ECMWF Feature article

from Newsletter Number 145 – Autumn 2015

METEOROLOGY

# An all-scale, finite-volume module for the IFS

Nicholashan/iStock/Thinkstock

www.ecmwf.int/en/about/news-centre/media-resources

doi:10.21957/f4gm176p

*This article appeared in the Meteorology section of ECMWF Newsletter No. 145 – Autumn 2015, pp. 24–29.*

# An all-scale, finite-volume module for the IFS

Piotr Smolarkiewicz, Willem Deconinck, Mats Hamrud, Christian Kühnlein, George Mozdzynski, Joanna Szmelter, Nils Wedi

ECMWF hosts the European Research Council-funded project PantaRhei, which explores novel numerical methods to complement existing, highly optimised numerical weather prediction (NWP) models. The need for such innovation stems from the fact that state-of-the-art global NWP models using the spectral transform method may become computationally inefficient at very fine resolutions due to the communication overhead associated with global spectral transformations.

As a first step, we have developed an autonomous, all-scale numerical module which uses the finite-volume method (Box A) to supplement ECMWF's Integrated Forecasting System (IFS). This module is compatible with emerging energy-efficient, heterogeneous hardware for high-performance computing (HPC), and it is able to represent elements of real weather on a large range of scales, including cloud-resolving scales.

## Motivation

The advance of massively parallel computing in the 1990s and beyond has encouraged finer grid intervals in NWP models. This has improved the spatial resolution of weather systems and enhanced the accuracy of forecasts, while stimulating the development of global non-hydrostatic models. Today many operational NWP models include non-hydrostatic options either for regional predictions or research. However, to date no NWP model runs globally in operations at resolutions where non-hydrostatic effects are important (*Wedi & Malardel*, 2010; *Wedi et al.*, 2012). Such high resolutions are still computationally unaffordable and too inefficient to meet the demands of the limited time window for distributing global forecasts to regional NWP recipients and, ultimately, the public.

Efforts to ensure the computational affordability of global non-hydrostatic forecasts face a twofold difficulty. On fine grids the spectral transform method becomes computationally inefficient because of the required global data-rich inter-processor communications (*Wedi et al.*, 2013). Therefore, simply scaling up the number of processors would be unaffordable, not least due to the huge increase in electric power consumption this would entail.

At the same time, replacing hydrostatic primitive equations (HPE) that have been central to the success of weather and climate prediction exacerbates the efficiency problem. In particular, with the simulated vertical extent of the atmosphere thin compared to its horizontal extent, the vertically propagating sound waves supported by the non-hydrostatic Euler equations, from which HPE derive, impose severe restrictions on the numerical algorithms. The hydrostatic balance assumption underlying HPE conveniently filters out vertically propagating sound waves, therefore permitting large time steps in the numerical integration. Moreover, HPE imply the separability of horizontal and vertical discretisation, thus facilitating the design of effective flow solvers, such as the semi-implicit semi-Lagrangian (SISL) time stepping combined with the spectral-transform spatial discretisation that is used today. Such separability does not apply in non-hydrostatic models.

While NWP strives to extend its skill towards finer scales, non-hydrostatic research models endeavour to extend their realm towards the global domain. The two routes of development must meet, but the way to merge the different areas of expertise is far from obvious. Altogether, NWP is at a crossroads. Although massively parallel computer technology promises continued advances in forecast quality, the latter cannot be achieved by simply applying the existing apparatus of NWP models to ever finer grids.

 doi:10.21957/f4gm176p

**A**

### Finite-volume method

The finite-volume method is an approach to the approximate integration of partial differential equations (PDEs) describing natural conservation laws. Similar to the finite difference method or finite element method, solutions are calculated at discrete places on a meshed geometry.

'Finite volume' refers to the small volume surrounding each node point on a mesh. In PDEs, integrals of the divergence terms over these finite volumes are converted into surface integrals using the Gauss divergence theorem. On a discrete mesh, these surface integrals are then evaluated as a sum of all fluxes through individual surfaces bounding each finite volume (Figure 1).

Because the flux entering a given volume is identical to that leaving the adjacent volume, these methods are conservative. An important advantage of the finite-volume method is that it can easily be formulated for unstructured meshes.

## A new way forward

Recognising the predictive skill of the IFS, we seek to address the challenges outlined above by supplying a complementary non-hydrostatic dynamical module with the capabilities of a cloud-resolving model, concurrently driven by large-scale IFS predictions based on the HPE. The first step towards this paradigm is the development of an autonomous, global, finite-volume, non-hydrostatic dynamical module capable of working on the IFS's reduced Gaussian grid and, in principle, on any horizontal mesh.

Partial differential equations (PDEs) require the calculation of differential operators. In the IFS, spatial differentiation is conducted in spectral space, and there are no practical means of calculating derivatives locally in the physical space. Moreover, the increased complexity of the currently available non-hydrostatic IFS option significantly increases the number of spectral transforms required, nearly doubling the cost compared to the hydrostatic forecast, which is already projected to be too slow on existing supercomputers at non-hydrostatic resolutions within the operational time window of 1 hour.

Supplementing the IFS with a complementary non-hydrostatic module makes it possible to use numerical procedures unavailable in the SISL spectral dynamical model with minimal disruption to its highly optimised code. In particular, such a module can replace global communication and computation with local equivalents, providing an effective testing ground for assessing the utility of innovations from outside the realm of spectral methods in the context of real weather. Realisations of this concept can be as simple as advecting selected critical fields in a conservative finite-volume fashion. Or they can be as complex as replacing the whole dynamical core with a cloud-resolving model, while driving the latter with a coarser solution of the spectral-transform hydrostatic model. Regardless of the envisaged realisation, having in place an alternative non-hydrostatic module capable of efficiently exchanging dependent variables with the IFS model, but also of providing complete solutions by itself, is a prerequisite for progress.

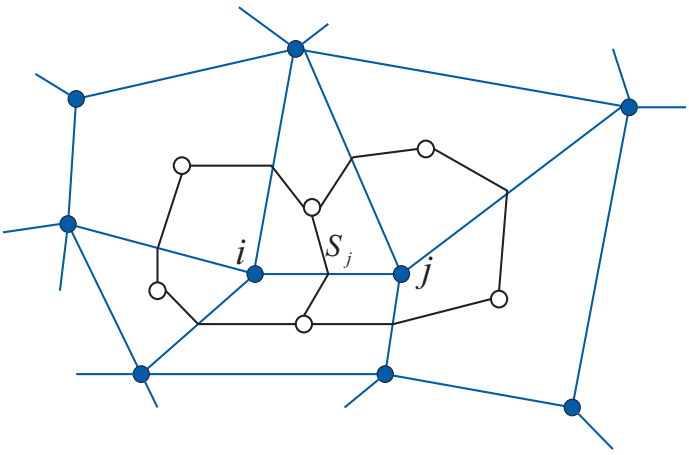


**Figure 1** An unstructured median-dual mesh in 2D. The edge connecting the nodes *i* and *j* of the primary polygonal mesh intersects, at its mid-point, the face *Sj* shared by computational dual cells surrounding the nodes *i* and *j*; the white circles represent the centres of mass of the primary mesh, while the blue and black lines belong to the primary and the dual mesh, respectively.

doi:10.21957/f4gm176p

## Discretisation

The new, autonomous, all-scale finite-volume module (hereafter FVM) is a hybrid descendant of the interdisciplinary all-scale Eulerian/semi-Lagrangian research model EULAG (*Smolarkiewicz et al.*, 2014) and the generalisation of its non-oscillatory forward-in-time numerics to unstructured meshes (*Szmelter & Smolarkiewicz*, 2010; *Szmelter et al.*, 2015). The default governing PDEs of the FVM are based on the fully compressible Euler equations, and simplified soundproof PDEs are included as an option. The standard latitude-longitude (lat-lon) spherical framework forms the basis of the analytical formulation, as it does in the IFS. The notorious stiffness of the global lat-lon system is circumvented by the flexibility of the finite-volume discretisation on an unstructured mesh. Figure 1 illustrates such a discretisation. Its key geometric objects are the *nodes* in which data reside and the solution is sought, the *edges* that connect nodes forming the primary mesh, and the *dual cells* over which differential operators are evaluated using vector calculus. In the FVM, unstructured meshes can be built around arbitrarily specified nodes. Of particular interest are meshes built around the grids admissible in the IFS, shown in Figure 2, as they provide support for spectral transforms.

In the FVM, all dependent variables are co-located in the nodes, accommodating both spectral transforms and grid-point solutions within the same code while facilitating time-stepping schemes with a large degree of implicitness. In three spatial dimensions, the mesh is prismatic, that is, the horizontal discretisation is common to all vertical levels of the computational space. A uniform finite-difference discretisation in the vertical is chosen to facilitate the solution of intricate elliptic problems in thin spherical shells. To accommodate planetary orography and a pliant height coordinate in physical space (such as hydrostatic isobars), the governing equations are cast in a generalised time-dependent curvilinear reference frame. To numerically solve the governing PDEs, the FVM employs proven semi-implicit non-oscillatory forward-in-time integrators of the governing PDE systems (*Smolarkiewicz et al.*, 2014) conceptually similar to, but more general than, those used in the IFS.

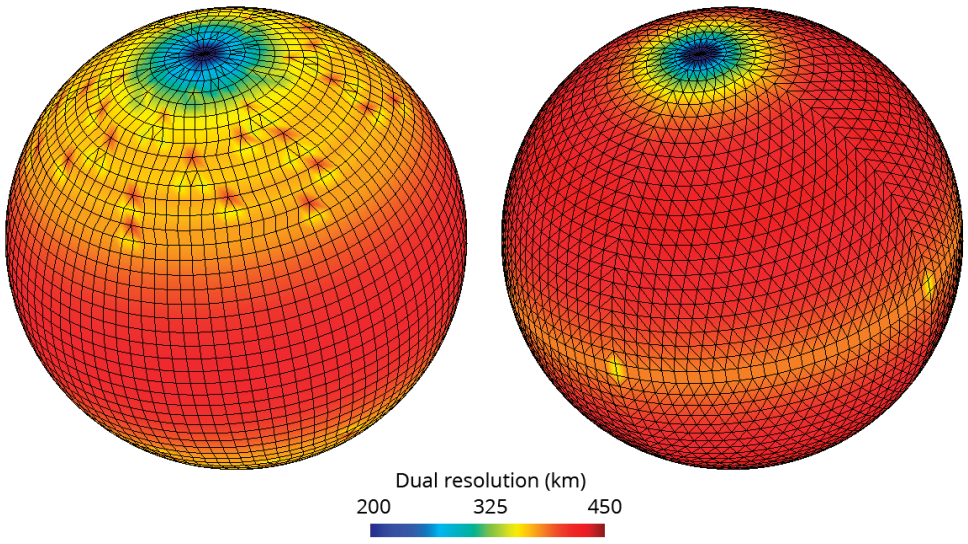


**Figure 2** Two primary meshes generated on N24 reduced Gaussian grid points with an approximate resolution of 3.75° (415 km). N24 indicates 24 latitudes between a pole and the equator. Gaussian grids are latitude-longitude grids. In reduced Gaussian grids, the number of grid points on latitudes near the poles is reduced to achieve a more even spacing of grid points. The shading represents the dual resolution, computed as the square root of the local dual volume.


doi:10.21957/f4gm176p

## Data structure and parallelisation

### *Flexible object-oriented data structure*

The conceptual building blocks of the FVM were established in EULAG, and they are well documented in the literature. Nonetheless, many aspects of the discrete apparatus had to be customised for the parallel implementation on the unstructured mesh constrained by the IFS grid. In particular, the FVM module is based on a newly developed framework called Atlas.

The Atlas framework provides parallel, distributed, flexible, object-oriented data structures for both structured and unstructured meshes on a sphere. It separates concerns of mathematical model formulation and numerical solutions from the cumbersome management of unstructured meshes, distributed memory parallelism, and input/output of data. While the FVM uses FORTRAN 2003, it is recognised that handling flexible structures and carefully controlled memory-management are not easily achievable using the FORTRAN language. Hence the language of choice for Atlas is C++, which provides excellent object-orientated support and builds upon C's strengths in memory management.
A FORTRAN 2003 interface exports all of Atlas's functionality to the FVM.

In principle, the FVM and Atlas can accommodate nearly any form of horizontal meshing. However, in the envisaged context of being able to use spectral transforms on the finite-volume mesh, the nodes have to satisfy certain constraints. They must be distributed in a structured way in a lat-lon domain, where latitudes are distributed as the roots of the Legendre polynomials, and longitudes are distributed uniformly on each latitude while reducing in number going from the equator towards the poles. Given these fixed node locations, Atlas is capable of generating a hybrid triangular/quadrilateral mesh. The IFS's set of reduced Gaussian grids satisfy these requirements, and the meshing approaches shown in Figure 2 illustrate Atlas's capabilities. These meshes are very coarse, for illustrative purposes only. For the target applications, the required meshes are N1280 and N2000, corresponding to grid increments of about 10 and 5 km. For such meshes the coordinates and connectivity tables forming the *Mesh* object in Atlas have an enormous memory footprint. This is why *Mesh* is distributed over parallel tasks.

### *Parallelisation scheme*

Because the FVM operates at the nodes of the IFS grid, it seamlessly inherits the load-balanced equal-regions domain decomposition parallelisation scheme of the IFS (*Mozdzynski et al.*, 2015). Like the IFS, the FVM is designed to run on HPC facilities which group computational cores into tasks. Computational cores in each task have access to shared memory among themselves, whereas tasks do not share memory and are connected by message passing.

The FVM hybridises two standards – the Message Passing Interface (MPI) and Open Multi-Processing (OpenMP) – for parallelisation, each with a different purpose. MPI is a standard handling the communication of data between tasks. It involves distributing data between tasks and requires information to be sent and received at regular intervals to advance a simulation. OpenMP is a standard for easily managing multiple computational threads working with shared memory. For example, several iterations in one *do* loop can

**Figure 3** Equal region domain decomposition with 1,600 partitions for the N1280 octahedral reduced Gaussian grid (see right-hand panel in Figure 2).

be executed simultaneously within an algorithm. Experience gained with the IFS shows that the hybrid MPI/OpenMP parallelisation delivers about 20% better performance than MPI alone. This is the result of an improved load balance and memory scalability for large grids.

The parallelisation of the FVM distributes the computational mesh with its unstructured horizontal index in such a way that structured vertical columns are always preserved in memory as a contiguous entity. The Atlas framework, upon which the FVM is built, is responsible for generating a distributed mesh, and it provides communication patterns using MPI to exchange information between the different partitions of the mesh. To minimise the cost of sending and receiving data, the distribution of the mesh is based on an equal-regions domain decomposition algorithm for a quasi-uniform node distribution on the globe.

As shown in Figure 3, such a decomposition divides the globe into bands oriented in zonal direction, and it subdivides each band into a number of regions so that globally each region has the same number of nodes. The bands covering the poles are not subdivided.

To evaluate finite-volume differential operators, the FVM requires only nearest-neighbour communications for halos of adjacent partitions. As shown in Figure 4, the halo thickness is limited to one element of the primary mesh. The nearest-neighbour communication available in the FVM contrasts with the global communication required for spectral transforms. It is an essential part of the computational arrangements in the FVM since global communication becomes a severe bottleneck as more and more cores are added to the HPC cluster. The underlying Atlas framework is used to create the internal halos and to manage communication patterns to exchange data in the halos, thus separating this concern from the FVM.

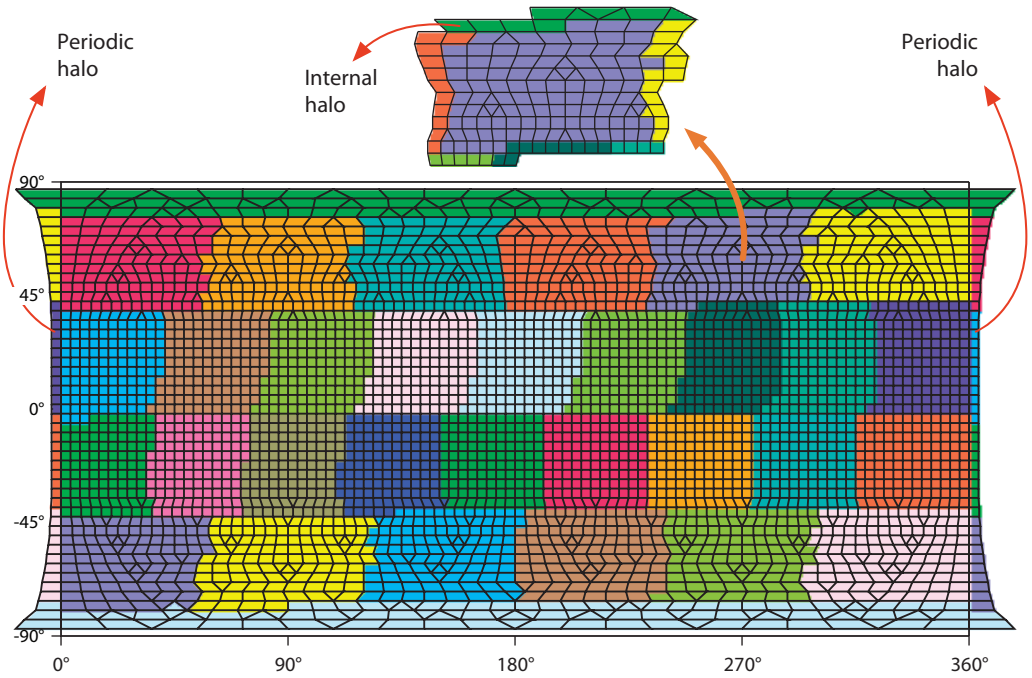


**Figure 4** The computational mesh shown in the left-hand panel of Figure 2 is here mapped onto tasks (32 partitions). Also shown is the internal halo of one partition, and the periodic halo responsible for the periodic boundary condition.


doi:10.21957/f4gm176p

## Examples

In the following, we discuss solutions generated with the compressible option of the FVM. In the course of the module's development, the acoustic and two soundproof options were also extensively used to benchmark and verify the module. These solutions are not shown here, as they corroborate the conclusions of *Smolarkiewicz et al*. (2014) regarding the relative comparability of compressible and soundproof results. Two physical problems are considered: the sheared orographic flow on a small planet with a 20.4 km radius, to highlight the non-hydrostatic quality of the solver; and baroclinic instability on an Earth-like planet, which concerns essentially hydrostatic global weather. The two problems serve to illustrate diverse aspects of the FVM.

### *Quasi-two-dimensional orographic flow with linear vertical shear*

This classical problem constitutes a particularly discriminating test of a non-hydrostatic model's capabilities, because in the presence of shear the non-hydrostatic and hydrostatic equations predict a fundamentally different propagation of orographically-forced gravity waves. While hydrostatic models produce a vertically propagating gravity wave, the correct solution is that of a trapped, horizontally propagating gravity wave.

The setup of the numerical experiment assumes a smooth mountain range centred at the equator of a tiny planet with a radius of about 20 km. The hill is only 500 m tall and has gentle slopes with a height-to-width aspect ratio of 20%. The assumed atmosphere has a simple structure with a constant Brunt-Väisälä frequency of 0.01 $s^{-1}$ (corresponding to a typical tropospheric value), and wind increasing linearly from 10 m/s at the surface at the equator to about 36 m/s at tropopause altitude (10.5 km), and remaining constant aloft.

The FVM domain is discretised in the horizontal using a finite-volume dual mesh generated around an N128 octahedral reduced Gaussian grid (see Figure 2), with a horizontal mesh spacing of about 250 m at the equator. In the vertical the model domain is resolved with 137 levels stretched smoothly, such that the vertical spacing increases from 70 metres near the ground to 1,400 metres near the top, set at 85 km. The integration time is 2 hours with a time step of 3 seconds.

Figure 5 shows isolines of the vertical velocity after two hours of wave evolution simulated with the FVM. Figure 5a shows the solution in the vertical cross section at the equator, and Figure 5b shows a horizontal cross section over the spherical surface at a height of 3 km. The key result is that there is a trapped lee wave behind the mountain, reminiscent of lee waves of clouds often seen downstream from mountains. Such clouds are a sign of wave updrafts, in which ascending moist air cools adiabatically so the water vapour condenses, before evaporating in the downdrafts. The importance of this result lies in the fact that it cannot be produced with hydrostatic models. If the same problem were solved with HPE, the solution would be a single vertically propagating mountain wave with consecutive lenticular clouds eventually forming above the mountain.

It should be noted that our solution agrees with the numerical results and linear analysis discussed in *Wedi & Smolarkiewicz* (2009). The lee wave has a dominant horizontal wavelength of about 14 km, corresponding to 40 degrees of longitude on the small planet. The wave energy leakage through the tropopause excites a weaker stratospheric wave with a dominant wavelength, supported by the ambient conditions, about twice that of the trapped wave below. Due to the domain periodicity, the upstream influence of the mountain can already be seen after 2 hours at the downstream end of the wave packet.

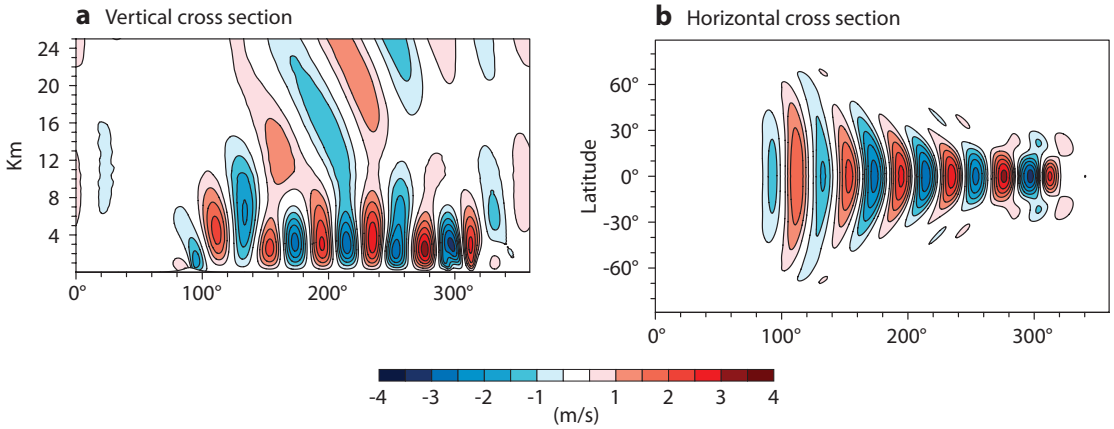


**Figure 5** Gravity wave developing in the lee of a hill placed on a tiny planet. The hill can be seen in panel (a), near the lower left corner between 80° and 100°. Red and blue patterns show updrafts and downdrafts, respectively, with contours of vertical velocity in m/s, in (a) vertical cross section along the equator, and (b) horizontal cross section at 3km above the surface.

### *Baroclinic instability*

While the preceding section addresses the FVM's non-hydrostatic performance, here we illustrate its skill in simulating essentially hydrostatic motions. To this end, we have repeated the inviscid baroclinic instability simulations used in *Smolarkiewicz et al*. (2014), where the authors discuss a series of experiments with soundproof and compressible equations using a coarse horizontal resolution of 128 by 64 nodes on a regular lat-lon grid. Here we show the corresponding compressible result generated with the FVM using an N800 octahedral reduced Gaussian grid with a horizontal resolution of 12 km (roughly corresponding to the current operational resolution of the IFS), and 61 stretched vertical levels with the resolution changing from 50 m near the ground, through 150 m at 2 km, to about 850 m at 24 km (the model top).

Figure 6 shows the baroclinic wave train at day 8 after evolving from a weakly perturbed unstable-equilibrium initial state consisting of the two planetary jets in the mid-latitudes. Figure 6a shows isolines of potential temperature (isentropes) overlaid with contours of the vertical velocity, in the vertical cross section through the centre of the northerly jet. Figure 6b shows isentropes and the meridional velocity at the sphere's surface. Together these figures illustrate the 3D structure of idealised frontogenesis and the formation of weather systems in the mid-latitudes. These results are consistent with those familiar from the literature. Notably, at the relatively high resolution used, the simulation begins to capture sharp frontal discontinuities and the associated radiation of mesoscale gravity waves (*Plougonven & Zhang*, 2013). The grid-scale features seen above the collapsing front are inevitable and are controlled by the model numerics.

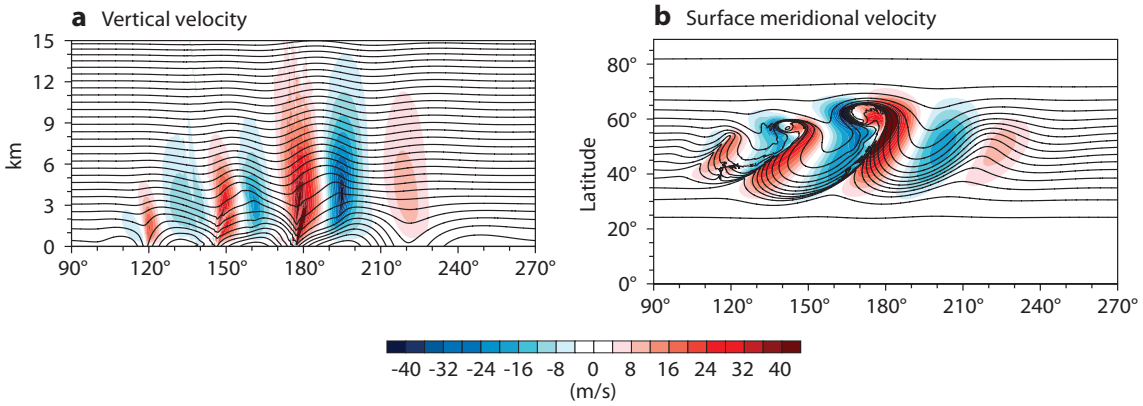


**Figure 6** Baroclinic instability, day 8, shown by (a) contours of vertical velocity (shading, in cm/s) overlaid with isentropes (shown with contour intervals of 5 °C) in the vertical cross section at 53°N, and (b) contours of surface meridional velocity (shading, in m/s) and the isentropes. The steepening of the isentropes to the vertical together with the radiation of mesoscale gravity waves and the generation of grid-scale features in (a) are indicative of the frontal collapse in (b).

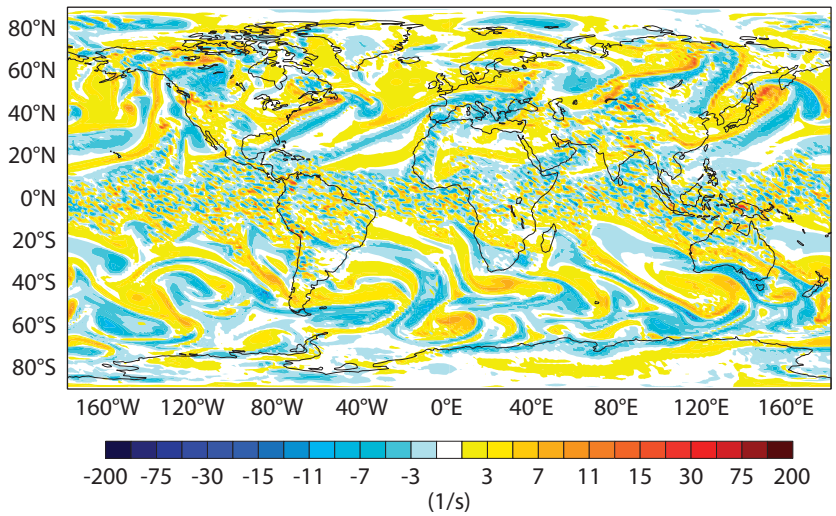


**Figure 7** FVM-produced simulation of a global circulation using realistic orography. The shading shows the vertical component of instantaneous relative vorticity (x $10^6$) at about 4 km above the surface, revealing baroclinic eddies in the mid-latitudes of both hemispheres and fine-scale features in the equatorial area and mountainous regions indicative of convection and gravity waves.


doi:10.21957/f4gm176p

## Summary and outlook

The PantaRhei project pursues an entirely flexible and hierarchical approach capable of satisfying the needs of current and future NWP. The first step is the development of the autonomous FVM supplementing the IFS. The FVM on its own can provide solutions representative of elements of real weather over a large range of scales. It does so using computation and communication patterns which are very different from spectral transform-based global NWP models. The idealised results shown in Figures 5 and 6 relate to key elements of natural weather at very different scales. The same elements can be identified in Figure 7, which shows a snapshot of the FVM simulation of global circulation on an N256 reduced Gaussian grid using the IFS orography together with an idealised frictional/diabatic forcing. The figure illustrates the importance of the FVM's capabilities in the broader NWP context.

The proposed concept offers exciting new opportunities for combining the strengths of spectral-transform models and cloud-resolving models. The FVM and Atlas provide a toolbox of methods previously inaccessible to spectral transform-based models, while the IFS provides the required advanced environment to explore alternative discretisation techniques in the context of operational weather forecasting.

The key to success are the advancement of numerical methods that can be applied to all-scale global flows, and the associated software development for large-scale high-performance computing. The FVM represents an essential algorithmic building block for future NWP and climate services and thus forms part of the future co-design process between ECMWF and leading HPC manufacturers.

## Further reading

**Mozdzynski G., M. Hamrud,** & **N.P. Wedi,** 2015: A PGAS implementation of the ECMWF Integrated Forecasting System (IFS). *Int. J. High Perform. C.*, **29,** 261–273.

**Plougonven R.,** & **F. Zhang,** 2013: Internal gravity waves from atmospheric jets and fronts. *Rev. Geophys.*, **52,** 33–76.

**Smolarkiewicz P.K., C. Kühnlein** & **N.P. Wedi,** 2014: A consistent framework for discrete integrations of soundproof and compressible PDEs of atmospheric dynamics. *J. Comput. Phys.*, **263,** 185–205.

**Szmelter, J.** & **P.K. Smolarkiewicz,** 2010: An edge-based unstructured mesh discretisation in geospherical framework. *J. Comput. Phys.*, **229,** 4980–4995.

**Szmelter, J., Z. Zhang** & **P.K. Smolarkiewicz,** 2015: An unstructured-mesh atmospheric model for nonhydrostatic dynamics: Towards optimal mesh resolution. *J. Comput. Phys.*, **294,** 363–381.

**Wedi, N.P.** & **P.K. Smolarkiewicz,** 2009: A framework for testing global non-hydrostatic models. *Q. J. R. Meteorol. Soc.*, **135,** 469–484.

**Wedi, N.P., M. Hamrud, G. Mozdzynski,** 2013: A fast spherical harmonics transform for global NWP and climate models. *Mon. Weather Rev.*, **141,** 3450–3461.

**Wedi N.P.** & **S. Malardel,** 2010: Non-hydrostatic modelling at ECMWF. *ECMWF Newsletter* **No. 125,** 17–21.

**Wedi N.P., M. Hamrud, G. Mozdzynski, G. Austad, S. Curic, J. Bidlot,** 2012: Global, non-hydrostatic, convection-permitting, medium-range forecasts: progress and challenges. *ECMWF Newsletter* **No. 133,** 17–22.



doi:10.21957/f4gm176p