Computer Science in Ocean and Climate Research

Lecture 8: Data Formats and Data Processing

Prof. Dr. Thomas Slawig

CAU Kiel
Dep. of Computer Science

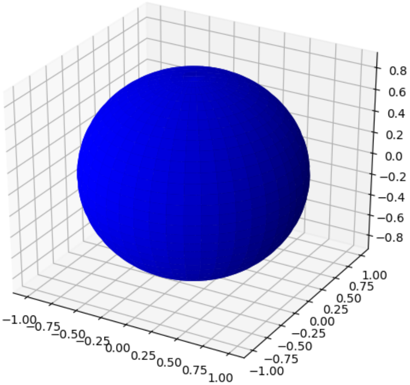
Summer 2020

Contents

Contents

Geometry

- Spatially resolved climate models (have to) take into account the Earth's special geometry.
- Earth's geometry is approximated by a sphere.
- Radius $r = 6371$km.
- Earth is not a perfect sphere.

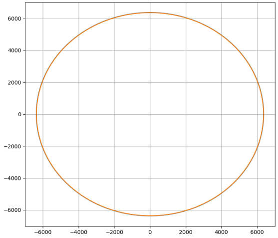


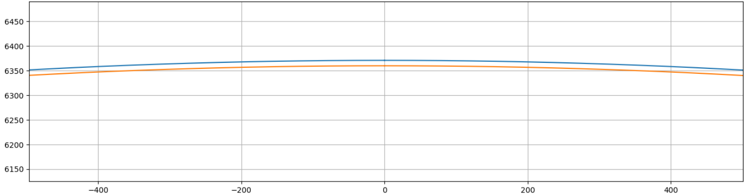
Geometry

- Processes in atmosphere and ocean take place in a spherical shell.
- Radius $r = 6371$km.
- Ocean depth max. ≈ 11km.

$\rightsquigarrow$ **Very thin** spherical shell.

$\rightsquigarrow$ Different scales in horizontal and vertical coordinates.





Visualization of the max. depth of ocean layer: part of spheres with $r = 6360$, $r = 6371$.

Contents

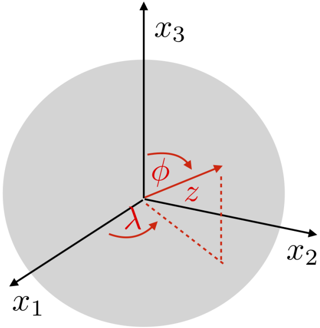
Coordinate system

- An appropriate coordinate system is obtained by using the spherical coordinates

$$(\lambda, \phi, z) \in [0, 2\pi) \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times [r_{min}, r_{max}],$$

  with **longitude** λ, **latitude** ϕ and vertical coordinate z.
- The cartesian coordinates have the representation

$$x(\lambda, \phi, z) := \begin{pmatrix} x_1(\lambda, \phi, z) \\ x_2(\lambda, \phi, z) \\ x_3(\lambda, \phi, z) \end{pmatrix} := \begin{pmatrix} z \sin \lambda \cos \phi \\ z \sin \lambda \sin \phi \\ z \cos \lambda \end{pmatrix}.$$



- The transformation is not unique at the poles ($\lambda = \pm\frac{\pi}{2}$).
- ⇝ Poles might be shifted over land (e.g., for ocean model).
- Sometimes, other symbols for the coordinates are used.
- Sometimes, the coordinate ϕ is defined in $[0, \pi]$.

Consequences of a coordinate transformation

- Climate models are discretized versions of differential equations.

⇝ The models include spatial derivatives.

- In a coordinate transformation, these derivatives are transformed using the chain rule.
- Let a quantity be given in cartesian coordinates:

$$q = q(x_1, x_2, x_3)$$

- We write it in spherical coordinates as

$$\tilde{q} = \tilde{q}(\lambda, \phi, z) := q\left(x_1(\lambda, \phi, z), x_2(\lambda, \phi, z), x_3(\lambda, \phi, z)\right).$$

- For the derivatives, we get the following type of relations (omitting the arguments):

$$\begin{aligned}\frac{\partial \tilde{q}}{\partial \lambda}(\lambda, \phi, z) &= \frac{\partial q}{\partial x_1}\frac{\partial x_1}{\partial \lambda} + \frac{\partial q}{\partial x_2}\frac{\partial x_2}{\partial \lambda} + \frac{\partial q}{\partial x_3}\frac{\partial x_3}{\partial \lambda} \\ &= \frac{\partial q}{\partial x_1} z \cos\lambda \cos\phi + \frac{\partial q}{\partial x_2} z \cos\lambda \sin\phi - \frac{\partial q}{\partial x_3} z \sin\lambda.\end{aligned}$$

⇝ Model equations are re-written in new coordinates.

Transformation of vector-valued quantities

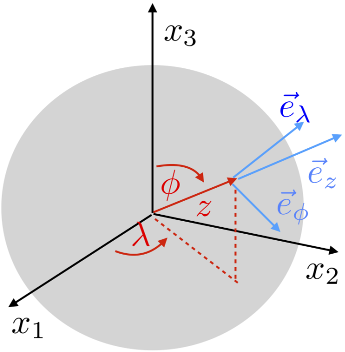
- Vector-valued quantities: velocity (water, air) and forces.
- Representation in cartesian coordinates:

$$\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + v_3\vec{e}_3$$

  with orthogonal unit coordinate vectors $\vec{e}_i, i = 1, 2, 3$.
- These basis vectors are the same at every point $x = (x_1, x_2, x_3)$ of the coordinate system.
- In spherical coordinates, we write the velocity as

$$\vec{v} = v_\lambda\vec{e}_\lambda + v_\phi\vec{e}_\phi + v_z\vec{e}_\phi,$$

  also with orthogonal unit coordinate vectors $\vec{e}_\lambda, \vec{e}_\phi, \vec{e}_z$.



- But these coordinate vectors now depend on the point (λ, ϕ, z).

$\rightsquigarrow$ When differentiating, also the coordinate vectors have to be differentiated w.r.t. space.

$\rightsquigarrow$ Additional terms appear in some model equations.

Special vertical coordinates

- Different types of vertical coordinates are used, e.g., in ocean models:
- Zero level $z = 0$ at surface, negative values for depth, $z \in [-h, 0]$.
- Inverted coordinate direction: zero level $z = 0$ at surface, $z \in [0, h]$ with h being maximal depth.
- Ocean surface varies with time: free surface, $z = 0$ refers to reference level (steady state), η deviation from it.
- Isopycnic coordinates: ocean is (approximately) stratified in layers of identical density

$\rightsquigarrow$ use these layers as coordinates.

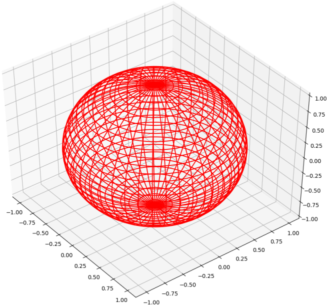
- Atmosphere: pressure-layer coordinates.
- σ-coordinates, following the terrain (orography).
- σ-hybrid coordinates, following the terrain in lower part, almost pressure coordinates in higher part.

Contents

Discretization and grids

- Most models have rectangular grids.
- Horizontal: equally spaced in the spherical coordinates

$\rightsquigarrow$ not equally spaced in cartesian space (smaller boxes at poles).



- Vertical: smaller grid-size close to ocean surface.
- Some atmosphere models use spectral methods for some processes (and grid-based methods for others):
  - No grid, but representation of variables as truncated series of special ansatz functions (Legendre polynomials, orthogonal functions on the sphere).
- Then the truncation level determines also horizontal resolution.
- Requires transformation of both formulations (Fast Fourier Transform FFT) in every step.

Discretization and grids

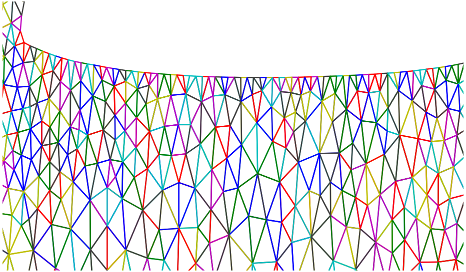
- Examples for resolutions in the atmosphere model ECHAM:

Version	truncation	spatial resolution	horizontal points	vertical levels	time step
CR	T031	3.8 deg	96×48	47	20 min
LR	T063	1.9 deg	96×192	47	10 min
MR	T063	1.9 deg	96×192	95	10 min
HR	T127	0.94 deg	384×192	95	5 min
XR	T255	0.5 deg	768×384	95	2 min

- Also used: Nested grids, higher resolution in relevant parts, embedded into grid with coarser resolution for the whole domain (e.g., ocean).

Newer development: unstructured grids

- Standard in applied mathematics in finite element/finite volume methods.
+ Better resolution of complicated-shaped geometry.
+ Local refinement possible.
- Used either in horizontal direction only, vertically still layers, ...
- ... or for full 3-D discretizations (tetrahedra, icosahedra)



Contents

Configuration files

- Used to specify parameters of the model and the simulation.
- Model parameters: physical, biological etc. constants.
- Parameters of the simulation setting: length.
- Numerical parameters: time-steps, numerical options.
- Parallelization options.
- Names and location of input data files ...
- ... for initial values and forcing domain and boundary data.
- Names and location of output files.
- Frequency of output.

Example: simple text file

- Metos3D option file:

```
# geometry
-Metos3DGeometryType                      Profile
-Metos3DProfileInputDirectory             data/TMM/2.8/Geometry/
-Metos3DProfileMaskFile                   landSeaMask.petsc
-Metos3DProfileVolumeFile                 volumes.petsc

# bgc tracer
-Metos3DTracerCount                       2
-Metos3DTracerName                        N,DOP
-Metos3DTracerInitValue                   2.17e+0,1.e-4
-Metos3DTracerOutputDirectory             work/
-Metos3DTracerOutputFile                  N.petsc,DOP.petsc
# weight with volumes and sum up
-Metos3DTracerMonitor

# diagnostic variables
-Metos3DDiagnosticCount                   0
# weight with volumes and sum up
#-Metos3DDiagnosticMonitor

# bgc parameter
-Metos3DParameterCount                    7
-Metos3DParameterValue                    0.02,2.0,0.5,30.0,0.67,0.5,0.858
```

Example: Fortran namelist file

- Fortran has the special namelist construct:

```
namelist/ namelist-group-name/ variable1-name ...
read(unit,nml=namelist-group-name)
```

- Namelist file:

```
! namelist for numerical simulation of the predator prey model
&model_parameters
npar = 6
alpha = 2.0
beta = 3.0
gamma = 1.0
delta = 3.0
lambda = 0.0
mu = 0.0
/
&spatial_parameters
kappa = 1.0d-3
```

Example: YAML-file

- “Yet Another Markup Language/YAML Ain’t Markup Language”
- Easily readable (better than XML).
- Nested list with pairs of properties and values.
- Example: Metos3D model description file:


```yaml
model:
    name:   N-DOP
    ny:     2
    nx:     52749
    nt:     2880
    y0:     2.17e+0,1.e-4
    yout:   N.petsc,DOP.petsc

parameter:
    nu:     7
    u0:     [0.04, 3.5, 0.3,  25.0, 0.4,  0.8,  0.78]
    ud:     [0.02, 2.0, 0.5,  30.0, 0.67, 0.5,  0.858]  # not used, info only
    lb:     [0.01, 1.0, 0.25, 15.0, 0.05, 0.25, 0.7]
    ub:     [0.05, 4.0, 1.0,  60.0, 0.95, 1.0,  1.5]
```

Input and output files with physical, biological etc. data

- When storing a variable or quantity, we also need the corresponding meta-data:
- Variable name.
- Most physical, biological quantities have units.
- They might be different, e.g., for length: m, km ...
- Dimension: 3-D variables, or 2-D on surfaces/layers
- Time-dependent?
- Spatial grid.

$\rightsquigarrow$ NetCDF data format and library.

Contents

NetCDF: Network Common Data Form

- Mostly used and standard data format in climate research/geoscience ...
- ... but not restricted to the field.
- File suffix `.nc`.
- Binary file.
- Contains data and metadata.
- Read- and writeable with libraries available in programming languages and in python, octave etc.
- Provider: *Unidata*: community of education and research institutions
- Webpage:

  `https://www.unidata.ucar.edu/software/netcdf/`

NetCDF: Design principles

Taken from `https://www.unidata.ucar.edu/software/netcdf/`:

- Self-Describing. A netCDF file includes information about the data it contains.
- Portable. A netCDF file can be accessed by computers with different ways of storing integers, characters, and floating-point numbers.
- Scalable. Small subsets of large datasets in various formats may be accessed efficiently through netCDF interfaces, even from remote servers.
- Appendable. Data may be appended to a properly structured netCDF file without copying the dataset or redefining its structure.
- Sharable. One writer and multiple readers may simultaneously access the same netCDF file.
- Archivable. Access to all earlier forms of netCDF data will be supported by current and future versions of the software.

Example: Content of a NetCDF file

- Meta information:

```
<class 'netCDF4._netCDF4.Dataset'>
root group (NETCDF3_64BIT_OFFSET data model, file format NETCDF3):
CDI: Climate Data Interface version 1.9.7.1 (http://mpimet.mpg.de/cdi)
Conventions: CF-1.6
history: Tue Nov 19 11:22:03 2019: cdo -f nc copy file1 file2.nc
institution: Max-Planck-Institute for Meteorology
CDO: Climate Data Operators version 1.9.7.1 (http://mpimet.mpg.de/cdo)
```

- Dimensions:

```
dimensions(sizes): time(516), lon(96), lat(48)
```

- Grid:

```
variables(dimensions):
float64 time(time), float64 lon(lon), float64 lat(lat)
```

- Actual data:

```
float32 varname(time,lat,lon) ...
```

Contents

Tools for post-processing, data analysis and visualization

- In general: every software with visualization functionality can be used, e.g., Python, Octave, R ...
- There are special programs designed by/for climate researchers:
  - Ferret: National Oceanic and Atmospheric Administration
    `https://ferret.pmel.noaa.gov/Ferret/`
  - Ncview: Scripps Institution of Oceanography
    `http://meteora.ucsd.edu/~pierce/ncview_home_page.html`
  - CDO (Climate Data Operators): MPI Hamburg
    `https://code.mpimet.mpg.de/projects/cdo`
- Often data analysis and visualization is combined.
- Different approaches: command line tools vs. GUIs
- On an HPC cluster with remote login, sometimes command line tools are helpful.

Example: working with the cdos

- Command line tools, available via `conda`.
- Working on NetCDF files: simple operations, statistics, generation of graphical output
- Store results in NetCDF file again.
- Example:

```
# select variable by name:
cdo selvar,var169 in.nc out.nc
# substract two fields:
cdo sub in1.nc in2.nc out.nc
# select years:
cdo selyear,1981/2010 in.nc out.nc
# average over time:
cdo timmean in.nc out.nc
# std deviation of annual means:
cdo timstd1 -yearmean in.nc out.nc
# plot as pdf:
cdo shaded,device=pdf,colour_min=violet,colour_max=red,colour_triad=cw in.nc out
```

What is important

- (Global) climate models have to take account the special, spherical geometry of the Earth.
- The physical model equations are transformed into this coordinate system.
- There is a big difference in the vertical and the horizontal scales of the computational domain.
- The standard grid of a climate model is a rectangular and regular, equally-spaced grid in horizontal direction.
- For the vertical coordinate, there exist different options, usually not equally-spaced.
- Parts of some models are computed with spectral methods that do not have a grid.
- The horizontal, vertical and temporal resolutions are interdependent.
- Data files are used to specify model and simulation parameters.
- NetCDF is the standard data format for in- and output data fields.
- There are several, specific data analysis and visualization tools designed for climate model output.