Numerical methods for Earth system modelling

Giovanni Tumolo [1]

[1] The Abdus Salam ICTP - Trieste, < gtumolo@ictp.it >

Pune, July 19, 2016

Outline

The heart of any Earth System Model (ESM) is represented by the dynamical core, where governing equations for the atmosphere and the ocean (fluid) dynamics are solved.

The main aspects to be taken into account:

- Governing equations
- Space dicretization schemes
- Time discretization schemes
- High performance computing

Governing equations for atmospheric flows

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{u}) = 0 \qquad \text{(conservation of mass)}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} = -\boldsymbol{\nabla} K - (2\boldsymbol{\Omega} + \boldsymbol{\zeta}) \times \boldsymbol{u} - \frac{1}{\rho}\boldsymbol{\nabla} p + \boldsymbol{\nabla}\Phi + \boldsymbol{F}^{\boldsymbol{u}}_{ext} + \boldsymbol{F}^{\boldsymbol{u}}_{par} \quad \text{(Newton 2nd law)}$$

$$\frac{\partial (\rho\epsilon)}{\partial t} + \boldsymbol{\nabla} \cdot [(\rho\epsilon + p + \boldsymbol{F}^{\epsilon}_{R}) \cdot \boldsymbol{u}] = 0 \qquad \text{(conservation of energy)}$$

$$p = \rho R T \qquad \text{(equation of state)}$$

where:
air considered as ideal gas with $\rho =$ mass density, $p =$ pressure, $T =$ temperature, $R =$ ideal gas constant for dry air
$\boldsymbol{u} =$ velocity
$K = \frac{1}{2}\boldsymbol{u} \cdot \boldsymbol{u} =$ kinetic energy per unit mass
$\boldsymbol{\zeta} = \boldsymbol{\nabla} \times \boldsymbol{u} =$ relative vorticity
$\boldsymbol{\Omega} =$ rotation velocity of the Earth
$\Phi =$ normal gravity potential
$\boldsymbol{F}^{\boldsymbol{u}}_{ext} =$ resultant of the external forces
$\boldsymbol{F}^{\boldsymbol{u}}_{par} =$ effect of parametrized processes
$\epsilon = e + K = c_p T + K =$ total (internal + kinetic) energy per unit mass
$\boldsymbol{F}^{\epsilon}_{R} =$ radiation energy flux

Governing equations for oceanic flows

$$\boldsymbol{\nabla} \cdot \boldsymbol{u} = 0 \qquad \text{(conservation of mass)}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} = -\boldsymbol{\nabla} K - (2\boldsymbol{\Omega} + \boldsymbol{\zeta}) \times \boldsymbol{u} - \frac{1}{\rho}\boldsymbol{\nabla} p + \boldsymbol{\nabla}\Phi + \boldsymbol{F}^{\boldsymbol{u}}_{ext} + \boldsymbol{F}^{\boldsymbol{u}}_{par} \qquad \text{(Newton 2nd law}$$

$$\frac{\partial \theta}{\partial t} + \boldsymbol{\nabla} \cdot (\theta \boldsymbol{u}) = \boldsymbol{F}^{\theta}_{par} \qquad \text{(conservation of energy)}$$

$$\frac{\partial S}{\partial t} + \boldsymbol{\nabla} \cdot (S \boldsymbol{u}) = \boldsymbol{F}^{S}_{par}$$

$$\rho = \rho(S, T, p) \qquad \text{(equation of state)}$$

where:

$\rho =$ mass density, $p =$ pressure, $T =$ temperature, for water

$S =$ salinity

$\boldsymbol{u} =$ velocity

$K = \frac{1}{2}\boldsymbol{u} \cdot \boldsymbol{u} =$ kinetic energy per unit mass

$\boldsymbol{\zeta} = \boldsymbol{\nabla} \times \boldsymbol{u} =$ relative vorticity

$\boldsymbol{\Omega} =$ rotation velocity of the Earth

$\Phi =$ normal gravity potential

$\boldsymbol{F}^{\boldsymbol{u}}_{ext} =$ resultant of the external forces

$\boldsymbol{F}^{\boldsymbol{u}}_{par}, \boldsymbol{F}^{\theta}_{par}, \boldsymbol{F}^{S}_{par} =$ effect of parametrized processes

$\theta =$ potential temperature

The problem to be solved

- The set of governing equations for the atmosphere and the ocean are examples of system of nonlinear partial differential equations (PDE's), to be solved on some spatial domain Ω and time interval $[0, t_f]$, given suitable boundary (BC) and initial (IC) conditions.

- Therefore the typical problem to be solved in ESM is an intial/boundary value problem of the (general) form:

$$\frac{\partial \psi}{\partial t} = \mathcal{L}(\psi)$$
$$\psi(0) = \psi_0$$
BC

where

- $\psi = \psi(\boldsymbol{x}, t)$ is a dependent variable, function of space and time $\boldsymbol{x} \in \Omega, t \in [0, t_f]$.
- $\mathcal{L}$ is a (generally nonlinear) differential operator.

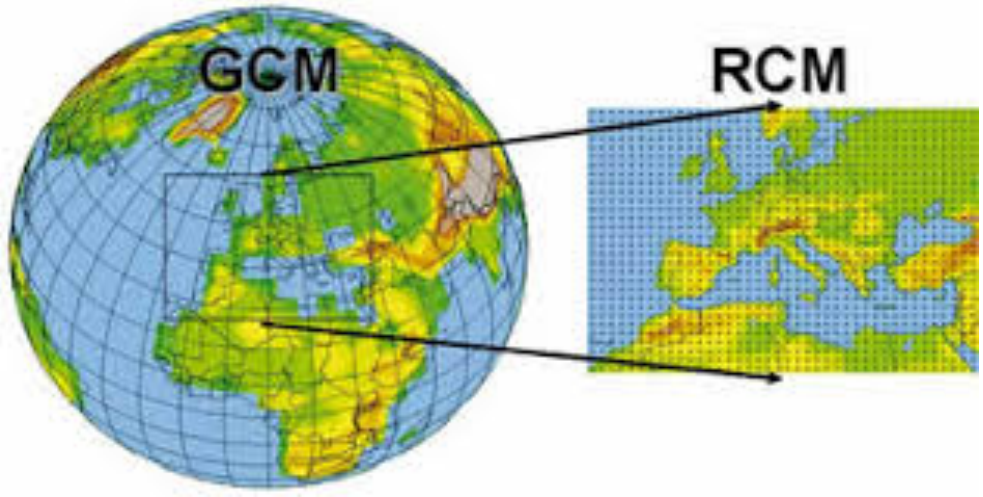
The problem to be solved: initial data

- ESM is an initial / boundary value problem: given an estimate of the present state of the atmosphere (initial conditions), and appropriate surface and boundary conditions, the ESM simulates the atmosphere / ocean evolution ($\frac{\partial}{\partial t} \Longrightarrow$ evolution PDE's)
- the more accurate the estimate of the initial conditions the better the quality of the forecast / simulation.
- currently initial conditions are produced in NWP centers through a statistical combination of observations and short range forecasts. This approach is called "data assimilation", uses all the available information to determine as accurately as possible the initial state of the atmospheric / oceanic flow, and consumes a relevant part of an EMS.
- data assimilation can take a relevant part of computational resources in a ESM.

The problem to be solved: boundary conditions

Boundary conditions: Global vs. Regional models.

- If Ω is a complete spherical shell sorrounding the Earth, ESMs are called Global circulation Models (GCMs) or global climate models (AGCMs vs. OGCMs).
- If Ω has the size of a continent, ESMs are called Limited Area Models (LAMs), or Regional models, or mesoscale models (ARCMs vs ORCM).



- AGCMs: no explicitly driven boundary conditions (no lateral boundary).
- RCMs need to be nested into GCMs, which provide the proper boundary conditions to the formers (the RCM is said is driven by a GCM).
- Surface boundary conditions: the coupling between different ESM components provides proper fluxes.

Consequences of the nonlinearity of the problem

The problem to be solved

$$\frac{\partial \psi}{\partial t} = \mathcal{L}(\psi)$$
$$\psi(0) = \psi_0$$
BC

is typically highly nonlinear. As a consequence:

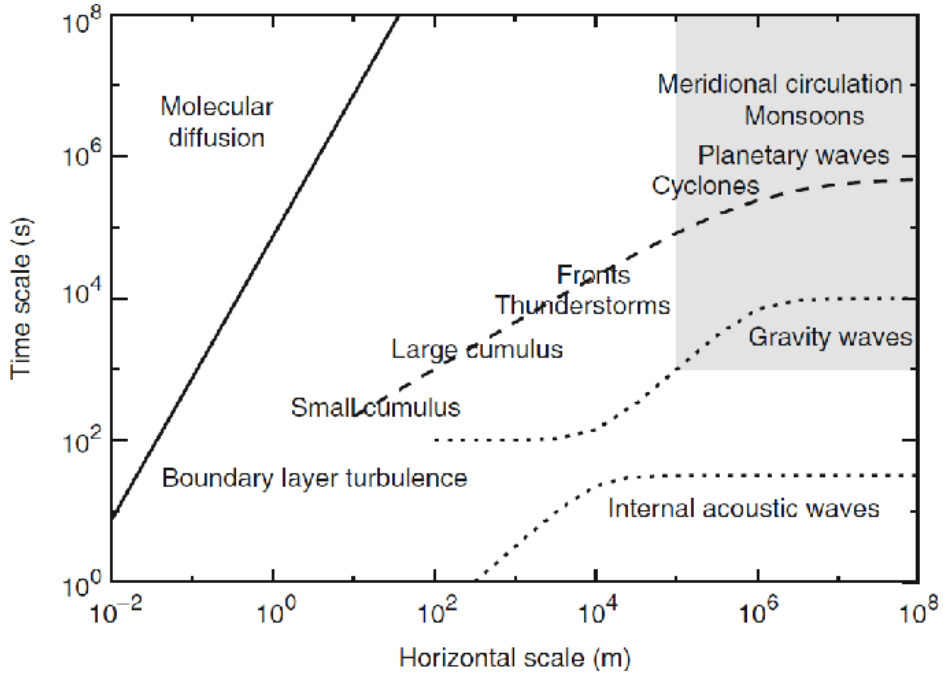
- even if you can prove that it is well posed, in general you are not able to find a representation formula for its solution ψ.

  $\Longrightarrow$ As a result, an approximation φ of the solution $\psi = \psi(\boldsymbol{x}, t)$ has to be searched, through the introduction of suitable
  - space discretization techniques
  - time integration techniques

- the system is chaotic, i.e. the solution is strongly dependent on the initial conditions. Therefore
  - ensamble simulations are considered for NWP
  - ergodic assumption is made for climate simulations

- there is an interaction btw. different space scales: possible source of ”nonlinear instabilities" in the numerical approximation φ.

Multiscale nature of the problem



- All the phenomena along the dashed line are important for weather and climate, and so need to be represented in numerical models.
- Important pahenomena occur at all scales (no spectral gap), and interact with eachother.
- Computer resources are finite and so numerical models must have a finite resolution h.
- Shaded region shows the resolved space and time scales in a typical current day climate model.
- The important unresolved processes cannot be neglected and so must be represented by sub-grid models or parametrizations.

Space discretization techniques

The original domain Ω is replaced by a spatially discretized domain Ω_h and the original problem is replaced by:

$$\frac{d\varphi_i}{dt} = \mathcal{L}_h(\varphi)_i, \quad i = 1, \ldots, m$$
$$\varphi_i(0) = (\varphi_0)_i, \quad i = 1, \ldots, m$$
$$\text{BC}_h$$

where
$\mathcal{L}_h$ denotes a discrete approximation of the continuous differential operator $\mathcal{L}$
h denotes the tipical size of the discrete spatial elements and determine the resolution of the spatial discretization.
This semi-discrete problem is a system of (generally nonlinear) Ordinary Differential Equations (ODE's) whose unknowns φ_i are approximations of the continuous solutions ψ.

Space discretization techniques

Main approaches:

1. Finite Differences (vertical discretization of IITM atmosphere component)

2. Spectral Transform (horizontal discretization of IITM atmosphere component)

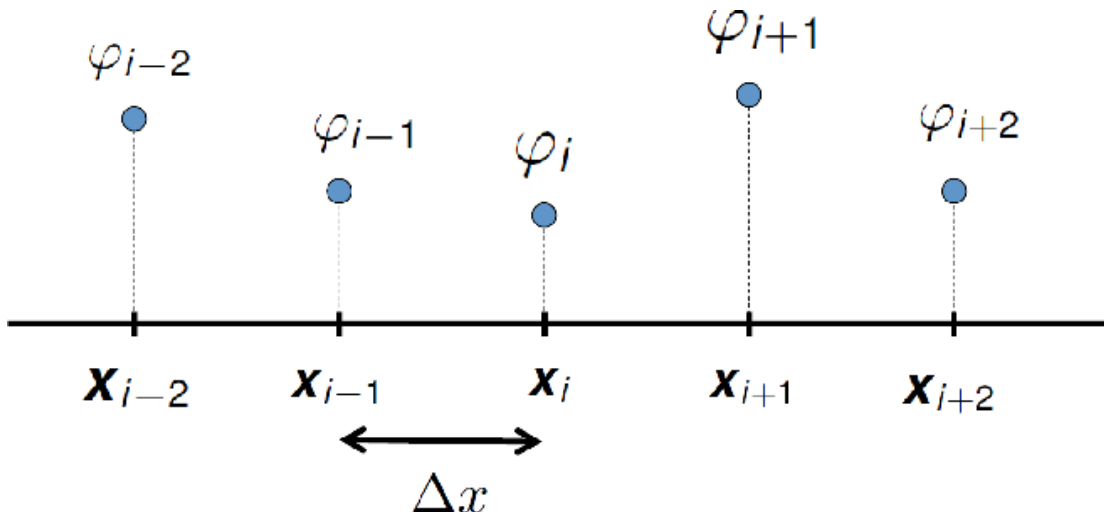
3. Local Galerkin Methods / Projection methods (Finite Element, Spectral Element, Discontinuous Galerkin)

4. Finite Volumes (space discretization of IITM ocean component)

Finite Differences

- $\Omega_h = \{\boldsymbol{x}_i, i = 0, \ldots, m\}$, called grid, is a regular array of discrete locations, called nodes. $h = \Delta x$ is the average spacing between nodes.
- $\varphi_i(t) \approx \psi(\boldsymbol{x}_i, t)$.
- $\mathcal{L}_h$ is obtained by replacing derivatives present in $\mathcal{L}$ with finite difference quotients.
- Example in 1D. If $\Omega = [0, L]$ and $\mathcal{L}(\psi) = \psi'$, then $\Omega_h = \{x_i = ih, \quad i = 0, \ldots, m\}$ with $h = \Delta x = L/m$, and, for example:

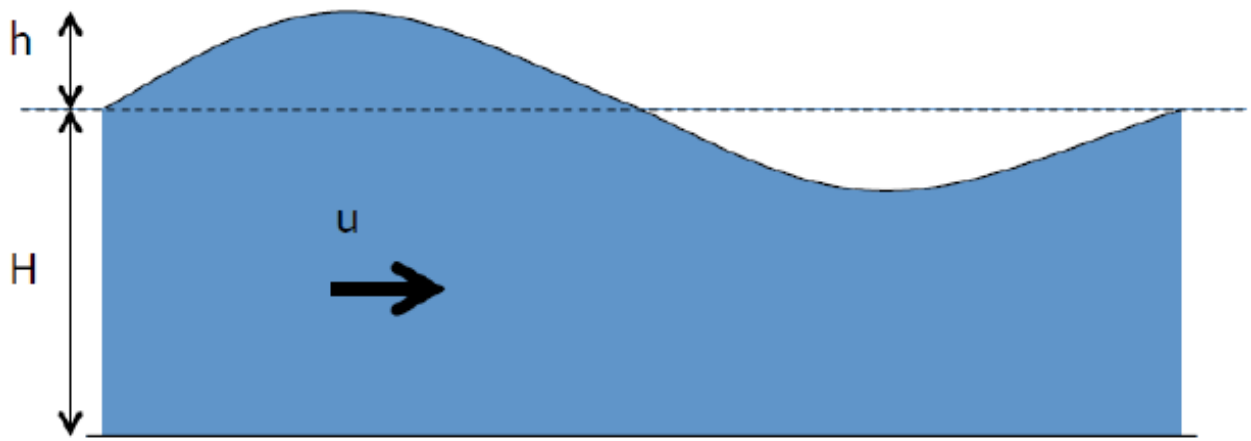
$$\mathcal{L}_h(\varphi)_i = \frac{\varphi_{i+1} - \varphi_{i-1}}{2h}$$



Example in 1D: Shallow Water Equations (SWE)

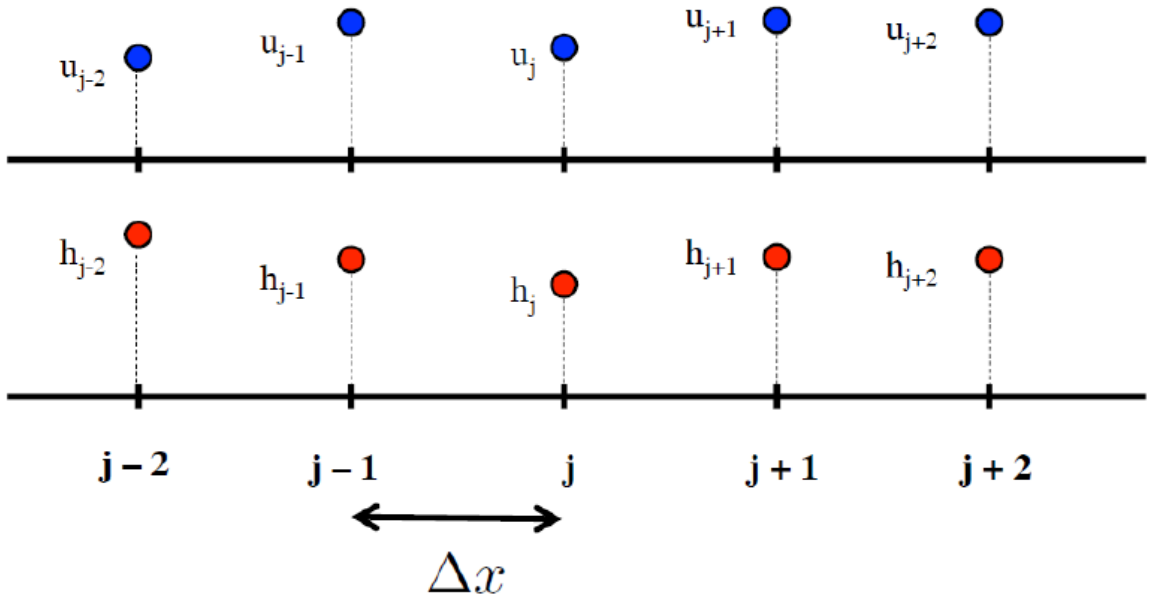
Shallow water equations, linearized about a state of rest, with Coriolis forces neglected: they are derived from depth-integrating the general governing equations, in the case where the horizontal length scale is much greater than the vertical length scale and the free surface perturbation is much smaller than the mean depth $h << H$:

$$\frac{\partial h}{\partial t} + H\frac{\partial u}{\partial x} = 0$$
$$\frac{\partial u}{\partial t} + g\frac{\partial h}{\partial x} = 0$$



Finite Differences: A staggering in 1D

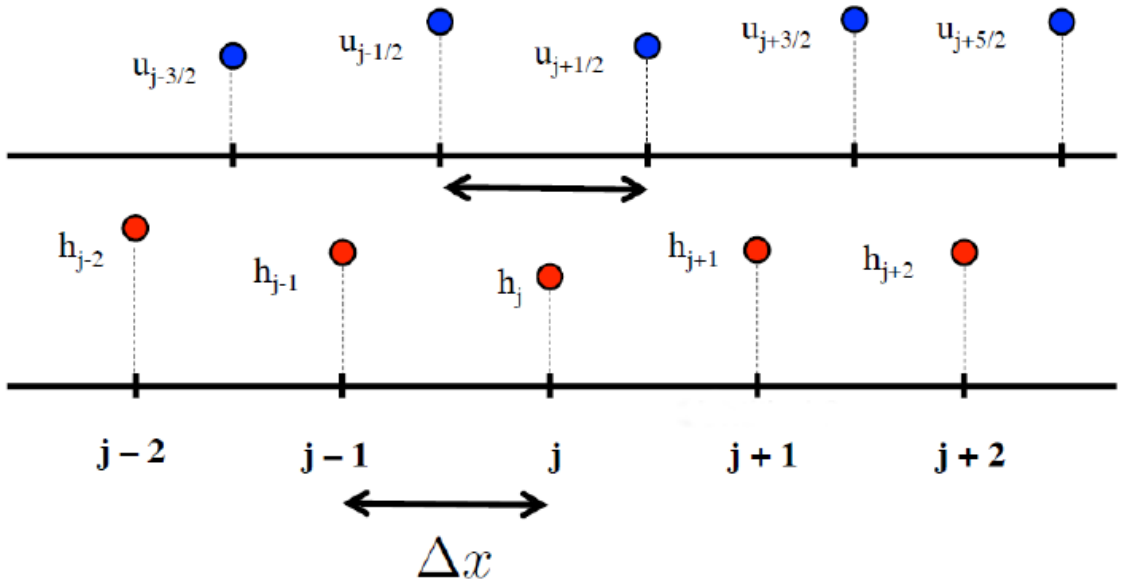
unknowns h and u are co-located at the same gridpoints (Arakawa A-grid):



$$\frac{\partial h_i}{\partial t} + H\frac{u_{i+1} - u_{i-1}}{2\Delta x} = 0$$

$$\frac{\partial u_i}{\partial t} + g\frac{h_{i+1} - h_{i-1}}{2\Delta x} = 0$$

Finite Differences: C staggering in 1D



unknowns h and u are located at staggered (shifted) gridpoints (Arakawa C-grid):

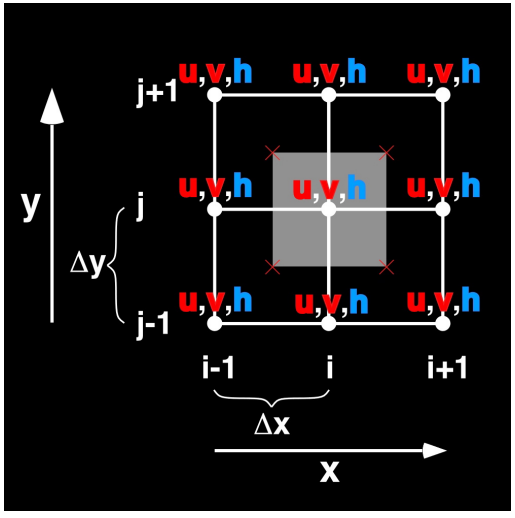
$$\frac{\partial h_i}{\partial t} + H \frac{u_{i+1/2} - u_{i-1/2}}{\Delta x} = 0$$

$$\frac{\partial u_{i+1/2}}{\partial t} + g \frac{h_{i+1} - h_i}{\Delta x} = 0$$

Finite Differences: Arakawa staggering in 2D

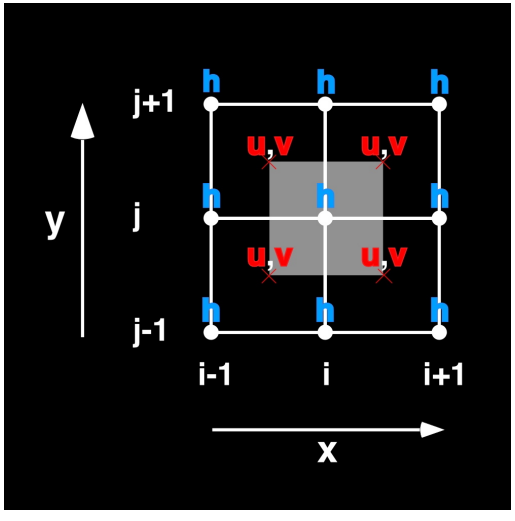
In two dimensions other arrangements for h and u, v nodes are possible:

Arakawa A grid:



- unstaggered grid - all variables defined everywhere;
- poor performances, first grid geometry employed in NWP models;
- noisy - large errors, short waves propagate energy in wrong direction, additional smoothing required;
- poorest at geostrophic adjustment - wave energy trapped;
- can use a 2x larger timestep than staggered grids.

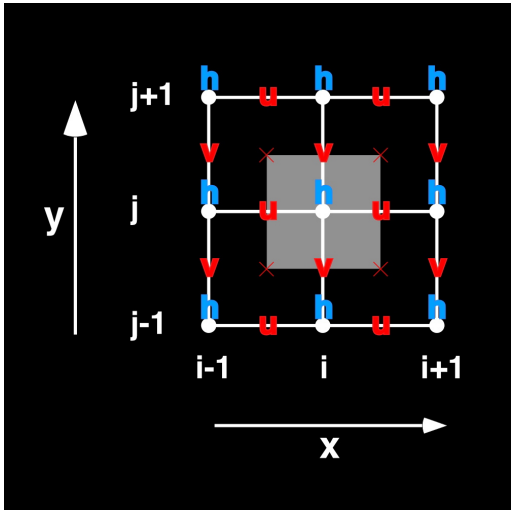
Arakawa B grid:



- staggered grid - velocities components co-located at corners;
- preferred at coarse resolution;
- superior for poorly resolved inertia-gravity waves;
- good for geostrophic adjustment and Rossby waves;
- bad for gravity waves: computational checkerboard mode;
- used by MM5 model (and hence by RegCM).

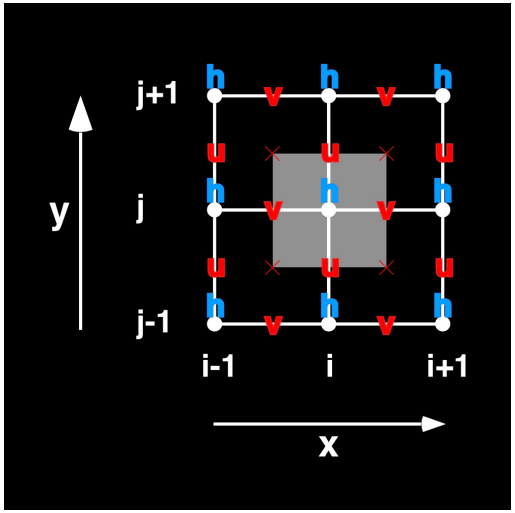
Finite Differences: Arakawa staggering in 2D

Arakawa C grid:



- staggered: pressure (h) at center, normal velocity at grid cell faces;
- preferred at fine resolution;
- superior for gravity waves;
- good for well resoved inertia-gravity waves;
- simulates Kelvin waves well;
- bad for poorly resolved waves: Rossby waves (computational checkerboard mode) and inertia-gravity waves due to averaging the Coriolis forces;
- used by WRF model.

Arakawa D grid:



- staggered: pressure (h) at center, tangential velocity at grid cell faces;
- poorest performances, worst dispersion properties, rarely used;
- noisy - large errors, short waves propagate energy in wrong direction.

Finite Differences Methods for vertical discretization

IITM-ESM atmospheric component uses finite differences for vertical discretization with 64 $\sigma-$ pressure hybrid layers:

- in the sigma system the vertical coordinate is the pressure normalized with the surface pressure:

$$\sigma := \frac{p}{p_S} \quad \text{(Phillips 1957)}$$

where $p_S\,(x, y, t)$ is the pressure at the surface.

- pro: simpler lower boundary condition;
- con: over steep orography problem of cancellation of two terms which comprise the pressure gradient.
- Alternative: hybrid sigma pressure coordinate (Sangster 1960, Simmons and Burridge 1981):

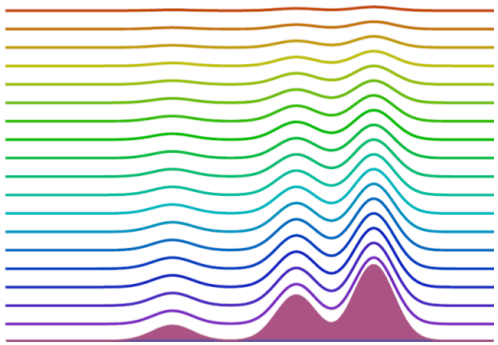
  - similar to the the usual sigma coordinate at lower level but
  - it reduces to pressure above a certain fixed level. For example:

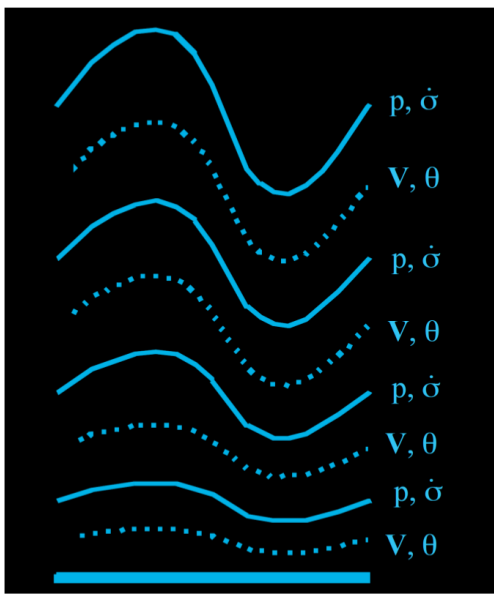
$$\sigma = \frac{p}{p_S} + \left(\frac{p}{p_S} - 1\right)\left(\frac{p}{p_S} - \frac{p}{p_0}\right), \quad p_0 \text{ a constant.}$$

$\sigma \to \frac{p}{p_S}$ if$p \to p_S$, while$\sigma \to \frac{p}{p_0}$ if$p \to 0$.

- the vertical velocity is $\dot{\sigma} = \frac{D\sigma}{Dt}$.
- Lorenz staggering: temperature staggered wrt vertical velocity. Pro: energetic consistency;

  con: vertical computational mode.



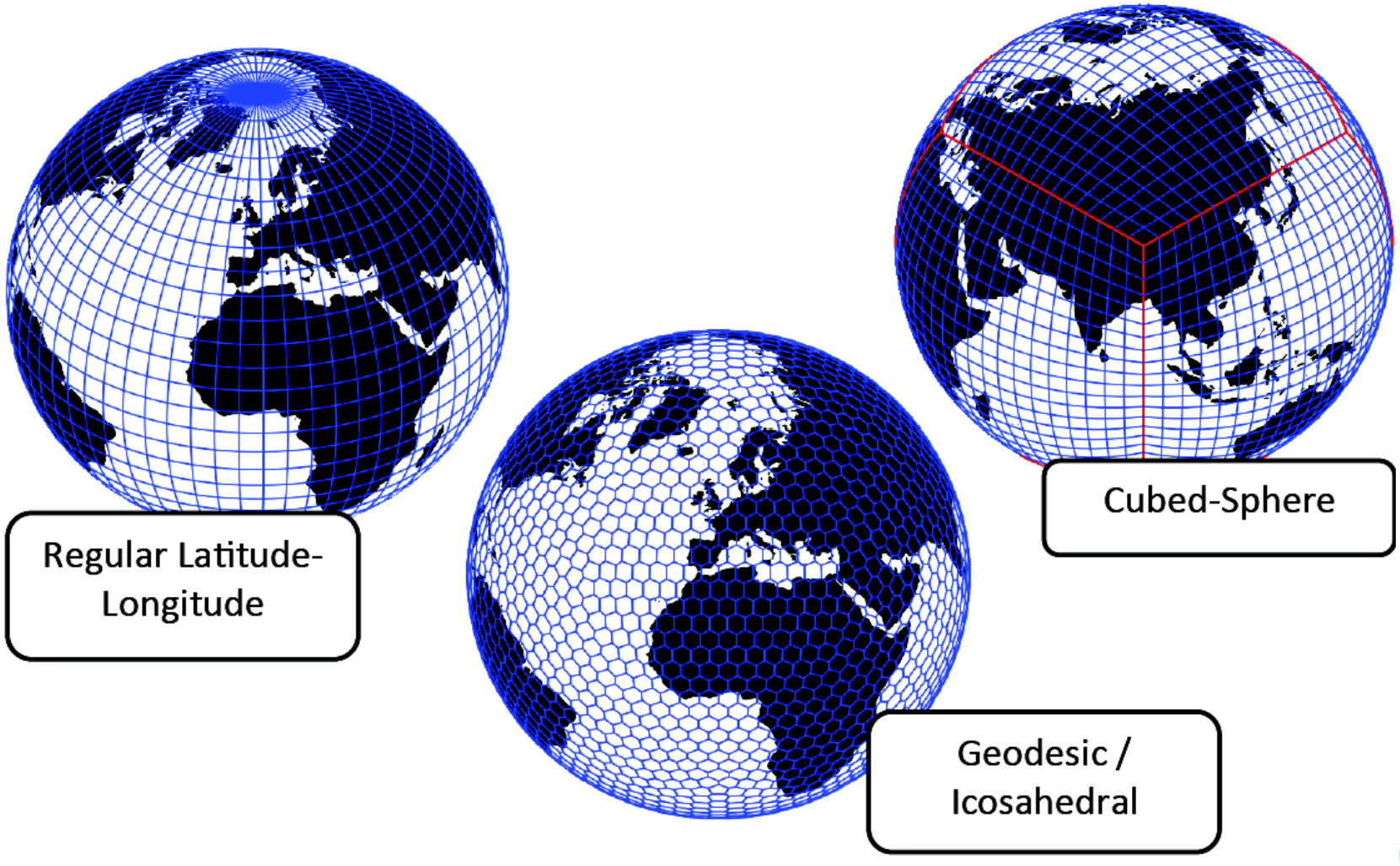


Finite Differences Methods: pros and cons

- easy to implement
- you need to be careful to ensure conservation and avoid spurious wave solutions (staggering, filtering ...)
- typically not high order accurate
- not easy to introduce adaptivity
- not flexible, for example not suited for unstructured grids.

Choice of the horizontal mesh

What is Ω_h ? The choice is not unique



Galerkin methods

- Consider a linear differential operator $\mathcal{P}$ and the corresponding PDE

$$\mathcal{P}u = f \quad \text{in} \quad \Omega$$

- Approximate the solution as linear combination of given basis functions $\Phi_i(\mathbf{x})$ (called trial functions)

$$u(\mathbf{x},t) \approx u^N(\mathbf{x},t) = \sum_{i=1}^{N} a_i(t)\Phi_i(\mathbf{x})$$

- minimize the residual $\mathcal{P}u^N - f$, another vector: look at its components in onother basis Ψ_k (called test functions)
- Trial (Φ_i) and test (Ψ_k) functions are from same basis and satisfy boundary conditions (BC)
- classification:
  - Φ_i, Ψ_k with global support: spectral method
  - Φ_i, Ψ_k with compact support: finite element / spectral element method

The Spectral Transform Method: historical introduction

- first introduced into meteorological modelling (Silberman, 1954), for the non-divergent barotropic vorticity equation in spherical geometry;
- for a balanced barotropic model, the spectral method could compete with the grid point method wrt performances and accuracy (Elsaesser, 1966);
- for high resolution complex non-adiabatic models, the spectral method was not considered a realistic alternative to the grid point method: nonlinear terms requiring computing and storing too many "interaction coefficients“, difficult incorporation of local physical processes;
- The advent of the Fast Fourier Transform (Cooley and Tukey, 1965) meant the introduction of the ”transform method”, independently by Eliasen et al., 1970 and Orszag, 1970: no interaction coefficients are involved, while at each timestep grid point values of dependent variables are computed in an auxiliary grid in the physical space;
- spectral transform models started to be used for routine NWP in Australia (Bourke et al., 1977), Canada (Daley et al., 1976), and today are used at ECMWF (Hortal 1991, Hortal and Simmons 1991), in the US at NCEP (Jouang, 2004) and IITM of course !

The Spectral Transform Method: basic principles

The complete set of equations used in any atmospheric model may be written in quite a general form as:

$$\frac{\partial}{\partial t}\mathcal{L}_i(\omega^i) = F_i(\omega^1, \omega^2, \ldots, \omega^I), \quad i = 1, 2, \ldots, I. \tag{1}$$

where the prognostic variables $\omega^i = \omega^i(\mathbf{r}, t), \quad i = 1, 2, \ldots, I$ are *scalar* functions of the space coordinates, given by $\mathbf{r}$, and of the time t.
$\mathcal{L}_i$ is a linear (space) differential operator, typically the laplacian $\mathcal{L}_i = \nabla^2$, but often reducing to the identity operator $\mathcal{L}_i(\omega^i) = \omega^i$.
F_i is a generally nonlinear space differential operator.
Any of the variables ω^i is approximated by a truncated series of the form:

$$\hat{\omega}^i(\mathbf{r}, t) = \sum_{n=1}^{N} \omega_n^i(t)\psi_n(\mathbf{r})$$

Remark: using such a representation, space derivatives can be evaluated analitically, without approximations.
Any series expansion method replaces in (1) ω^i with $\hat{\omega}^i$, transforming a system of PDEs for the prognostic variables into a system of ODE's for their expansion coefficients $\omega_n^i = \omega_n^i(t)$ only.
Replacing ω^i with $\hat{\omega}^i$, then (1) is satisfied up to a residual
$R^i = \frac{\partial}{\partial t}\mathcal{L}_i(\hat{\omega}^i) - F_i(\hat{\omega}^1, \hat{\omega}^2, \ldots, \hat{\omega}^I)$

The Spectral Transform Method: basic principles 2

And considering the case of a scalar equation: $R = \frac{\partial}{\partial t}\mathcal{L}(\hat{\omega}) - F(\hat{\omega})$.
N equations for the expansion coefficients $\omega_n(t)$ are obtained by imposing to the residual to be orthogonal to the test functions ψ_n (Galerkin projection):

$$\int_S R(\hat{\omega})\psi_n dS = 0, \quad n = 1, 2, \ldots, N. \tag{2}$$

Assuming the expansion functions ψ_n to be *eigenfunctions* of the operator $\mathcal{L}$

$$\mathcal{L}(\psi_n) + \epsilon_n \psi_n = 0, \quad n = 1, 2, N \tag{3}$$

then equations (3) take the form:

$$\sum_{n'=1}^{N} \epsilon_{n'} \int_S \psi_n \psi_{n'} dS \frac{d\omega_{n'}}{dt} = \int_S F(\sum_{n'=1}^{N} \omega_{n'} \psi_{n'}) \psi_n dS, \quad n = 1, \ldots, N \tag{4}$$

which is a system of N linear algebraic equations in the tendencies $\frac{d\omega_{n'}}{dt}$, to be solved with a proper time integrator (see second part of the lecture).
Remark: if the basis functions $\psi_n, \quad n = 1, \ldots, N$ are chosen to be *orthogonal*, then the matrix at the left bhand side becomes *diagonal*:

$$\int_S \psi_n \psi_{n'} dS = \delta_{n,n'} \tag{5}$$

so that no matrix inversions are required for the estimation of the tendencies $\frac{d\omega_{n'}}{dt}$.

The Spectral Transform Method on a toy problem: linear advection eq.

The linear advection equation may be written in 1D as:

$$\frac{\partial \omega}{\partial t} + \gamma \frac{\partial \omega}{\partial \lambda} = 0, \quad \text{in } \mathbb{T}^1 \times (0,t), \quad \gamma = \text{constant advection angular velocity.}$$

Clearly it is: $\omega(\lambda + 2\pi p, t) = \omega(\lambda, t), \quad \forall p \in \mathbb{N}$.

A natural choice for the expansion functions is given by the trigonometric ones:

$$\hat{\omega}(\lambda, t) = \sum_{m=-M}^{M} \omega_m(t) e^{im\lambda},$$

where being $\hat{\omega}(\lambda, t)$ a real valued function, the complex valued expansion coefficients must satisfy $\omega_{-m}(t) = (\omega_m(t))^*$, with $()^*$ indicating the complex conjugate. Therefore $\hat{\omega}$ is completely specified is coefficients $\omega_m(t)$ are given for $0 \leq m \leq M$.

For such expansion functions following orthogonality relation holds:

$$\frac{1}{2\pi} \int_0^{2\pi} e^{im\lambda} e^{-im'\lambda} = \delta_{m,m'}$$

giving the following expression for the expansion coefficients $\omega_m(t)$:

$$\omega(t) = \frac{1}{2\pi} \int_0^{2\pi} \hat{\omega}(\lambda, t) e^{-im\lambda} \, d\lambda.$$

The Spectral Transform Method on a toy problem: linear advection eq.

Replacing in the advection equation the unknown $\omega(\lambda, t)$ with its truncated expansion $\hat{\omega}(\lambda, t)$, we get following spectral truncated equation:

$$\sum_{m=-M}^{M} (\frac{d\omega_m}{dt} + im\gamma\omega_m)e^{im\lambda} = 0.$$

As the expansion functions are linearly independent this equation is equivalent to following equations for the expansion coefficients only:

$$\frac{d\omega_m}{dt} + im\gamma\omega_m = 0 \quad \text{for} - M \leq m \leq M,$$

whose solution is:

$$\omega_m(t) = \omega_m(0)e^{-im\gamma t}$$

Substituing this expression for the expansion coefficients $\omega_m(t)$ into the truncated series for $\hat{\omega}$, we get the spectral solution:

$$\hat{\omega}(\lambda, t) = \sum_{m=-M}^{M} \omega_m(0)e^{im(\lambda-\gamma t)} \equiv \hat{\omega}(\lambda - \gamma t, 0)$$

Remark: since the chosen expansion functions $e^i m\lambda$ are eigenfunctions of the space differential operator $F = \frac{\partial}{\partial\lambda}$ involved in the linear advection equation, then no minimization of the residual is required and the truncated series exactly satisfies the equation (way different from grid point methods).

The Spectral Transform Method on the sphere: spherical harmonics

For representing fields defined on the sphere surface by mean of truncated series expansions we use surface spherical harmonics denoted by $Y_n^m(\lambda, \mu)$. These complex valued functions of variables $\lambda \in [0, 2\pi]$ and $\mu \in [-1, 1]$, are defined as eigenfunctions of the Laplace operator on the sphere, i.e. as solutions of the equation:

$$\nabla^2 \psi = k\psi \quad \text{on the sphere surface, } k = \text{constant.}$$

Consider the Laplace operator in spherical coordinates (λ, θ, r) :

$$\nabla^2 \psi = \frac{1}{r^2}\left[\frac{\partial}{\partial r}\left(r^2 \frac{\partial \psi}{\partial r}\right) + \frac{1}{\cos\theta}\frac{\partial}{\partial \theta}\left(\cos\theta \frac{\partial \psi}{\partial \theta}\right) + \frac{1}{\cos^2\theta}\frac{\partial^2}{\partial \lambda^2}\right],$$

then its restriction to the spherical shell $r = 1$ is:

$$\nabla^2 \psi|_{r=1} = \frac{1}{\cos\theta}\frac{\partial}{\partial \theta}\left(\cos\theta \frac{\partial \psi}{\partial \theta}\right) + \frac{1}{\cos^2\theta}\frac{\partial^2}{\partial \lambda^2}$$

and consider the change of variables from θ to $\mu = \mu(\theta) = \sin\theta$:

$$\frac{\partial}{\partial \theta} = \frac{\partial}{\partial \mu}\frac{d\mu}{d\theta} \quad \Longrightarrow \quad \frac{\partial}{\partial \theta} = \cos\theta \frac{\partial}{\partial \mu} \quad \Longrightarrow \quad \frac{\partial}{\partial \mu} = \frac{1}{\cos\theta}\frac{\partial}{\partial \theta} \tag{6}$$

and hence:

$$\nabla^2 \psi|_{r=1} = \frac{\partial}{\partial \mu}\left((1-\mu^2)\frac{\partial \psi}{\partial \mu}\right) + \frac{1}{1-\mu^2}\frac{\partial^2 \psi}{\partial \lambda^2} \tag{7}$$

The Spectral Transform Method on the sphere: spherical harmonics

$\nabla^2\psi = k\psi$ requires k to have the form: $k = -n(n+1), \quad n \in \mathbb{N}.$
so have to solve:

$$\frac{\partial}{\partial\mu}\left((1-\mu^2)\frac{\partial\psi}{\partial\mu}\right) + \frac{1}{1-\mu^2}\frac{\partial^2\psi}{\partial\lambda^2} + n(n+1)\psi = 0$$

Using the method of separation of variables, we assume $\psi(\lambda,\mu) = L(\lambda)P(\mu)$, insert it into the previous eq. and multilpy by $(1-\mu^2)/(PL)$:

$$\frac{1-\mu^2}{P}\frac{d}{d\mu}((1-\mu^2)P') + n(n+1)(1-\mu^2) = -\frac{L''}{L}$$

lhs depends on μ only, while rhs depends on θ only, so they have to be constant:

$$-\frac{L''}{L} = m^2, \quad m \in \mathbb{N} \quad \Longrightarrow \quad L'' + m^2 L = 0 \quad \Longrightarrow \quad L(\lambda) = e^{\mp im\lambda}.$$

$$\frac{d}{d\mu}((1-\mu^2)P') + \left(n(n+1) - \frac{m^2}{1-\mu^2}\right)P = 0 \quad \Longrightarrow \quad P = P_n^m(\mu)$$

which is known as associated Legendre equation, whose solutions $P_n^m(\mu)$ are called Legendre functions of the first kind of order m and degree n.
So finally our spherical harmonics expansion functions are defined as:

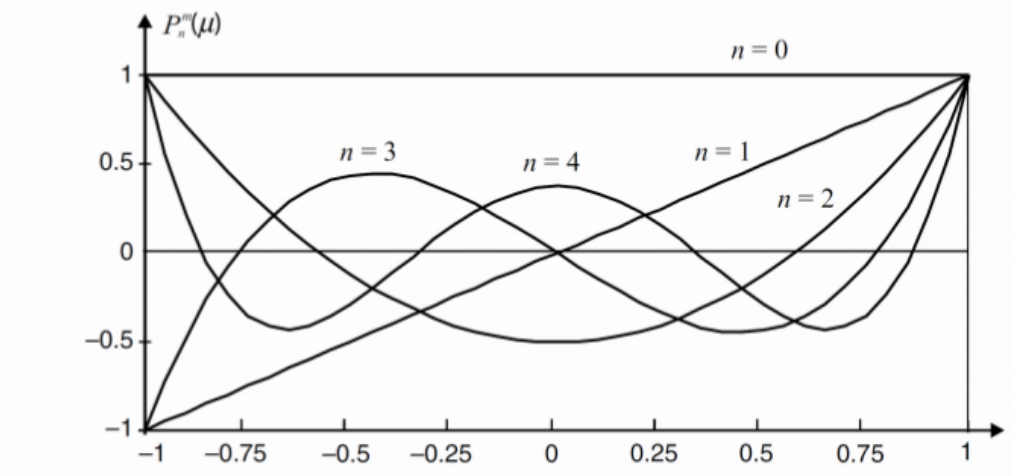
$$\psi(\lambda,\mu) = P(\mu)L(\lambda) = P_n^m(\mu)e^{im\lambda} = Y_n^m(\lambda,\mu)$$

The Spectral Transform Method on the sphere: spherical harmonics

Properties of spherical harmonics $Y_n^m(\lambda,\mu) = P_n^m(\mu)e^{im\lambda}$:
$e^{im\lambda}$ describes the zonal variation of the spherical harmonic wave;
$P_n^m(\mu)$ describes the meridional variation of the spherical harmonic wave.
$P_n^m(\mu)$ are expressed analitically by the Rodrigues formula:

$$P_n^m(\mu) = \frac{\left(1-\mu^2\right)^{\frac{m}{2}}}{2^n n!}\frac{d^{n+m}}{d\mu^{n+m}}(\mu^2-1)^n$$

From this representation formula for P_n^m follow these properties:
$P_n^m(\mu)$ is a polynomial of degree n s.t. $P_n^m = 0 (\Rightarrow Y_n^m = 0\ )$ if $n+m > 2n$
i.e. if $n < |m|$, therefore it make sense to consider only the spectral harmonics Y_n^m with $n \geq |m|$.



The first (non normalized) Legendre functions for $m = 0$.

The Spectral Transform Method on the sphere: spherical harmonics

From the Rodrigues formula:

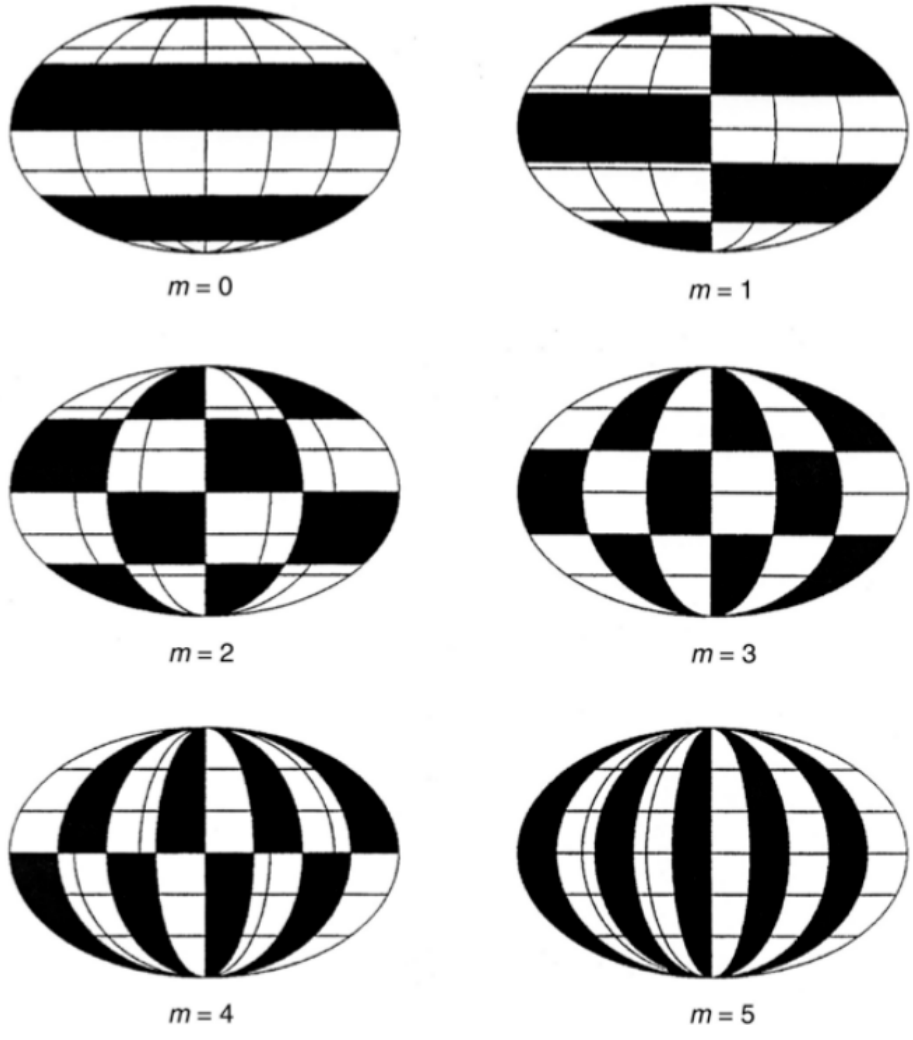
$$P_n^m(\mu) = \frac{\left(1-\mu^2\right)^{\frac{m}{2}}}{2^n n!} \frac{d^{n+m}}{d\mu^{n+m}} (\mu^2 - 1)^n$$

we deduce:

$$P_n^m(\mu) = (1-\mu^2)^{\frac{|m|}{2}} Q_{n-|m|}(\mu),$$

$Q_{n-|m|}(\mu)$ being a polynomial of degree n-|m| and parity $n - |m|$, i.e.:

- $P_n^m(-\mu) = P_n^m(\mu)$ if $n - |m|$ is even, $\Rightarrow Y_n^m(-\mu) = Y_n^m(\mu)$ (simmetric wrt equator)
- $P_n^m(-\mu) = -P_n^m(\mu)$ if $n - |m|$ is odd, $\Rightarrow Y_n^m(-\mu) = -Y_n^m(\mu)$ (antisymmetric wrt equator).
- n-|m| corresponds to the number of zeros of P_n^m in $]-1, 1[$, and hence it can be interpreted as meridional wavenumber of Y_n^m, with n being called total wavenumber.



$Y_n^m(\lambda, \mu), n = 5.$

The Spectral Transform Method on the sphere: spherical harmonics

orthogonality properties of spherical harmonics:

$$\int_{-1}^{1} P_{n_1}^{m_1}(\mu) P_{n_2}^{m_2}(\mu)\, d\mu = \begin{cases} \frac{(n+m)!}{(n-m!)} \frac{2}{2n+1} & \text{if } m_1 = m_2 = m \text{ and } n_1 = n_2 = n \\ 0 & \text{if } m_1 \neq m_2 \text{ or } n_1 \neq n_2 \end{cases}$$

Recurrence relations

taking a derivative in zonal direction of a spherical harmonic is easy

$$\frac{\partial}{\partial \lambda} Y_n^m = \frac{\partial}{\partial \lambda} P_n^m(\mu) e^{im\lambda} = imP_n^m(\mu) e^{im\lambda} = imY_n^m$$

Taking a derivative in meridional direction is more complicated and recurrence relations are needed. Starting with $P_0^0(\mu)$ they allow the computation of $P_n^m(\mu)$ and its derivatives for any given m and n ($n > m$):

$$\mu P_n^m(\mu) = \epsilon_{n+1}^m Pm_{n+1}(\mu) + \epsilon_n^m Pm_{n-1}(\mu)$$

$$(1-\mu^2) dP_n^m/d\mu(\mu) = -n\epsilon_{n+1}^m P_{n+1}^m(\mu) + (n+1)\epsilon_n^m P_{n-1}^m(\mu)$$

$$(1-\mu^2) dP_n^m/d\mu(\mu) = (2n+1)\epsilon_n^m P_{n-1}^m(\mu) - n\mu P_n^m(\mu)$$

$$(1-\mu^2)^{1/2} P_n^m(\mu) = g_n^m P_{n+1}^{m+1}(\mu) - h_n^m P_{n-1}^{m+1}(\mu)$$

where

$\epsilon_n^m = \left(\frac{n^2-m^2}{4n^2-1}\right)^{1/2}$, $g_n^m = \left(\frac{(n+m+1)(n+m+2)}{(2n+1)(2n+3)}\right)^{1/2}$, $h_n^m = \left(\frac{(n-m-1)(n-m)}{(2n+1)(2n-1)}\right)^{1/2}$

Remark: similar relations may be obtained for Y_n^m by multiplying the previous ones by $e^{im\lambda}$

Expanding a spherical field

A field $A(\lambda, \mu)$ on the sphere can be written as:

$$A(\lambda, \mu) \sum_{m=-\infty}^{+\infty} \sum_{n=|m|}^{+\infty} A_n^m Y_n^m(\lambda, \mu)$$

Spectral coefficients A_n^m are complex numbers that can be computed, thanks to orthogonality relation, via:

$$A_n^m = \frac{1}{4\pi} \int_{-1}^{1} \int_{0}^{2\pi} A(\lambda, \mu) \overline{Y_n^m} d\lambda d\mu$$

Expansion coefficients verify the Parseval-Plancherel relation (again thanks to orthogonality):

$$\frac{1}{4\pi} \int_{-1}^{1} \int_{0}^{2\pi} [A(\lambda, \mu)]^2 d\lambda d\mu = \sum_{m=-\infty}^{+\infty} |A_n^m|^2,$$

hence global quadratic quantities (like kinetic energy or enstrophy) can be computed directly in spectral space
If $A(\mu, \lambda)$ is real, then it must hold: $A_n^{-m} = \overline{A_n^m}$

Truncating the expansion

Due to limited memory capacity, spherical harmonics expansions need to be truncated, which leads to a truncation error.
The expansion truncation is defined as the set T of wavenumbers m and n used in the expansion:

$$A(\lambda,\mu) = \sum_{(m,n)\in T} A_n^m Y_n^m(\lambda,\mu)$$

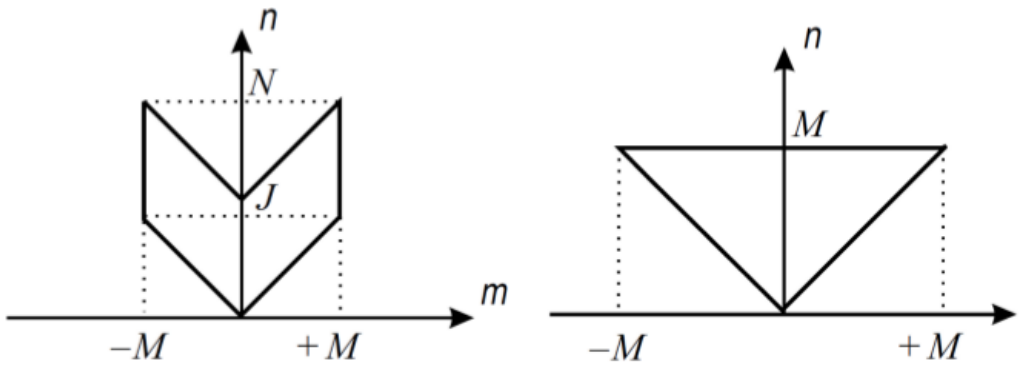
Several types of truncation are used, the main being:

(a) rhomboidal truncation

$$T = \{(n,m) : 0 \le |m| \le M, 0 \le |m| \le n \le |m| + N - M\}$$

(b) triangular truncation

$$T = \{(n,m) : 0 \le |m| \le M, 0 \le |m| \le n \le M\}$$



a - Rhomboidal truncation b - Triangular truncation

Remark: the main advantage of the triangular truncation is that it is isotropic, i.e. the resolution is uniform whatever the position of the poles.

Truncating the expansion

- Most used truncation is the triangular one:

$$A(\lambda,\mu) = \sum_{m=-M}^{M} \sum_{n=|m|}^{M} A_n^m Y_n^m(\lambda,\mu)$$

  e.g. T42 means triangular truncation with M=42.

- The total number of Y_n^m for triangular truncation TM is $(M+1)(M+2)/2$ Hence a T42 spectral model contains $(43 \times 44)/2 = 946$ spectral components.

- An equivalence can be established btw. the truncation M of a spectral model and the mesh size Δx of a gridpoint model:
  the smallest wavelength represented by a spectral model is $2\pi a/M$,
  while a gridpoint model takes into account at most $2\Delta x$ wavelength.
  Therefore we have:

$$2\Delta x \approx 2\pi a/M \quad \Longleftrightarrow \quad \Delta x_{(km)} \approx 20000/M$$

- The atmospheric component of the NCEP-ESM is based on the NCEP-GFS model with a spectral triangular truncation T126 wich is corresponding to a resolution of $0.9^\circ (\approx 120km)$.

Calculation of spectral coefficients of linear terms: wind components

In spherical coordinates, velocity components u, v are discontinuous across poles, so they cannot be expanded in spehrical harmonics.
Instead Robert functions are considered:

$$U = u\cos\theta, \qquad V = v\cos\theta$$

with Helmoltz decomposition (ψ =streamfunction, χ = velocity potential):

$$U = \frac{\partial\chi}{\partial\lambda} - \cos\theta\frac{\partial\psi}{\partial\theta} = \frac{\partial\chi}{\partial\lambda} - (1-\mu^2)\frac{\partial\psi}{\partial\mu}$$

$$V = \frac{\partial\psi}{\partial\lambda} + \cos\theta\frac{\partial\chi}{\partial\theta} = \frac{\partial\psi}{\partial\lambda} + (1-\mu^2)\frac{\partial\chi}{\partial\mu}$$

expanding χ and ψ in spherical harmonics with truncation T:

$$\psi(\lambda,\mu) = \sum_{(m,n)\in T} \psi_n^m Y_n^m, \qquad \chi(\lambda,\mu) = \sum_{(m,n)\in T} \chi_n^m Y_n^m$$

and remembering:

$$\frac{\partial}{\partial\lambda}Y_n^m = imY_n^m, \qquad (1-\mu^2)\frac{\partial P_n^m}{\partial\mu} = -n\epsilon_{n+1}^m Y_{n+1}^m(\mu) + (n+1)\epsilon_n^m Y_{n-1}^m(\mu)$$

we get:

$$U_n^m = im\chi_n^m + (n-1)\epsilon_n^m\psi_{n-1}^m - (n+2)\epsilon_{n+1}^m\psi_{n+1}^m,$$
$$V_n^m = im\psi_n^m - (n-1)\epsilon_n^m\chi_{n-1}^m + (n+2)\epsilon_{n+1}^m\chi_{n+1}^m$$

Calculation of spectral coefficients of linear terms: vorticity and divergence

The vorticity ζ and the divergence D can be written as Laplacian of the streamfunction ψ and of the velocity potential χ respectively:

$$\zeta = \nabla^2 \psi$$
$$D = \nabla^2 \chi$$

but, by definition, spherical harmonics are the eigenfunctions of the Laplace operator on the sphere:

$$\nabla^2 Y_n^m = -n(n+1) Y_n^m$$

therefore it is trivial to get spectral coefficients of ζ and D from those of ψ and χ :

$$\zeta_n^m = -n(n+1)\psi_n^m$$
$$D_n^m = -n(n+1)\chi_n^m$$

Remark: computation of spectral coefficients of linear terms is simple, but what about nonlinear terms? They represent the main weakpoint of the spectral method, making mandatory the shift to the spectral *transform* method.

Calculation of spectral coefficients of nonlinear terms: interaction coefficients method

Given two fields on the sphere $A(\lambda,\mu)$ and $B(\lambda,\mu)$, how to compute the spectral coefficients C_n^m of their product $C(\lambda,\mu) = A(\lambda,\mu)B(\lambda,\mu)$?
If

$$A(\lambda,\mu) = \sum_{m_1}\sum_{n_1} A_n^m Y_n^m, \quad B(\lambda,\mu) = \sum_{m_1}\sum_{n_1} B_n^m Y_n^m, \quad C(\lambda,\mu) = \sum_{m_1}\sum_{n_1} C_n^m Y_n^m$$

then

$$C_n^m = \frac{1}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi} C(\lambda,\mu)\overline{Y_n^m}\,d\lambda d\mu$$

$$= \frac{1}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi} \left(\sum_{m_1}\sum_{n_1} A_{n_1}^{m_1} e^{im_1\lambda} P_{n_1}^{m_1}\right)\left(\sum_{m_2}\sum_{n_2} B_{n_2}^{m_2} e^{im_2\lambda} P_{n_2}^{m_2}\right) e^{-im\lambda} P_n^m\, d\lambda d\mu$$

$$= \sum_{m_1}\sum_{n_1}\sum_{m_2}\sum_{n_2}\left[\frac{1}{2\pi}\int_{0}^{2\pi} e^{i(m_1+m_2-m)\lambda} d\lambda \frac{1}{2}\int_{-1}^{1} P_{n_1}^{m_1} P_{n_2}^{m_2} P_n^m\, d\mu\right] A_{n_1}^{m_1} B_{n_2}^{m_2}$$

Remarks:

- If $A(\lambda,\mu), B(\lambda,\mu)$ are defined with truncation M, the product $C(\lambda,\mu)$ is defined with truncation $2M$, for which only the terms included in truncation M need to be kept.
- spectral coefficients C_n^m are expressed as a weighted sum of the product of coefficients A_n^m and B_n^m, the weights being constituted by the integrals $[\quad]$ termed *interaction coefficients*.
- This method is not computationally efficient unless M is small ($M \leq 6$): however, when dealing with a large number of spectral components M, the number of interaction coefficients becomes too large ($M = 126 \Longrightarrow \#[\,] \approx 5.4 \times 10^{11}$), and their storage, bookkeeping and off and on retrival times become cumbersome.

Calculation of spectral coefficients of nonlinear terms: transform method

Spectral coefficients C_n^m of the product $C = AB$ can be computed more efficiently not directly from A_n^m, B_n^m (through interaction coefficients) but instead by running through these steps:

1. *from spectral space to grid space*: we compute the values of the factors A and B at points (λ_j, μ_k) of a transformation grid $G = \{(\lambda_j, \mu_k), \quad j = 1, \ldots, J, \quad k = 1, \ldots, K\}$ i.e. we evaluate $A(\lambda_j, \mu_k), B(\lambda_j, \mu_k)$.
2. *we compute the product in the grid space*:

$$C(\lambda_j, \mu_k) = A(\lambda_j, \mu_k) B(\lambda_j, \mu_k)$$

3. *from grid space to spectral space* : $C(\lambda_j, \mu_k)$ are transformed back to C_n^m by computing the integral (spectral transform)

$$C_n^m = \frac{1}{2} \int_{-1}^{1} \left[\frac{1}{2\pi} \int_0^{2\pi} C(\lambda, \mu) e^{-im\lambda} d\lambda \right] P_n^m(\mu) d\mu$$

through proper numerical quadrature formulae which are using only prevoiusly computed grid point values $C(\lambda_j, \mu_k)$.

Calculation of spectral coefficients of nonlinear terms: transform method

The spectral transform

$$C_n^m = \frac{1}{2}\int_{-1}^{1}\left[\frac{1}{2\pi}\int_0^{2\pi} C(\lambda,\mu)e^{-im\lambda}d\lambda\right]P_n^m(\mu)d\mu$$

can be viewed as the composition of a Fourier transform,

$$C_m(\mu_k) = \frac{1}{2\pi}\int_0^{2\pi} C(\lambda,\mu_k)e^{-im\lambda}d\lambda$$

giving the Fourier coefficients of $C(\lambda,\mu)$ at fixed latitude $\mu=\mu_k$ and computed through a trapezoidal quadrature

$$C_m(\mu_k) = \frac{1}{J}\sum_{j=1}^{J} C(\lambda_j,\mu_k)e^{-im\lambda_j}$$

which is exact if at least $J-1 = 2M+M \Longrightarrow J = 3M+1$,
and a Legendre transform

$$C_n^m = \frac{1}{2}\int_{-1}^{1} C_m(\mu)P_n^m(\mu)d\mu$$

computed through a Gaussian quadrature (w_k = Gaussian weights, μ_k Gaussian nodes)

$$C_n^m = \sum_{k=1}^{K} w_k C_m(\mu_k)P_n^m(\mu_k)$$

which is exact if at least $2K-1 = 3M \Longrightarrow K = (3M+1)/2$
In conclusion: by using triangular truncation T_M and choosing the number of points of the *Gaussian* grid G so as to comply with the constraints:

$$J \geq 3M+1, \quad K \geq \frac{3M+1}{2},$$

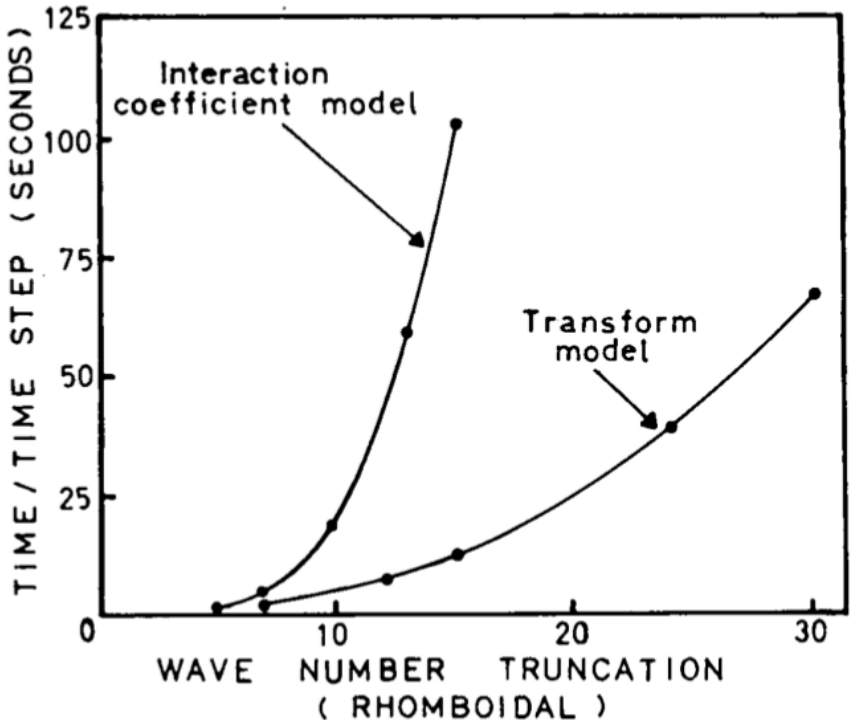
the spectral coefficients C_n^m m of the product $C=AB$ belonging to truncation T_M can be calculated exactly using the quadrature indicated.

Spectral transform method: Gaussian grids

- The Gaussian grids are defined by the quadrature points used to facilitate the accurate numerical computation of the integrals involved in the Fourier and Legendre transforms.
- The grids are labelled by N where N is the number of latitude lines between the pole and the equator: for example, for the N=640 Gaussian grid, there are 640 lines of latitude between the pole and the equator giving 1280 latitude lines in total.
- The grid points in latitude, θ_k , are given by the zeros of the Legendre polynomial of order $2N$ (i.e., the total number of latitude lines from pole to pole): $P_{2N}^0(\mu_k = \sin\theta_k) = 0$. A consequence of this is that a Gaussian grid has:
  - latitude lines which are not equally spaced;
  - no latitude points at the poles;
  - no line of latitude at the equator;
  - latitude lines which are symmetric about the equator.

Spectral transform method vs. spectral method

Computation time per time step as a function of spectral resolution.
Integrations of a global spectral model employing a transform method and employing the interaction coefficient method are compared
(from Bourke 1972):



This is the reason why you will find no spectral model nowadays but only spectral *transform* models.

Shallow water equation spectral transform model

Let us consider the shallow water equation (SWE) system (it contains all the horizontal operators of a full 3d model), being ϕ the geopotential and $\mathbf{u}_H$ the horizontal velocity :

$$\frac{\partial \phi}{\partial t} = -\nabla \cdot (\phi \mathbf{u}_H)$$
$$\frac{\partial \mathbf{u}_H}{\partial t} = -(\mathbf{u}_H \cdot \nabla)\mathbf{u}_H - f\mathbf{k} \times \mathbf{u}_H - \nabla\phi$$

using the identity: $(\mathbf{u}_H \cdot \nabla)\mathbf{u}_H = \frac{1}{2}\nabla\left(\mathbf{u}_H \cdot \mathbf{u}_H\right) + \zeta\mathbf{k} \times \mathbf{u}_H$, ($\zeta$ =vorticity) the momentum eq. is:

$$\frac{\partial \mathbf{u}_H}{\partial t} = -(\zeta + f)\mathbf{k} \times \mathbf{u}_H - \nabla\left(\phi + \frac{1}{2}\mathbf{u}_H \cdot \mathbf{u}_H\right)$$

a spectral model requires momentum eq. to be written in terms of vorticity and divergence equations, so that we take $\mathbf{k} \cdot \nabla\times$ and $\nabla\cdot$ of prevoius eq., obtaining:

$$\frac{\partial \zeta}{\partial t} = -\nabla \cdot (\zeta + f)\mathbf{u}_H$$
$$\frac{\partial}{\partial t}(\nabla \cdot \mathbf{u}_H) = \frac{1}{a\cos\theta}\frac{\partial}{\partial \lambda}[v(\zeta + f)] - \frac{1}{a\cos\theta}\frac{\partial}{\partial \theta}[u\cos\theta(\zeta + f)] - \nabla^2\left(\phi + \frac{1}{2}\mathbf{u}_H \cdot \mathbf{u}_H\right)$$

to remove discontinuities at poles we replace u, v with $U = \frac{u\cos\theta}{a}$, $V = \frac{v\sin\theta}{a}$, so that our SWE system takes the form ($D = \nabla \cdot \mathbf{u}_H$):

$$\frac{\partial \phi}{\partial t} = -\frac{1}{\cos^2\theta}\left(\frac{\partial (U\phi)}{\partial \lambda} + \cos\theta\frac{(\partial V\phi)}{\partial \theta}\right) - \phi D$$
$$\frac{\partial \zeta}{\partial t} = -\frac{1}{\cos^2\theta}\left(\frac{\partial}{\partial \lambda}(U(\zeta + f)) + \cos\theta\frac{\partial}{\partial \theta}(V(\zeta + f))\right)$$
$$\frac{\partial D}{\partial t} = \frac{1}{\cos^2\theta}\left(\frac{\partial}{\partial \lambda}[V(\zeta + f)] - \cos\theta\frac{\partial}{\partial \theta}[U(\zeta + f)]\right) - \nabla^2\left(\frac{U^2 + V^2}{2\cos^2\theta} + \phi\right)$$

Shallow water equation spectral transform model

SWE equations can be written in terms of streamfunction ψ, velocity potential χ (remember that $\zeta = \nabla^2\psi,\ D = \nabla^2\chi$):

$$\frac{\partial\phi}{\partial t} = -\frac{1}{\cos^2\theta}\left(\frac{\partial(U\phi)}{\partial\lambda} + \cos\theta\frac{(\partial V\phi)}{\partial\theta}\right) - \phi\nabla^2\chi$$

$$\frac{\partial\nabla^2\psi}{\partial t} = -\frac{1}{\cos^2\theta}\left(\frac{\partial}{\partial\lambda}(U(\nabla^2\psi + f)) + s : \cos\theta\frac{\partial}{\partial\theta}(V(\nabla^2 + f))\right)$$

$$\frac{\partial\nabla^2\chi}{\partial t} = \frac{1}{\cos^2\theta}\left(\frac{\partial}{\partial\lambda}[V(\nabla^2\psi + f)] - \cos\theta\frac{\partial}{\partial\theta}[U(\nabla^2\psi + f)]\right) - \nabla^2\left(\frac{U^2 + V^2}{2\cos^2\theta} + \phi\right)$$

Now considering a spectral expansion of ψ, χ, ϕ, U, V, i.e. $\psi = \sum_m \sum_n \psi_n^m Y_n^m$ and similarly for χ, ϕ, U, V, after substitution of these spectral expansions in previous equations, the spectral formulation of SWE takes the form:

$$\frac{d}{dt}\phi_n^m = -[\alpha(U\phi, V\phi)]_n^m - \phi D_n^m$$

$$-n(n+1)\frac{d}{dt}\psi_n^m = -[\alpha(A, B)]_n^m$$

$$-n(n+1)\frac{d}{dt}\chi_n^m = [\alpha(B, -A)]_n^m - \left[\nabla^2\left(\frac{U^2 + V^2}{2\cos^2\theta} + \phi\right)\right]_n^m$$

where $A = U(\nabla^2\psi + f), \quad B = V(\nabla^2\psi + f), \quad \alpha(A, B) = \frac{1}{\cos^2\theta}\left(\frac{\partial A}{\partial\lambda} + \cos\theta\frac{\partial B}{\partial\theta}\right)$

Remark: spectral coefficients of nonlinear terms at rhs are calculeted through the spectral transform method, i.e. by:

- first projecting ψ, χ, ϕ, U, V, from spectral to the grid-space domain on a Gaussian grid,
- multiplying these grid point values to obtain the grid-point values of the nonlinear terms,
- Fourier-Legendre transforming them to convert grid-point values of the nonlinear terms into their spectral coefficients.

Shallow water equation spectral transform model

- The sum of linear and nonlinear terms at right hand side forms the tendencies $\frac{d\phi_n^m}{dt}(t)$, $\frac{d\psi_n^m}{dt}(t)$, $\frac{d\chi_n^m}{dt}(t)$ of the various spectral coefficients.
- These tendencies along with the spectral coefficients $\psi_n^m(t-\Delta t), \chi_n^m(t-\Delta t), \phi_n^m(t-\Delta t)$ at the previous timestep can be used to obtain the future values of the spectral coefficients by using an explicit finite difference time integration like leapfrog:

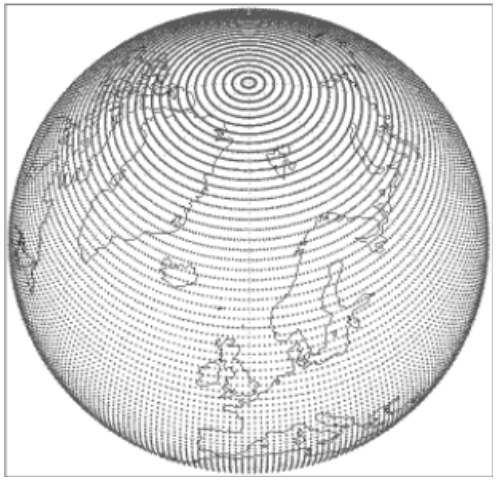
$$\begin{aligned}
\phi_n^m(t+\Delta t) &= \phi_n^m(t-\Delta t) + 2\Delta t \frac{d\phi_n^m}{dt}, \\
\psi_n^m(t+\Delta t) &= \psi_n^m(t-\Delta t) + 2\Delta t \frac{d\psi_n^m}{dt}, \\
\chi_n^m(t+\Delta t) &= \chi_n^m(t-\Delta t) + 2\Delta t \frac{d\chi_n^m}{dt}.
\end{aligned} \tag{8}$$

- Finally the new spectral coefficients $\phi_n^m(t+\Delta t), \psi_n^m(t+\Delta t), \chi_n^m(t+\Delta t)$ are transformede back to grid-point domain to obtain the forecast fields of the streamfunction ψ (or vorticity ζ), of the velocity potential χ (or divergence), and of the geopotential ϕ (or height) at the free surface.

How to further improve efficiency of spectral transform methods: reduced Gaussian grids (Hortal 1991, Hortal and Simmons 1991)

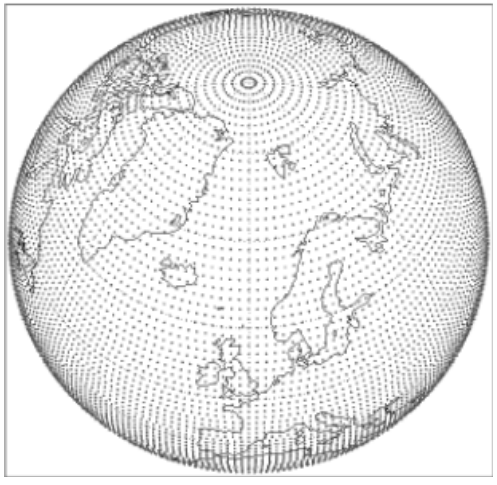
Regular (or full) Gaussian grid
A regular Gaussian grid has the following characteristics:

- there are $4N$ longitude points along each latitude circle;
- each latitude circle has a grid point at 0° longitude;
- the longitudinal resolution in degrees is $90^\circ/N$;
- the points get closer together (i.e. more crowded) as the latitude increases towards the poles;
- the total number of grid points is $8N^2$.



Reduced (or quasi-regular) Gaussian grid
A reduced Gaussian grid:

- has the same number of latitude lines ($2N$) as the corresponding regular Gaussian grid;
- has a grid point at 0° longitude on each latitude circle;
- has a decreasing number of longitude points towards the poles;
- has a quasi-regular grid spacing in distance at each latitude;
- provides a uniform CFL (Courant-Friedrichs-Lewy) condition.



How to further improve efficiency of spectral transform methods: reduced spectral transform (Jouang, 2004)

The amplitude $log_{10}|P_n^m|$ of the associated Legendre functions is the major key to the application of a reduced spectral transformation.

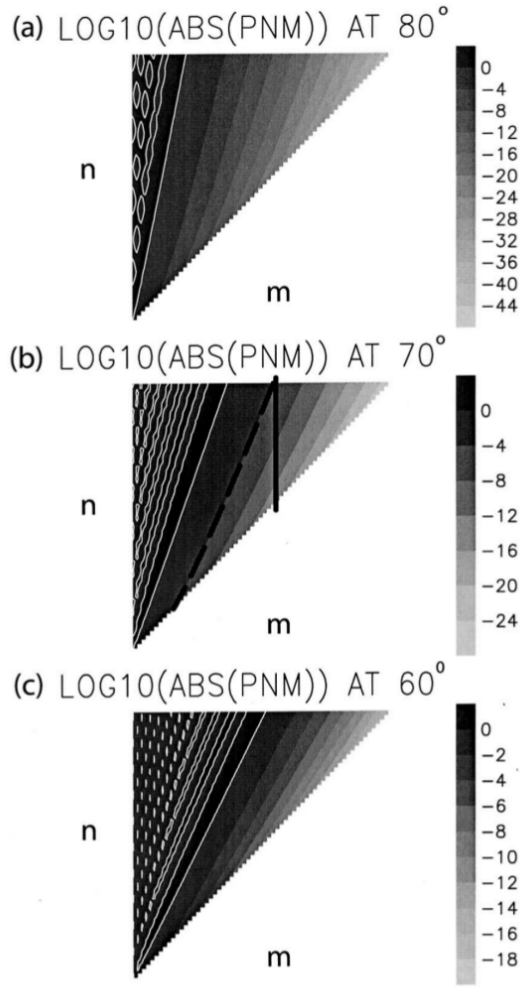


FIG. 1. Logarithm scale with base 10 of the absolute amplitude of the associated Legendre function, for T62 near the latitudes of (a) 80°, (b) 70°, and (c) 60°N. The black solid line in (b) indicates the trapezoidal spectral summation, and the dashed line indicates the scalenelike spectral summation. The white contours are zero values that indicate the unit of magnitude of the associated Legendre coefficient.

Features of $log_{10}|P_n^m|$:

- beams of local maxima that rotate from a lower zonal wave number m near the pole to a higher zonal wavenumber near the equator;
- at any given latitude $log_{10}|P_n^m|$ shows two regions separated by the rightmost white line with significant values of $|P_n^m|$ to the left and and dramatically monotonically decreasing values to the right;
- then the idea is that the contributions to the spectral expansion coming from P_n^m to the right of the rightmost white line can be neglected.

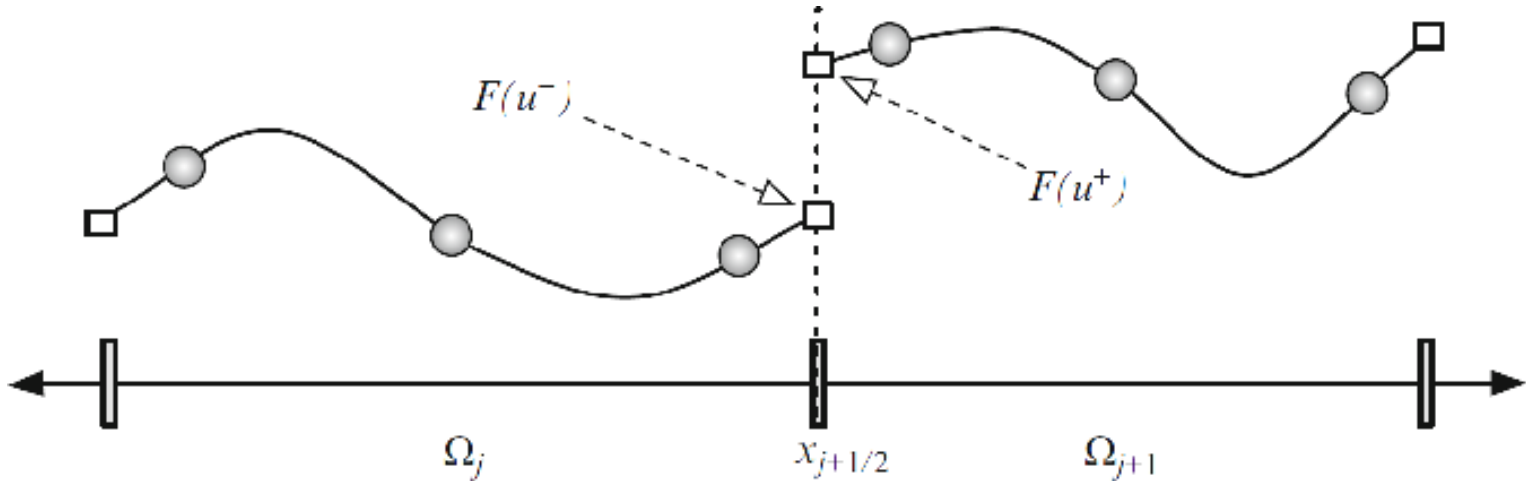
Spectral transform methods: pros and cons

- Advantages:
  - Fourier representation allows to compute horizontal derivatives exactly / analytically;
  - nonlinear quadratic terms calculated without aliasing (if computed in spectral space or using quadratic grids);
  - for a given accuracy fewer degrees of freedom are required than in a grid-point model;
  - Spherical harmonics are eigenfunctions of the Laplace operator in spherical coordinates: simpler solution of Helmoltz problem, and hence easy to construct semi-implicit schemes;
  - Spectral transform mtds are suited for GCMs: no pole problem;
  - Phase lag errors of mid-latitude synoptic disturbances are reduced w.r.t. grid-point models;
  - The use of staggered grids is avoided.

- Disadvantages:
  - The schemes appear complicated, even if they are relatively easy to implement;
  - Spectral harmonics are inerhently not suitable for limited-area models (RCMs);
  - the computation of nonlinear terms takes a long time unless the transform method is used;
  - physical processes effects cannot be included unless the transform method is used;
  - Spectral transform methods require global communication to accurately handle nonlinear terms, this decreases parallel performances.

Local Galerkin methods

- Same idea of spectral methods but now basis functions are not global, instead they have compact support ($\Longrightarrow$ fully discrete systems have sparse matrices);
- in finite elements (low order) and spectral elements (high order), basis functions are globally continuous;
- in Discontinuous Galerkin methods basis functions may jump across inter-element boundaries (enhanced flexibility).



DG methods for conservation laws

- Start from strong form of governing equations

$$\frac{\partial \mathbf{u}}{\partial t} + \boldsymbol{\nabla} \cdot \mathcal{F}(\mathbf{u}) = 0 \tag{9}$$

with proper initial and boundary conditions, to be solved in a domain $\Omega \subset \mathbb{R}^2$

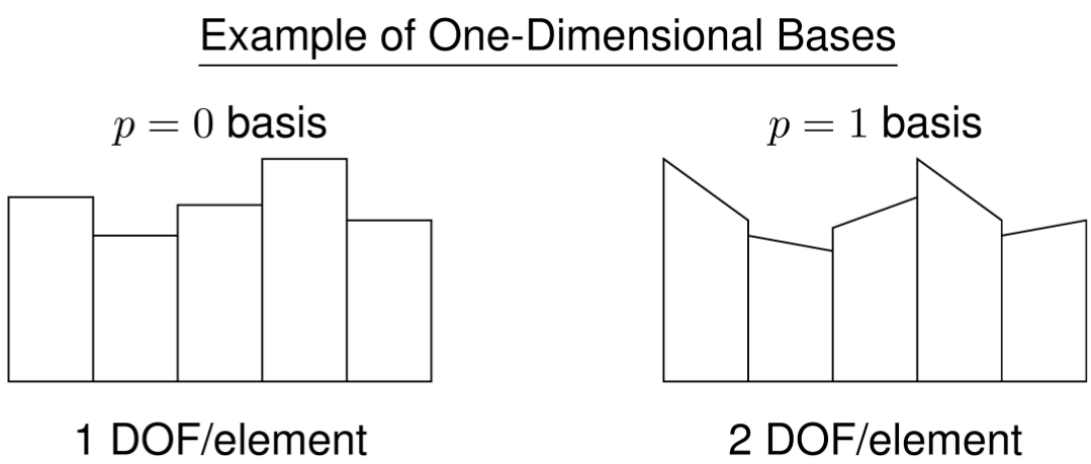
- Define a partition $\mathcal{T}_h = \{K_I\}_{I=1}^{N}$ of domain Ω and choose a finite dimensional subspace where we are looking for approximate solution :

$$\mathbf{u}_h(\cdot, t) \in V_h := \left\{ f \in L^2(\Omega) \,:\, f|_{K_I} \in \mathbb{Q}_{p_I}(K_I) \right\},$$

i.e. approximate the solution over the element K_I as the expansion over a finite set of given basis functions $\psi_{I,j}, j = 1, \ldots, (p_I + 1)^2$

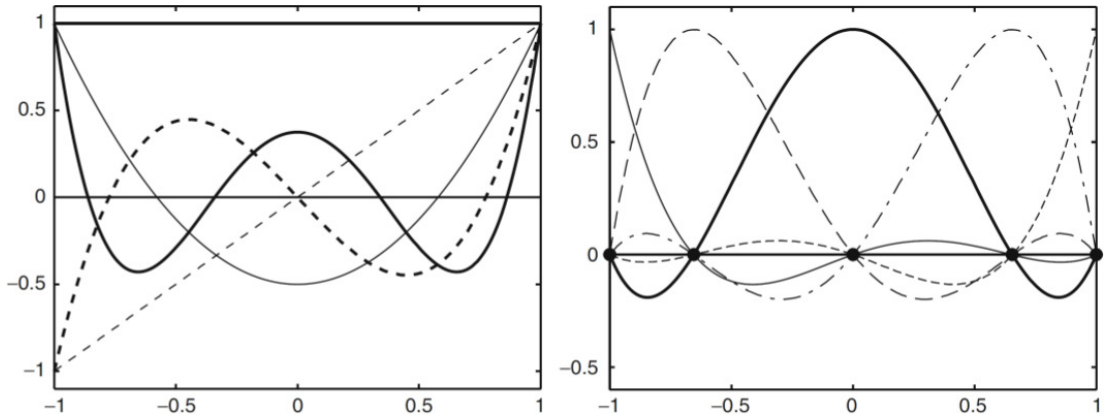
$$\mathbf{u}_h(\mathbf{x}, t)\big|_{K_I} = \sum_{j=1}^{(p_I+1)^2} \mathbf{u}_{I,j}(t)\psi_{I,j}(\mathbf{x}). \tag{10}$$

-



DG methods for conservation laws

- Modal vs nodal bases:



- multiply governing equations by a test function $\varphi_k \in V_h$ and integrate over element $K_I \in \mathcal{T}_h$:

$$\int_{K_I} \varphi_k \left[\frac{\partial \mathbf{u}_h}{\partial t} + \boldsymbol{\nabla} \cdot \mathcal{F}(\mathbf{u}_h) \right] d\mathbf{x} = 0,$$

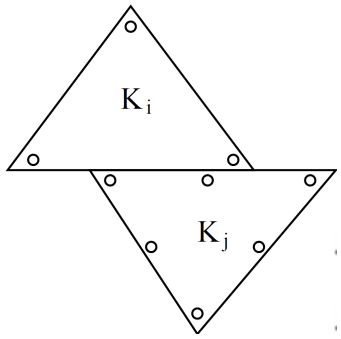
- integrate second term by parts (assuming interior element):

$$\int_{K_I} \varphi_k \frac{\partial \mathbf{u}_h}{\partial t}\, d\mathbf{x} - \int_{K_I} \boldsymbol{\nabla}\varphi_k \cdot \mathcal{F}\, d\mathbf{x} + \int_{\partial K_I} \varphi_k \hat{\mathcal{F}}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n})\, ds = 0, \quad (11)$$

- define proper numerical fluxes $\hat{\mathcal{F}}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n})$,
- substitute expansion (10) into (11) and obtain a ODE system for the expansion coefficients $\mathbf{u}_{I,j}^n, j = 1, \ldots, (p_I + 1)^2$.

DG advantages

- DG methods exhibit many interesting features, as:
  - Wide range of PDE's treated within the same unified framework;
  - high order accuracy;
  - flexibility in the mesh and finite element space design:



    - nonstructured and nonconforming meshes (hanging nodes)
    - nonuniform approximation degree $p = p_I$

  - freedom of choice of basis functions,
  - orthogonal bases can be easily constructed,
  - locality (elemental formulation),
  - high parallel scalability,

- therefore DG are appealing for design of new generation dycores but ....

DG challenging issues

- ... when coupled to explicit time stepping, DG methods are affected by severe stability restrictions as polynomial order increases:

  "The RKDG algorithm is stable provided the following condition holds:

  $$u\frac{\Delta t}{h} < \frac{1}{2p+1}$$

  where p is the polynomial degree; (for the linear case this implies a CFL limit $\frac{1}{3}$)"

  Cockburn-Shu, Math. Comp. 1989

- ... moreover DG requires more degrees of freedom per element than Continuous Galerkin (CG) approach, thus more expensive.

How to increase computational efficiency of DG ?

- coupling DG to semi-implicit semi-Lagrangian (SI-SL) technique (no CFL)

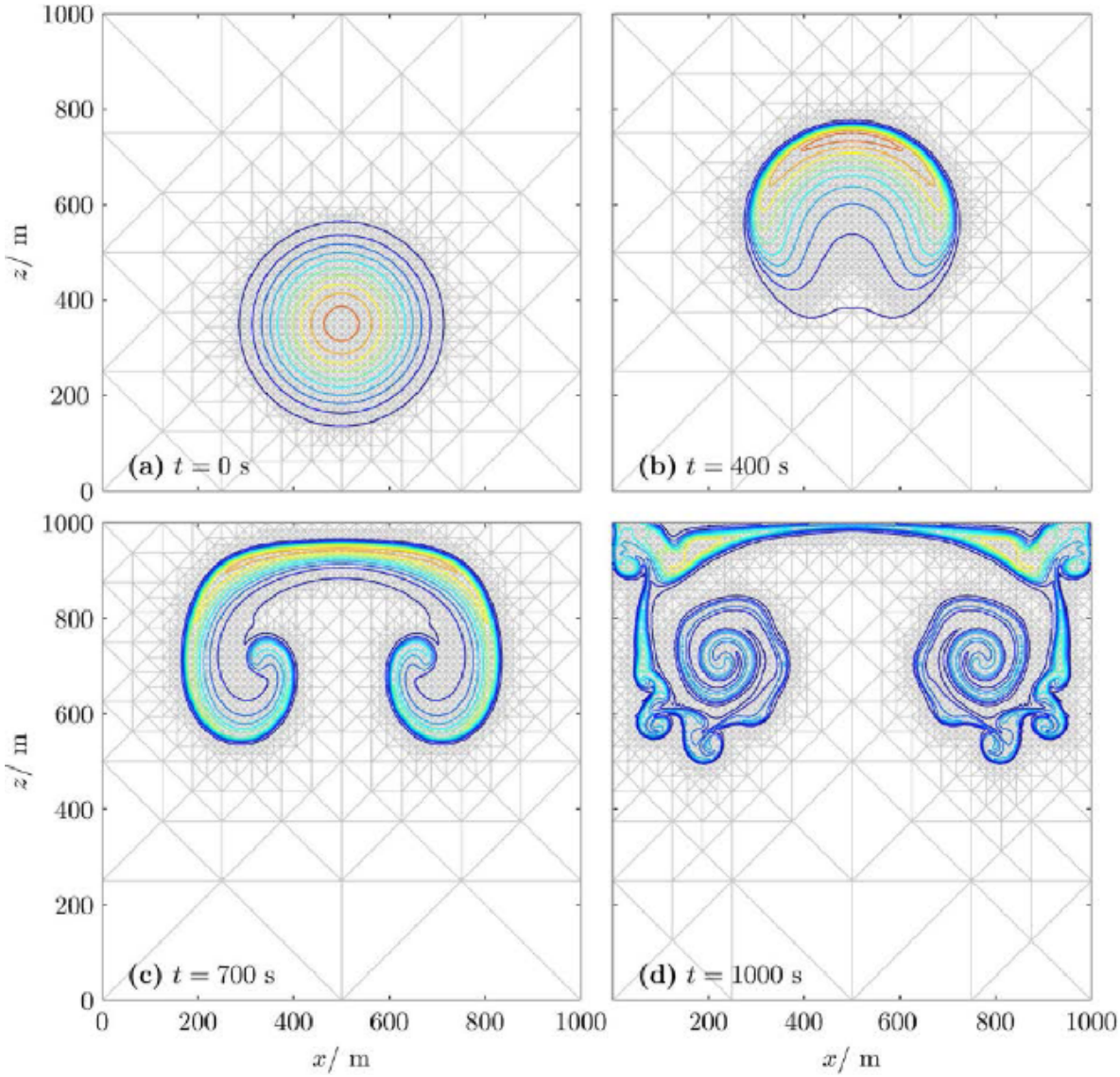
- introduction of $p-$ adaptivity (flexible degrees of freedom)

$\Longrightarrow$ p-SISLDG:

G. Tumolo, L. Bonaventura, M. Restelli, *J. Comput. Phys.* , 2013

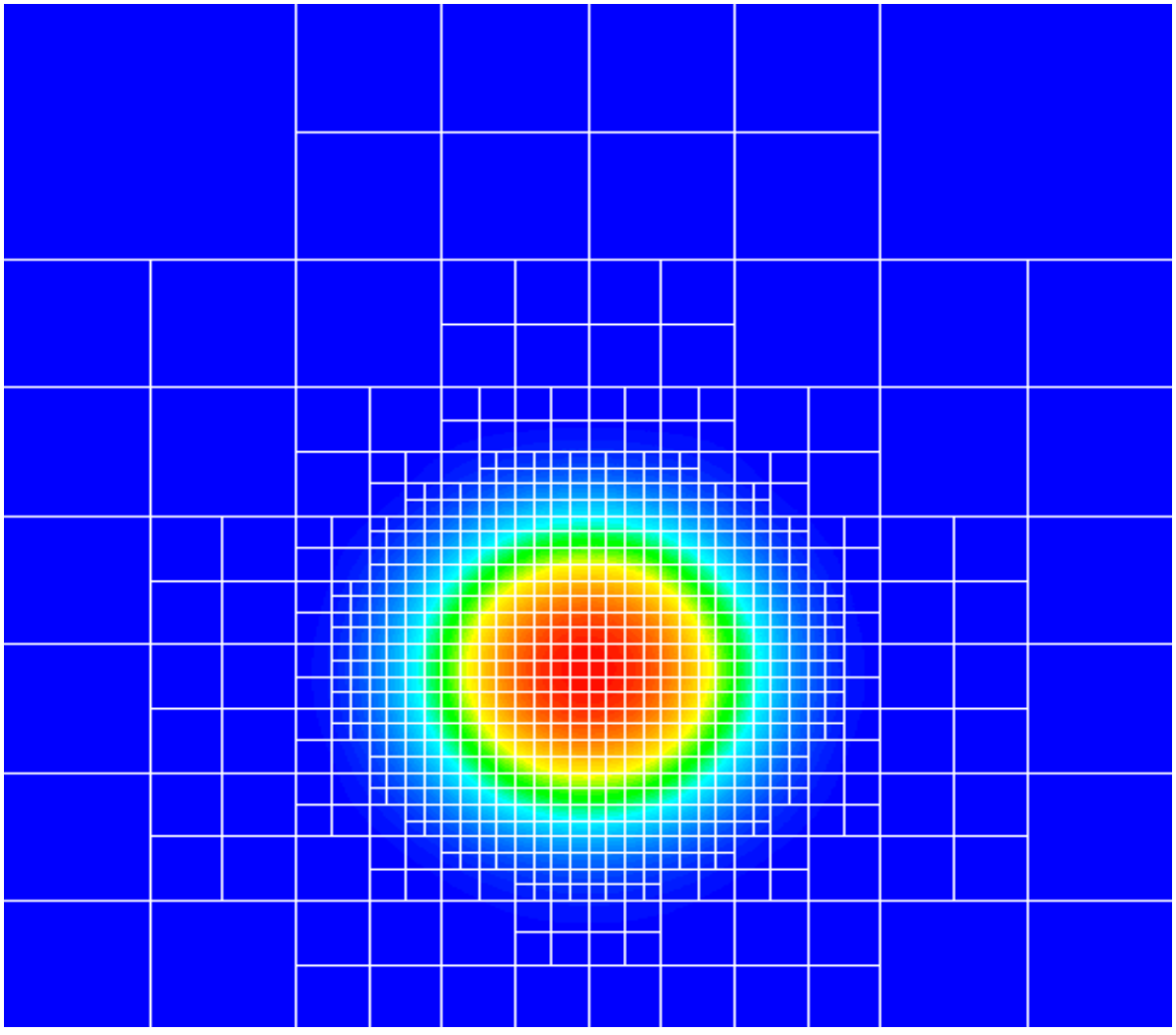
G. Tumolo, L. Bonaventura, *Quarterly J. Royal Met. Soc.*, 2015

Local Galerkin methods: dynamic $h-$ adaptivity / adaptive mesh refinement (AMR)



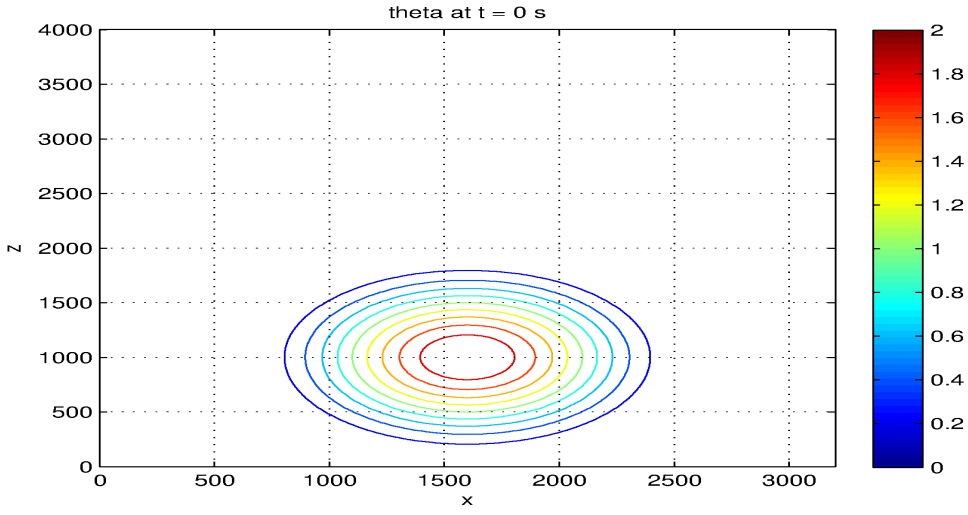
Giraldo and Restelli 2008

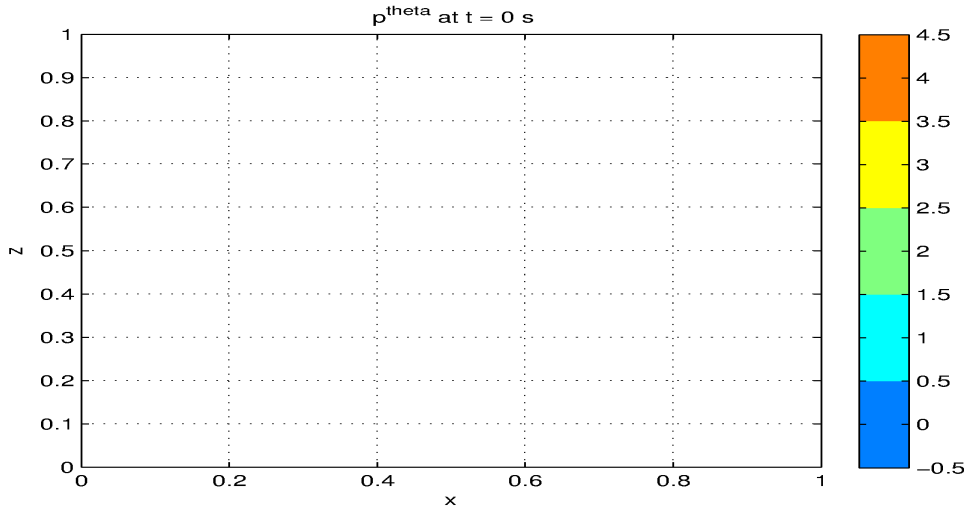
Galerkin methods: AMR

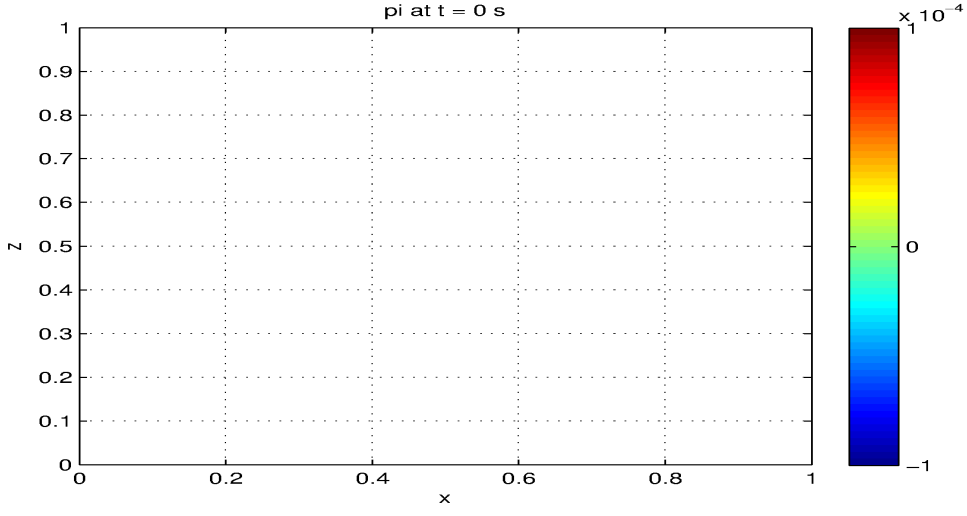


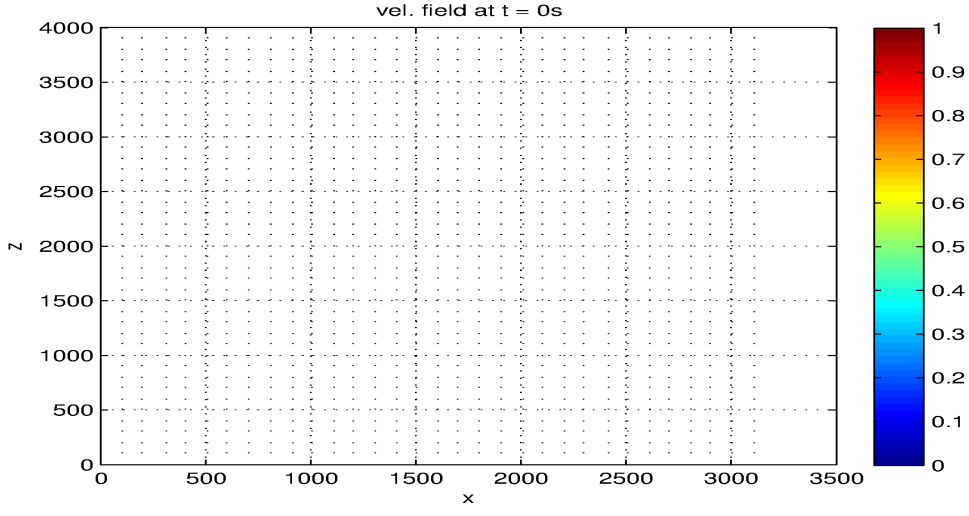
Giraldo and Restelli 2008

Galerkin methods: dynamic $p-$ adaptivity





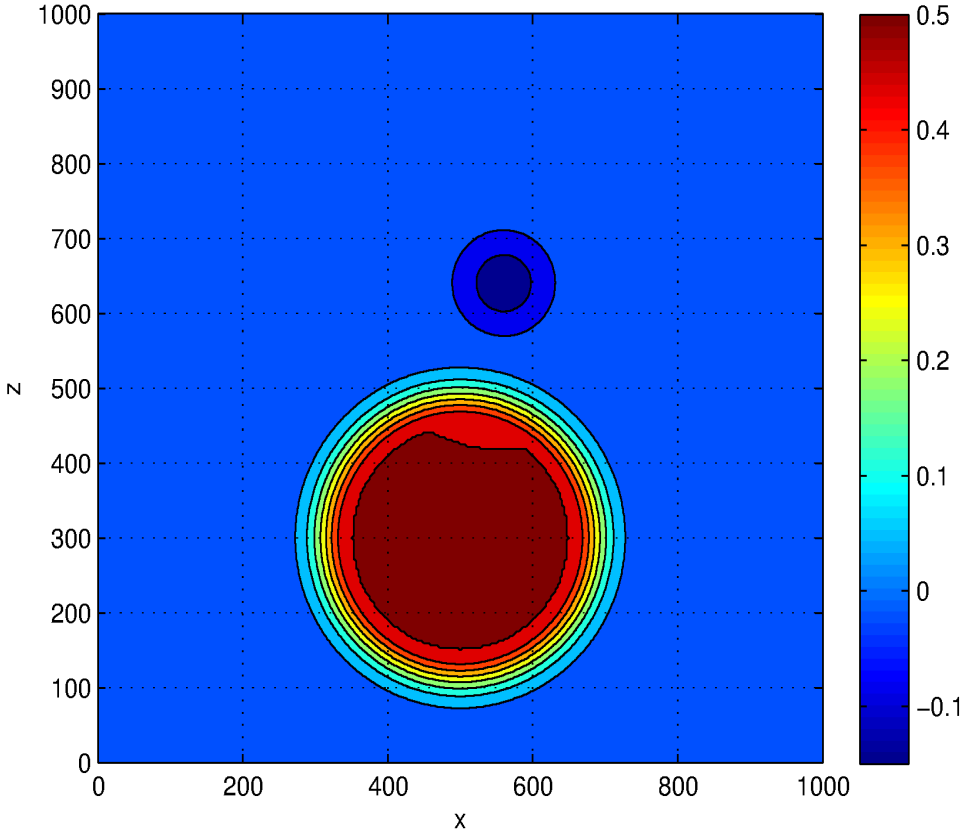


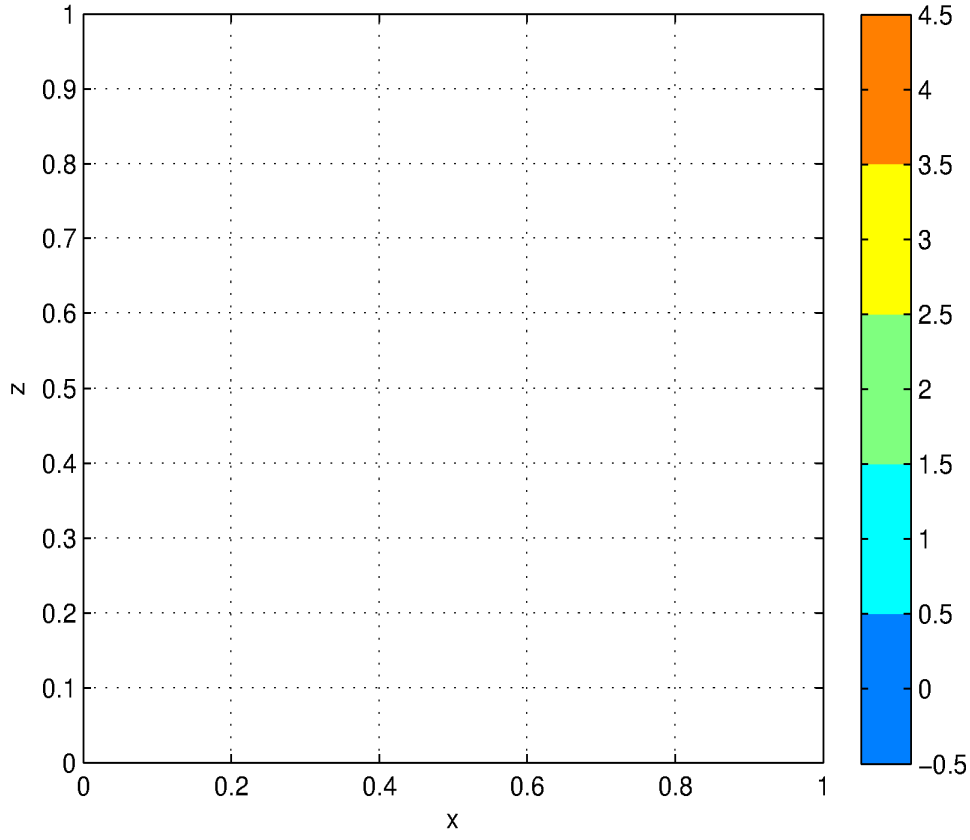


Tumolo and Bonaventura, Q. J. R. Meteorol. Soc. 2015

Interacting bubbles test (Robert, 1993)

50×50 elements, $p^{\pi} = 4$, $p^{u} = 5$, $\Delta t = 1$ s, $C \approx 87$.





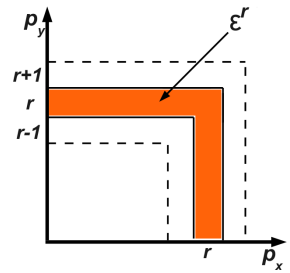
Dynamic p-adaptation: the strategy (Tumolo et al. âĂŐJ. Comput. Phys 2013)

- p-adaptivity easier by the use of modal bases: here tensor products of Legendre polynomials;
- hence, the representation for a model variable α becomes ($I = (I_x, I_y)$ multi-index):

$$\alpha(\mathbf{x})\big|_{K_I} = \sum_{k=1}^{p_I^\alpha+1} \sum_{l=1}^{p_I^\alpha+1} \alpha_{I,k,l} \psi_{I_x,k}(x) \psi_{I_y,l}(y).$$

- and its 2-norm is given by (in planar geometry):

$$\mathcal{E}_I^{tot} = \sum_{k,l=1}^{p_I^\alpha+1} \alpha_{I,k,l}^2 = \sum_{r=1}^{p_I^\alpha+1} \mathcal{E}_I^r, \quad \mathcal{E}_I^r := \sum_{max(k,l)=r} \alpha_{I,k,l}^2,$$



- while the quantity $w_I^r = \sqrt{\frac{\mathcal{E}_I^r}{\mathcal{E}_I^{tot}}}$ will measure the relative 'weight' of the $r-$ degree modes
- *Given an error tolerance $\epsilon_I > 0$ for all $I = 1, \ldots, N$, at each time step repeat following steps:*

1) *compute w_{p_i}*

2.1) *if $w_{p_i} \geq \epsilon_i$, then*
- 2.1.1) *set $p_i(\alpha) := p_i(\alpha) + 1$*
- 2.1.2) *set $\alpha_{i,p_i} = 0$, exit the loop and go the next element*

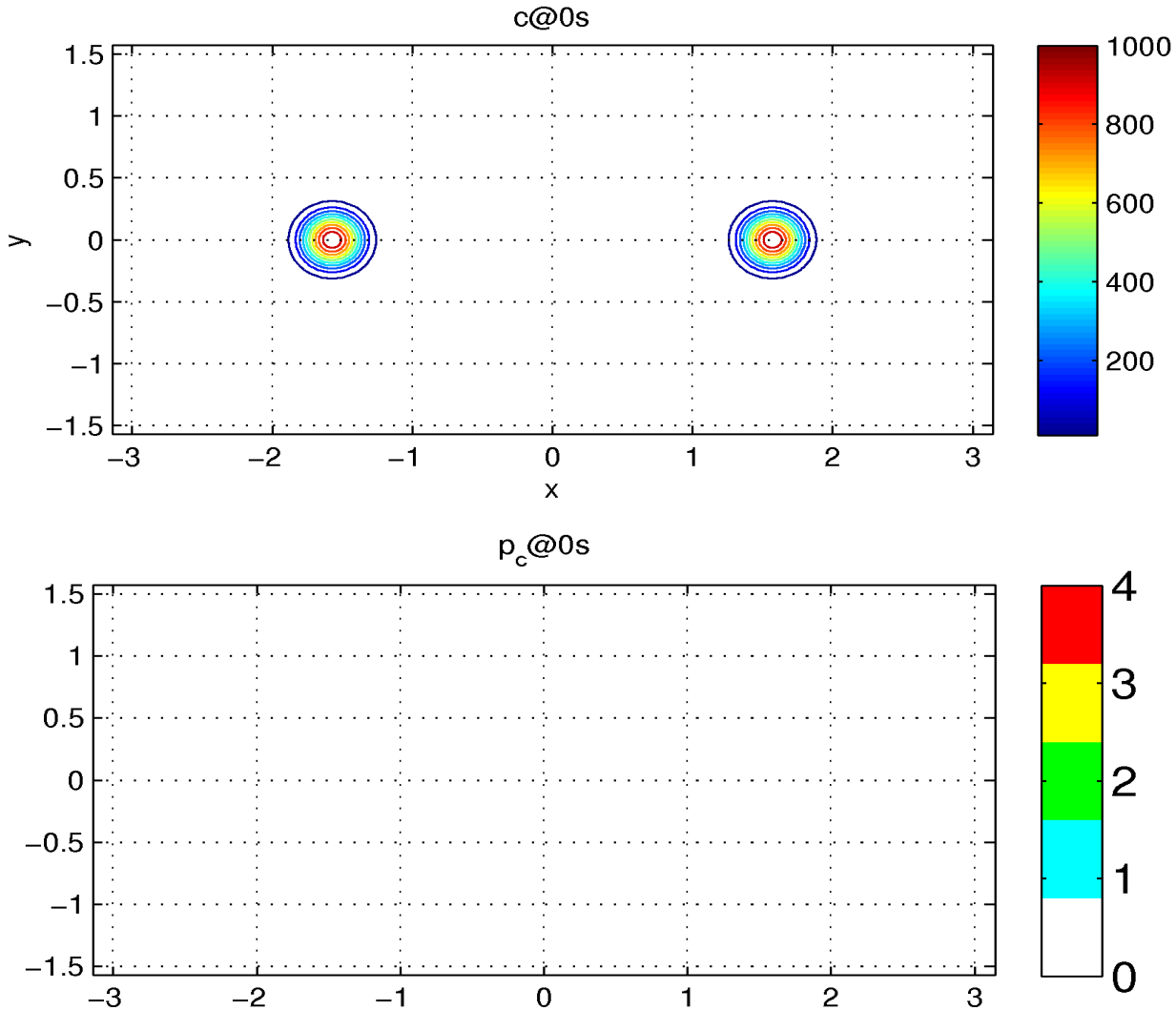
2.2) *if instead $w_{p_i} < \epsilon_i$, then*
- 2.2.1) *compute w_{p_i-1}*
- 2.2.2) *if $w_{p_i-1} \geq \epsilon_i$, exit the loop and go the next element*

Potential of $p-$adaptivity for atmospheric modelling applications

- No remeshing required of many physical quantities like orography profiles, data on land use and soil type, land-sea masks.
- Completely independent resolution for each single model variable.
- Easier coupling with SL technique, especially on unstructured meshes (no need to store two meshes).
- Possibility also of static p-adaptation: e.g. reduced p as counterpart of reduced grid, i.e. locally imposed p controlling the local Courant number ($\Longrightarrow$ significant $\#$gmres-iterations reduction).
- Main critical issue: dynamic load balancing is mandatory for massively parallel implementations.

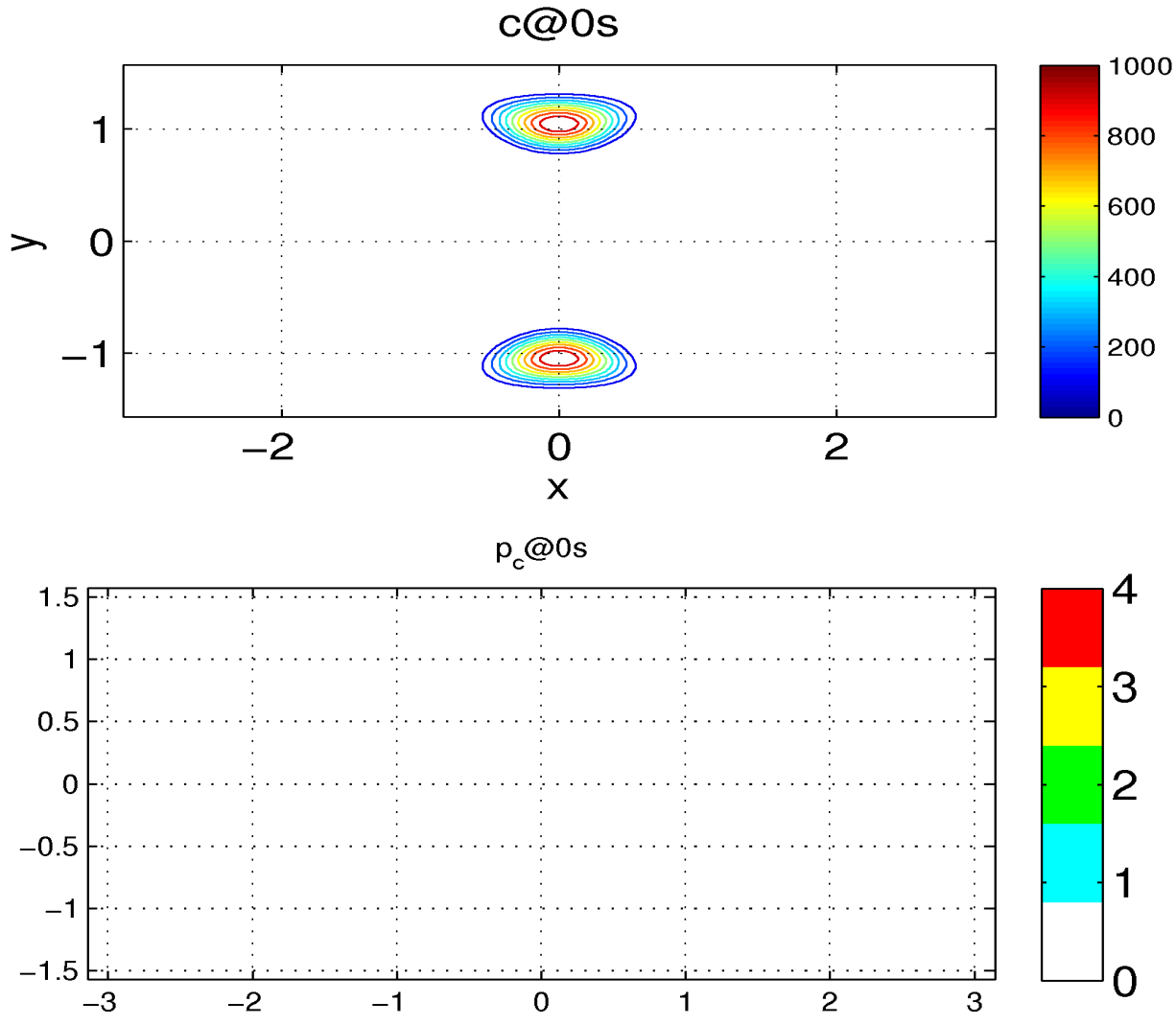
Solid body rotation on the sphere

120×60 elements, max $p^c = 4$, $\Delta t = 7200$s, $C_{vel,x} \approx 400$, $C_{vel,y} \approx 4$



Deformational flow on the sphere (adapted from Nair, Lauritzen 2010)

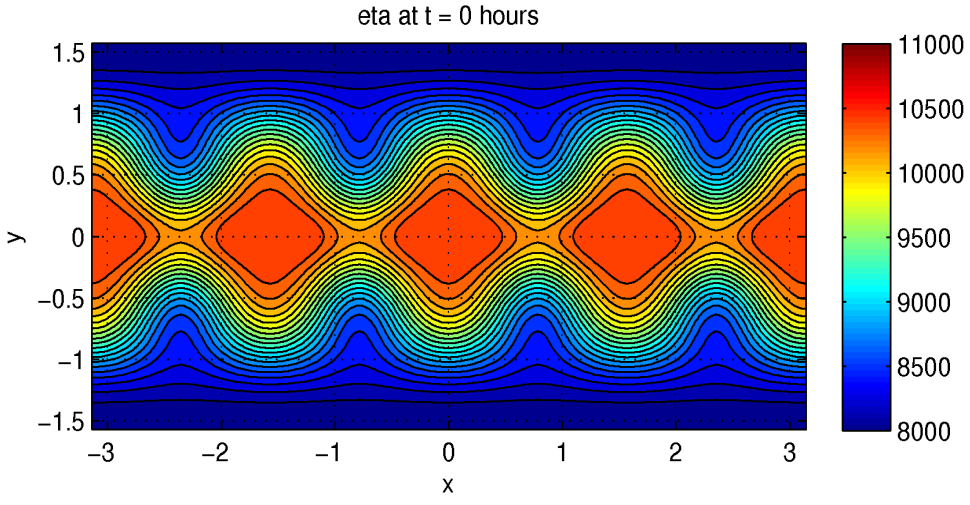
80×40 elements, max $p^c = 4$, $\Delta t = 1800$s

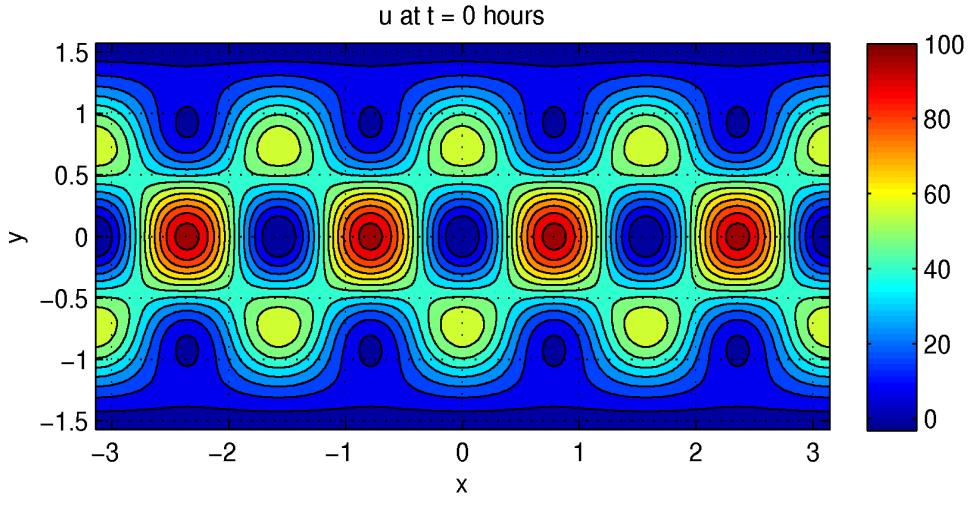


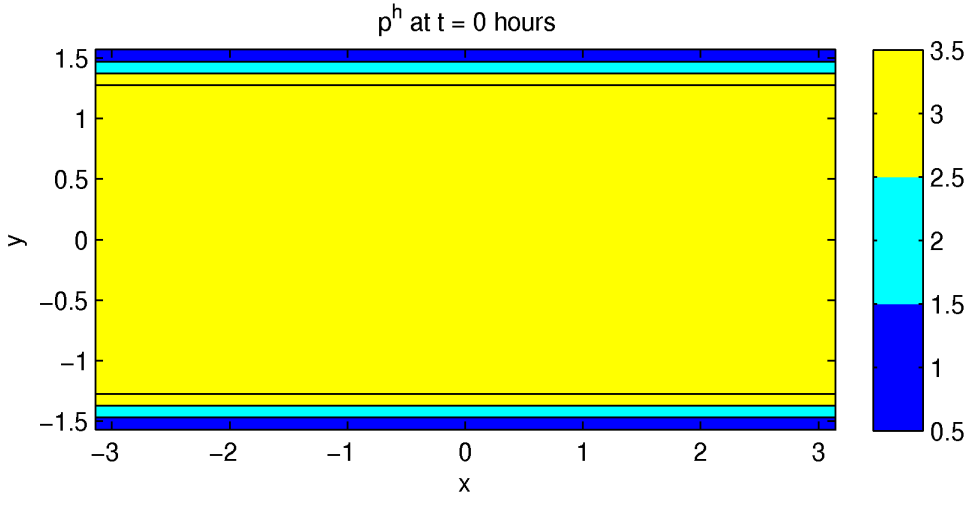
Combination of static + dynamic p-adaptation: Williamson's test 6

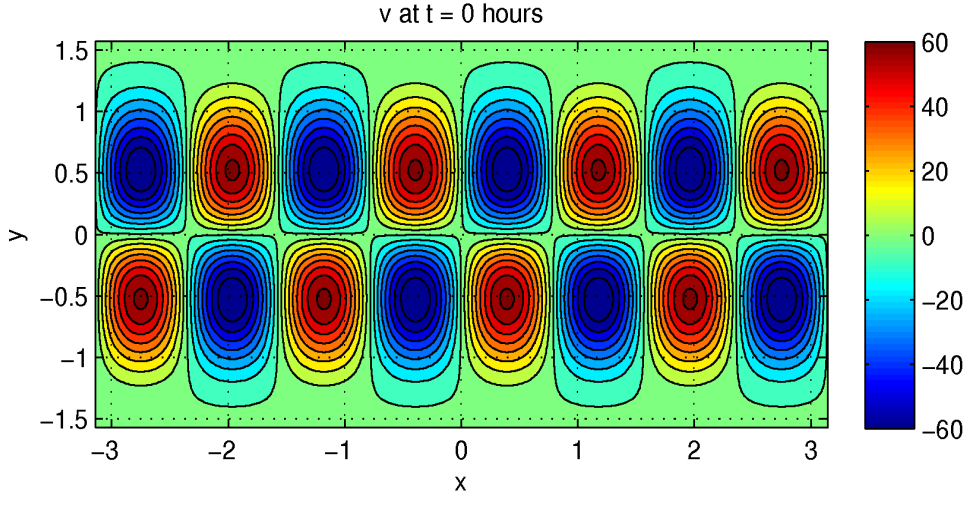
64×32 elements, $\max p^h = 4$, $\Delta t = 900s$ ($C_{cel} \approx 83$ without adaptivity).

$$\frac{\#\text{gmres-iterations}(p^h = \text{adapted})}{\#\text{gmres-iterations}(p^h = \text{uniform})} \approx 13\%, \qquad \Delta^n_{dof} = \frac{\sum_{I=1}^{N}(p^n_I + 1)^2}{N(p_{max}+1)^2} \approx 45\%.$$









Williamson's test 6: time convergence rate and p-adaptation efficiency

- Relative errors at $t_f = 15$ days for different number of elements, with respect to NCAR spectral model solution at resolution T511:

$N_x \times N_y$	Δt[min]	$l_1(h)$	$l_2(h)$	$l_\infty(h)$	q_2^{emp}
10×5	60	2.92×10^{-2}	3.82×10^{-2}	6.75×10^{-2}	-
20×10	30	5.50×10^{-3}	6.80×10^{-3}	1.11×10^{-2}	2.4
40×20	15	1.40×10^{-3}	1.80×10^{-3}	3.20×10^{-3}	2.0

- Relative differences btw adaptive (tol. $\epsilon = 10^{-2}$) and nonadaptive solution at $t_f = 15$ days:

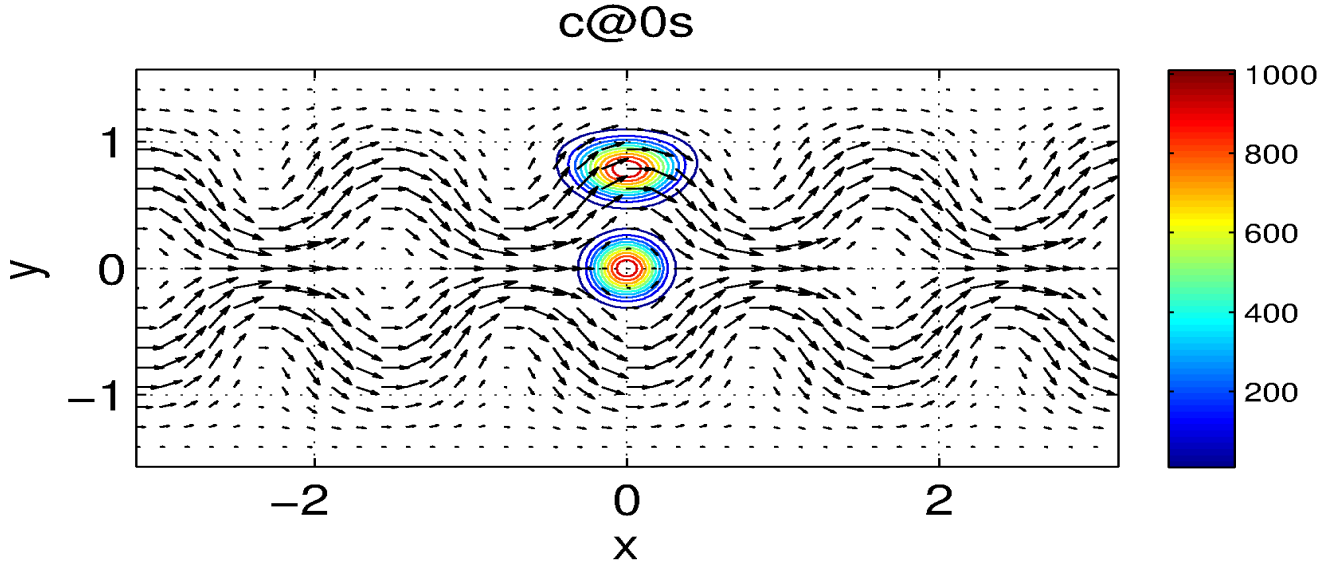
adaptivity	$l_1(h)$	$l_2(h)$	$l_\infty(h)$
static	2.182×10^{-4}	3.434×10^{-4}	2.856×10^{-4}
static + dynamic	3.407×10^{-4}	4.301×10^{-4}	7.484×10^{-4}

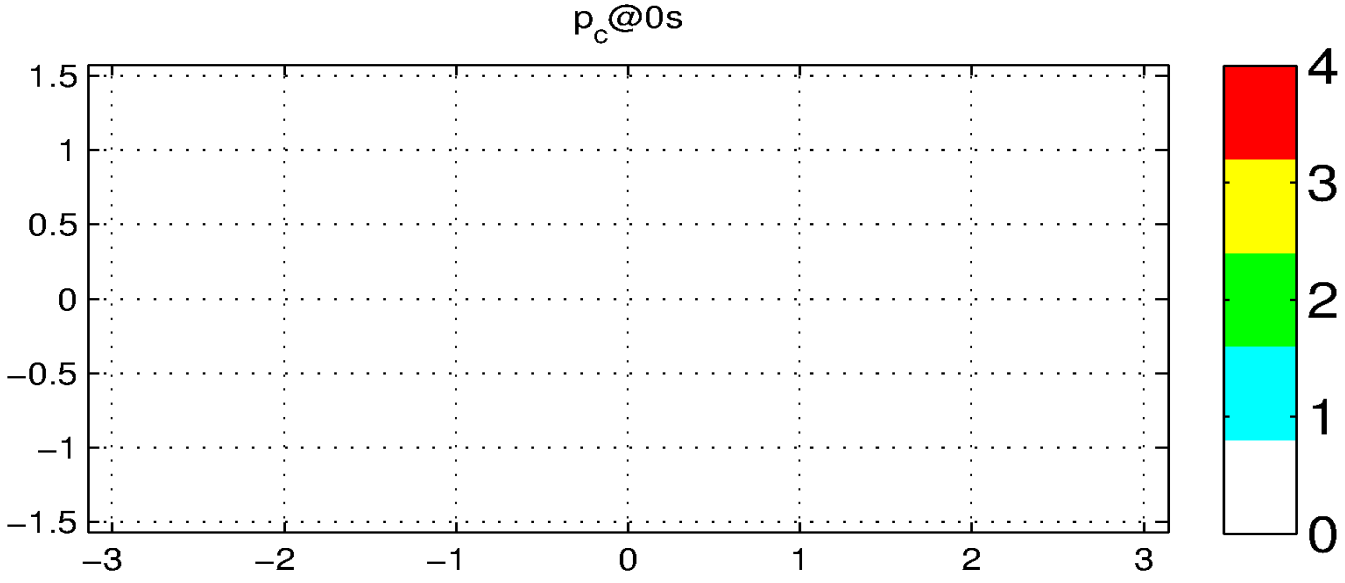
- CPU time: static and dynamic p-adaptive solution execution time is around 24% of that for nonadaptive solution.

- for details on these p-adaptive experiments G. Tumolo, L. Bonaventura, *Quarterly J. Royal Met. Soc.*, 2015.

Rossby Haurwitz wave velocity field

120×60 elements, max $p^c = 4$, $\Delta t = 900s$, $C_{vel,x} \approx 1$





Finite Volumes

- $\Omega_h = \{K_i, i = 1, \ldots, m\}$, called mesh, is a partition of Ω in non-overlapping control volumes K_I such that $\Omega = \bigcup_{i=1}^{m} K_i$;
  $h = max_i diam(K_i)$ is a measure of the size of the elements of Ω_h.

- in order to derive a discretized problem, the original continuous problem is rewritten in divergence (conservation) form:

$$\frac{\partial \psi}{\partial t} + \nabla \cdot \boldsymbol{F}(\psi) = S$$

  and integrated over each control volume K_i,

$$\int_{K_i} \frac{\partial \psi}{\partial t}\, d\boldsymbol{x} + \int_{K_i} \nabla \cdot \boldsymbol{F}(\psi)\, d\boldsymbol{x} = \int_{K_i} S\, d\boldsymbol{x}$$

- and the Gauss theorem is applied, to obtain:

$$\frac{d}{dt} \int_{K_i} \psi\, d\boldsymbol{x} + \int_{\partial K_i} \boldsymbol{n} \cdot \boldsymbol{F}(\psi)\, d\boldsymbol{x} = \int_{K_i} S\, d\boldsymbol{x}$$

- if $\varphi_i(t) \approx \frac{1}{|K_i|} \int_{K_i} \psi(\boldsymbol{x}, t)\, d\boldsymbol{x}$ and numerical fluxes $\hat{\boldsymbol{F}}$ are introduced in order to recover approximated fluxes at the cell edges from the cell averaged values $\varphi_i(t)$ (flux reconstruction), then our finite volume scheme for the evolution of the cell averages $\varphi_i(t)$ in terms of their numerical fluxes across the control volume edges is:

$$|K_i| \frac{d\varphi_i(t)}{dt} + \int_{\partial K_i} \boldsymbol{n} \cdot \hat{\boldsymbol{F}}(\boldsymbol{\varphi})\, d\boldsymbol{x} = \int_{K_i} S\, d\boldsymbol{x}, \qquad \text{where} \quad \boldsymbol{\varphi} = (\varphi_i)_{i=1}^{m}.$$

Finite Volumes: pros and cons

- inherently conservative;
- can be easily made to satisfy monotonicity and positivity constraints (i.e. avoid negative tracer densities): they can incorporate slope limiters for controlling spurious oscillations in the solution;
- Among others, MIT General Circulation Model and GFDL dynamical core are based on finite volume approach;
- they are very robust and are heavily used in other fields (e.g. industrial applications);
- mostly low order accurate (third order or less): to have high order finite volume schemes, i.e. high order flux reconstruction, large stencils are needed.

Time discretization techniques

- The time continuous solution $\varphi_i(t), \quad t \in [0, t_f]$ is approximated through Finite Differences, by introducing a timestep $\Delta t = t_f/n$ and a set of discrete time levels $t_k = k\Delta t, \quad k = 0, \ldots, n$
- numerical methods for ODE's provide an approximated solution (fully discrete!) $\varphi_i^k, \quad k = 0, \ldots, n$
- main approches include:
  - explicit schemes
  - semi-implicit schemes
  - semi-Lagrangian shemes
- the choice of Δt cannot generally be made based only on accuracy or efficiency considerations, but must also comply with numerical stability criteria, such as the CFL condition

Explicit schemes

- pro: they do not require the solution of a system at each timestep to update the solution from one discrete time level to the next;
- con: they are affected by severe stability conditions on the size of Δt
- example: one of the most popular explicit time discretization is the leapfrog. It is a multistep method (three time levels scheme) derived by approximating the time derivative by a centered difference approximation:

$$\frac{\varphi_i^{k+1} - \varphi_i^{k-1}}{2\Delta t} = \mathcal{L}_h(\varphi^k)_i$$

  presence of a computational mode, to be properly filtered (Asslin filter)

- efficiency of simple explicit schemes can be improved through the "split-explicit“ or ”mode splitting" technique: terms responsible for the fastest waves motion are treated separately, still explicitly, but emplying a substepping procedure with a smaller timestep.

Semi-implicit schemes

- an implicit discretization is selectively applied only to the terms responsible for the fastest waves motion. Remaining terms are treated explicitly.
- pro: only mild stability restrictions
- con: they do require the solution of a (generally linear) system at each timestep to update the solution from one discrete time level to the next. This can affect scalability.
- one of the most widely used implicit methods is the Crank-Nicolson scheme, defined as:

$$\frac{\varphi_i^{k+1} - \varphi_i^k}{\Delta t} = \alpha \mathcal{L}_h(\varphi^{k+1})_i + (1-\alpha)\mathcal{L}_h(\varphi^k)_i$$

  with $\alpha \in [0,1]$ averaging parameter: stability is guaranteed for $\alpha \in [1/2, 1]$,
  second order accuracy for $\alpha = 1/2$.
- also known as IMEX (implicit- explicit) especially in the Runge-Kutta framework.
- possible variant are HEVI methods (Horizontally Explicit Vertically Implicit).

Semi-Lagrangian schemes

- SL method is a discretization approach that links the spatial and time discretization for advection equations.
- the governing equations can be written in advective form:

$$\frac{d\psi}{dt} = \hat{\mathcal{L}}(\psi)$$

where $\frac{dc}{dt} = \frac{\partial c}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla} c$ is the Lagrangian derivative of $c = c(\boldsymbol{x}, t)$

- then its SL discretization takes the form

$$\frac{\varphi_i^{k+1} - \varphi_{i,*}^k}{\Delta t} = \mathcal{L}_h(\varphi^k)_{i,*}$$

where $\varphi_{i,*}^k$ denote an approximation of $\psi(\boldsymbol{x}_*, t_k)$
and $\boldsymbol{x}_* = \boldsymbol{X}(t^k; t^{k+1}, \boldsymbol{x}_i)$ denotes the solution at $t = t^k$ of

$$\frac{d\boldsymbol{X}(t; t^{k+1}, \boldsymbol{x}_i)}{dt} = \boldsymbol{u}(\boldsymbol{X}(t; t^{k+1}, \boldsymbol{x}_i), t) \tag{12}$$

with initial datum $\mathbf{X}(t^{k+1}; t^{k+1}, \boldsymbol{x}_i) = \boldsymbol{x}_i$ at time $t = t^{k+1}$.