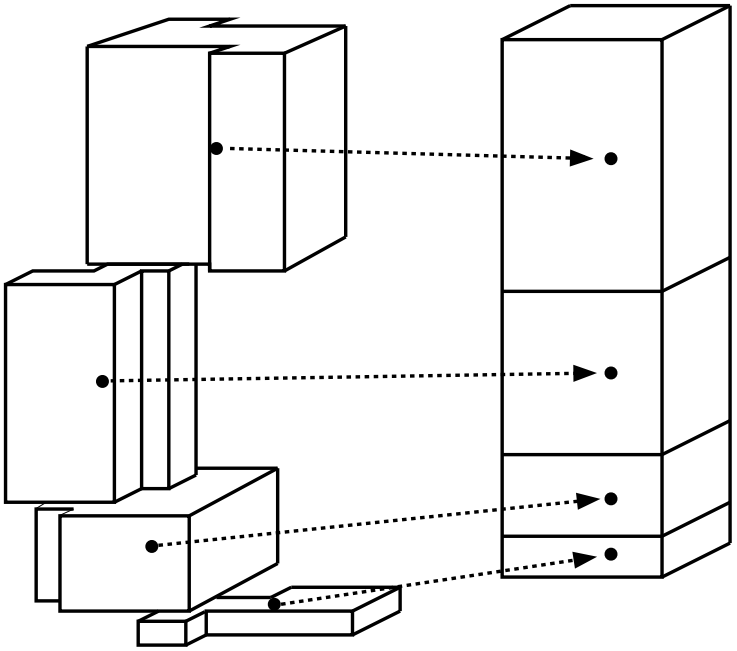
An inherently mass-conservative semi-implicit semi-Lagrangian model

by

Peter Hjort Lauritzen



ph.d. thesis

Department of Geophysics
University of Copenhagen
Denmark

September, 2005

Abstract

A locally mass-conservative dynamical core using a two-time-level, semi-Lagrangian semi-implicit integration scheme is presented. First a shallow water model is developed and tested, where after the approach is extended to a three-dimensional hydrostatic model.

The momentum equations are solved with the traditional semi-Lagrangian grid-point form. The explicit continuity equation is solved using a cell-integrated semi-Lagrangian (CISL) scheme and the semi-implicit part is designed such that the resulting elliptic equation is on the same form as for the traditional semi-Lagrangian grid-point system.

The accuracy of the shallow water model is assessed by running standard test cases adapted to a limited area domain. The accuracy and efficiency of the new model is comparable to traditional semi-Lagrangian methods and is not susceptible to noise problems for high Courant number flow over orography.

The shallow water model is extended to a baroclinic model by using a hybrid trajectory algorithm which is backward in the horizontal and forward in the vertical. In addition, the vertical part of the trajectory scheme is consistent with the discretized explicit CISL continuity equation. Since the vertical part of the trajectory is forward, the cells depart from model levels making the upstream integral two-dimensional and existing CISL schemes directly applicable. A price to pay is that the prognostic variables must be mapped back to the model grid at each time step. The problem is, however, one-dimensional.

Two model versions are derived. In the first version the thermodynamic equation is discretized as in traditional semi-Lagrangian models (only adapted to the hybrid trajectory). In the other model version the conversion term in the thermodynamic equation is discretized consistently with the semi-implicit CISL continuity equation. In both model versions the momentum equations are discretized in grid-point form.

The new dynamical cores are inherently mass conservative and can perform consistent online transport of tracers. They are implemented within the framework of HIRLAM and are tested using the Jablonowski-Williamson idealized baroclinic wave test case. The new dynamical cores run stably with long time steps and without the need for decentering or filtering of the non-linear terms in time as is needed in HIRLAM. In less active parts of the domain HIRLAM is noisy whereas the cell-integrated models produce smooth solutions. Compared to HIRLAM the baroclinic development is equally or more intense with the cell-integrated models. The CISL model version using a consistent energy conversion term has the strongest baroclinic development.

Table of Contents

Chapter II

Overview of the semi-Lagrangian method

The equations of motion for the weather and climate system can be derived in either a Lagrangian form or an Eulerian form. In an Eulerian system the observer describes the evolution of the flow from a fixed point in the coordinate system, whereas Lagrangian equations describe the evolution of the flow that would be observed following the motion of each individual fluid parcel. The Eulerian method has the advantageous property of having a regular mesh throughout the integration, but it often suffers from overly restrictive time step limitations. The Lagrangian method is normally not subject to strict time step limitations, but introduces another problem: an initially regular mesh will quickly evolve into an irregular mesh with high concentration of mesh points in convergent areas and low concentration of mesh points in divergent areas. Thus the accuracy of the method will be lower in divergent areas compared to convergent areas of the domain. In order to avoid the problems of non-uniform meshes and at the same time allow long time steps, one can periodically map the distribution from the irregular (Lagrangian) mesh back to the regular (Eulerian) grid and then start all over. Instead of following the same set of fluid parcels during the entire integration, a new set of fluid parcels are chosen after each time step. Hereby the distribution of parcels can be kept quasi uniform throughout the integration. This method is called the *semi-Lagrangian* method.

2.1 Trajectory determination

The first step of the semi-Lagrangian method is to determine the parcel trajectories. Mathematically this corresponds to the integration of the first-order ordinary differential equation

$$\frac{d\mathbf{r}}{dt} = \mathbf{v}(\mathbf{r}, t), \tag{1}$$

where $\mathbf{v}$ is the velocity vector $\mathbf{v} = (u, v)$ and $\mathbf{r}$ is the displacement vector. The kinematic relation (1) is among the prognostic equations of a semi-Lagrangian model. The accuracy of the method used for computing trajectories is crucial for the overall accuracy of the model. In particular, the use of long time steps emphasizes the need for reducing time-truncation errors associated with the scheme.

The discussion here is restricted to two-time-level schemes and only backward trajectories are discussed, i.e. consider parcels which depart from time level (n) and arrive at grid points at time level $(n + 1)$. For simplicity assume Cartesian geometry. The spatial location of the departure point is specified with subscript $*$ so $\mathbf{r}_*^n$ denotes the location of the departure point. The departure point is given by

$$\mathbf{r}_*^n = \mathbf{r}^{n+1} - \int_{n\,\Delta t}^{(n+1)\,\Delta t} \mathbf{v}(\mathbf{r}, t)\, dt. \tag{2}$$

The integral on the right-hand side of (2) is approximated using velocities at time levels (n) and $(n - 1)$.

The simplest algorithm for finding the backward trajectory is Euler's method

$$\mathbf{r}_*^n = \mathbf{r}^{n+1} - \mathbf{v}^n(\mathbf{r}^{n+1})\, \Delta t, \tag{3}$$

where $\mathbf{v}^n(\mathbf{r}^{n+1})$ refers to the velocity at time level (n) evaluated at the arrival point. Euler's method is first-order accurate in time, which can lead to large truncation errors especially when used with large time steps. (Robert 1981) found that the time truncation error should be no worse than $\mathcal{O}(\Delta t^2)$ to keep the overall errors at an acceptable level.

A variety of $\mathcal{O}(\Delta t^2)$ schemes have been proposed in the literature. A very popular scheme is the second-order implicit midpoint method where

$$\mathbf{r}_*^n = \mathbf{r}^{n+1} - \tilde{\mathbf{v}}^{n+1/2}\left(\frac{\mathbf{r}^{n+1} + \mathbf{r}_*^n}{2}\right)\Delta t, \tag{4}$$

is iterated with (3) as a first guess (McDonald and Bates 1987, Temperton and Staniforth 1987). The velocity field is first extrapolated to time level $(n+1/2)$, $\tilde{\mathbf{v}}^{n+1/2}$, and thereafter interpolated to the approximate midpoint of the trajectory, $(\mathbf{r}^{n+1} + \mathbf{r}_*^n)/2$. Linear interpolation in space is found to be sufficient and typically a few iterations of (4) are needed (e.g. Staniforth and Côté 1991).

The extrapolation in time for obtaining the middle point of the semi-Lagrangian trajectory has, however, been identified as a potential source of instability. Non-meteorological noise has been observed in forecasts using two-time-level semi-Lagrangian models at a number of meteorological centers, e.g., in the Aire Limité Adaptation dynamique Développement InterNational/Limited Area Central European (ALADIN/LACE) model (Gospodinov et al. 2001), HIRLAM (McDonald 1999) and IFS model (Hortal 2002). This has forced modelers to rethink the design of 2TLSL models. Here we focus on the trajectory computations.

McDonald (1999) successfully removed the noise in forecasts made with HIRLAM by using an alternative approximation to the extrapolated velocity on the right-hand side of (4). The velocity extrapolated to time level $(n+1/2)$, $\tilde{\mathbf{v}}^{(n+1/2)}$, can be written as a linear combination of the known velocities at time levels (n) and $(n-1)$ evaluated at the arrival point $\mathbf{r}^{n+1}$, departure point $\mathbf{r}_*^n$ and $\mathbf{r}^{n+1} - 2\,\Delta t\,\mathbf{v}_*^n$, respectively:

$$\tilde{\mathbf{v}}^{(n+1/2)} = \sum_{\mu=1}^{6} \mathbf{w}_\mu\,\chi_\mu, \tag{5}$$

where

$$\begin{aligned}
\chi_1 &= \mathbf{v}^n(\mathbf{r}^{n+1}),\\
\chi_2 &= \mathbf{v}^{n-1}(\mathbf{r}^{n+1}),\\
\chi_3 &= \mathbf{v}^n(\mathbf{r}_*^n),\\
\chi_4 &= \mathbf{v}^{n-1}(\mathbf{r}_*^n),\\
\chi_5 &= \mathbf{v}^n\left[\mathbf{r}^{n+1} - 2\,\Delta t\mathbf{v}^n(\mathbf{r}^{(n+1)})\right],\\
\chi_6 &= \mathbf{v}^{n-1}\left[\mathbf{r}^{n+1} - 2\,\Delta t\mathbf{v}^n(\mathbf{r}^{(n+1)})\right]
\end{aligned}$$

and $\mathbf{w}_i$ are weights. Requiring second-order temporal accuracy imposes constraints on the weights $\mathbf{w}_i$. Three free parameters result. Several schemes proposed in the literature belong to this family

of schemes, for example the *extrapolation along the trajectory* scheme[5] (equation 42 in Temperton and Staniforth 1987) and the more economical scheme described in Hortal (1998)[6]. By running the forecast model and measuring the level of noise McDonald found that the first *extrapolation along the trajectory* scheme produces almost noise-free forecasts, while the Hortal scheme only reduces the noise. The noise could be reduced even further by an "optimal" choice of the free parameters. The optimal scheme, measured in terms of noise, used χ_1, χ_3, χ_4, χ_5 and χ_6 for the approximation of the trajectory.

By using more values for the approximation of $\tilde{\mathbf{v}}^{n+1/2}$ without increasing the formal accuracy, the approximation for $\tilde{\mathbf{v}}^{n+1}$ is somewhat smoothed. McDonald's scheme seems to remove the noise because of the smoothing. On the other hand a smoothing could lead to a decrease in accuracy. McDonald considered only noise levels in his study and did not report the forecast accuracy in terms of verification scores. So it is an open question how the overall accuracy of the forecast is affected by this trajectory scheme that produces practically noise-free forecasts. For example, Hortal (2002) reported no noise problems when using the *extrapolation along the trajectory* scheme, but got much worse verification scores.

All departure point algorithms mentioned so far assume straight line trajectories in a "space-time diagram", i.e. they do not take the acceleration into account. When using long time steps it would be desirable to include the accelearation in the trajecory estimation. Several trajectory schemes have been proposed in the literature.

The location of the departure point $\mathbf{r}_*^n$ can be written in terms of a Taylor series expansion about the departure point

$$\mathbf{r}_*^n = \mathbf{r}^{n+1} - \sum_{\nu=1}^{N} \frac{(\Delta t)^\nu}{\nu!} \frac{d^\nu \mathbf{r}_*^n}{dt^\nu}, \tag{6}$$

or about the arrival point

$$\mathbf{r}_*^n = \mathbf{r}^{n+1} + \sum_{\nu=1}^{N} \frac{(-\Delta t)^\nu}{\nu!} \frac{d^\nu \mathbf{r}^{n+1}}{dt^\nu}, \tag{7}$$

where d/dt is the material derivative (also called total or Lagrangian derivative)

$$\frac{d}{dt} = \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y}. \tag{8}$$

By including more terms in the Taylor series expansion the trajectory is no longer a straight line in a "space-time diagram" where time is plotted on the y axis and departure point distance on the x axis. The question is how to approximate the derivatives $\frac{d^\nu \mathbf{r}}{dt^\nu}$, $\nu = 2, .., N$.

Motivated by noise problems in the operational two-time-level semi-Lagrangian (2TLSL) model at the European Center for Medium-Range Weather Forecasts (ECMWF), Hortal (2002) derived

[5] $\mathbf{w}_1 = 0$, $\mathbf{w}_2 = 0$, $\mathbf{w}_3 = \frac{3}{2}$, $\mathbf{w}_4 = 0$, $\mathbf{w}_5 = 0$ and $\mathbf{w}_6 = -\frac{1}{2}$.
[6] $\mathbf{w}_1 = \frac{1}{2}$, $\mathbf{w}_2 = 0$, $\mathbf{w}_3 = 1$, $\mathbf{w}_4 = -\frac{1}{2}$, $\mathbf{w}_5 = 0$ and $\mathbf{w}_6 = 0$.

an improved trajectory scheme called 'Stable Extrapolation Two-Time-Level Scheme' (SETTLS). (6) was chosen as the basis for the trajectory scheme with

$$\frac{d\mathbf{r}_*^n}{dt} = \mathbf{v}_*^n,$$
$$\frac{d^2\mathbf{r}_*^n}{dt} = \frac{\mathbf{v}^n(\mathbf{r}^{n+1}) - \mathbf{v}_*^{n-1}}{\Delta t}.$$

The choice for the approximation for the acceleration was chosen after exploring many possibilities and so that the scheme could also be used for the other prognostic equations without complicating the elliptic equation associated with the semi-implicit system. The formal temporal accuracy of the scheme is second order and the computational cost of the scheme is comparable to that of the classical mid-point method (4). The treatment of the right-hand side of the trajectory equation was also used for the non-linear terms on the right-hand side of the remaining equations of motion. The forecast skill was not effected negatively when switching from the mid-point method to the SETTLS and noise problems were reduced. During sudden stratospheric warmings, however, Hortal (2004) report "new" noise problems.

McGregor (1993) proposed to discard the Eulerian velocity change in the approximation of the derivative

$$\frac{d}{dt} \approx \mathbf{v} \cdot \nabla, \tag{9}$$

and to use the formula in which the Taylor series expansion is about the arrival point (7). By using (7) for the approximation the need for spatial interpolation is eliminated. Using McGregor's scheme the location of the departure point is given by

$$\mathbf{r}_*^n = \mathbf{r}^{n+1} - \Delta t\, \tilde{\mathbf{v}}^{n+1} + \sum_{\nu=1}^{N-1} \frac{(-\Delta t)^{\nu+1}}{(\nu+1)!} \frac{d^\nu}{dt^\nu} \left(\tilde{\mathbf{v}}^{n+1}\right), \tag{10}$$

where

$$\frac{d}{dt} \approx \tilde{\mathbf{v}}^{n+1/2} \cdot \nabla, \tag{11}$$
$$\widetilde{(\cdot)}^{n+1} = 2\,\widetilde{(\cdot)}^{n} - \widetilde{(\cdot)}^{n-1}, \tag{12}$$

and the higher-order derivatives are defined recursively

$$\frac{d^\nu \mathbf{v}}{dt^\nu} = \frac{d}{dt}\left(\frac{d^{\nu-1}\mathbf{v}}{dt^{\nu-1}}\right) \quad \nu = 2, 3, .., N-1. \tag{13}$$

Note that the scheme does not involve iterations. McGregor's formulation is general in the sense that high-order terms can easily be included since they are defined recursively. The order of accuracy can therefore easily be increased. Note that the formal order of accuracy of iterative methods does not increase when increasing the number of iterations. Nair et al. (2003) reported better results with McGregor's trajectory algorithm compared to a Runge-Kutta scheme for advection

experiments in Cartesian and spherical geometry. The scheme has also been extended to spherical geodesic grids, where better results were obtained compared to the iterative midpoint method (Giraldo 1999). It is, however, not know if it alleviates noise problems in baroclinic models. A critic of the scheme is that it assumes $\partial \mathbf{v}/\partial t = 0$. That assumption is also used in the midpoint method and other iterative methods. The SETTLS scheme, however, does not assume that $\frac{\partial \mathbf{v}}{\partial t}$ is zero.

Instead of using higher-order trajectory schemes one could alternatively split the trajectory into k segments so that each sub trajectory takes $\Delta t/k$ (D.L. Williamson 2004, personal communication). For each segment a low order efficient scheme could be used. As far as the author is aware the method has not been tested in baroclinic models so far.

The noise problems encountered in 2TLSL models have emphasized the fact that the computation of the trajectories can not be studied as an isolated problem. The differential equation for the trajectories is part of the prognostic equations of the full model and it is important that it "interacts well" with the remaining model equations. For example, the SETTLS scheme was designed so that the right-hand side of the equations of motion was treated in exactly the same way as the right-hand side of the trajectory equation. There is a need to incorporate trajectories in a more consistent manner in semi-Lagrangian models instead of treating trajectory determination and the solution to the remaining equations of motion as two separate tasks. Trajectories are always computed using explicit schemes. In a SISL model one should ideally solve the trajectory equation using the same semi-implicit scheme,i.e. the divergence should be averaged along the trajectory. It does, however, seem difficult to design a method where the trajectory algorithm is included in the elliptic system of a SISL model. The issue of incorporating trajectories more into the dynamics and studying the effect of trajectories on model properties other than stability, has been very little discussed in the literature. Exceptions are (Staniforth et al. 2003) and Cordero et al. (2005) who studied the impact of trajectories on the dynamical equivalence between momentum and angular momentum formulations of the equations of motion, and the impact on trajectories on the vertical models in a non-hydrostatic column model, respectively. Clearly, more research is desirable in this area.

Having discussed trajectory algorithms published in the meteorological literature and their use in baroclinic models, the next step in the semi-Lagrangian method is the solution to the equations of motion. In the following the focus is on the continuity equation, but the treatment of the the other equations of motion is similar and is discussed in Chapter III and IV in connection with the shallow water equations and the primitive equations, respectively.

2.2 The form of the continuous equations

Consider the continuity equation on flux form and advective form

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \, \mathbf{v}) = 0, \tag{14}$$

$$\frac{d \rho}{d t} + \rho \, \nabla \cdot \mathbf{v} = 0, \tag{15}$$

respectively, where $\mathbf{v}$ is the velocity field, and ρ is some mass-specific quantity. The continuity equation is alternatively called the transport or advection equation. The numerical solution to this type of equation is probably the most studied problem in computational fluid dynamics since it represents a fundamental property of fluid flow. The flux-form equation for the transport of a passive scalar is closely associated to (14)

$$\frac{\partial}{\partial t} (\rho \, q) + \nabla \cdot (\rho \, q \, \mathbf{v}) = 0, \tag{16}$$

where q is the concentration of the tracer per unit mass (also referred to as mixing ratio). The two flux-form equations (14) and (16) imply an advective form of the tracer transport equation

$$\frac{d \, q}{d t} = 0, \tag{17}$$

which simply states that q is conserved along characteristics of the flow. Both the flux form and advective form of the continuous equations are differential forms. Alternatively, the equations of motions can be written on integro-differential form which is the partial differential equation integrated over an infinitesimal volume. If the volume is stationary the form is Eulerian, while a Lagrangian form results when the volume moves with the flow. Integrating the flux-form continuity (14) and tracer advection equation (16) over a stationary volume δV, the Eulerian integro-differential form results

$$\frac{\partial}{\partial t} \left(\overline{\psi} \, \delta V \right) + \oint \oint \oint_{\partial V} \psi \cdot \mathbf{n} \, dS = 0, \quad \psi = \rho, \, \rho \, q, \tag{18}$$

where

$$\overline{\psi} = \frac{1}{\delta V} \iiint_{\delta V} \psi \, dV \tag{19}$$

is the average of ψ over δV, ∂V is the boundary of δV and $\mathbf{n}$ the normal vector to ∂V. The divergence theorem, also known as Gauss's theorem, has been used to convert the volume integral of the divergence term to a surface integral. In a Lagrangian method, where the volume moves with the flow $\delta V = \delta V(t)$, there is no flux through ∂V and hence the Lagrangian version of (18) is given by

$$\frac{d}{d t} \left(\overline{\psi} \, \delta V \right) = 0, \tag{20}$$

which simply states that $\overline{\psi}$ is conserved for a volume moving with the flow.

Numerical methods based on the integral of a conservation law over a volume are called *finite volume* (FV) methods. An alternative and perhaps more descriptive name also used in the literature is *cell-integrated* methods. Eulerian and Lagrangian FV discretizations of the continuity equation are based on (18) and (20), respectively. Since FV or *cell-integrated* methods are based on tracking the integral of ψ, in this case mass, they are locally conservative. In Eulerian FV methods the fluxes through the boundaries of cells are tracked and mass flowing out of a cell wall is gained in the neighboring cell, thus guaranteeing that no mass is lost or added. In a Lagrangian FV method the mass is tracked as it moves with the flow, and as long as all the volumes span the entire domain the mass is conserved both globally and locally. Here only semi-Lagrangian methods are reviewed. An updated review of Eulerian, as well as semi-Lagrangian, finite-volume methods is currently under way (Machenhauer et al. 2005 in prep.).

2.3 Requirements for transport schemes

Before discussing the many finite-volume schemes used in the atmospheric sciences, it is important to realize which properties a transport scheme ideally should posses. The equation subject to the toughest requirements is probably the continuity equation for noisy tracers such as moisture. Rasch and Williamson (1990) have defined seven widely accepted desirable properties for transport schemes: accurate, stable, computationally reasonable, transportive, local, conservative and shape-preserving.

Accurate The high accuracy property is, of course, the primary aim for any numerical method and all but the efficiency requirement are part of the overall accuracy. Note that for a flow with shocks or sharp gradients the formal order of accuracy in terms of Taylor series expansions does not necessarily guarantee a high level of accuracy. Part of the accuracy is also the rate of convergence of the numerical algorithm.

Widely used measures of accuracy in the meteorological community for idealized test cases, are the standard error measures l_1, l_2 and l_∞ (e.g., Williamson et al. 1992):

$$l_1(\psi) = \frac{I(|\psi - \psi_E|)}{I(|\psi_T|)}, \tag{21}$$

$$l_2(\psi) = \frac{\left\{I\left[(\psi - \psi_E)^2\right]\right\}^{1/2}}{\left\{I\left[(\psi_T)^2\right]\right\}^{1/2}}, \tag{22}$$

$$l_\infty(\psi) = \frac{\max\left[I(|\psi - \psi_E|)\right]}{\max\left[I(|\psi_T|)\right]}, \tag{23}$$

where $I(\cdot)$ denotes the integral over the entire domain, ψ is the numerical solution and ψ_E is the exact solution if it exists. In case an exact solution does not exist ψ_E is a high resolution

reference solution. l_1 and l_2 are measures for the global "distance" between ψ and ψ_E, and l_∞ is the maximum deviation of ψ from ψ_E over the entire domain. In addition to the measures l_1, l_2 and l_∞, the normalized maximum and minimum values are also used to indicate errors related to overshooting and undershooting.

Stable The stability property ensures that the solution does not "blow up" during the time of integration. Usually the stability of Eulerian methods is governed by the Courant-Friedrichs-Levy (CFL) condition, which in one dimension is given by

$$\max \left| \frac{u \, \Delta t}{\Delta x} \right| \leq 1, \tag{24}$$

where u is the velocity, Δt the time step and Δx the grid interval. Hence a fluid parcel may not travel more than one grid interval during one time step. This overly restrictive time step limitation is usually alleviated in Lagrangian methods and can be replaced by the less severe Lipschitz criterion for stability

$$\left| \frac{\partial u}{\partial x} \right| \, \Delta t < 1, \tag{25}$$

(Smolarkiewicz and Pudykiewicz 1992), which guarantees that parcel trajectories do not cross during one time step. Hence in semi-Lagrangian models the time step can be chosen for accuracy and not for stability.

Note that for global models the efficiency and stability of the schemes are often challenged by the convergence of the meridians near the poles. If conventional latitude-longitude grids are used, special care must be taken in the vicinity of the poles. The problem can also be tackled by using other types of grids that do not have these singularities or at least reduce the effect of them, for example the icosahedral-hexagonal grid used operationally by the Deutscher Wetterdienst (e.g., Sadourny et al. 1968; Williamson 1968; Thuburn 1997; Majewski et al. 2002) and the cubed sphere approach (Sadourny 1972). These grids are more isotropic than conventional latitude-longitude grids, i.e. all cells have nearly the same size contrary to latitude-longitude grids where the areas decrease as aspect ratios increase toward the poles. This effect can be alleviated by using a Gaussian reduced grid in which the number of longitudes decrease toward the poles. It is an important part of accuracy that the advection schemes can transport distributions across the poles without distorting them and without imposing severe time-step limitations.

Computationally reasonable Computing resources are not unlimited and, given the complexity of geophysical fluid dynamics, the algorithms should be computationally efficient in order to allow for high resolution runs and/or a large number of prognostic variables. Efficiency is, however, hard to measure objectively. One measure for the efficiency of an algorithm is the number of elementary mathematical operations or the total number of floating-point operations per second (FLOPS) used by the algorithm. The advantage of counting FLOPS is that it can be done without turning

the computer on and is therefore a machine independent measure. But the number of FLOPS only captures one of several dimensions of the efficiency issue. The actual program execution involves subscripting, memory traffic and countless other overheads. In addition different computer architectures favor different kinds of algorithms and compilers optimize code differently. Measuring efficiency in terms of the execution time on a specific platform can be misleading for a user on another computer platform. Weather prediction and climate models are executed on massively parallel computersm wherefore the efficiency is partly determined by the amount of communication between the nodes. Hence the parallel programmer is concerned about algorithms being local thus minimizing the need for communication between the nodes. Nevertheless the most important measure of efficiency is probably the level of simplicity of the algorithm.

Transportive, local, conservative and shape-preserving The transportive and local property guarantee that information is transported with the characteristics and that only adjacent grid values affect the forecast at a given point. Integral invariants of the corresponding continuous problem as well as the shape of the distribution for non-divergent flows should ideally be preserved in the numerical solution. If the velocity field is divergent the shape of the distribution may be altered in the form of new extrema. So in the divergent case the numerical scheme should reproduce the physical extrema without creating spurious numerical extrema.

Integral invariants should, of course, be conserved for any kind of flow. For long simulations the conservation properties become increasingly important as numerical sources and sinks can degrade the accuracy significantly over time (e.g., Moorthi et al. 1995). Hence for climate models the finite-volume methods are very attractive given their inherent conservation properties.

In addition to the seven desirable properties of Rasch and Williamson (1990) even more desirable properties have emerged in the literature.

Consistency The consistency property is less frequently discussed in the literature. Notable exceptions are Jöckel et al. (2001) and Byun (1999). If $q = 1$ in the tracer transport equation (16), it mathematically degenerates to the continuity equation (14). This should ideally be the case numerically as well. If the two equations are solved using the same numerical method on the same grid and using the same time step, the consistency is, of course, guaranteed. However, in a realistic and practical setting found in many atmospheric models, the consistency is harder to achieve. The consistency property, or rather the lack of it, has been discussed in detail in the Introduction.

Preservation of constancy Another desirable property is the ability of the scheme to preserve a constant tracer field for a non-divergent flow.

2.4 Grid-point semi-Lagrangian transport schemes

Grid-point semi-Lagrangian schemes (also referred to as traditional semi-Lagrangian schemes) have been thoroughly reviewed in Staniforth and Côté (1991) and hence the method will only be discussed briefly here with updates.

The traditional semi-Lagrangian scheme is based on the discretization of (17)

$$\psi^{n+1} = \psi^n_*, \tag{26}$$

where the superscript refers to the time level and the subscript $*$ refers to the evaluation of ψ at the upstream departure point. The generic variable ψ is used (which for the continuity equation for a tracer is q) since the method also applies to the other equations of motion. Since the departure point does not necessarily coincide with a grid point, some kind of interpolation must be invoked in order to approximate ψ^n_*. In principle, any kind of interpolation may be used.

Probably the most widely used interpolator for two-dimensional problems is bicubic Lagrange interpolation or quasi biparabolic interpolation. It is a good compromise between accuracy and computational efficiency (Bates and McDonald 1982). Since it is such a well established and widely used method, it will be used as a reference throughout this study. Another popular interpolation method in semi-Lagrangian schemes is cubic splines (e.g, Purnell 1976;Riishøjgaard et al. 1998). To reduce the computational cost of fully two-dimensional methods, cascade methods have been introduced (Purser and Leslie 1991). Here the problem is reduced to two one-dimensional problems. This is done by defining a Lagrangian mesh (the regular mesh advected one time step) and then performing the interpolation along the Lagrangian longitudes and latitudes. Several schemes published in the meteorological literature are based on this method (e.g., Nair et al. 1999; Laprise and Plante 1995; Sun and Yeh 1997), which provides an economical and accurate alternative to fully two-dimensional methods.

2.4.1 *Enforcing monotonicity*

Monotonic filter The interpolators mentioned so far do not preserve the shape of the distribution unless special measures are taken. Using the ideas of flux-corrected transport (FCT; Zalesak (1979)), undershots and overshoots can be eliminated by using a simple and efficient filter. The resulting scheme is called *quasi-monotone semi-Lagrangian* scheme (QMSL) (Bermejo and Staniforth 1992). The filter works as follows. Whenever the upstream interpolated value is greater/(smaller) than the surrounding grid point values, the value is set equal to the maximum/(minimum) of the surrounding grid-point values. So the scheme reduces to low order whenever the high-order method overshoots or undershoots. The algorithm is currently used operationally in the IFS. The monotonicity is, however, enforced at the expense of accuracy. E.g. the mass is increased and there is a significant decrease in accuracy in terms of standard error measures (see table 1). The QMSL filter clips local extrema that in some situations should be retained (e.g.,

Method	$\int \psi / \int \psi_0$	$\int \psi^2 / \int \psi_0^2$	$Max(\psi)$	$Min(\psi)$	E_{DISS}	E_{DISP}
SL	1.00	0.90	4.51	-0.62	2.32(-3)	5.71(-2)
QMSL	1.01	0.82	3.99	0.00	0.96(-2)	0.95(-1)
CQMSL	1.00	0.81	3.99	0.00	1.00(-4)	6.46(-2)

Table 1: Error measures after six revolutions of a rotating slotted cylinder (see e.g., Bermejo and Staniforth 1992 for details) using cubic-spline interpolation with no "correction", with a quasi-monotone semi-Lagrangian (QMSL) filter (Bermejo and Staniforth 1992) and with the quasi-monotone semi-Lagrangian (CQMSL) algorithm of Priestley (1993), respectively. The error measures are (left to right) the mass, square of the mass, minimum value, maximum value and the dissipation and dispersion error measures introduced by Takacs (1985), respectively. All values are from the publications referred above.

see Fig. 2c in Nair et al. 1999).

More advanced filters in two dimensions are computationally expensive, but in one dimension the filter can be improved to deal with situations where the local extrema are physical, and then be applied to each one-dimensional sub-problem in cascade schemes (Sun and Yeh 1997; Nair et al. 1999).

Splines Alternative approaches for doing shape-preserving semi-Lagrangian advection is to use splines. The methods can be efficiently extended to two and three dimensions using a tensor product approach. Monotonocity is enforced by constraining derivative estimates (e.g., Williamson and Rasch 1989; Holnicki 1995). The unmodified splines tend to have better conservation properties than schemes using conventional interpolation. When modifying the derivative estimates in order to ensure monotonicity the scheme conserves mass less well as the peaks of small-scale structures are clipped. Hence, as for the schemes using Lagrange interpolation, also splines enforce monotone at the expense of conservation and vice versa.

2.4.2 *Mass fixers*

For general flows the traditional schemes do not conserve mass[7]. In order to restore mass one must periodically add or remove mass to the system. The question is how to do that without altering the shape of the distribution and without degrading the local property. The simplest procedure for restoring mass conservation is to add mass gained or lost to the average. Thereby a long-range transport of mass is introduced into the system and the numerical method is no longer local. This

[7] for divergence-free fbws the cubic spline method does conserve mass (Bermejo 1990)

method is therefore not desirable although widely used.

Priestley (1993), Bermejo and Conde (2002) and Sun and Sun (2004) have developed *ad hoc* mass-restoration algorithms which attempt to restore mass conservation while not violating monotonicity and locality requirements. Although the papers solve the problems differently, the basic idea is the same. In general the high order approximation to ψ_*^n can be written as a weighted sum of the surrounding grid-point values. Mass conservation can be regained by carefully choosing the weights. In Bermejo and Conde (2002) a new set of weights are computed by a minimization process. The monotone and conservative solution ψ is sought by minimizing the squared difference between ψ and the known non conservative but monotone solution with global mass conservation as a constraint. Mathematically it is a classical minimization problem that can be solved using Lagrange multipliers. A similar approach was taken by Sun and Sun (2004). In Priestley (1993) ψ_*^n is written as a linear combination of a high-order and low-order solution. The two solutions are weighted so that mass is conserved.

All the mass-fixing methods do, however, not guarantee entirely local mass conservation. The scheme of Priestley (1993) is less accurate with respect to the mean square error and the dispersion error compared to the unaltered solution, but slightly more accurate with respect to the dissipation error (see table 1). The scheme is, however, from 25% to 76% more expensive than the unaltered scheme, depending on which computer architecture is used (Table 4 in Priestley 1993). Similar conclusions hold for the scheme of Bermejo and Conde (2002). As already discussed in detail in the Introduction, all mass-fixing algorithms are non-local.

2.4.3 *Summary*

The traditional semi-Lagrangian schemes have several advantages over other numerical methods. They are economical in terms of CPU time and memory usage, are unconditionally stable and have small amplitude and phase errors for smooth flows. The efficiency is increased even further if a large number of tracers are transported simultaneously.

Common to all the traditional semi-Lagrangian schemes is that they are all cast in non-conservative form, and consequently have intrinsic difficulties in conserving mass. This can result in a significant drift in the global mass fields. To restore global mass conservation *ad hoc a-posteriori* algorithms must be employed. There is a degree of arbitrariness in these "mass-fixing" algorithms and they may degrade preexisting desirable properties. It is therefore not only mathematically more rigorous, but also desirable to use inherently conserving methods. During the last decade it has been demonstrated that it is possible to design semi-Lagrangian schemes which are inherently mass conservative. These schemes are the subject of the next section.

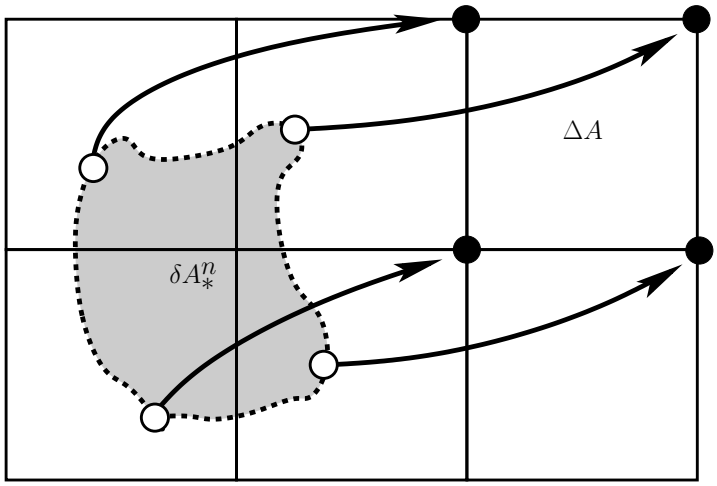


Fig. 2: The regular arrival cell with area ΔA and the irregular departure cell (shaded region) with area δA_*^n in the continuous case. The arrows are the parcel trajectories from the departure points (open circles) which arrive at the regular cell vertices (filled circles).

2.5 Cell-integrated semi-Lagrangian transport schemes

The semi-Lagrangian scheme can either be based on backward or forward trajectories, i.e. by considering parcels arriving or departing from a regular grid, respectively. The majority of semi-Lagrangian schemes are based on backward trajectories because it is usually simpler to remap from a regular to a distorted mesh. The deformed grid resulting from tracking the parcels moving with the flow is referred to as the *Lagrangian grid* while the stationary and regular grid is referred to as the *Eulerian grid*. The curve resulting from tracking a set of points along a latitude is referred to as a *Lagrangian latitude*. Similarly for *Lagrangian longitudes*.

Using backward trajectories the two-dimensional discretization of (20) leads to the CISL scheme

$$\overline{\psi}_{exp}^{n+1} \Delta A = \overline{\psi}_*^n \, \delta A_*^n \tag{27}$$

where

$$\overline{\psi}_*^n = \frac{1}{\delta A_*^n} \iint_{\delta A_*^n} \psi^n \, dA \tag{28}$$

is the integral mean value of ψ over the irregular departure cell area δA_*^n and $\overline{\psi}_{exp}^{n+1}$ is the mean value of ψ over the regular arrival cell area ΔA (see Fig. 2). The approximation of the integral on the right-hand side of (27) employs two steps. Firstly, defining the geometry of the departure cell. Secondly, performing the remapping, i.e. computing the integral over the departure cell using some reconstruction of the sub-grid distribution at the previous time step. The geometrical definition of the departure cell and the complexity of the sub-grid-scale distribution are crucial for the efficiency

and accuracy of the scheme. Before describing how the upstream integral can be approximated, the reconstruction of the sub-grid distribution is discussed.

2.5.1 *Sub-grid representation*

One-dimensional reconstructions Several one-dimensional methods for reconstructing the sub-grid distribution have been published in the literature. The simplest sub-grid representation is a piecewise constant function followed, in complexity, by a piecewise linear representation (van Leer 1977). Both methods are computationally cheap, monotonic and positive definite, but on the other hand excessively damping and therefore not suited for long runs. To reduce the dissipation to a tolerable level, the sub-grid-cell representation must be polynomials of at least second degree. Requirements of computational efficiency puts an upper limit to the order of the polynomials used, which explains why the predominant choice is second order.

The coefficients of the polynomials are determined by imposing constraints. Apart from the basic requirement of mass conservation within each grid cell, the choice of constraints is not trivial. Probably the simplest parabolic fit is obtained by requiring that the polynomial

$$p_i(x) = (a_0)_i + (a_1)_i\, x + (a_2)_i\, x^2, \quad x \in \left[x_{i-1/2}, x_{i+1/2}\right[\tag{29}$$

not only conserves mass in the ith grid cell

$$\int_{x_{i-1/2}}^{x_{i+1/2}} p_i(x) = \Delta x_i\, \overline{\psi}_i \tag{30}$$

but also in the two adjacent cells:

$$\int_{x_{i+1/2}}^{x_{i+3/2}} p_i(x) = \Delta x_{i+1}\, \overline{\psi}_{i+1}, \tag{31}$$

and

$$\int_{x_{i-3/2}}^{x_{i-1/2}} p_i(x) = \Delta x_{i-1}\, \overline{\psi}_{i-1}, \tag{32}$$

(Laprise and Plante 1995). Here $x_{i+1/2}$ and $x_{i-1/2}$ refer to the position of the ith cell border. Substituting (29) into (30), (31) and (32), and evaluate the analytic integrals, result in a linear system that can easily be solved for the three unknown coefficients $(a_0)_i$,$(a_1)_i$ and $(a_2)_i$. Performing this operation for all cells, a global piecewise-parabolic representation is obtained. Note that the method is only locally of second order since it is not necessarily continuous across cell borders. This method is referred to as the piecewise parabolic method 1 (PPM1).

An alternative way of constructing the parabolas, which ensures a globally continuous distribution, is the piecewise-parabolic method of Colella and Woodward (1984) (hereafter referred to as PPM2)

which has been reviewed in the context of meteorological modeling in Carpenter et al. (1990). It is convenient to use the cell average, $\overline{\psi}_i$ and the value of p_i at the left and right cell border, $(a_L)_i = p_i(x_{i-1/2})$ and $(a_R)_i = p_i(x_{i+1/2})$, respectively, instead of using $(a_0)_i$, $(a_1)_i$ and $(a_2)_i$ to define the parabolas. The equivalent formula for $p_i(x)$ is given by

$$p_i(x) = \overline{\psi}_i + (\delta a)_i\, x + (a_6)_i\, \left(\tfrac{1}{12} - x^2\right), \tag{33}$$

where $(\delta a)_i$ is the mean slope of p_i

$$(\delta a)_i = (a_R)_i - (a_L)_i,$$

and $(a_6)_i$ is the "curvature" of the parabola

$$(a_6)_i = 6\,\overline{\psi}_i - 3\,\left[(a_L)_i + (a_R)_i\right].$$

The first constraint is, of course, mass conservation (30). Secondly, the value of p_i at the left cell border, $(a_L)_i$, is fitted with a cubic polynomial using surrounding cell average values. Similarly, for the right border value, $(a_R)_i$. The result is

$$\begin{aligned}(a_L)_i &= \tfrac{1}{2}(\overline{\psi}_{i+1} + \overline{\psi}_i) + \tfrac{1}{6}\left(\Delta a_i + \Delta a_{i-1}\right),\\ (a_R)_i &= (a_L)_{i+1}\end{aligned}$$

for an equidistant grid (for a non-equidistant grid see Colella and Woodward 1984). This uniquely defines the parabolas and guarantees that the global sub-grid distribution is continuous across cell borders. Zerroukat et al. (2002) found in passive advection tests with their scheme that, when using the PPM2 for the sub-grid cell reconstructions (where the parabolas were continuous across cell borders), more accurate solutions were obtained compared to PPM1 (in which the distribution is not necessarily continuous across cell borders).

Instead of using the PPM1/2 Zerroukat et al. (2002) used a cubic generalization of the piecewise parabolic method for the reconstruction of the sub-grid-cell distributions. Of course any kind of reconstruction which is mass conserving can be used, for example rational functions as used in the transport scheme of Xiao et al. (2002) or parabolic splines. At present the most wide-spread sub-grid cell reconstruction method is PPM2.

Note that without further constraining the coefficients of the parabolas they do not guarantee monotonicity or positive definiteness. Standard filters (or limiters) that accommodate these requirements are given in Colella and Woodward (1984) and Lin and Rood (1996). The monotonic filter proposed in Colella and Woodward (1984) is very damping since it reduces the sub-grid distribution to a constant when the polynomial has monotonicity violating variation (that is when $p_i(x)$ takes values outside the range of a_L and a_R). Hence there are two situations in which the sub-grid scale distribution is modified: when $\overline{\psi}_i$ is a local extremum and when $\overline{\psi}_i$ is in between a_R and a_L, but sufficiently close to one of the values so that the parabola takes values outside the range of the edge values. The clipping proposed by Colella and Woodward (1984) can significantly reduce the accuracy in idealized advection tests (see Table 2). An alternative monotonicity filter has been

designed by Zerroukat et al. (2004a). It is formulated for cubic polynomials but can be adapted to the piecewise parabolic method (Dr. M. Zerroukat 2004, personal communication). It first detects if the monotonicity violating behavior is for an extremum or not. If $\overline{\psi}_i$ is an extremum then the high-order parabola is retained otherwise the order of the fitting polynomial is consecutively reduced until the variation is monotone. Then the severe clipping of "peaks" is eliminated and grid-scale noise is removed but without excessive damping. Note that the PPM2 (which unmodified is globally continuous) will be rendered discontinuous at some cell borders after the application of filters.

Fully two-dimensional methods The PPM in one dimension can be directly extended to two dimensions as has been done by Rančić (1992). This fully biparabolic fit involves the computation of nine coefficients, which makes the method computationally expensive. The computational cost can be reduced significantly by using a quasi-biparabolic sub-grid cell representation. Contrary to fully biparabolic fits, the quasi-biparabolic representation does not include the "diagonal" terms and simply consists of the sum of two one-dimensional parabolas, one in each coordinate direction. Using the form (33) for the parabolas, the quasi-biparabolic sub-grid-cell representation is given by

$$p_{ij}(x,y) = \overline{\psi}_{ij} + a^x_{ij}\,x + b^x_{ij}\,\left(\frac{1}{12} - x^2\right) + a^y_{ij}\,x + b^y_{ij}\,\left(\frac{1}{12} - y^2\right),$$

where a^x,b^y and a^y, b^y are the coefficients of the parabolic functions in each coordinate direction (Machenhauer and Olk 1998). This representation significantly reduces the computational cost of the sub-grid-cell reconstruction but, of course, does not include variation along the diagonals of the cells.

By using one-dimensional filters that prevent undershoots and overshoots to the parabolas in each coordinate direction, monotonicity violating behavior can be reduced but not strictly eliminated. In case of negative values at the cell boundaries of both unfiltered one-dimensional parabolic representations, even larger negative values may be present in one or more of the cell corners when the 1D representations are added. The monotone and positive definite filters eliminate only the negative values at the boundaries and not the larger negative corner values.

2.5.2 *Remapping*

For realistic flows the upstream cells deform into non-rectangular and possibly locally concave shapes (see Fig. 2). The question is how to integrate efficiently over a complex area. Several approaches have been suggested in the literature.

Fully two-dimensional schemes In Fig. 3 the departure cells of four different CISL schemes are shown. The simplest departure cell approximation in terms of geometry is the configuration of Laprise and Plante (1995); the departure cell is defined as a rectangle where the edges have the

same orientation as the arrival cell (see Fig. 3a). This is done by tracing the traverse motion of cell edges A, B, C and D, and not the cell vertices. Hereby the upstream cell retain orthogonality, which simplifies the upstream integral. On the other hand mass conservation is lost since the cells can overlap and have gaps between them. The cell geometry must be more advanced in order to obtain mass conservation, which is the primary design goal of CISL schemes.

Rančić (1992) defines the departure cell as a quadrilateral by tracking backward the cell vertices E, F, G, and H, and connecting them with straight lines (Fig. 3b). The vertices are not necessarily aligned with the coordinate axis which leads to some algorithmic complexity. In addition, the sub-grid-scale distribution used by Rančić was a piecewise biparabolic representation which in itself is quite expensive to compute. The combination of the complex geometry of the departure cell and the fully two-dimensional sub-grid-cell representation makes the scheme approx. 2.5 times less efficient than the traditional semi-Lagrangian advection scheme. This has hindered the scheme for use in full models. In order to speed up the remapping process, Machenhauer and Olk (1998) simplified both the geometry of the departure cell and the sub-grid-scale distribution. The departure cell is defined as a polygon with sides parallel to the coordinate axis (Fig. 3c). The sides parallel to the x-axis are at the y-values of the departure points, and the sides parallel to the y-axis pass through I, J, K, and L, located halfway between the departure points. With this form of departure cell and by using the pseudo-biparabolic sub-grid-scale distribution, the integral over the departure cell can be computed much more efficiently compared to the approach taken by Rančić (1992). For advection in Cartesian geometry NM02 reported a 10% overhead compared to the traditional semi-Lagrangian scheme. The NM02 scheme is hereafter referred to as the pseudo-biparabolic CISL (PB-CISL) scheme.

Cascade schemes The remapping can also be performed by splitting it into two one-dimensional remapping steps using the so-called cascade approach (Purser and Leslie 1991; Leslie and Purser 1995; Rančić 1995; Nair et al. 1999).

One such approach is the conservative cascade CISL scheme (CC-CISL) of Nair et al. (2002). In this scheme the departure cells are also defined as polygons with sides parallel to the coordinate axis and in each one-dimensional cascade step the PPM2 is used. Compared to the PB-CISL scheme the departure cell geometry in the CC-CISL scheme is defined somewhat differently (see Fig. 3d). Two of the sides parallel to the y-axis, $x = x(E)$ and $x = x(G)$, are defined as in the PB-CISL scheme and the remaining two sides are at the Eulerian longitude $x = x_i$. The sides parallel to the x-axis are determined from the intermediate Lagrangian grid points M, N, O, P, Q, and R defined as $y = 1/2\,\{y(\mathrm{M}) + y(\mathrm{N})\}$, $y = 1/2\,\{y(\mathrm{O}) + y(\mathrm{P})\}$, $y = 1/2\,\{y(\mathrm{P}) + y(\mathrm{Q})\}$, and $y = 1/2\,\{y(\mathrm{R}) + y(\mathrm{M})\}$, respectively. The y-values of the intermediate points are determined by cubic Lagrange interpolation between the y-values of four adjacent departure points along the *Lagrangian latitude* (dashed line on Fig. 3d). The upstream integral is computed by a remapping in the north-south direction from the Eulerian cells to the intermediate cells (crosshatched rectangular regions on Fig. 3d) followed by a remapping along the *Lagrangian latitudes* from the intermediate cells to the departure cells. Since the two remappings are one dimensional the scheme is more than

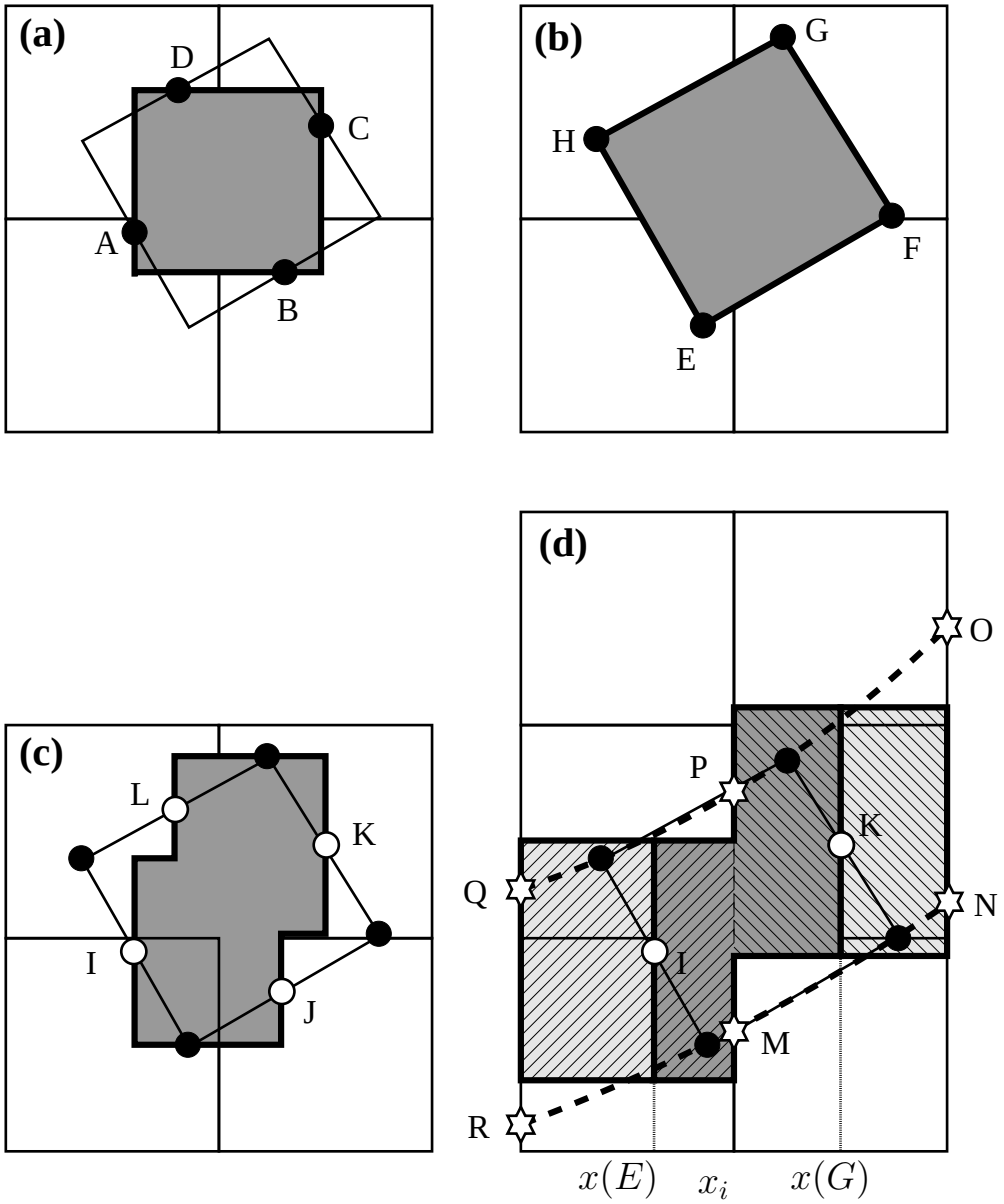


Fig. 3: The departure cells (shaded area) when using the (a) scheme of Laprise and Plante (1995), (b) scheme of Rančić (1992), (c) PB-CISL scheme and (d) CC-CISL scheme, respectively. The filled circles are the departure points, open circles the midpoints between the departure points, asterisks are the intermediate grid points which are used to define the intermediate cells in the cascade scheme (crosshatched area).

twice as efficient as the PB-CISL scheme (Nair et al. 2002).

Note that the departure areas in Fig. 3b, c and d completely cover the entire integration area without overlaps or cracks. Hence the total mass is conserved exactly. The size of the PB-CISL area

Scheme	l_1	l_2	l_∞	Min	Max
PB-CISL	0.075	0.051	0.083	-0.0088	-0.083
CC-CISL	0.051	0.039	0.076	-0.0070	-0.076
PB-CISL-M	0.077	0.089	0.18	-0.0038	-0.18
CC-CISL-M	0.070	0.086	0.186	0.0	-0.186

Table 2: Standard non-dimensional error measures for solid body rotation of a cosine bell for the PB-CISL scheme and the CC-CISL, respectively. The angle between the axis of solid body rotation and the polar axis is 30°, the resolution is 2.8125° and one revolution is completed in 256 time steps. The letter M denote the monotonic option (Colella and Woodward 1984). Additional details are in Chapter III.

(Fig. 3c) is equal to that of the Rančić area (Fig. 3a) whereas the CC-CISL area (Fig. 3d) only approximately so. Therefore, accepting the Rančić definition of the departure area as the most accurate, the PB-CISL area is the most accurate one of the two other schemes. The "jump" in the north and south walls in the CC-CISL scheme are not necessarily midway between the departure points, and hence the CC-CISL scheme may conserve mass locally less accurately than the PB-CISL scheme. On the other hand the CC-CISL scheme may in certain cases obtain a more accurate sub-grid-scale representation, namely when significant variations are along the *Lagrangian latitudes* and these are sloping toward north-east and south-east. In such situations the PB-CISL sub-grid-scale representation becomes less accurate due to the missing "diagonal" terms (see Table 2). Generally the most accurate sub-grid-scale representation of the three schemes is obtained with the full biparabolic representation of Rančić (1992).

The schemes discussed so far are based on integrating over areas which, given the departure cell topology for complex flows, can lead to some algorithmic complexity. Unlike the "area approach" Laprise and Plante (1995), Rančić (1995) and Zerroukat et al. (2002) have developed conservative cascade schemes which do not make explicit reference to departure areas during the remapping procedure. Hence they are not cell-integrated schemes in the sense considered so far where the "actual" integral over an area is computed.

The scheme of Laprise and Plante (1995) is not based on the finite-volume approach, but mass conservation is obtained globally anyway by differentiating a cumulative mass function. Rančić (1995) assigned mass to nodes (or "mass points") and used a cascade scheme similar to the one of Purser and Leslie (1991) based on forward trajectories. Instead of remapping along Lagrangian latitude and longitudes using Lagrange interpolation, a finite-volume method with PPM2 was used for the remapping. However, the application of the scheme for non-uniform grids and the extension to spherical geometry is not obvious.

A scheme that has also been extended to spherical geometry is the one of Zerroukat et al. (2002) called the SLICE (Semi-Lagrangian Inherently Conserving and Efficient) scheme. The scheme

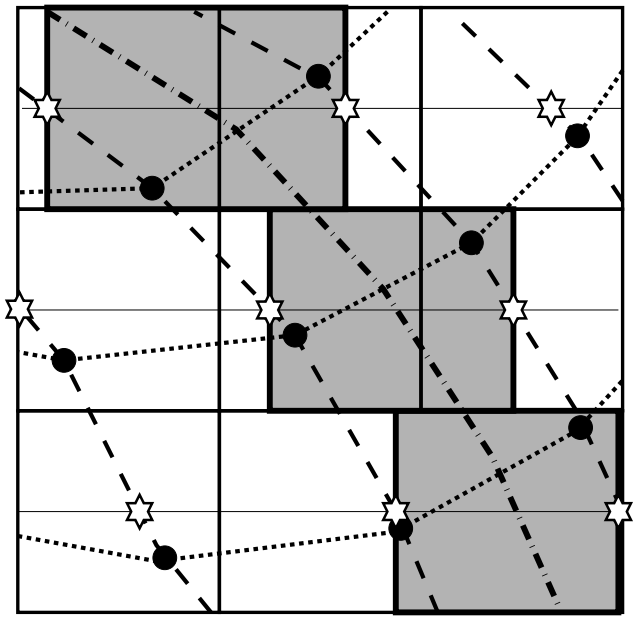


Fig. 4: Graphical illustration of the remappings in the SLICE scheme. The filled circles are the departure points corresponding to the cell vertices. The dotted/dashed lines are the Lagrangian latitudes/longitudes defined by connecting the departure points which arrive along the same latitude/longitude with straight line segments. The shaded areas are the intermediate Eulerian areas which are defined by the crossings between the Lagrangian longitudes and the Eulerian latitudes passing through the center of the Eulerian cells (thin lines). The crossings are marked with asterisk. The dash-dotted line is the line along which the cumulative distance function is defined and is used for the second remapping.

does not make explicit reference to areas during the remapping but is, however, similar to the CC-CISL scheme. As in the CC-CISL scheme the problem is divided into two one-dimensional remappings; one along an Eulerian coordinate direction and then one along a Lagrangian curve corresponding to the translation by the winds of one of the coordinate isolines. The remapping procedure is graphically illustrated on Fig. 4. As in the PB-CISL and CC-CISL schemes the cell vertices are tracked backward. The corresponding departure points are connected with straight lines to define *Lagrangian longitudes* and *latitudes*. Regular intermediate cells are defined by the intersections between the *Lagrangian longitudes* and the *Eulerian latitudes* that pass through the center of the cells. Similarly to the CC-CISL scheme the cell averages are mapped from the Eulerian cells to the regular intermediate cells defined by the intersections. The remap from the intermediate cells to the departure cells is quite different from the CC-CISL scheme. A cumulative distance function is defined in order to define the north-south cell borders along the *Lagrangian longitudes* passing through the center of the cells. The second remapping is performed along the *Lagrangian longitude* and by using the cumulative distance function to indicate the location of the cell "walls" along the *Lagrangian longitudes*. Hence the method makes no explicit reference to areas during the remapping procedure.

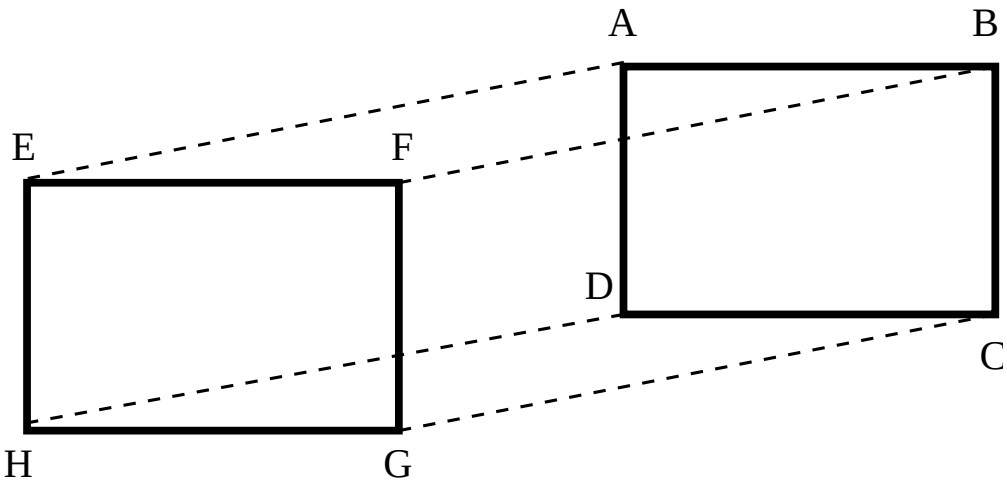


Fig. 5: The figure illustrates the equivalence between Eulerian flux-form and CISL schemes. See text for details.

A great potential of cascade schemes is that they may be extended to three dimensions without excessive computational cost and algorithmic complexity. For example, three-dimensional Lagrangian interpolation requires $\mathcal{O}(p^3)$ operations, where p is the formal order of accuracy of the interpolator, while cascade schemes require $\mathcal{O}(p)$ operations (e.g., Purser and Leslie 1991). Fully higher-dimensional cell-integrated schemes, on the other hand, increase rapidly in complexity as the number of dimensions is increased.

2.5.3 *A flux-form semi-Lagrangian scheme*

Lin and Rood (1996) and Leonard et al. (1996) have developed a scheme referred to as a finite-volume semi-Lagrangian scheme although it is quite different from the CISL schemes discussed so far. The approximation of the upstream integral can also be formulated in terms of fluxes instead of an integral over the departure area. The equivalence is illustrated on Fig. 5 for a uniform velocity field. The regular arrival cell is the rectangle ABCD and the corresponding departure cell is EFGH. The trajectories are the dashed lines connecting vertices of the departure and arrival cell. In the continuous case the integral over the departure cell is identical to the sum of the fluxes through the arrival cell walls, i.e. by adding the integrals over ABFE, DCGH, BCGF and ADHE computed with a positive sign for inflow and negative for outflow. For a "geometrical" CISL scheme using long time-steps, the integrals approximating the fluxes will stretch over several Eulerian cells and it is therefore more economical to integrate over the departure cell.

The fluxes can, however, be approximated using a less conceptional but more economical approach as done in Lin and Rood (1996) and Leonard et al. (1996). The respective schemes are identical, except for implementation details. The schemes approximate the fluxes using a combination of

one-dimensional flux-form and advective form operators. Thereby not only the transport components along coordinate axis are included as in conventional operator splitting schemes, but also the traverse motion is included. In fact for a uniform advecting velocity the method is equivalent to a cell-based semi-Lagrangian scheme in terms of the equivalence described above (Leonard et al. 1996). Contrary to the CISL scheme the flux-form schemes of Lin and Rood (1996) and Leonard et al. (1996) preserve a constant for non-divergent flow fields. They are extended to long time-steps by using integer advection and are therefore not semi-Lagrangian schemes in the traditional sense. The schemes are widely used in operational models for tracer advection. It is, however, not clear how the schemes could be incorporated in the dynamical core of existing semi-Lagrangian models since they have not been formulated with semi-implicit time stepping. For explicit time-stepping a dynamical core has been designed based on the Lin and Rood (1996) scheme (Lin 2004).

2.5.4 *Extensions to spherical geometry*

The singularities on the sphere are one of the main challenges for transport schemes. The number of schemes developed in Cartesian geometry is significantly larger than the number of schemes formulated for spherical geometry and that can transport quantities accurately and without excessive time-step limitations over the poles. For example the CISL schemes of Rančić (1992), Rančić (1995) and Laprise and Plante (1995) discussed in the previous sections have not been extended to spherical geometry.

Most algorithms require a certain amount of "engineering" to tackle the pole problem that often reduces the efficiency and simplicity of the algorithms. In Cartesian geometry the most accurate approximation to a departure cell, given the departure points, is the quadrilateral resulting from connecting the departure points with straight lines. Similarly, in spherical geometry the cells defined by connecting the departure points with great circle arcs are the optimal choice. But as in Cartesian geometry integrating along the optimal curves leads to complicated and computationally expensive algorithms. Therefore as in the Cartesian case the area approximation must be simplified.

The PB-CISL and CC-CISL schemes are extended to spherical geometry by using the μ grid, i.e. a latitude-longitude grid in which the latitude θ is replaced by $\mu = \sin\theta$. This transformation is invariant in the sense that the departure cells and corresponding upstream integrals take exactly the same form as in Cartesian geometry. In the vicinity of the poles, however, the cells on the μ-grid are a poor representation of the cells on the spherical latitude-longitude grid and some "engineering" is needed. In the PB-CISL scheme local area-preserving tangent planes at the poles are introduced. The areas in which the tangent planes are used is referred to as the *polar cap*. It is assumed that the east and west walls of the cells are straight lines in the tangent coordinate system. Using that assumption, the remapping procedure on the *polar cap* is formally equivalent to the Cartesian case. However, the east and west sides have a strong curvature when viewed in the μ-coordinate system. Hence extra latitudes are introduced in the *polar cap* so that the assumption of straight east and west walls in the tangent plane does not significantly degrade the accuracy. In the Lagrangian belt

Scheme	l_1	l_2	l_∞
PB-CISL	0.063	0.046	0.048
CC-CISL	0.054	0.042	0.065
SLICE-S	0.079	0.049	0.042

Table 3: Standard non-dimensional error measures for solid body rotation of a cosine bell over the pole for the PB-CISL scheme, the CC-CISL and SLICE-S scheme, respectively. The resolution is 2.8125° and one revolution is completed in 256 time steps. For additional details see e.g. NM02.

containing the pole point (referred to as the *singular belt*) the algorithm breaks down since the Lagrangian cell containing the Eulerian pole is not well defined. The total mass inside the *singular belt* can, however, easily be computed. The mass inside the *singular belt* is then distributed among the cells using Lagrange weights.

In cascade schemes the pole problem manifest itself by the fact that some Lagrangian latitudes may not cross a given Eulerian longitude. The CC-CISL scheme was first formulated for meridional Courant numbers less than unity, $C_\theta < 1$ (Nair et al. 2002), and later extended to longer time steps (Nair 2004). The schemes were extended to spherical geometry using the μ-grid which, as for PB-CISL, means that, away from the polar caps, the algorithm takes the same form as in the Cartesian case. In the general CC-CISL scheme the two-dimensional PB-CISL scheme was used over the polar caps and the unmodified CC-CISL scheme elsewhere (Nair 2004). Hereby, most of the efficiency of the CC-CISL scheme was retained while allowing for long time-steps over the *polar caps*.

The SLICE-S scheme uses a regular latitude-longitude grid (Zerroukat et al. 2004b). The cascade approach is based on the crossings of Lagrangian longitudes and Eulerian latitudes. Near the poles it can not be guaranteed that all Lagrangian longitudes cross a specific Eulerian latitude. In such cases the mass is simply distributed to the closest crossings based on distance-dependent weights. In principle the algorithm as described so far handles the pole problem, but the grid distortion near the poles can make the cascade axis nearly parallel instead of nearly orthogonal, making the remapping inaccurate. Therefore the mass is redistributed with a "post fix" procedure based on Lagrange weights similarly to the procedure in the PB-CISL scheme.

Of the three cascade schemes described here the SLICE-S scheme is the simplest regarding algorithmic complexity. It does not use high resolution *polar cap* belts and tangent planes. Since the CC-CISL and PB-CISL increase the resolution over the polar belts the accuracy of these schemes tends to be higher for transport over the poles compared to the SLICE-S scheme, but at an extra computational cost (see Table 3).

2.5.5 *Summary*

A review of the CISL schemes published in the literature has been given. These schemes can be divided into two categories: Fully two-dimensional methods and cascade schemes which split the problem into two one-dimensional ones. In general the CISL schemes perform much better than the traditional non-conservative schemes: they conserve mass exactly, permit long time steps, have the option of being monotone or simply positive definite, and have small phase and amplitude errors. For passive advection the two-dimensional schemes have a small computational overhead compared to traditional schemes, while cascade schemes are slightly more efficient than traditional methods. The accuracy of cascade and fully two-dimensional schemes are comparable in idealized advection tests. As discussed in this review of CISL schemes, the cascade approach may have an advantage when the flow is along cell diagonals. In this case the remapping is performed along the characteristics of the flow and thereby "catch" more of the diagonal variation.

The main criticism of CISL schemes is the amount of "engineering" needed near the poles. This leads to some algorithmic complexity. Another potential problem with CISL schemes is their extension to three dimensions. A fully three-dimensional extension of the fully two-dimensional scheme would be very expensive and complex. The cascade approach, on the other hand, seems straight forward to extend to three dimensions at little extra computational cost.

So far CISL schemes have only been applied to advection of passive tracers and have not been applied to full model systems. As discussed in the Introduction, accurate tracer advection requires that the mass fields and wind fields supplied to the advection scheme are consistent. The only way to ensure this (without *ad hoc* "fixing" algorithms) is to use the CISL scheme for the continuity equation of the dynamical core as for tracer advection.

To exploit the efficiency of semi-Lagrangian methods in full models it is, however, important to use long time steps which requires a special treatment of the equation terms involved in gravity wave motions, for example treating the fast waves semi-implicitly. So far CISL schemes have not been applied to full models. The flux-form semi-Lagrangian scheme of Lin and Rood (1996), which is an Eulerian scheme extended to long time steps by integer advection, has been applied in an hydrostatic dynamical core using explicit time stepping (e.g., Lin 2004). For the application of CISL schemes in complete models they must be coupled with semi-implicit time stepping in order to remain efficient.