\ \tilde{b}_\nu \end{array}} \quad \begin{array}{c} c | A \\ b \end{array} = \frac{\begin{array}{c} c_1 \\ \vdots \\ c_\nu \end{array} \left| \begin{array}{ccc} \alpha_{11} & \cdots & \alpha_{1\nu} \\ \vdots & & \vdots \\ \alpha_{\nu 1} & \cdots & \alpha_{\nu\nu} \end{array} \right.}{\begin{array}{c} b_1 \\ \cdots \\ b_\nu \end{array}} \\ \tilde{\alpha}_{ij} = 0, \text{ for } j \geq i \text{ (explicit)} \quad \alpha_{ij} = 0, \text{ for } j > i \text{ (diagonally implicit),}$$

combinations of schemes—both through linear analyses [23, 40, 63] and numerical simulations [23, 40, 106]. The analyses focus on the accuracy and stability implied for components of simple linear systems (acoustic and gravity waves, and advection); and the performance of numerical implementations for idealised (dry) atmospheric tests, often

where $\tilde{c}_i = \sum_{j=1}^{\nu} \tilde{\alpha}_{ij}$ and $c_i = \sum_{j=1}^{\nu} \alpha_{ij}$, and $\sum \tilde{b}_j = 1$ and $\sum b_j = 1$. Applied to the atmospheric system of interest, subject to a HEVI-based discretization, this notation leads to:

$$\mathbf{Y}^{(j)} = \mathbf{y}^n + \Delta t \sum_{\ell=1}^{j-1} \tilde{\alpha}_{j\ell} \mathcal{R}_g \left(t^n + \tilde{c}_j \Delta t, \mathbf{Y}^{(\ell)} \right) + \sum_{\ell=1}^j \alpha_{j\ell} \mathcal{R}_f \left(t^n + c_j \Delta t, \mathbf{Y}^{(\ell)} \right), \tag{9}$$

$$\mathbf{y}^{n+1} = \mathbf{y}^n + \Delta t \sum_{j=1}^{\nu} \tilde{b}_j \mathcal{R}_g \left(t^n + \tilde{c}_j \Delta t, \mathbf{Y}^{(j)} \right) + \sum_{j=1}^{\nu} b_j \mathcal{R}_f \left(t^n + c_j \Delta t, \mathbf{Y}^{(j)} \right), \tag{10}$$

where the explicit scheme is used to integrate the horizontally-propagating terms in \mathcal{R}_g , and the implicit scheme is used for the vertically-propagating terms in \mathcal{R}_f .

Similar to the SE approach, the aim is to optimize the computational cost by selecting a relatively cheap (due to its local nature) but appropriately accurate (say, 3rd order) conditionally stable explicit scheme, which places a limit on the time-step Δt ; and a less accurate (say, 2nd order) unconditionally stable implicit scheme to handle the large Courant number vertical processes. The loss of accuracy in computing the vertical processes is offset by their lesser physical importance. As for the SE case, the implicit problem is also cheap (e.g., a tridiagonal system), since vertical grid columns remain complete on each compute node.

Expressing the RK-HEVI approach as a double Butcher tableau makes it efficient to explore many alternative

compared to solutions from a very high-resolution high-order explicit RK method. Substantially different approaches have been recommended: some completely splitting the vertical and horizontal integrations using Strang-type splitting [8, 97, 100]; others proposing schemes that keep the vertical and horizontal solutions balanced in time, by integrating over the same time-interval, at each predictor-stage [40, 63, 106]. In keeping with the semi-implicit approach (Sect. 3.2), [106] stresses the importance of ending the integration with a stage that includes an implicit integration, thereby ensuring a balanced final solution.

Colavolpe et al. [23] proposes a further extension to the double Butcher tableau approach—a quadruple Butcher tableau, whereby the horizontal pressure-gradient and divergence terms are treated separately to the horizontal advection. Under their scheme, all four solutions are balanced in time at each predictor-stage, but in addition, a forward–backward type operation is introduced for the pressure-gradient and divergence terms (based on [68]) that alternates the forward/backward operations for the

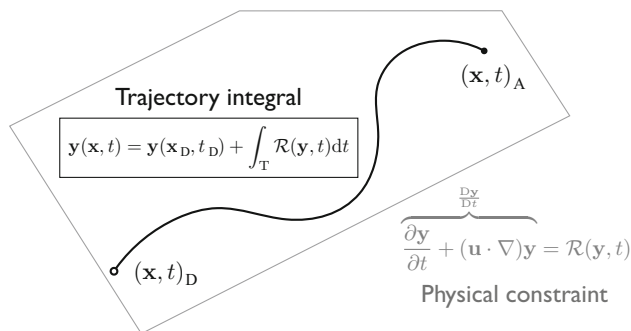


Fig. 4 Conceptual schematic of PBTI schemes

windspeed/pressure solutions at two predictor stages. They demonstrate that the additional splitting brings greater stability and accuracy at no extra computational cost.

Only one operational global NWP model but a number of global non-hydrostatic research models (cf. Table 1) have adopted HEVI time-integration methods. This trend seems to indicate that HEVI schemes are being considered as a valuable alternative to more commonly adopted time-integration strategies (such as the semi-implicit semi-Lagrangian method), especially for high-resolution models.

3.2 Path-based time-integration (PBTI)

PBTI schemes known also as *Lagrangian* methods, solve the original IBVP simultaneously in space and time without separating the BVP from the IVP. The PDEs in this case are seen as physical constraints on the path that can be followed to connect two states in the four-dimensional time-space continuum [93], as depicted in Fig. 4. The connection between two arbitrary states is obtained through a trajectory integral applied to the equation

$$\overbrace{\frac{\partial \mathbf{y}}{\partial t}}^{\frac{D\mathbf{y}}{Dt}} + (\mathbf{u} \cdot \nabla)\mathbf{y} = \mathcal{R}(\mathbf{y}, t). \tag{11}$$

In Eq. (11), the advective term (where \mathbf{u} is the advection velocity that can eventually depend on \mathbf{y}) on the left-hand side is absorbed into the material (or path) derivative $\frac{D\mathbf{y}}{Dt}$ and the right-hand side \mathcal{R} consists of forcing terms, e.g., the pressure-gradient term, the Coriolis terms and the source terms arising from the parametrization of the sub-grid physical processes in an NWP model.

The trajectory integral of \mathcal{R} can be approximated using a weighted average of the integrand values from the physical space at a past time t_D , the “departure time” where the state of the system is known, and the values from the physical space at a future time t_A , the “arrival time”, where the solution is sought. In the PBTI approach when the approximation solely relies on the past values, the integration scheme is *explicit*, while when the approximation depends on the future (unknown) values of the integrand \mathcal{R} , the integration scheme is *implicit*. In the case of explicit integration schemes, the solution of the system is fairly straightforward but the approximation can become numerically unstable, for time-steps exceeding the Eulerian CFL condition of the fast processes, leading to failure of the simulation. On the other hand, in the case of implicit integration schemes, such as the commonly used trapezoidal rule (semi-implicit Crank–Nicholson), the approximation is guaranteed to be unconditionally stable but the resulting system of equations, usually in the form of a

BVP, becomes more complicated and its numerical solution more difficult.

PBTI techniques have been very successful in NWP (as indicated in Table 1 with adoption of the technique as late as 2014). The most common PBTI strategy employed in NWP is the Semi-Lagrangian (SL) scheme that revolutionized the field two decades ago [94, 98]. Pure Lagrangian approaches, where the exact solution at the next time step is sought by translating the flow information on the mesh at the current time level along the trajectory integrals, with remapping only for postprocessing purposes, have never been adopted into operational NWP models. This has been mainly due to the initial mesh being significantly deformed in a few time-steps which results in large spatial truncation errors. Having a mesh with vertically aligned grid-points is essential for resolving complex diabatic processes and vertically-propagating gravity waves in a NWP model. In contrast to the pure Lagrangian approach, there is no mesh deformation in a semi-Lagrangian scheme. The reason is that backward trajectories are calculated at each time-step; these end at the model grid-points while they start from locations between mesh grid-points that must be determined.

More recently, forward-in-time finite volume (FTFV) integrators, that can be written in a congruent manner as the SL scheme, have also emerged [91] with applications in NWP and climate. Furthermore, vertical Lagrangian coordinates have been successfully applied in hydrostatic models [46].

3.2.1 The semi-implicit semi-Lagrangian scheme (SISL)

The semi-Lagrangian (SL) method [74, 94] is an unconditionally stable scheme for solving the generic transport equation

$$\frac{Dy}{Dt} = S, \quad \frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla, \quad \mathbf{u} = (u, v, w) \tag{12}$$

where \mathbf{u} denotes a wind vector, y a transported variable and S a source term. Beyond stability, an additional strength of the SL numerical technique is that it exhibits very good phase speeds and little numerical dispersion (see, e.g., [38, 94]). Because of these properties, SL solvers can integrate the prognostic equation sets of atmospheric models stably with long time-steps at Courant numbers much larger than unity, without distorting the important atmospheric Rossby waves. When a SL scheme is coupled with a semi-implicit (SI) time discretization, long time-steps can be used in realistic atmospheric flow conditions where a multitude of fast and slow processes coexist. In semi-implicit semi-Lagrangian (SISL) schemes the high-speed gravity waves associated with high-frequency fluctuations in the wind divergence are mitigated, where the

terms responsible for the gravity waves are identified and treated in an implicit manner, thereby slowing down the fastest gravity waves.

The SISL approach is currently the most popular option for operational global NWP models while it is also often used in limited area modeling. As shown in Table 1 the vast majority of the listed global NWP centers are using a model with a SISL dynamical core. A typical example is the ECMWF forecast model IFS, which has used a SISL approach since 1991. As discussed in [86], the change from Eulerian to semi-Lagrangian numerics improved the efficiency of IFS by a factor of six thus enabling a significant resolution upgrade at that time. Since 1991, further successful upgrades followed and currently the (high resolution) global forecast model is run at 9 km resolution in grid-point space, to this date the highest in the world.

To explain how a SISL method works we shall write the prognostic equations of the atmosphere in the compact form:

$$\frac{D\mathbf{y}}{Dt} = \mathcal{R}(\mathbf{y}), \tag{13}$$

where $\mathbf{y} = (y_i)$, $i = 1, 2, \dots, N$ is a vector of N three-dimensional prognostic scalar fields y_i (such as the wind components, temperature, density, water vapour and other tracers) and $\mathcal{R} = (R_i)$ is the corresponding forcing term. Integrating (13) along a trajectory, which starts at a point in space D, the departure point, and terminates at a point in space A, the arrival point

$$\frac{\mathbf{y}_A^{t+\Delta t} - \mathbf{y}_D^t}{\Delta t} = \int_t^{t+\Delta t} \mathcal{R}(\mathbf{y}(t))dt \tag{14}$$

and approximating the right-hand side integral using the second order trapezoidal scheme yields the following SISL discretization

$$\frac{\mathbf{y}_A^{t+\Delta t} - \mathbf{y}_D^t}{\Delta t} = \frac{1}{2}(\mathcal{R}_D^t + \mathcal{R}_A^{t+\Delta t}). \tag{15}$$

In any SISL scheme there are three crucial steps that influence the numerical properties of the discretization, namely (i) the calculation of the departure point locations and the related interpolation of the prognostic variables at these points, (ii) the semi-implicit time discretization of the nonlinear forcing terms and (iii) the solution of the final semi-implicit system reduced in the form of a Helmholtz elliptic equation. We detail each of these steps in the following.

(i) *SL advection and calculation of the departure points*

All operational SL codes work “backwards” in the sense that at a given discrete point in time t and with a model time-step of Δt an air-parcel will start from a point in space between grid-points and will terminate at a given mesh

grid-point. The latter are called “arrival points” and coincide with the model mesh grid points while the former are called “departure points” and they must be found as they are not known a priori. There is a unique departure point associated with each grid-point to be computed (this assumes that characteristics do not intersect, i.e., no discontinuities are permitted). Therefore, for a simple passive scalar advection of a generic field y without forcing, the solution at a new time step is:

$$y_A^{t+\Delta t} = y_D^t. \tag{16}$$

This means that to compute the field y values at the new time-step $t + \Delta t$, it suffices to compute a departure point “D” for each model grid point and then interpolate the transported field y at these departure points. The interpolation method uses the known y -values at time t , at a set of grid points nearest to “D”; the number and location of these grid points depend on the order of interpolation method used. For the more general problem (13), the forcing terms should also be interpolated at the departure point. To compute the location of the departure points the following trajectory equation must be solved:

$$\frac{D\mathbf{r}}{Dt} = \mathbf{u}(\mathbf{r}, t), \tag{17}$$

where \mathbf{r} denotes the coordinates of a moving fluid parcel, for example $\mathbf{r} = (x, y, z)$ if a Cartesian system is used. By integrating Eq. (17), we obtain:

$$\mathbf{r}_A - \mathbf{r}_D = \int_t^{t+\Delta t} \mathbf{u}(\mathbf{r}, t)dt. \tag{18}$$

The right-hand side integral of (18) is usually approximated using a 2nd order scheme such as the midpoint rule, resulting in an implicit equation of the form

$$\mathbf{r} - \mathbf{r}_D = \Delta t \mathbf{u}\left(\frac{\mathbf{r} + \mathbf{r}_D}{2}, t + \frac{\Delta t}{2}\right), \tag{19}$$

which is solved iteratively (for details see [27]). The accuracy with which the departure points are computed influences greatly the overall accuracy of the model as shown in [27].

In addition, the method employed to interpolate the terms of Eq. (13) to the departure points has also important implications in the model accuracy. From this perspective, it is common practice in operational SISL models to use a cubic interpolation formula most often based on tri-cubic Lagrange interpolation followed by formulae based on cubic Hermite or cubic spline polynomials. The interpolation is directional, i.e., it is performed separately in each of the three spatial coordinates. There is an intriguing interplay between the spatial and time truncation error in the SL advection method. Following the convergence analysis in [31], verified experimentally in [112] using the Navier–Stokes system, the leading order truncation error term for a SL

method solving a 1D constant wind advection equation with an interpolation formula of order p on a grid with constant spacing Δx and a time-integration method for the departure point of order k with time-step Δt is $\mathcal{O}(\Delta t^k + \Delta x^{p+1}/\Delta t)$. This suggests that reducing the time-step only without refining the mesh resolution may not improve the overall solution accuracy as it increases the contribution from the error term which has Δt in the denominator. However, with a shorter time-step the accuracy of the departure point calculation improves and a higher order interpolation scheme improves the accuracy of spatial structures such as waves [29].

(ii) *Semi-implicit time-discretization of forcing terms* Eq. (15) is expensive and complex to solve due to its large dimension, its implicitness and in general its nonlinear form (right hand-side \mathcal{R} includes nonlinear terms). For this reason, an approach commonly used in NWP is to extract fast terms from the right-hand side and linearise them around a constant reference profile. For example, in the IFS model the right-hand forcing term is split as follows:

$$\mathcal{R} = \mathcal{N} + \mathcal{L}$$

where \mathcal{L} contains the linear and linearised fast terms which are integrated implicitly and \mathcal{N} the remaining nonlinear terms $\mathcal{N} = \mathcal{R} - \mathcal{L}$ which are integrated explicitly. A two-time-level second order SISL discretization of (15) can be written as follows:

$$\frac{\mathbf{y}_A^{t+\Delta t} - \mathbf{y}_D^t}{\Delta t} = \frac{1}{2}(\mathcal{L}_D^t + \mathcal{L}_A^{t+\Delta t}) + \frac{1}{2}(\mathcal{N}_D^{t+\Delta t/2} + \mathcal{N}_A^{t+\Delta t/2}). \quad (20)$$

The slowly varying nonlinear terms at $t + \Delta t/2$ can be “safely” approximated by a second order extrapolation formula such as

$$\mathcal{N}^{t+\Delta t/2} = \frac{3}{2}\mathcal{N}^t - \frac{1}{2}\mathcal{N}^{t-\Delta t}$$

or alternatives such as SETTLS [49], which are less prone to generate numerical noise. The latter aspect—i.e., the numerical noise issue—is particularly relevant in the stratosphere where large vertically stable areas occur and any small scale oscillations appearing due to the three time-level form of the extrapolation formula used may be amplified. Using “iterative semi-implicit” schemes [11, 26, 111], in which a future model state is predicted with 2-iterations (or more) with the first serving as a predictor and the second as a corrector, is the most effective method for solving the noise issue. However, it is more costly due to its iterative nature.

There is considerable variation in the implementation of the SI time-stepping by different models. An alternative, iterative, approach to the SI method (20) is followed by the UK Met Office (UKMO) Unified Model, where there is no separate treatment between linear and nonlinear terms.

Here, the standard off-centered semi-implicit discretization is used:

$$\frac{\mathbf{y}_A^{t+\Delta t} - \mathbf{y}_D^t}{\Delta t} = (1 - \alpha)\mathcal{R}_D^t + \alpha\mathcal{R}_A^{t+\Delta t}. \quad (21)$$

The weight α is either 0.5 or slightly larger to avoid non-physical numerical oscillations (noise) which may arise due to spurious orographic resonance [76]. To tackle the implicitness of Eq. (21) an iterative method with an outer and inner loop is used. This functions as a predictor-corrector two-time-level scheme. As stated in [66], in the outer iteration loop, the departure-point locations are updated using the latest available estimates of the winds at the next time step. In the inner loop the nonlinear terms, together with the Coriolis terms, are evaluated using estimates of the prognostic variables obtained at the previous iteration. Further details on the iterative approach followed by the UM can be found in [66, 111]. Iterative SI schemes are expensive algorithms, however, they are used by most non-hydrostatic SISL global models as in practice the cheaper non-iterative schemes based on time-extrapolation become unstable when long time steps are used.

(iii) *Helmholtz solver* Once the right-hand side of (20) has been evaluated the semi-implicit system can be solved. To avoid solving simultaneously all implicit equations in (20), it is common practice to derive a Helmholtz equation from these. The form of the Helmholtz equation depends on the type of space discretization. In spectral transform methods, such as the one used in IFS [104], the specific form of the semi-implicit system is derived from subtracting a system of equations linearised around a horizontally homogeneous reference state. The solution of this system is greatly accelerated by the separation of the horizontal and the vertical part, which matches the large anisotropy of horizontal to vertical grid dimensions prevalent in atmospheric models. In spectral transform methods, one uses the special property of the horizontal Laplacian operator in spectral space on the sphere

$$\nabla^2 \psi_n^m = \frac{n(n+1)}{a^2} \psi_n^m, \quad (22)$$

where ψ symbolises a prognostic variable, a is the Earth radius, and (n, m) are the total and zonal wavenumbers of the spectral discretization [104]. This conveniently transforms the 3D Helmholtz problem into a 2D matrix operator inversion with dimension of the vertical levels only, resulting in a very cheap direct solve [75]. Even in the non-hydrostatic context, formulated in mass-based vertical coordinates [60], only the solution of essentially two coupled Helmholtz problems allow the reduction of the system in a similar way [12, 14, 115]. This technique requires a transformation from grid-point space to spectral space and vice versa at each time-step, an aspect that increases the

associated computational cost although the spectral space computations are based on FFTs and matrix–matrix multiplications that are well suited for modern computing architectures. One disadvantage of this technique is the need for a somewhat simple reference state that does not allow, by definition, the inclusion of horizontal variability (as would be desirable for terms involving orography). The relaxation of this constraint and some alternatives are discussed, for example, in [17].

For grid point models using finite differences, such as the UKMO Unified Model, a variable coefficient 3D Helmholtz problem is solved using an iterative Krylov subspace linear solver (e.g., BiGCstab, GCR(k), GMRES etc.) [111]. This type of solver is generally more expensive despite grid point models not requiring transformations from spectral to grid point space and vice versa, which offsets some of the extra cost. However, typically up to 80 percent of computations are spent in the solver in grid-point based semi-implicit methods, compared to 10–40% in spectral transforms (depending on the resolution and the number of MPI communications involved) on today's high performance computing (HPC) architectures. For emerging and future architectures, that may heavily penalize global communication patterns moving to high throughput capabilities, this is a serious concern that needs to be properly investigated and addressed.

3.3 Summary

In this section, we categorized time-integration schemes into two classes Eulerian-based, EBTI, and path-based, PBTI. The former discretizes the original PDE problem in space first, thus obtaining an ODE, and subsequently in time through a suitable time-integration strategy. The PBTI class, instead, solves the PDE problem in a single step, where the advection term is adsorbed into the path (or material) derivative and the right-hand side is formed by forcing terms only. In this case, the system of PDEs can be seen as a physical constraint on the path that can be followed to link two states in the four-dimensional continuum constituted by space and time.

For each of the two categories, EBTI and PBTI, we outlined the most prominent time-integration schemes adopted in the NWP and climate communities. In particular, SE and HEVI for the EBTI class, and the SISL approach for the PBTI class. The latter was the most widely adopted in the past few decades thanks to the hydrostatic approximation ubiquitously used in global weather and climate models. Indeed, PBTI strategies and SISL in particular, were extremely competitive given the large time-steps they allow and their extremely favorable dispersion properties, which yield the correct representation of wave-like solutions—e.g., Rossby and gravity

waves. EBTI strategies are instead now emerging as a potential alternative to PBTI in non-hydrostatic models. This is mainly because they can be constructed to ‘filter’ the fast and atmospherically irrelevant acoustic modes that propagate vertically. In fact, SE and HEVI schemes that had been mainly used in the context of limited-area models are now being taken into consideration in global weather models, as they seem to represent an attractive compromise between solution accuracy, time- and energy-to-solution and reliability. In addition, they can address the non-conservation issues typical of PBTI-based approaches, a feature that is particularly relevant for climate simulations. Some additional strategies, beyond HEVI, SE and SISL methods are also under investigation within global NWP and climate simulations, namely IMEX schemes, fully implicit methods and conservative semi-Lagrangian schemes—the latter addressing the conservation issues of traditional SISL schemes. These additional strategies will be briefly discussed in Sect. 5.

Note that while the time-integration approaches described in this section might differ in their implementation details from one model to another, the general concepts and properties, which motivate their use, hold true across all the models adopting each strategy. In the next Sect. 4, we will highlight the time-stepping strategies employed by the main operational global NWP and climate centres and emphasize in more detail why these were selected. In addition, we will outline the implications these choices may have in the context of the changing hardware and weather modeling landscape. Finally, we will introduce some of the (several) projects undertaken within the weather and climate industry to address the computational challenges of the incoming decades.

4 Overview of Operational Time-Integration Strategies Adopted by the NWP and Climate Centers

The time-stepping practices adopted by the leading NWP and climate operational centers are reported in Table 1, where we also indicate the model name, the application type, the set of equations solved—i.e., either hydrostatic (H) or non-hydrostatic (NH), the specific time-integration scheme used and its class—i.e., either EBTI or PBTI, the approximate date of operational activity or documented performance, and the country of origin. All these models vary widely in respect to the choices of the physical parameterization and spatial discretization. However, here we concentrate on the time-integration schemes.

From this point of view, Table 1 clearly shows that nearly all operational NWP and climate models use PBTI (SISL) approaches and the majority use the hydrostatic

approximation—see top part of the Table, where, apart from the German model ICON that uses a non-hydrostatic model and a HEVI scheme, all the others use PBTI schemes, and SISL in particular. These choices are dictated by a desire to maximize time-to-solution performance, that has been (and still is) one of the main objectives in the industry. It is also worth noting that, despite the continuous upgrade of the operational models—see the column “Date” in the Table—all of them keep adopting SISL, even if they transitioned to a non-hydrostatic approximation. This might be an indicator of the inevitable algorithmic inertia within the weather and climate industry due to the strict operational constraints and the relatively complex and large code frameworks developed over several decades.

On the other hand, if we look at research models (bottom part of Table 1, denoted with RD), they are predominantly non-hydrostatic and use EBTI approaches with substantially smaller permitted time-steps. The use of EBTI schemes is justified by their better parallel efficiency (than SISL) that may compensate for the shorter forward-in-time stepping of the dynamical core. Indeed, there seems to be a trend in the weather and climate industry to favour EBTI approaches over PBTI for next-generation non-hydrostatic models, as briefly mentioned in Sect. 3. The associated spatial discretization for the spatial derivatives arising on the right-hand side of each time-stepping strategy can assume various forms that include global spherical harmonics bases, finite-differences, finite-volumes, and spectral element methods (e.g., continuous and discontinuous Galerkin).

Given what we discussed above, it should be clear that the choice of the time-stepping strategy is strongly influenced by the choice of equations, as most non-hydrostatic equations require numerical control of vertical acoustic wave propagation, whereas hydrostatic models filter these modes *a priori*. The stiffness of the problem arising from the fast propagating acoustic waves needs to be handled carefully. Options include SE methods, HEVI schemes, and, more recently, IMEX methods (discussed in Sect. 5). In essence, many of the approaches originally used in limited-area modeling have been applied in non-hydrostatic global models. It should be also noted that these schemes are favourable in terms of parallel communication patterns, as they require a reduced amount of data to be transferred at each time-step, being mostly nearest-neighbour communication algorithms, and have a higher throughput (flops/communication) than PBTI approaches. This attractive feature is particularly suited to emerging computing architectures that are commonly communication bounded. However, the larger number of time-steps (due to shorter permitted time-steps), can counter balance

this advantage in favour of larger-time-step algorithms, including SISL.

Overall, the suitability to emerging hardware architectures of a given time-stepping strategy is of crucial importance for planning a sustainable path to future high-resolution weather and climate prediction systems, as there will be the pressing need to mitigate the exponentially increasing running costs of operational centres. In fact, the use of emerging many-core computing architectures (e.g., GPUs and MICs) help reducing the power consumption of the overall HPC centre where the simulations are run (the power required grows exponentially with the clock-speed of the processor; therefore the use of many-core computing technologies that have a lower clock rate mitigate the growth in energy consumption—see e.g., [9]). With the increased model resolutions envisioned in the next few decades, another critical aspect is the efficient treatment of a massive amount of gridded data (the so-called *data tsunami problem*). In this case, the adopted time-stepping strategy has important implications, since a larger number of time-steps could imply a larger amount of gridded data produced. This obviously has direct consequences on the costs to maintain and power the servers where the data are stored, and might affect the efficient post-processing and dissemination (to the clients or member states) of the results.

Indeed, these issues are being closely evaluated by several weather and climate agencies through various projects, including the European Exascale Software Initiative (EESI), Energy efficient SCalable Algorithms for weather Prediction at Exascale (ESCAPE) and Excellence in Simulation of Weather And Climate in Europe (ESi-WACE). All these initiatives have the general recommendation for research in the areas of development of efficient numerical methods and solvers capable of complying with exabyte data sets and exascale computational efficiency—i.e. next-generation HPC systems.

The efforts being spent to address the computational efficiency issues of weather and climate algorithms run in parallel to the development of the high-resolution medium-range global and nested NWP models, as many government agencies are extensively researching new candidates for their operational services. In fact, these new models must be able to make a quantum leap forward in forecast skill, by being able to efficiently utilize the latest developments in data assimilation (e.g., 4D-En-Var [45]), scale-aware physical parameterizations, and, as just mentioned, modern HPC technologies (e.g., GPU/MIC). From this perspective, several international and interagency activities have been organized to select best candidates for next-generation models. For example, the NOAA High Impact Weather Prediction Project (HIWPP), seeks to improve hydrostatic-scale global modeling systems and demonstrate their skill

at the coarsest resolution down to ~ 10 km. They are also trying to accelerate the development and evaluation of higher-resolution non-hydrostatic, global modeling systems at cloud-resolving (~ 3 km) scales. As a part of Research to Operations (R2O) Initiative, the NOAA / National Weather Service (NWS) led the inter-agency effort to develop a unified Next Generation Global Prediction System (NGGPS) for 0–100 days predictions to be used for the next 10–20 years. The new prediction model system designed to upgrade the current operational system, GFS, has to be adaptable to and scalable on evolving HPC architectures. The research and development efforts included the U.S. Navy, NOAA, NCAR and university partners. A similar effort has been undertaken in Japan with the Japan Meteorological Agency (JMA), which has been exploring its own next generation non-hydrostatic global numerical weather prediction model since 2009. The UK MetOffice is also pursuing new modeling efforts through its development of a scalable dynamical core “GungHo” (Globally Uniform, Next Generation, Highly Optimised) [95]. The dynamical cores are also a subject of many intercomparison projects, such as the Dynamical Core Model Intercomparison Project (DCMIP) [101], which in a few editions has hosted more than 20 different models. In addition to the current operational models, several research models may offer good candidates for next generation operational systems replacing the current state-of-the-art solutions.

The overarching objective is to increase the computational efficiency of the models, thereby reducing the operating costs (e.g., energy-to-solution) and the time-to-solution performance, in order to increase the accuracy and resolution of the models at a sustainable economic cost. This aspect is intimately related to the overall co-design of the algorithms underlying weather and climate models where current and emerging hardware and the adopted time-integration method will play a major role. Specifically, the amount of data communicated during each simulation, alongside the length of permitted time-step and the percentage of peak performance achieved on a given computing machine, are a direct consequence of the chosen numerical discretization, and in particular the time-integration. Emerging and future time-integration strategies should take into account all these factors to provide an effective path to sustainable NWP and climate simulations.

5 Emerging Alternatives

For both EBTI and PBTI, we outline some emerging alternatives that are being considered within the community, namely IMEX and fully implicit schemes for the EBTI class, and conservative semi-Lagrangian schemes for

the PBTI class. The investigation of these time-stepping schemes for weather and climate is aligned with the general guidelines provided in Sect. 4, where co-design of software and hardware, energy-to- and time-to-solution are key players. In fact, in terms of EBTI strategies, there seems to be a trend to explore (semi-)implicit solutions, that eventually maximize the time-to-solution but would require the development of efficient algorithms to achieve increased percentage of peak performance (of a given machine) and better scalability properties. On the other hand, for PBTI strategies, efforts have been spent on developing conservative SISL schemes, thereby addressing the concerns of lack of accuracy for long-range weather and climate simulations.

5.1 EBTI Strategies

Both the strategies described below aim to improve the time-to-solution requirement of weather and climate simulations. In addition, if coupled with compact-stencil spatial discretizations, including finite volume and spectral element methods, they can be efficiently used on emerging computing architectures, due to the reduced parallel communication costs required. However, the solution procedure for both will require iterative elliptic solvers and iterative Newton-type methods, with all the associated numerical complexities that are to be addressed to make them a suitable alternative for operational NWP and climate.

5.1.1 Implicit–Explicit (IMEX) Methods

Although not as prevalent as HEVI schemes, IMEX methods are gaining interest in many fields (including the geosciences) for evolving PDEs forward in time. We can outline IMEX methods starting from the scale-separated problem described earlier:

$$\frac{\partial \mathbf{y}}{\partial t} = \mathcal{R}_g(t, \mathbf{y}) + \mathcal{R}_f(t, \mathbf{y}), \quad \|\mathcal{R}_f\| \gg \|\mathcal{R}_g\|,$$

where, for global atmospheric models, the scale-separation occurs due to the physical processes contained within \mathcal{R}_f and \mathcal{R}_g . For instance, \mathcal{R}_f contains a linearized form of not just the vertically-propagating acoustic waves but also the horizontally propagating ones as well, while \mathcal{R}_g contains the remainder of the nonlinear processes. Therefore, we would apply a double Butcher tableaux to this form of the equations, where we use the explicit tableau for \mathcal{R}_g and the implicit tableau for \mathcal{R}_f . This would result in the exact same form given in Eqs. (9)–(10). The difference between the IMEX method described here and the HEVI scheme described previously is that the resulting linear implicit problem is now fully three-dimensional; whereas in the

HEVI case, it is only one-dimensional along the vertical direction. What this means for the IMEX method is that the solution procedure will require iterative (elliptic) solvers which are usually handled via matrix-free approaches since the resulting system of equations will be too large to store in memory—for HEVI methods, the resulting system is quite small and hence direct solvers are appropriate. Note that the HEVI and IMEX methods can be written within the same unified temporal discretization as described in [40, 106] where the HEVI scheme is denoted as the 1d-IMEX method, meaning that HEVI is simply an IMEX method that has been partitioned to be implicit in only one of the spatial directions. Further note that whereas HEVI schemes only circumvent the CFL condition related to the vertically-propagating acoustic waves, IMEX methods entirely circumvent the CFL condition related to all acoustic waves. However, IMEX methods still must adhere to the explicit CFL condition due to the nonlinear wind field. IMEX methods become competitive with HEVI methods only when the vertical to horizontal grid aspect ratio, z_h/s_h , approaches unity, when the stiffness of the problem is no longer dominated by the vertical direction due to the anisotropy of the grid.

5.1.2 Fully-Implicit Methods

While the most common option remains HEVI schemes, many research groups are exploring variants of fully-implicit methods. The fully-implicit approach begins in a similar fashion to that for the explicit problem

$$\frac{\partial \mathbf{y}}{\partial t} = \mathcal{R}(t, \mathbf{y}),$$

where \mathcal{R} contains the full nonlinear right-hand-side operator. Next, we apply the implicit Butcher tableau (using the same notation as previously for the IMEX methods) to \mathcal{R} , resulting in the following fully-discrete problem:

$$\mathbf{Y}^{(j)} = \mathbf{y}^n + \sum_{\ell=1}^j \alpha_{j\ell} \mathcal{R}(t^n + c_j \Delta t, \mathbf{Y}^{(\ell)}) \quad (23)$$

$$\mathbf{y}^{n+1} = \mathbf{y}^n + \sum_{j=1}^v b_j \mathcal{R}(t^n + c_j \Delta t, \mathbf{Y}^{(j)}), \quad (24)$$

which looks conceptually simpler than the IMEX problem given by Eqs. (9)–(10). However, the complexity comes in through the vector R which is a fully three-dimensional nonlinear function of \mathbf{y} . Consequently, we must solve this problem iteratively using Newton-type methods. To simplify the exposition, let us assume that we are using a 1st-order implicit RK method (i.e., backward Euler, which is strong-stability preserving and highly stable) as follows:

$$\mathbf{y}^{n+1} = \mathbf{y}^n + \Delta t \mathcal{R}(t^{n+1}, \mathbf{y}^{n+1}). \quad (25)$$

Before employing Newton's method, we first write Eq. (25) as the functional

$$\mathbf{F}^{n+1} = \mathbf{y}^{n+1} - \mathbf{y}^n - \Delta t \mathcal{R}(t^{n+1}, \mathbf{y}^{n+1}) \equiv 0$$

where we note that a 2nd order Taylor series expansion yields

$$\mathbf{F}^{n+1} = \mathbf{F}^n + \frac{\partial \mathbf{F}^n}{\partial \mathbf{y}} (\mathbf{y}^{n+1} - \mathbf{y}^n) \equiv 0$$

and, after rearranging, results in the classical Newton's method

$$\frac{\partial \mathbf{F}^n}{\partial \mathbf{y}} (\mathbf{y}^{n+1} - \mathbf{y}^n) = -\mathbf{F}^n. \quad (26)$$

Note that Eq. (26) is solved iteratively whereby we replace $n \rightarrow k$ where k is the iteration counter such that $k = 0$ implies starting from the previous time-step (t^n) as follows:

$$\frac{\partial \mathbf{F}^{(k)}}{\partial \mathbf{y}} \Delta \mathbf{y} = -\mathbf{F}^{(k)}, \quad \mathbf{y}^{(0)} = \mathbf{y}^n \quad (27)$$

with $\Delta \mathbf{y} = (\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)})$. Equation (27) represents a linear fully three-dimensional matrix problem that must be solved until an acceptable stopping criterion is reached (e.g., $\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|_p < \epsilon_{stop}$ where p is a selected norm). In classical Newton-type methods, one of the largest costs is due to the formation of the Jacobian matrix $\mathbf{J} = \frac{\partial \mathbf{F}}{\partial \mathbf{y}}$. However, this cost can be substantially mitigated by the introduction of Jacobian-free Newton-Krylov methods (see, e.g., [58]) whereby we recognize that the Jacobian can be approximated as follows

$$\mathbf{J}^{(k)} = \frac{\mathbf{F}(\mathbf{y}^{(k)} + \epsilon \Delta \mathbf{y}) - \mathbf{F}(\mathbf{y}^{(k)})}{\epsilon \Delta \mathbf{y}}$$

(where ϵ is, e.g., machine zero) and direct substitution into Eq. (27) leads to

$$\frac{\mathbf{F}(\mathbf{y}^{(k)} + \epsilon \Delta \mathbf{y}) - \mathbf{F}(\mathbf{y}^{(k)})}{\epsilon} = -\mathbf{F}^{(k)}, \quad (28)$$

which we can write in the residual form

$$\mathbf{Res} = \frac{\mathbf{F}(\mathbf{y}^{(k)} + \epsilon \Delta \mathbf{y}) - \mathbf{F}(\mathbf{y}^{(k)})}{\epsilon} + \mathbf{F}(\mathbf{y}^{(k)}). \quad (29)$$

Equation (29) is a linear fully three-dimensional problem that has been written in a matrix-free (Jacobian-free) form since we only need to construct the vectors \mathbf{F} and then solve the residual problem only approximately by any Krylov subspace method (e.g., GMRES, BiCGStab, etc.). One of the advantages of the Jacobian-free Newton-Krylov (JFNK) method described is that we can exploit the iterative nature of both the Krylov and Newton methods,

meaning that, typically, the Krylov stopping criterion need not be as stringent as it would be for an IMEX method as long as the final solution in the Newton solver satisfies certain physical conditions.

Fully-implicit methods are unconditionally stable, like the PBTI methods described previously. However, this flexibility in taking any length time-step size (restricted only by accuracy considerations) is in practice off-set by the prohibitive cost of the iterative solution of both the inner (Krylov solution) and outer (Newton solution) loops of the JFNK method. The condition number of the linear Krylov solution is proportional to the time-step size so taking a large time-step size translates into an increase in the number of Krylov iterations. For this reason it is especially important to choose the proper Krylov method (e.g., the cost of GMRES increases quadratically with the number of iterations/Krylov vectors). Preconditioners become all the more important for this class of time-integration methods if one wishes to build competitive (using the time-to-solution metric) strategies. In fact, in the past two Supercomputing conferences, the Gordon Bell prizes have been awarded to geoscience models using fully-implicit time-integrators—in 2015 for the simulation of the Earth’s mantle [78] and in 2016 for the (dry) simulation of atmospheric flows [113].

5.2 PBTI Strategies

The main drawback of SISL schemes is their inability to conserve, e.g. mass and scalar tracers. Conservative semi-Lagrangian methods exist and are being developed, with the main issue that they usually require computationally demanding re-meshing at each time-step. This is due to the varying-control-volumes in time required to maintain the exact conservative nature of the method. Therefore, their use in the context of weather and climate is limited and currently being investigated.

5.2.1 Conservative Semi-Lagrangian Methods

In Eq. (11), we described the classical semi-Lagrangian methods that are typically used in operational NWP and climate models and we stated that they do not formally conserve, e.g., mass, although in practice they conserve it to within an acceptable level. However, conservative semi-Lagrangian methods do exist and we describe a specific class of them below.

Instead of writing the original problem in Eq. (11) as follows

$$\frac{\partial \mathbf{y}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{y} = \mathcal{R}(\mathbf{y}, t), \tag{30}$$

we can write it in conservation form, that is

$$\frac{\partial \mathbf{Y}}{\partial t} + \nabla \cdot (\mathbf{Y}\mathbf{u}) = \mathcal{R}(\mathbf{Y}, t), \quad \mathbf{Y} = \mathbf{y}\rho \tag{31}$$

where ρ is a mass variable (e.g., density) and \mathbf{Y} is now a conservation variable.

Integrating Eq. (31) in space yields

$$\int_{\Omega_e} \left(\frac{\partial \mathbf{Y}}{\partial t} + \nabla \cdot (\mathbf{Y}\mathbf{u}) \right) d\Omega_e = \int_{\Omega_e} \mathcal{R}(\mathbf{Y}, t) d\Omega_e \tag{32}$$

where Ω_e is a control volume. Using the Reynolds transport theorem allows us to rewrite Eq. (32) as follows

$$\frac{d}{dt} \int_{\Omega_e(t)} \mathbf{Y} d\Omega_e = \int_{\Omega_e(t)} \mathcal{R}(\mathbf{Y}, t) d\Omega_e, \tag{33}$$

where $\Omega_e(t)$ is meant to imply that the shape of the control volume changes in time and, thereby, reveals the Lagrangian nature of the approach. This approach is exactly conservative provided that the integrals at the departure control volume are solved exactly (see, e.g., [37, 39, 61, 117]). The difficulty with this approach is that it requires tracking back the control volumes at the departure points which essentially amounts to the construction of the mesh defined by the departure points.

6 Discussion and Concluding Remarks

There are two key aspects that the weather and climate industry need to face in the oncoming years in terms of time-integration strategies: (i) the continuous increment in spatial resolution that is moving the equations of choice towards non-hydrostatic approximations, such that the vertical acoustic modes are not filtered a priori and a separation of horizontal and vertical motions is no longer readily achieved, placing substantially increased constraints on the time-stepping methods; (ii) the emerging computing technologies, that are significantly changing the paradigms traditionally adopted in parallel programming, thus demanding a review of algorithms and numerical solutions to maintain and possibly improve computational efficiency in next-generation HPC systems. In addition to these two points, the overall approach to the time-integration of weather and climate models should consider the operational constraints, time-to-solution (requirement to deliver a 10-days global forecast in 1 h real-time) and cost-to-solution (the latter usually translated into energy-to-solution), accuracy and robustness of the entire simulation framework.

Taking into account all these factors, it is clear that high efficiency time-stepping algorithms are required to allow operational NWP centers to complete extensive simulations with limited computer resources and satisfy the strict operational bounds. For this reason, due to the good

efficiency of semi-Lagrangian advection schemes at high advective Courant numbers, the SISL methods have been at the heart of the most successful operational NWP and climate systems, e.g. IFS, UM, GSM, GFS. SISL-based schemes guarantee boundedness of the solution and unconditional stability [96], which have the advantage—compared to explicit schemes—that they permit a relatively large time-step and a very competitive time-to-solution performance [24, 73, 105, 111, 114].

The efficiency and robustness achieved with SISL is not easily replaced by alternative choices [110], although the better scalability that can be exploited on future HPC architectures, due to less reliance on communication exchanges beyond nearest neighborhoods will work in favor of techniques with compact stencils, e.g. finite volume (FV) and spectral element methods (SEM), coupled with EBTI-based approaches. The additional problem of the convergence of meridians at the globe's poles in classical latitude-longitude grids, which results in very small simulated distances of the grid in the zonal direction (shorter than the physical distance between compute processors in some of today's very high resolution models) and, thus, extremely small time-steps, has been widely addressed. In particular, the use of reduced quasi-homogeneous grids increases the stability of the solution and icosahedral (cf. ICON, MPAS, NICAM) or cubed sphere (cf. FV3, CAM-SE, NUMA) grids improve the computational efficiency [62, 65, 72]. Despite these improvements, compact-stencil EBTI-based techniques need to overcome additional issues in order to become competitive with PBTI schemes, especially in terms of time-to-solution. From this perspective, the development of efficient parallel preconditioners and their co-design with the underlying hardware is key, especially for IMEX, semi- and fully-implicit methods. The use of compact-stencil EBTI techniques, including FV and SEM, can also allow local grid-refinement, that in conjunction with local sub-time-stepping, can allow for improved resolution in the vicinity of steep topographic slopes at the lower boundary of models (up to 70 degrees in high-resolution global models). Such features can impose severe restrictions on the time-step and subsequently undermine numerical stability [116]. Also, the conservation of important quantities, e.g. mass and scalar tracers, that is of critical importance for climate simulations, is favored in EBTI schemes compared to classical SISL (although note the recent developments in conservative semi-Lagrangian schemes outlined in Sect. 5.2).

Given all the aspects discussed, it is clear that the choice of the time-integration scheme is a constrained multifold problem, for which a single and unified answer might not exist. However, the most promising future directions in terms of time-stepping should involve three key points:

- (a) to overcome the bottlenecks of today's highly efficient SISL schemes and the associated cost of the solver by overlapping communications and computations [70, 81]; and to overcome accuracy drawbacks related to the large time-step choice while still correctly simulating all relevant wave dispersion relations. Promising approaches for satisfying the latter condition are exponential time integrators [36, 47];
- (b) to overcome the overly restrictive time-step limitations of EBTI schemes combined with highly scalable horizontal discretizations, either through horizontal/vertical splitting (HEVI) [2, 8, 40] or through combining SISL PBTI methods with discontinuous Galerkin (DG) discretization [99]; and
- (c) to further the scalability and the adaptation of algorithms to emerging HPC architectures involving SE [32] or fully-implicit time-stepping approaches [113], and further through exploiting additional parallelism with time-parallel algorithms [33].

The success of any of these approaches may depend on the closer integration of software and hardware development (co-design), with dedicated hardware features accelerating specific aspects of a given algorithm. From the algorithm developer's perspective, expressing algorithms through Domain Specific Language (DSL) concepts can lead to enhanced flexibility and modularity [25], thereby shortening the deployment to potentially novel and disruptive hardware technologies (e.g., optical processors and quantum computing). Also, hardware development currently is (and will be in the next decade) strongly driven by the requirements of artificial intelligence. From this point of view, efforts are being undertaken to adapt solution algorithms to fit this new programming paradigm. This implies the use of technology based on training sets (determining weights of associated neural networks) computed outside critical time windows, that are successively applied to problems (including PDEs) described by approximated models constructed with the neural network weights (see, e.g., [77]). The weather and climate community should be aware of the factors driving hardware vendors and facilitate co-design of the underlying algorithms, in order to exploit new generation computing machines at their best. This, in turn, might help to increase percentage of machine use with respect to peak performance of the simulations, and might significantly improve time-to- and cost-to-solution, while increasing the resolution and accuracy of the models.

Acknowledgements This work was possible thanks to the ESCAPE project that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 671627. FXG gratefully acknowledges the support of the U.S. Office

of Naval Research through PE-0602435N and the U.S. Air Force Office of Scientific Research, Computational Mathematics program.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Proceedings of the ECMWF Workshop on Non-hydrostatic Modelling, 8–10 November 2010. European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire, UK (2011). <https://books.google.com/books?id=iHE2AwEACAAJ>
2. Abdi DS, Giraldo FX, Constantinescu EM, Carr III LE, Wilcox LC, Warburton TC (2017) Acceleration of the implicit-explicit non-hydrostatic unified model of the atmosphere (NUMA) on manycore processors. *Int J High Perform Comput Appl* (in press)
3. Alexander R (1977) Diagonally implicit Runge–Kutta methods for stiff ODEs. *SIAM J Numer Anal* 14(6):1006–1021
4. Arakawa A, Konor CS (2009) Unification of the anelastic and quasi-hydrostatic systems of equations. *Mon Weather Rev* 137:710–726
5. Ascher U, Ruuth S, Wetton B (1995) Implicit-explicit methods for time-dependent partial differential equations. *SIAM J Numer Anal* 32(3):797–823
6. Baldauf M (2010) Linear stability analysis of Runge–Kutta-based partial time-splitting schemes for the Euler equations. *Mon Weather Rev* 138:4475–4496
7. Baldauf M (2013) A new fast-waves solver for the Runge–Kutta dynamical core. Technical Report 21, Consortium for Small-Scale Modelling. <http://www.cosmo-model.org>
8. Bao L, Klöforn R, Nair R (2015) Horizontally explicit and vertically implicit (HEVI) time discretization scheme for a discontinuous Galerkin nonhydrostatic model. *Mon Weather Rev* 143(3):972–990
9. Bauer P, Thorpe A, Brunet G (2015) The quiet revolution of numerical weather prediction. *Nature* 525(7567):47–55
10. Benacchio T, O’Neill W, Klein R (2014) A blended soundproof-to-compressible numerical model for small-to mesoscale atmospheric dynamics. *Mon Weather Rev* 142(12):4416–4438
11. Bénard P (2003) Stability of semi-implicit and iterative centered-implicit time discretizations for various equation systems used in NWP. *Mon Weather Rev* 131:2479–2491
12. Bénard P, Vivoda J, Mašek J, Smolíkova P, Yessad K, Smith C, Brozková R, Geleyn JF (2010) Dynamical kernel of the Aladin-NH spectral limited-area model: Revised formulation and sensitivity experiments. *Q J R Meteorol Soc* 136:155–169
13. Brayton R, Gustavson F, Hachtel G (1972) A new efficient algorithm for solving differential-algebraic systems using implicit backward differentiation formulas. *Proc IEEE* 60(1):98–108
14. Bubnová R, Hello G, Bénard P, Geleyn JF (1995) Integration of the fully elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/Aladin NWP system. *Mon Weather Rev* 123:515–535
15. Butcher J (1987) The numerical analysis of ordinary differential equations: Runge–Kutta and general linear methods. Wiley-Interscience, New York
16. Butcher J (2006) General linear methods. *Acta Numer* 15:157–256
17. Caluwaerts S, Degrauwe D, Voitus F, Termonia P (2016) Discretization in numerical weather prediction: a modular approach to investigate spectral and local SISL methods. In: *Mathematical problems in meteorological modelling, mathematics in industry*, Springer, Berlin, vol 24, pp 19–46
18. Carpenter M, Kennedy C, Bijl H, Viken S, Vatsa V (2005) Fourth-order Runge-Kutta schemes for fluid mechanics applications. *J Sci Comput* 25(1–2):157–194
19. Cash J (1979) Diagonally implicit Runge–Kutta formulae with error estimates. *IMA J Appl Math* 24(3):293–301
20. Cash J (1980) On the integration of stiff systems of ODEs using extended backward differentiation formulae. *Numer Math* 34(3):235–246
21. Cash J (1983) The integration of stiff initial value problems in ODEs using modified extended backward differentiation formulae. *Comput Math Appl* 9(5):645–657
22. Ceschino F, Kuntzmann J (1966) Numerical solution of initial value problems. Prentice-Hall, Upper Saddle River
23. Colavolpe C, Voitus F, Bénard P (2017) RK-IMEX HEVI schemes for fully compressible atmospheric models with advection: analyses and numerical testing. *Q J R Meteorol Soc* 143:1336–1350
24. Davies T, Cullen M, Malcolm A, Mawson M, Staniforth A, White A, Wood N (2005) A new dynamical core for the Met Office’s global and regional modelling of the atmosphere. *Q J R Meteorol Soc* 131:1759–1782
25. Deconinck W, Bauer P, Diamantakis M, Hamrud M, Kühnlein C, Maciel P, Mengaldo G, Quintino T, Raoult B, Smolarkiewicz P, Wedi N (2017) Atlas: A library for numerical weather prediction and climate modelling. *Comput Phys Commun* 220:188–204
26. Diamantakis M, Davies T, Wood N (2007) An iterative time-stepping scheme for the Met Office’s semi-implicit semi-Lagrangian non-hydrostatic model. *Q J R Meteorol Soc* 133:997–1011
27. Diamantakis M, Magnusson L (2016) Sensitivity of the ECMWF model to semi-Lagrangian departure point iterations. *Mon Weather Rev* 144:3233–3250
28. Durran D (1991) The third-order Adams–Bashforth method: an attractive alternative to leapfrog time differencing. *Mon Weather Rev* 119(3):702–720
29. Durran D (1999) Numerical methods for wave equations in geophysical fluid dynamics. Springer, Berlin
30. Durran D (2008) A physically motivated approach for filtering acoustic waves from the equations governing compressible stratified flow. *J Fluid Mech* 601:365–379
31. Falcone M, Ferretti R (1998) Convergence analysis for a class of high-order semi-Lagrangian advection schemes. *SIAM J Numer Anal* 35:909–940
32. Fuhrer O, Chadha T, Hoefler T, Kwasniewski G, Lapillonne X, Leutwyler D, Osuna Schulthess T, Vogt H (2017) Near-global climate simulation at 1 km resolution: establishing a performance baseline on 4888 GPUs with COSMO 5.0. *Geoscientific Model Development Discussions*. In review
33. Gander M (2013) Multiple shooting and time domain decomposition methods, chap. 50 years of time parallel time integration. *MuS-TDD, Heidelberg*, pp 69–113
34. Gassmann A (2013) A global hexagonal C-grid non-hydrostatic dynamical core (ICON-IAP) designed for energetic consistency. *Q J R Meteorol Soc* 139(670):152–175
35. Gassmann A, Herzog HJ (2007) A consistent time-split numerical scheme applied to the nonhydrostatic compressible equations. *Mon Weather Rev* 135:20–36

36. Gaudreault S, Pudykiewicz J (2016) An efficient exponential time integration method for the numerical solution of the shallow water equations on the sphere. *J Comput Phys* 322:827–848
37. Giraldo FX (1997) Lagrange-galerkin methods on spherical geodesic grids. *J Comput Phys* 136(1):197–213
38. Giraldo FX (1998) The Lagrange-Galerkin spectral element method on unstructured quadrilateral grids. *J Comput Phys* 147(1):114–146
39. Giraldo FX (2000) Lagrange-galerkin methods on spherical geodesic grids: the shallow water equations. *J Comput Phys* 160(1):336–368
40. Giraldo FX, Kelly JF, Constantinescu EM (2013) Implicit-explicit formulations of a three-dimensional nonhydrostatic unified model of the atmosphere (NUMA). *SIAM J Sci Comput* 35(5):B1162–B1194
41. Giraldo FX, Restelli M (2008) A study of spectral element and discontinuous Galerkin methods for the Navier–Stokes equations in nonhydrostatic mesoscale atmospheric modeling: Equation sets and test cases. *J Comput Phys* 227(8):3849–3877
42. Giraldo FX, Restelli M, Läuter M (2010) Semi-implicit formulations of the Navier–Stokes equations: application to non-hydrostatic atmospheric modeling. *SIAM J Sci Comput* 32(6):3394–3425
43. Godel N, Schomann S, Warburton T, Clemens M (2010) GPU accelerated Adams–Bashforth multirate discontinuous Galerkin FEM simulation of high-frequency electromagnetic fields. *IEEE Trans Magn* 46(8):2735–2738
44. Grabowski W (2016) Towards global large eddy simulation: super-parameterization revisited. *J Meteorol Soc Jpn. Ser. II* 94(4):327–344
45. Gustafsson N, Bojarova J (2014) Four-dimensional ensemble variational (4D-En-Var) data assimilation for the high resolution limited area model (HIRLAM). *Nonlinear Process Geophys* 21(4):745–762
46. Harris L, Lin SJ (2013) A two-way nested global-regional dynamical core on the cubed-sphere grid. *Mon Weather Rev* 141(1):283–306
47. Hochbruck M, Ostermann A (2010) Exponential integrators. *Acta Numer* 19:209–286
48. Holton J, Haynes P, McIntyre M, Douglass A, Rood R, Pfister L (1995) Stratosphere-troposphere exchange. *Rev Geophys* 33(4):403–439
49. Hortal M (2002) The development and testing of a new two-time-level semi-Lagrangian scheme (SETTLS) in the ECMWF forecast model. *Q J R Meteorol Soc* 128:1671–1687
50. Ineson S, Scaife A (2009) The role of the stratosphere in the European climate response to El Niño. *Nat Geosci* 2(1):32–36
51. Jackiewicz Z (2009) General linear methods for ordinary differential equations. Wiley, New York
52. Jameson A, Schmidt W, Turkel E (1981) Numerical solutions of the euler equations by finite volume methods using Runge–Kutta time-stepping schemes. AIAA paper 1259
53. Jeevanjee N (2017) Vertical velocity in the gray zone. *J Adv Model Earth Syst* 9:2304–2316
54. Kassam AK, Trefethen L (2005) Fourth-order time-stepping for stiff PDEs. *SIAM J Sci Comput* 26(4):1214–1233
55. Kennedy C, Carpenter M, Lewis R (2000) Low-storage, explicit Runge–Kutta schemes for the compressible Navier–Stokes equations. *Appl Numer Math* 35(3):177–219
56. Klemp J, Skamarock W, Dudhia J (2007) Conservative split-explicit time integration methods for the compressible nonhydrostatic equations. *Mon Weather Rev* 135:2897–2913
57. Klemp J, Wilhelmson R (1978) The simulation of three-dimensional convective storm dynamics. *J Atmos Sci* 35(6):1070–1096
58. Knoll D, Keyes D (2004) Jacobian-free Newton–Krylov methods: a survey of approaches and applications. *J Comput Phys* 193(2):357–397
59. Kværnø A (2004) Singly diagonally implicit Runge–Kutta methods with an explicit first stage. *BIT Numer Math* 44(3):489–502
60. Laprise R (1992) The Euler equations of motion with hydrostatic pressure as an independent variable. *Mon Weather Rev* 120(1):197–207
61. Lauritzen P, Nair R, Ullrich P (2010) A conservative semi-Lagrangian multi-tracer transport scheme (CSLAM) on the cubed-sphere grid. *J Comput Phys* 229(5):1401–1424
62. Lin SJ (2004) A “vertically Lagrangian” finite-volume dynamical core for global models. *Mon Weather Rev* 132(10):2293–2307
63. Lock SJ, Wood N, Weller H (2014) Numerical analyses of Runge–Kutta implicit-explicit schemes for horizontally explicit, vertically implicit solutions of atmospheric models. *Q J R Meteorol Soc* 140(682):1654–1669
64. Malardel S, Wedi N (2016) How does subgrid-scale parametrization influence nonlinear spectral energy fluxes in global NWP models? *J Geophys Res: Atmos* 121:5395–5410
65. McGregor J, Dix M (2001) The CSIRO conformal-cubic atmospheric GCM. In: IUTAM symposium on advances in mathematical modelling of atmosphere and ocean dynamics, vol 61, pp 197–202. Springer, Berlin
66. Melvin T, Dubal M, Wood N, Staniforth A, Zerroukat M (2010) An inherently mass-conserving iterative semi-implicit semi-Lagrangian discretization of the non-hydrostatic vertical-slice equations. *Q J R Meteorol Soc* 136:799–814
67. Mengaldo G (2015) Discontinuous spectral/hp element methods: development, analysis and applications to compressible flows. Ph.D. thesis, Imperial College London
68. Mesinger F (1977) Forward-backward scheme, and its use in a limited area model. *Contrib Atmos Phys* 50:200–210
69. Miller M, Smolarkiewicz P (2008) Predicting weather, climate and extreme events. *J Comput Phys* 7(227):3429–3430
70. Mozdzyński G, Hamrud M, Wedi N (2015) A partitioned global address space implementation of the European centre for medium range weather forecasts integrated forecasting system. *Int J High Perform Comput Appl* 29:261–273
71. Polvani L, Kushner P (2002) Tropospheric response to stratospheric perturbations in a relatively simple general circulation model. *Geophys Res Lett* 29(7):18–21
72. Putman W, Suarez M (2011) Cloud-system resolving simulations with the NASA Goddard Earth Observing System global atmospheric model (GEOS-5). *Geophys Res Lett* 38(16):L16809
73. Qian JH, Semazzi F, Scroggs J (1998) A global nonhydrostatic semi-Lagrangian atmospheric model with orography. *Mon Weather Rev* 126(3):747–771
74. Ritchie H (1987) Semi-Lagrangian advection on a gaussian grid. *Mon Weather Rev* 115:608–619
75. Ritchie H, Temperton C, Simmons A, Hortal M, Davies T, Dent D, Hamrud M (1995) Implementation of the semi-Lagrangian method in a high-resolution version of the ECMWF forecast model. *Mon Weather Rev* 123:489–514
76. Rivest C, Staniforth A, Robert A (1994) Spurious resonant response of semi-Lagrangian discretizations to orographic forcing: diagnosis and solution. *Mon Weather Rev* 122:366–376
77. Rudd K, Ferrari S (2015) A constrained integration (CINT) approach to solving partial differential equations using artificial neural networks. *Neurocomputing* 155:277–285
78. Rudi J, Malossi ACI, Isaac T, Stadler G, Gurnis M, Staar P, Ineichen Y, Bekas C, Curioni A, Ghattas O (2015) An extreme-scale implicit solver for complex PDEs: highly heterogeneous flow in Earth’s mantle. In: Proceedings of the international

- conference for high performance computing, networking, storage and analysis, SC '15, ACM, pp 5:1–5:12
79. Saito K (2003) Time-splitting of advection in the JMA nonhydrostatic model. *CAS/JSC WGNE Res Activ Oceanic Modell* 33:315–316
 80. Saito K, Fujita T, Yamada Y, Ishida JI, Kumagai Y, Aranami K, Ohmori S, Nagasawa R, Kumagai S, Muroi C, Kato T, Eito H, Yamazaki Y (2006) The operational JMA nonhydrostatic mesoscale model. *Mon Weather Rev* 134(4):1266–1298
 81. Sanan P, Schnepf S, May D (2016) Pipelined, flexible Krylov subspace methods. *SIAM J Sci Comput* 38(5):C441–C470
 82. Satoh M, Matsuno T, Tomita H, Miura H, Nasuno T, Iga S (2008) Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations. *J Comput Phys* 227:3486–3514
 83. Satoh M, Tomita H, Yashiro H, Miura H, Kodama C, Seiki T, Noda A, Yamada Y, Goto D, Sawada M, Miyoshi T, Niwa Y, Hara M, Ohno T, Iga S, Arakawa T, Inoue T, Kubokawa H (2014) The non-hydrostatic icosahedral atmospheric model: description and development. *Progress Earth Planet Sci* 1:1–18
 84. Schiesser W (2012) The numerical method of lines: integration of partial differential equations. Elsevier, Amsterdam
 85. Schneider T, Lan S, Stuart A, Teixeira J (2017) Earth system modeling 2.0: a blueprint for models that learn from observations and targeted high-resolution simulations. arXiv preprint [arXiv:1709.00037](https://arxiv.org/abs/1709.00037)
 86. Simmons A (1991) Development of a high resolution, semi-Lagrangian version of the ECMWF forecast model. In: Seminar on numerical methods in atmospheric models. ECMWF, Reading
 87. Skamarock W, Klemp J (1992) The stability of time-split numerical methods for the hydrostatic and the nonhydrostatic elastic equations. *Mon Weather Rev* 120(9):2109–2127
 88. Skamarock W, Klemp J (2008) A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J Comput Phys* 227(7):3465–3485
 89. Skamarock W, Klemp J, Duda M, Fowler L, Park SH, Ringer T (2012) A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tessellations and C-grid staggering. *Mon Weather Rev* 140:3090–3105
 90. Smolarkiewicz P (2011) Modeling atmospheric circulations with soundproof equations. In: Proceedings of the ECMWF workshop on nonhydrostatic modelling, 8–10 November, 2010, Reading, pp. 1–15
 91. Smolarkiewicz P, Deconinck W, Hamrud M, Kühnlein C, Mozdzyński G, Szmelter J, Wedi N (2016) A finite-volume module for simulating global all-scale atmospheric flows. *J Comput Phys* 314:287–304
 92. Smolarkiewicz P, Kühnlein C, Wedi N (2014) A consistent framework for discrete integrations of soundproof and compressible PDEs of atmospheric dynamics. *J Comput Phys* 263:185–205
 93. Smolarkiewicz P, Pudykiewicz J (1992) A class of semi-Lagrangian approximations for fluids. *J Atmos Sci* 49(22):2082–2096
 94. Staniforth A, Côté J (1991) Semi-Lagrangian integration schemes for atmospheric models—a review. *Mon Weather Rev* 119(9):2206–2223
 95. Staniforth A, Melvin T, Wood N (2013) Gungho! a new dynamical core for the unified model. In: Proceeding of the ECMWF workshop on recent developments in numerical methods for atmosphere and ocean modelling
 96. Staniforth A, Wood N (2008) Aspects of the dynamical core of a nonhydrostatic, deep-atmosphere, unified weather and climate-prediction model. *J Comput Phys* 227(7):3445–3464
 97. Strang G (1968) On the construction and comparison of difference schemes. *SIAM J Numer Anal* 5(3):506–517
 98. Temperton C, Staniforth A (1987) An efficient two-time-level semi-Lagrangian semi-implicit integration scheme. *Q J R Meteorol Soc* 113(477):1025–1039
 99. Tumolo G, Bonaventura L (2015) A semi-implicit, semi-Lagrangian discontinuous Galerkin framework for adaptive numerical weather prediction. *Q J R Meteorol Soc* 141:2582–2601
 100. Ullrich P, Jablonowski C (2012) Operator-split Runge–Kutta–Rosenbrock methods for nonhydrostatic atmospheric models. *Mon Weather Rev* 140(4):1257–1284
 101. Ullrich P, Jablonowski C, Kent J, Lauritzen P, Nair R, Reed K, Zarzycki C, Hall D, Dazlich D, Heikes R et al (2017) DCMIP2016: a review of non-hydrostatic dynamical core design and intercomparison of participating models. *Geosci Model Dev* 10(12):4477–4509
 102. Vos P, Eskilsson C, Bolis A, Chun S, Kirby R, Sherwin S (2011) A generic framework for time-stepping partial differential equations (PDEs): general linear methods, object-oriented implementation and application to fluid problems. *Int J Comput Fluid Dyn* 25(3):107–125
 103. Wedi N, Bauer P, Deconinck W, Diamantakis M, Hamrud M, Kühnlein C, Malardel S, Mogensen K, Mozdzyński G, Smolarkiewicz P (2015) The modelling infrastructure of the integrated forecasting system: recent advances and future challenges. ECMWF Research Department Technical Memorandum 760
 104. Wedi N, Hamrud M, Mozdzyński G (2013) A fast spherical harmonics transform for global NWP and climate models. *Mon Weather Rev* 141:3450–3461
 105. Wedi N, Smolarkiewicz P (2009) A framework for testing global non-hydrostatic models. *Q J R Meteorol Soc* 135(639):469–484
 106. Weller H, Lock SJ, Wood N (2013) Runge–Kutta IMEX schemes for the horizontally explicit/vertically implicit (HEVI) solution of wave equations. *J Comput Phys* 252:365–381
 107. Wicker L (2009) A two-step Adams–Bashforth–Moulton split-explicit integrator for compressible atmospheric models. *Mon Weather Rev* 137(10):3588–3595
 108. Wicker L, Skamarock W (1998) A time-splitting scheme for the elastic equations incorporating second-order Runge–Kutta time differencing. *Mon Weather Rev* 126:1992–1999
 109. Wicker L, Skamarock W (2002) Time-splitting methods for elastic models using forward time schemes. *Mon Weather Rev* 130(8):2088–2097
 110. Williamson D (2007) The evolution of dynamical cores for global atmospheric models. *J Meteorol Soc Jpn* 85B:241–269
 111. Wood N, Staniforth A, White A, Allen T, Diamantakis M, Gross M, Melvin T, Smith C, Vosper S, Zerroukat M, Thurn J (2014) An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Q J R Meteorol Soc* 140(682):1505–1520
 112. Xiu D, Karniadakis G (2001) A semi-Lagrangian high-order method for Navier–Stokes equations. *J Comput Phys* 172:658–684
 113. Yang C, Xue W, Fu H, You H, Wang X, Ao Y, Liu F, Gan L, Xu P, Wang L, Yang G, Zheng W (2016) 10m-core scalable fully-implicit solver for nonhydrostatic atmospheric dynamics. In: SC 2016: Proceedings of the international conference for high performance computing, networking, storage and analysis, IEEE, pp 57–68
 114. Yeh KS, Côté J, Gravel S, Méthot A, Patoine A, Roch M, Staniforth A (2002) The CMC-MRB global environmental multiscale (GEM) model. Part III: nonhydrostatic formulation. *Mon Weather Rev* 130(2):339–356

115. Yessad K, Wedi N (2011) The hydrostatic and non-hydrostatic global model IFS/ARPEGE: deep-layer model formulation and testing. Technical Report 657
116. Zängl G (2012) Extending the numerical stability limit of terrain-following coordinate models over steep slopes. *Mon Weather Rev* 140:3722–3733
117. Zerroukat M, Wood N, Staniforth A (2002) SLICE: A semi-Lagrangian inherently conserving and efficient scheme for transport problems. *Q J R Meteorol Soc* 128(586):2801–2820