The way out of the scalability gridlock: ESCAPing dwarfs

Daan Degrauwe, Steven Caluwaerts, Piet Termonia

RMI Belgium
Why it’s bad (recap from last year)
- scalability of spectral methods
- non-spectral methods have their own problems

Why it’s even worse
- some 3D results

The way out
- The ESCAPE project

Conclusions
From the *accuracy* point of view, spectral methods are unsurpassable: their order of accuracy is infinite!

Moreover, the calculation of derivatives and solving the Helmholtz equation are trivial. This allows for (semi-)implicit timestepping and large timesteps. So our spectral dynamics are also quite *efficient*. 
From the accuracy point of view, spectral methods are unsurpassable: their order of accuracy is infinite!

Moreover, the calculation of derivatives and solving the Helmholtz equation are trivial. This allows for (semi-)implicit time-stepping and large timesteps. So our spectral dynamics are also quite efficient.

... but they require spectral transforms (FFT or Legendre transform for the global). These are nonlocal, i.e. they require domain-wide communication. This makes their use problematic on massively parallel machines.

Another disadvantage of a spectral model is the requirement of a homogeneous reference state for the semi-implicit time-stepping. e.g. orography cannot be treated implicitly.

But at what point do the costs no longer justify the accuracy?

– to answer this question, we must closely investigate the alternatives.
When considering alternatives for the spectral horizontal discretization, we try to keep as much as possible of the model intact:

- only way to make a clean comparison
- limited development cost (no need to modify physics, . . .)

So for the time being, we stick to a semi-implicit time discretization and a semi-Lagrangian advection scheme.
When considering alternatives for the spectral horizontal discretization, we try to keep as much as possible of the model intact:

- only way to make a clean comparison
- limited development cost (no need to modify physics, ...)

So for the time being, we stick to a semi-implicit time discretization and a semi-Lagrangian advection scheme.

Finite-difference discretizations are considered on the following grids:

- **A-grid**
  - wind speed
  - not staggered

- **C-grid**
  - wind speed
  - staggered

- **Z-grid**
  - vorticity/divergence
  - not staggered
Shallow-water toy model tests show that each grid has its own problems:

- C-grid is staggered and requires heavy modifications
- A-grid has bad propagation of short waves (negative group velocity)
- Z-grid has bad projection of short waves, resulting in very noisy wind fields
Shallow-water toy model tests show that each grid has its own problems:

- C-grid is staggered and requires heavy modifications
- A-grid has bad propagation of short waves (negative group velocity)
- Z-grid has bad projection of short waves, resulting in very noisy wind fields
- Shallow-water toy model tests show that each grid has its own problems:
  - C-grid is staggered and requires heavy modifications
  - A-grid has bad propagation of short waves (negative group velocity)
  - Z-grid has bad projection of short waves, resulting in very noisy wind fields
Shallow-water toy model tests show that each grid has its own problems:

- C-grid is staggered and requires heavy modifications
- A-grid has bad propagation of short waves (negative group velocity)
- Z-grid has bad projection of short waves, resulting in very noisy wind fields

But how representative are these toy-model results for a real atmospheric model?
Some results with ALADIN

- It’s possible to emulate a gridpoint model with our spectral ALADIN model.
  
The trick is to modify the spectral response of differential operators, e.g.

\[
\frac{\partial \hat{\psi}_k}{\partial x} \quad i k \hat{\psi}_k \quad \rightarrow \quad i \frac{\sin k \Delta x}{\Delta x} \hat{\psi}_k
\]

- (of course, this doesn’t allow to test the scalability.)
Adiabatic run, $\Delta x = 7$ km, quadratic truncation, no DFI

RMS w.r.t. spectral

RMS error geopotential at 500 hPa

RMS error temperature at 500 hPa

The way out of the scalability gridlock: ESCAPing dwarfs
Some results with ALADIN

- Adiabatic run, $\Delta x = 7$ km, quadratic truncation, no DFI

500 hPa geopotential after 1 timestep

The way out of the scalability gridlock: ESCAPing dwarfs
Some results with ALADIN

- Adiabatic run, $\Delta x = 7$ km, quadratic truncation, no DFI

500 hPa geopotential after 12h

FD2 A-GRID

FD2 Z-GRID

The way out of the scalability gridlock: ESCAPing dwarfs
Some results with ALADIN

- Adiabatic run, $\Delta x = 7$ km, quadratic truncation, no DFI
- Run with ALARO physics

**RMS w.r.t. spectral**

### RMS error geopotential at 500 hPa

- $Z$ SPEC
- $Z$ FD2
- $Z$ FD4
- $Z$ FD6
- $Z$ FD8
- A SPECA
- A FD2
- A FD4
- A FD6
- A FD8

### RMS error temperature at 500 hPa

- $Z$ SPEC
- $Z$ FD2
- $Z$ FD4
- $Z$ FD6
- $Z$ FD8
- A SPECA
- A FD2
- A FD4
- A FD6
- A FD8
Some results with ALADIN

- The footprint of the faulty projection in the Z-grid discretization is clearly visible!
- The fields remain noisy during the entire forecast, so a simple DFI will not solve the problem.
- Why is this so bad? It turns out that short scales do matter, even in a 3D model with diffusion and quadratic spectral truncation.
Some results with ALADIN

- The footprint of the faulty projection in the Z-grid discretization is clearly visible!
- The fields remain noisy during the entire forecast, so a simple DFI will not solve the problem.
- Why is this so bad? It turns out that short scales do matter, even in a 3D model with diffusion and quadratic spectral truncation.
- On the sunny side: the A-grid scheme performs quite okay, especially the higher-order variants.
So we have a giant (IFS/ARPEGE/ALADIN/HARMONIE), who is restrained by a grid.

**Question**: How to break him free?
So we have a giant (IFS/ARPEGE/ALADIN/HARMONIE), who is restrained by a grid.

**Question**: How to break him free?  
**Answer**: you cut him into pieces.
So we have a giant (IFS/ARPEGE/ALADIN/HARMONIE), who is restrained by a grid.

**Question**: How to break him free?  
**Answer**: you cut him into pieces.
ESCAPE (Energy-efficient Scalable Algorithms for Weather Prediction at Exascale) is a H2020 project, coordinated by ECMWF, and involving HIRLAM and ALADIN members, HPC hardware manufacturers, universities and supercomputing centers.

The core of ESCAPE is the identification of fundamental algorithm building blocks. These are the so-called NWP dwarfs (an idea based on the Berkeley Dwarfs).

So far, 4 dwarfs have been defined:
- Global spectral transform
- (2D) elliptic solver with iterative method
- Cloud scheme (column physics)
- Bifourier spectral transform

Candidates for more dwarfs:
- Advection schemes (semi-Lagrangian or Eulerian)
- 3D Helmholtz sparse solver
- Time integration schemes
- Radiation scheme
The concept of the NWP dwarfs offers the modularity that is necessary to cleanly compare alternatives.

Further steps in the ESCAPE project are
- Code adaptation: accelerator directives, interoperate with a domain-specific language
- Hybrid computing: optimize dwarfs for accelerators
- Benchmarking and diagnostics: time-to-solution, energy consumption, determine NWP benchmark references.

The common basis for the dwarfs is the ATLAS library, which supports structured and unstructured grids.
Be careful when relying on ‘common knowledge’
- ‘A-grid has problem at the shortest scales’
- ‘Formulation in vorticity/divergence (Z-grid) solves this’

Every method has its drawbacks; only clean testing can bring answers

Clean testing means modularity

Modularity (dwarfs) is the central idea of the ESCAPE project
Conclusions

- Be careful when relying on ‘common knowledge’
  - ‘A-grid has problem at the shortest scales’
  - ‘Formulation in vorticity/divergence (Z-grid) solves this’
- Every method has its drawbacks; only clean testing can bring answers
- Clean testing means modularity
- Modularity (dwarfs) is the central idea of the ESCAPE project
- Big challenges ahead:
  - Quoting Enda O’Brien:
    
    *How to build a giant from several dwarfs?*

  - Don’t forget there’s also a scientific impact of switching methods!
Thank you!

And many thanks and congratulations to Steven!