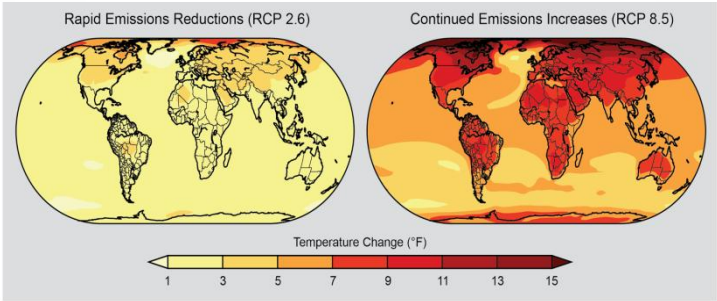
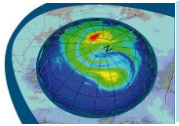
PRESENTATION OF THE NETCDF FILE FORMAT AND TOOLS FOR PROCESSING AND VISUALIZATION



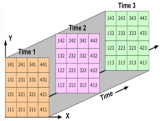
Presented by
Dr. Charlène GABA

The CODATA-RDA Research Data Science Advanced Workshop: Climate Data Sciences
Trieste, Italy. 19 August 2019

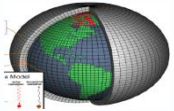
Outline



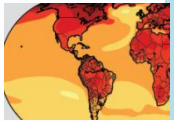
1 Background: Climate Modelling



2 netCDF format for climate data

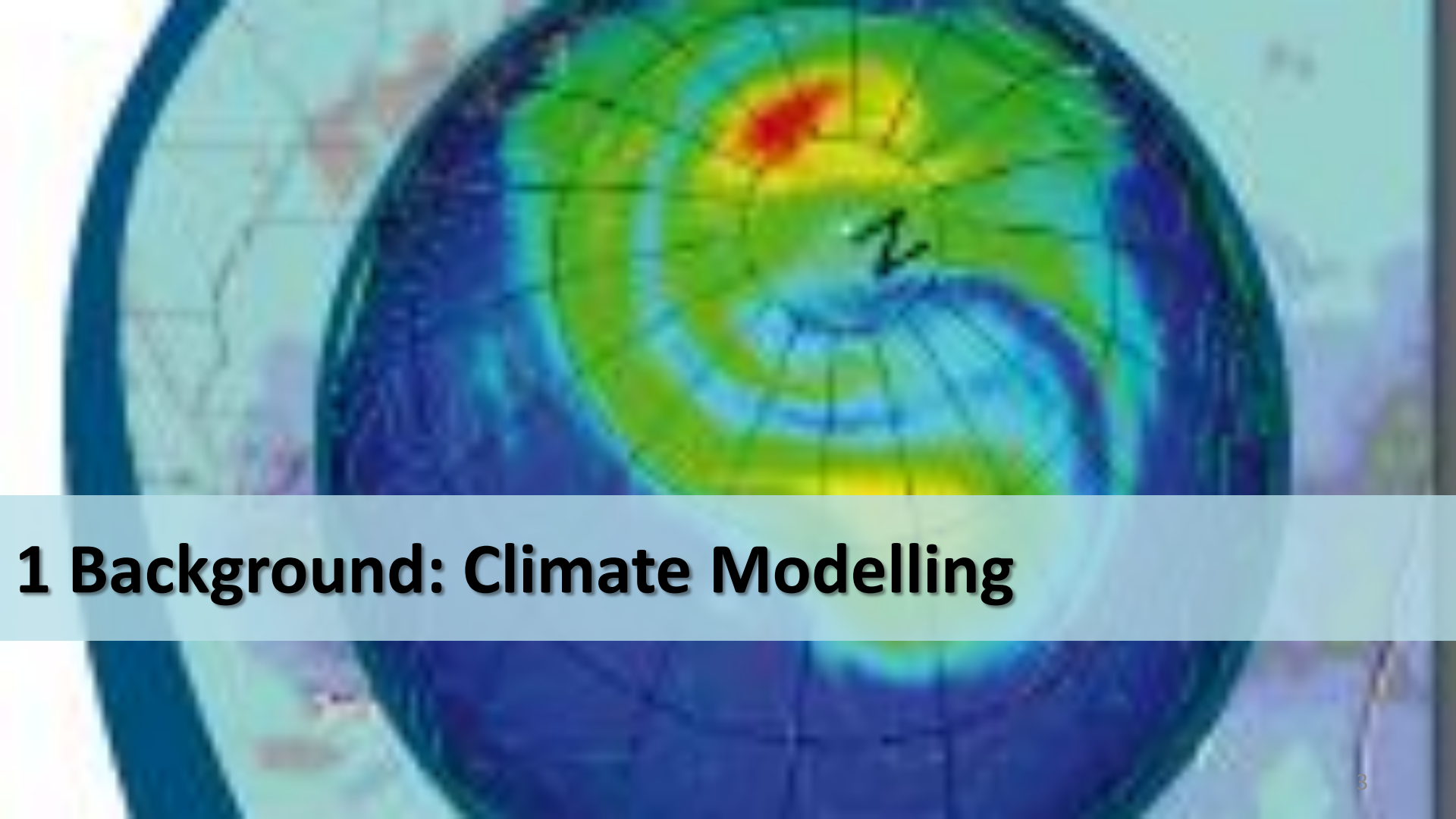


3 Analysis and Processing of netCDF data



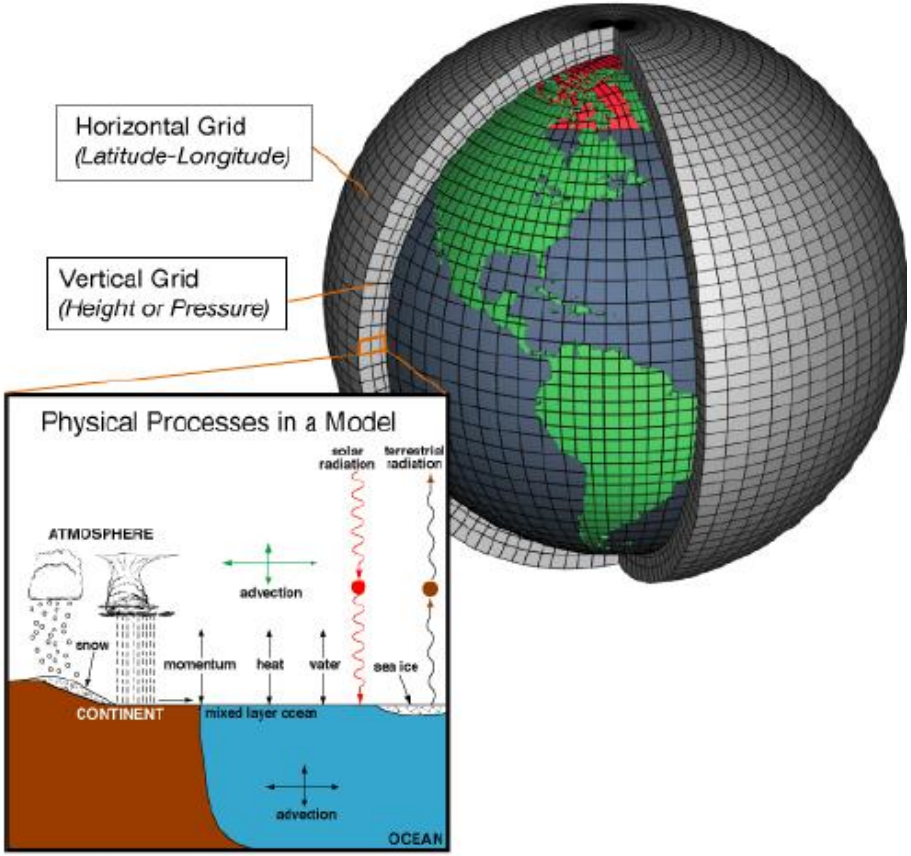
4 Visualisation of netCDF data

"We do not inherit the Earth from our Ancestors, we borrow it from our Children"



1 Background: Climate Modelling

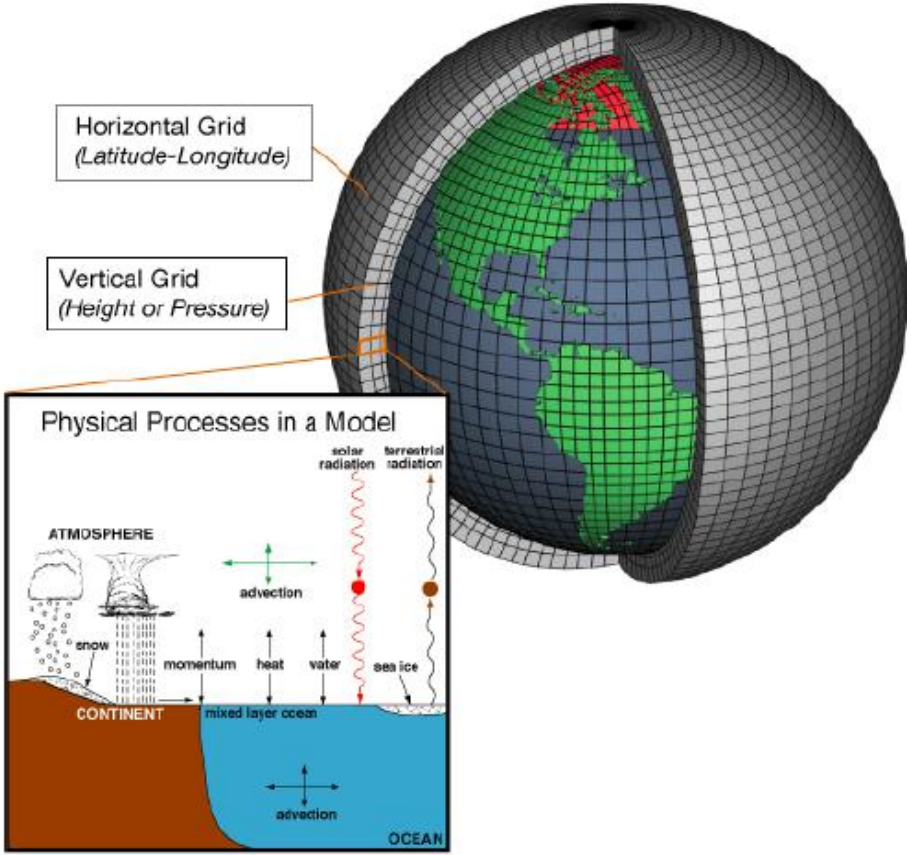
1- Background: Climate Modelling



https://upload.wikimedia.org/wikipedia/commons/7/73/AtmosphericModelSchematic.png

Special demands for data storage
- large data sets (100s of MByte per simulation year)
-data sets to be merged / split into subsets
- gridded data
many physical quantities → meta-data becomes of relevance

1- Background: Climate Modelling



https://upload.wikimedia.org/wikipedia/commons/7/73/AtmosphericModelSchematic.png

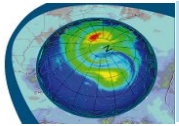
Classical ASCII data: <u>not</u> a suitable file format

- input / output relatively slow

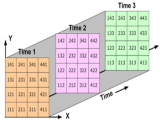
-storage of numerical data via characters inefficient

- data structure difficult to represent

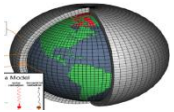
- handling of metadata difficult



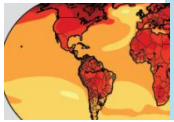
1 Background: Climate Modelling



2 netCDF format for climate data

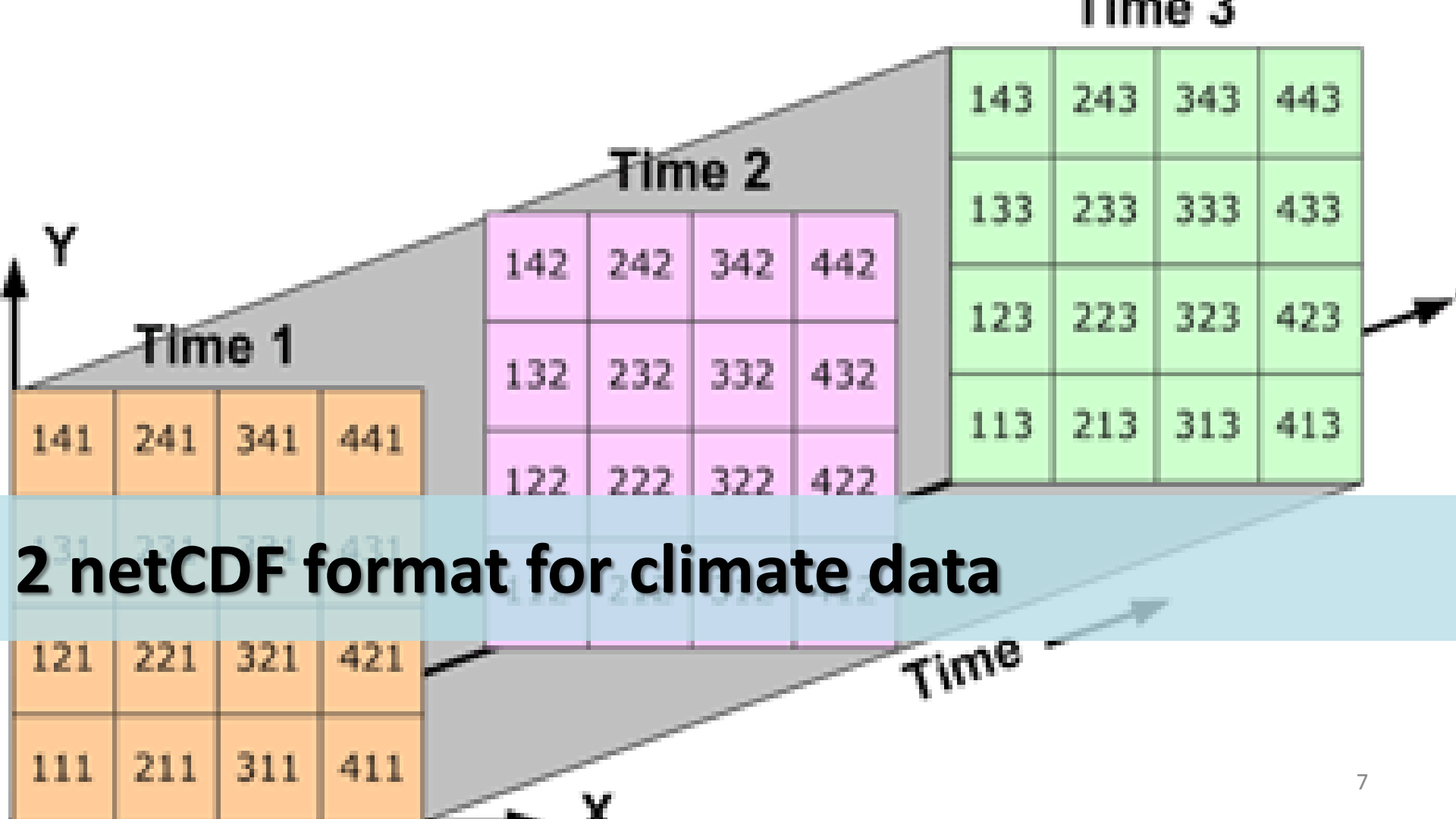


3 Analysis and Processing of netCDF data



4 Visualisation of netCDF data

“We do not inherit the Earth from our Ancestors, we borrow it from our Children”

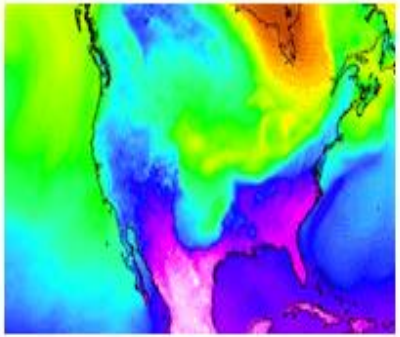


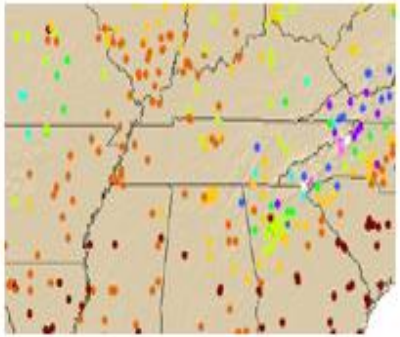
2 netCDF format for climate data

2- netCDF format for climate data

2-a) What are netCDF data?

Network Common Data Form (NetCDF) is a file format that stores multidimensional (variable) scientific data, such as temperature, humidity, pressure, wind speed and direction. Each of these variables can be displayed via a dimension (for example, time)





Examples of netCDF data: left (temperature); right (pressure at specific locations)

2- netCDF format for climate data

2-b) How to learn more about netCDF?

The first source of information about netCDF data is the Unidata community: https://www.unidata.ucar.edu/software/netcdf/

Unidata is a diverse community of education and research institutions with the common goal of sharing geoscience data and the tools to access and visualize that data.

2- netCDF format for climate data

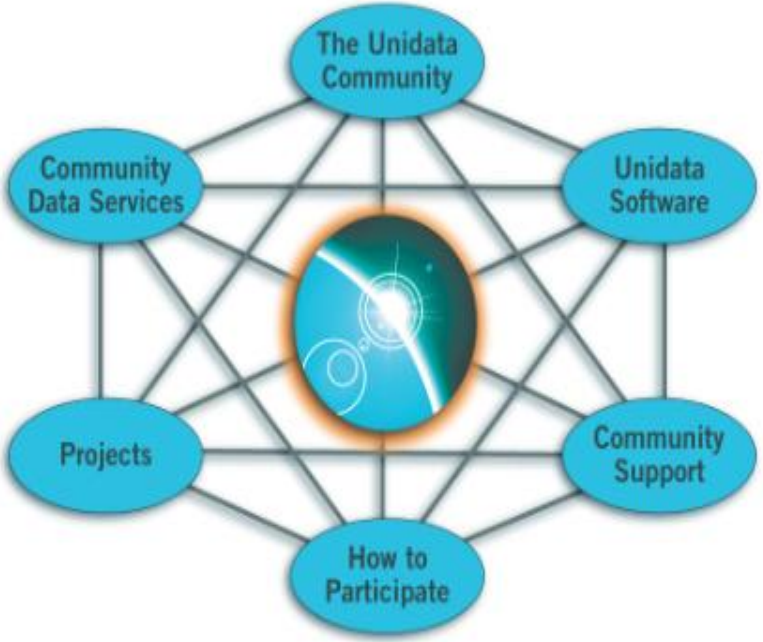


Image courtesy of UCAR/Unidata

For more than 30 years, Unidata has been providing data, software tools, and support to enhance Earth-system education and research. The Unidata Program Center in Boulder, Colorado is the nexus of program activities.

2- netCDF format for climate data

Unidata is primarily sponsored by the National Science Foundation (NSF) and managed by the University Corporation for Atmospheric Research(UCAR).

Several organizations and groups of scientists from different countries have adopted netCDF as the standard method for representing certain scientific data (https://www.unidata.ucar.edu/software/netcdf/conventions.html).

ICTP is one of the organizations using netCDF for archiving and accessing some of their data.

2- netCDF format for climate data

2-c) Presentation of netCDF data

According to Unidata:
“ NetCDF (network Common Data Form) is a set of interfaces for array-oriented data access and a freely distributed collection of data access libraries for C, Fortran, C++, Java, and other languages. The netCDF libraries support a machine-independent format for representing scientific data. Together, the interfaces, libraries, and format support the creation, access, and sharing of scientific data.”

2- netCDF format for climate data

In conclusion, NetCDF is more than just a file format. In the simple view, netCDF is a:

- File format
- Application programming interface (API)
- Data model
- Library implementing the API

NetCDF (Network Common Data Form) is a file format designed to support the creation of scientific data and the access to and sharing of such data. It is widely used among oceanographic and atmospheric communities to store variables such as temperature, pressure, wind speed and wave height.

2- netCDF format for climate data

NetCDF data is: (extension **.nc**)

- ***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***. A small subset of a large dataset may be accessed efficiently.

- ***Appendable***. Data may be appended to a properly structured netCDF file without copying the dataset or redefining its structure.

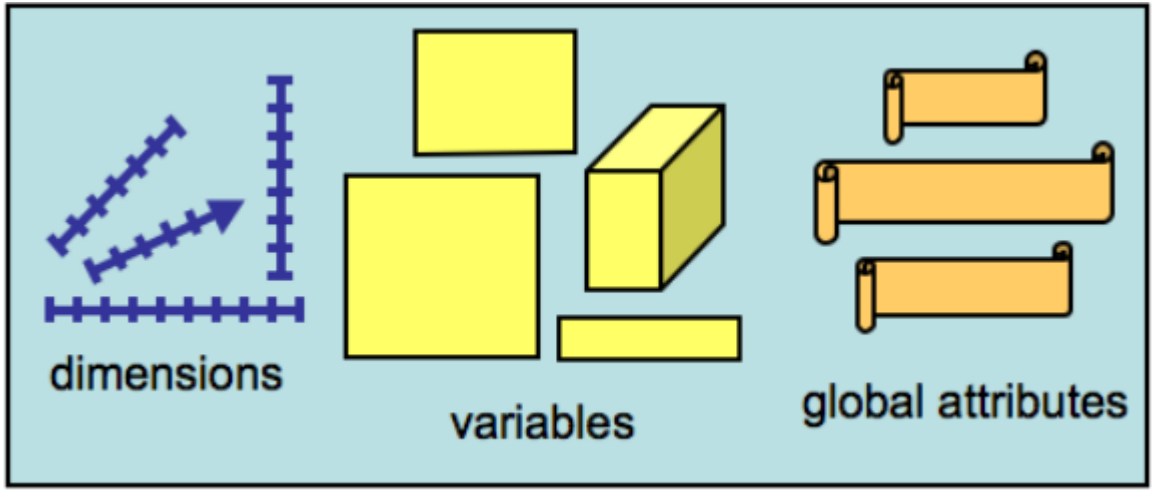
- ***Sharable***. One writer and multiple readers may simultaneously access the same netCDF file.

- ***Archivable***. Access to all earlier forms of netCDF data will be supported by current and future versions of the software.

2- netCDF format for climate data

2-d) The Structure of NetCDF files (based on the "Classic" format)

NetCDF files are containers for Dimensions, Variables, and Global Attributes

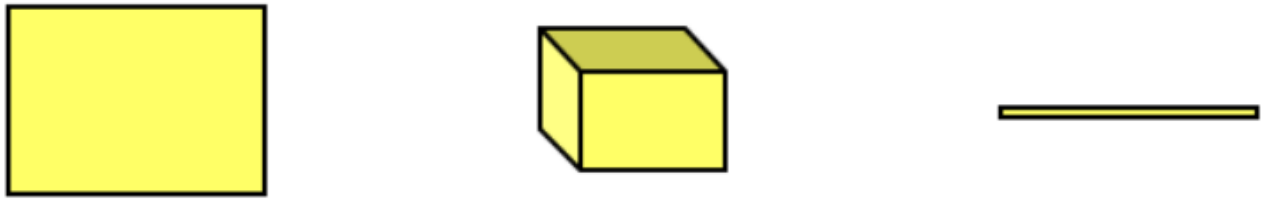


A netCDF file has a **path name** and possibly some **dimensions, variables, global** (file-level) **attributes**, and **data values** associated with the variables. Sometimes we refer to netCDF files more abstractly as datasets.

2- netCDF format for climate data

Variables

Variables hold data values. In the classic netCDF data model, a variable can hold a multidimensional array of values of the same type.



2- netCDF format for climate data

NetCDF Variables

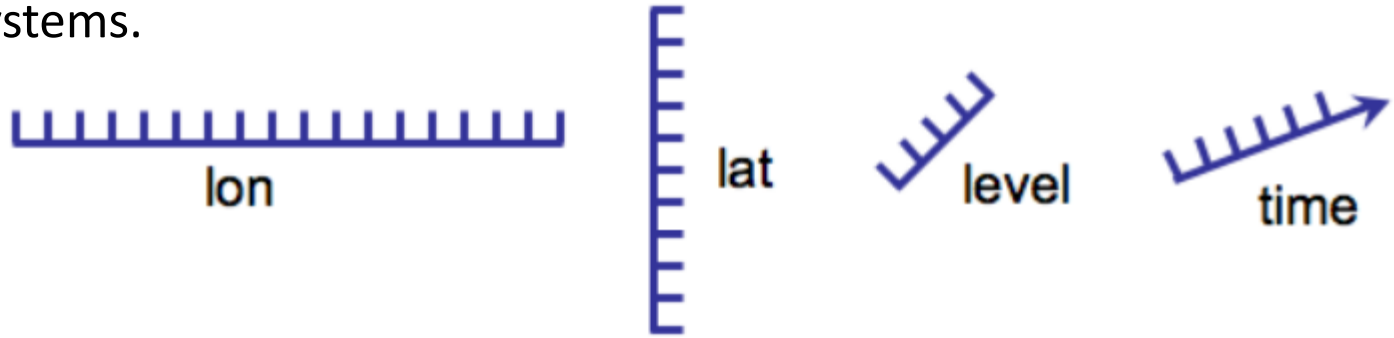
NetCDF Variables have:

- A **type**, e.g. char (text character), byte (8 bits) or float (32 bits)
- A **shape**, specified by a list of dimensions, e.g.:
    - 1 dimension: a 1-D (vector) variable, such as time
    - 2 dimensions: a 2-D (grid or matrix) variable, such as surface_pressure
- **Attributes** (optionally) – specifying properties such as long name and units.
- **Values** – the actual data values.

2- netCDF format for climate data

Dimensions

Dimensions are used to specify variable shapes, common grids, and coordinate systems.

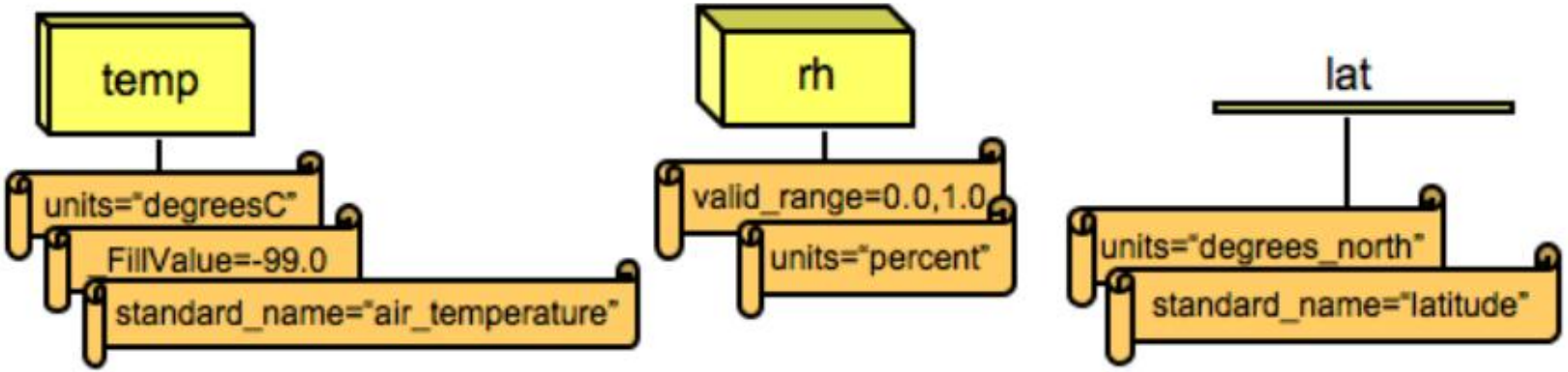


A dimension has a name and a length. Dimensions are used to define the shape of one or more variables in a netCDF file. In the classic netCDF data model, at most one dimension can have the unlimited length, which means variables can grow along that dimension. Record dimension is another term for an unlimited dimension.

2- netCDF format for climate data

Attributes

Attributes hold metadata (data about data). An attribute contains information about properties of a variable or dataset.



Attributes can be “global” (applying to the whole file) or “variable attributes” (applying only to a specified variable).

2- netCDF format for climate data

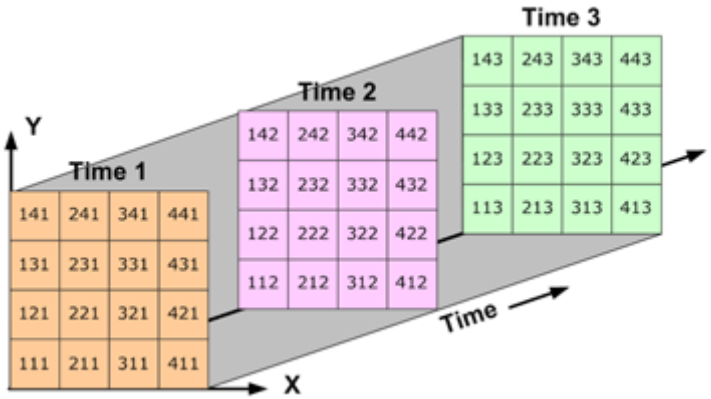
NetCDF data storage

The data in a NetCDF file is stored in table form.
For example, the variation of temperature over time at a location is stored as a one-dimensional array. The temperature above an area at a given time is stored as a two-dimensional array.

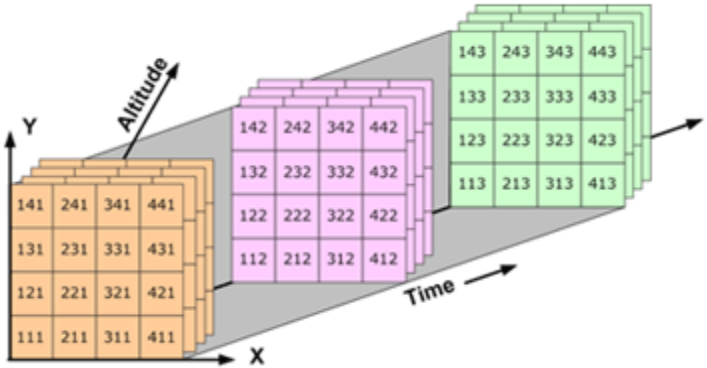
Three-dimensional (3D) data, such as the temperature over a region that varies over time, or four-dimensional (4D) (temperature over an area that varies over time and depending on the altitude) are stored as a series of two-dimensional arrays.

2- netCDF format for climate data

NetCDF data storage



Three-dimensional data: data over an area that varies over time.



Four-dimensional data: data over an area that varies over time and according to altitude.

2- netCDF format for climate data

An easier way to view NetCDF: CDL

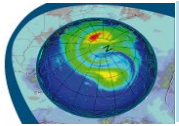
CDL (network Common Data form Language) is a human readable notation for netCDF objects and data.

```
netcdf filename {
dimensions:
        lat = 3 ;
        lon = 4 ;
        time = UNLIMITED ; // (2 currently)

variables:
        float lat(lat) ;
                lat:long_name = "Latitude" ;
                lat:units = "degrees_north" ;
        float lon(lon) ;
                lon:long_name = "Longitude" ;
                lon:units = "degrees_east" ;
        int time(time) ;
                time:long_name = "Time" ;
                time:units = "days since 1895-01-01" ;
                time:calendar = "gregorian" ;
        float rainfall(time, lat, lon) ;
                rainfall:long_name = "Precipitation" ;
                rainfall:units = "mm yr-1" ;
                rainfall:missing_value = -9999.f ;

// global attributes:
                :title = "Historical Climate Scenarios" ;
                :Conventions = "CF-1.0" ;

data:
 lat = 48.75, 48.25, 47.75;
 lon = -124.25, -123.75, -123.25, -122.75;
 time = 364, 730;
 rainfall =
   761, 1265, 2184, 1812, 1405, 688, 366, 269, 328, 455, 524, 877,
   1019, 714, 865, 697, 927, 926, 1452, 626, 275, 221, 196, 223;
}
```

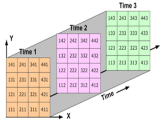
Coordinate variable

Variable attribute

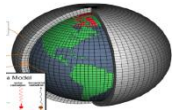
Global attribute



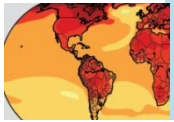
1 Background: Climate Modelling



2 netCDF format for climate data

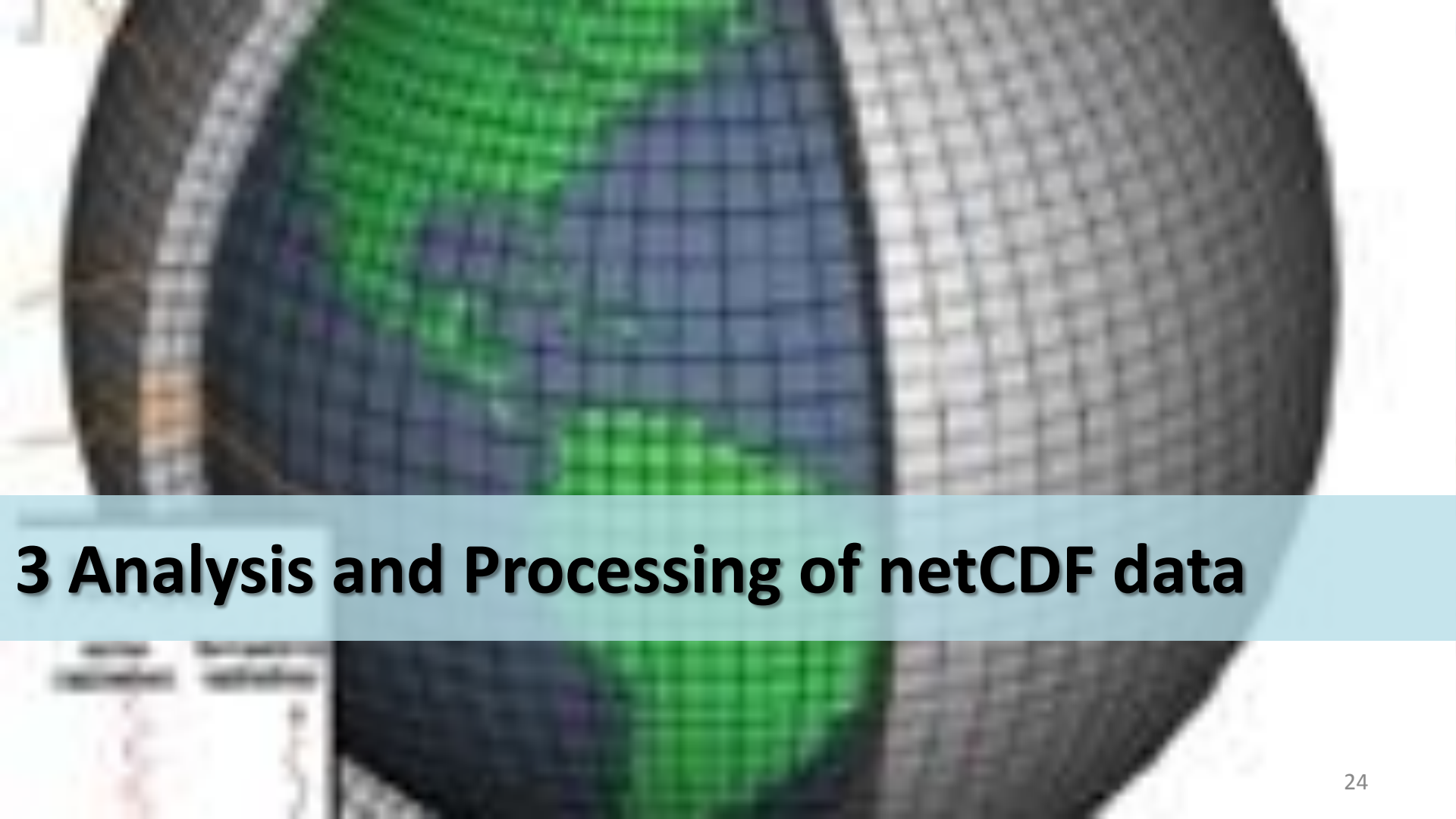


3 Analysis and Processing of netCDF data



4 Visualisation of netCDF data

“We do not inherit the Earth from our Ancestors, we borrow it from our Children”



3 Analysis and Processing of netCDF data

3- Analysis and Processing of netCDF data

Operating on a NetCDF file. When working with a netCDF file you can:

- Create a new file, given its path name and whether to overwrite or not.
- Open an existing file for access, given dataset name and read or write intent.
- Add dimensions, variables, or attributes.
- Close a file, writing to disk if required.
- Get the number of dimensions, variables or global attributes.
- Get the unlimited dimension, if present.

3- Analysis and Processing of netCDF data

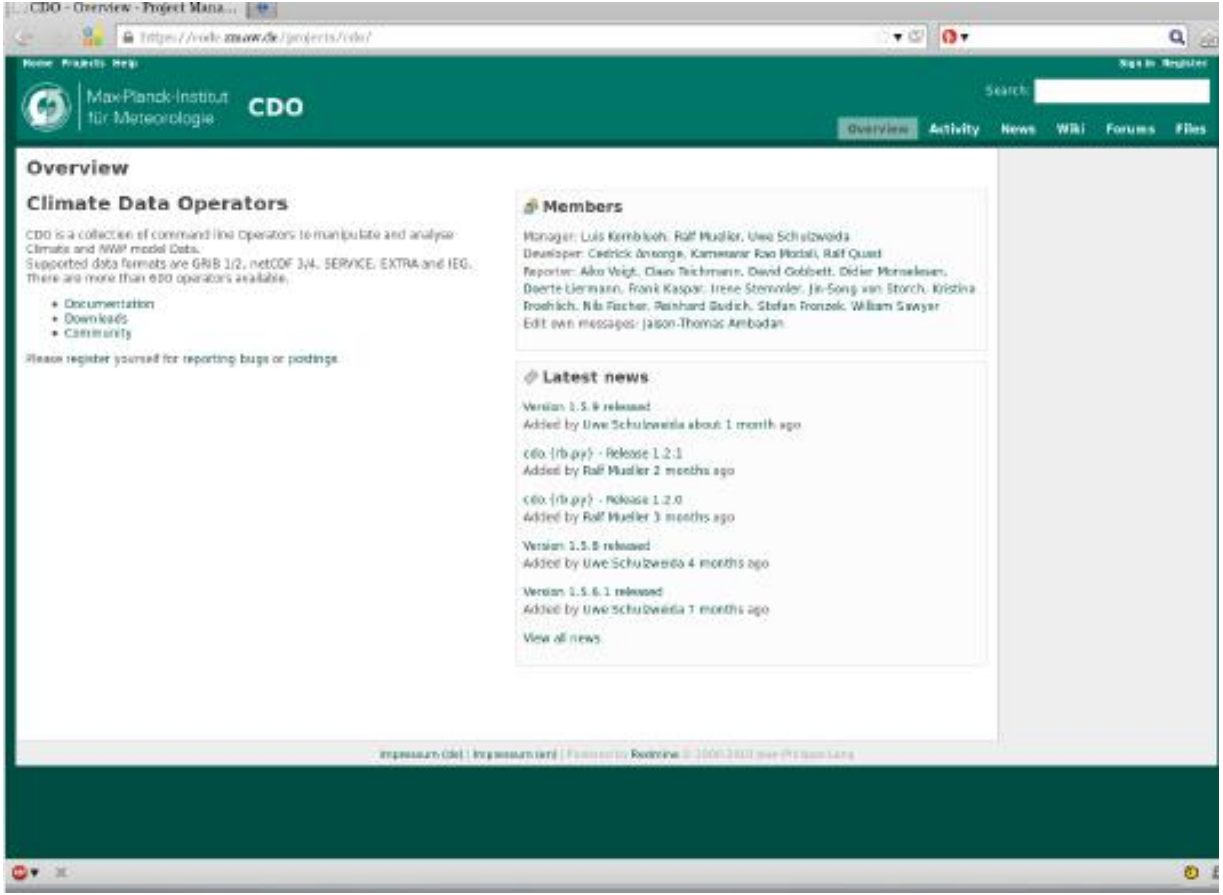
cdo stands for “Climate Data Operators”

It is an extremely useful tool for both meteorologist and oceanographers and for everyone who uses .grib or netcdf files.

Cdo developed at Max-Planck Institut fűr Meteorologie, and
It can be downloaded from --- hSps://code.zmaw.de/ projects/cdo/
In the same site you can find detailed documentation and usage examples.

Also contained in most Linux distribution software.

3- Analysis and Processing of netCDF data



3- Analysis and Processing of netCDF data

A single command with hundreds of operators

- CDO was inspired by NCO – providing a range of climate data-related operations through the command-line
- Designed to operate on netCDF3/4, GRIB1/2 primarily
- Much functionality can be used for any NetCDF/gridded data
- Very efficient for specific tasks
- Manages memory effectively

3- Analysis and Processing of netCDF data

CDO

- installation (ubuntu): `sudo apt-get install cdo`
- windows version (limited functionality): `https://code.zmaw.de/attachments/download/8378/cdo-1.6.4-win32.zip`
- comprehensive documentation of the CDO available at: `https://code.zmaw.de/projects/cdo/embedded/index.html`
- looks complex, but easy to use with basic understanding of CDO's functionality

3- Analysis and Processing of netCDF data

CDO: a command-line tool

CDO is a single command-line tool. It is used as follows:

```
$ cdo  <operator>  [options]  <files>
```

3- Analysis and Processing of netCDF data

Types of CDO Operator

File information and file operations
Selection and Comparison
Modification of metadata
Arithmetic operations
Statistical analysis
Regression and Interpolation
Vector and spectral Transformations
Formatted I/O
Climate indices

3- Analysis and Processing of netCDF data

Many, many operators

CDO breaks down its functionality into individual operators.

There are over 650 at the time of writing this. For example:

`showstdname` Show standard names

`sellonlatbox` Select a longitude/latitude box

`setmissval` Set a new missing value

`monadd` Add monthly time series

`zonstd` Zonal standard deviation

`eca_hd` Heating degree days per time period

3- Analysis and Processing of netCDF data

CDO – information operators

Quantities (variables) contained in a file:

```
$ cdo pardes INIOM_PD_echam5_main_mm_3901-4000_climatological_mean.nc

130  t            temperature [K]
131  u            u-velocity [m/s]
132  v            v-velocity [m/s]
133  q            specific humidity [kg/kg]
135  omega        vertical velocity [Pa/s]
155  sd           divergence [1/s]
156  geopoth      geopotential height [m]
157  rhumidity    relative humidity
 85  tradl        net LW radiation 200mb [W/m^2]
 86  sradl        net SW radiation 200mb [W/m^2]
                   .
                   .
                   .
```

3- Analysis and Processing of netCDF data

File Information

```
$ cdo infov ifile
```

This is an example result of a dataset with one 2D variable over 3 time steps:

```
-1 : Date Time           Varname Level Size Miss : Minimum Mean   Maximum
1  : 1987-01-31 12:00:00 SST     0     2048 1361 : 232.77  266.65 305.31
2  : 1987-02-28 12:00:00 SST     0     2048 1361 : 233.64  267.11 307.15
3  : 1987-03-31 12:00:00 SST     0     2048 1361 : 225.31  267.52 307.67
```

3- Analysis and Processing of netCDF data

File Operations The full list of file operations covers:

Copying and concatenating

Merging fields/times

Splitting fields by variable/level/grid/time Some examples follow

3- Analysis and Processing of netCDF data

To copy a file and convert the output to NetCDF:

```
$ cdo -f nc copy ifile ofile.nc
```

To merge all files along the time axis:

```
$ cdo mergetime ifile1 ifile2 ifile3 ofile
```

3- Analysis and Processing of netCDF data

Selection

Data can be selected by a number of different methods including:

Variable code, name or attribute

Levels

Time

Spatial region (latitude/longitude)

3- Analysis and Processing of netCDF data

Selection

To select variables with ids: "SSTK" and "CI":

```
$ cdo selname,SSTK,CI infile.nc outfile.nc
```

To select a lat/lon bounding box:

```
$ cdo sellonlatbox,120,-90,20,-20 infile.nc
outfile.nc
```

To select a date/time range:

```
$ cdo seldate,2014-12-12T12:00:00, 2015-01-
31T18:00:00 infile.nc outfile.nc
```

3- Analysis and Processing of netCDF data

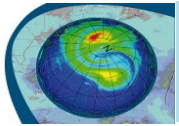
Combining operators: example 1

For example, we can replace:

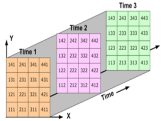
```
$ cdo timavg ifile1 tmp1
$ cdo dayavg ifile2 tmp2
$ cdo sub tmp2 tmp1 ofile
$ rm tmp1 tmp2
```

With...

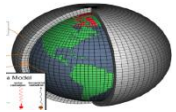
```
$ cdo sub -dayavg ifile2 -timavg ifile1 ofile
```



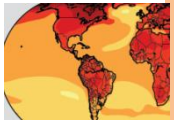
1 Background: Climate Modelling



2 netCDF format for climate data

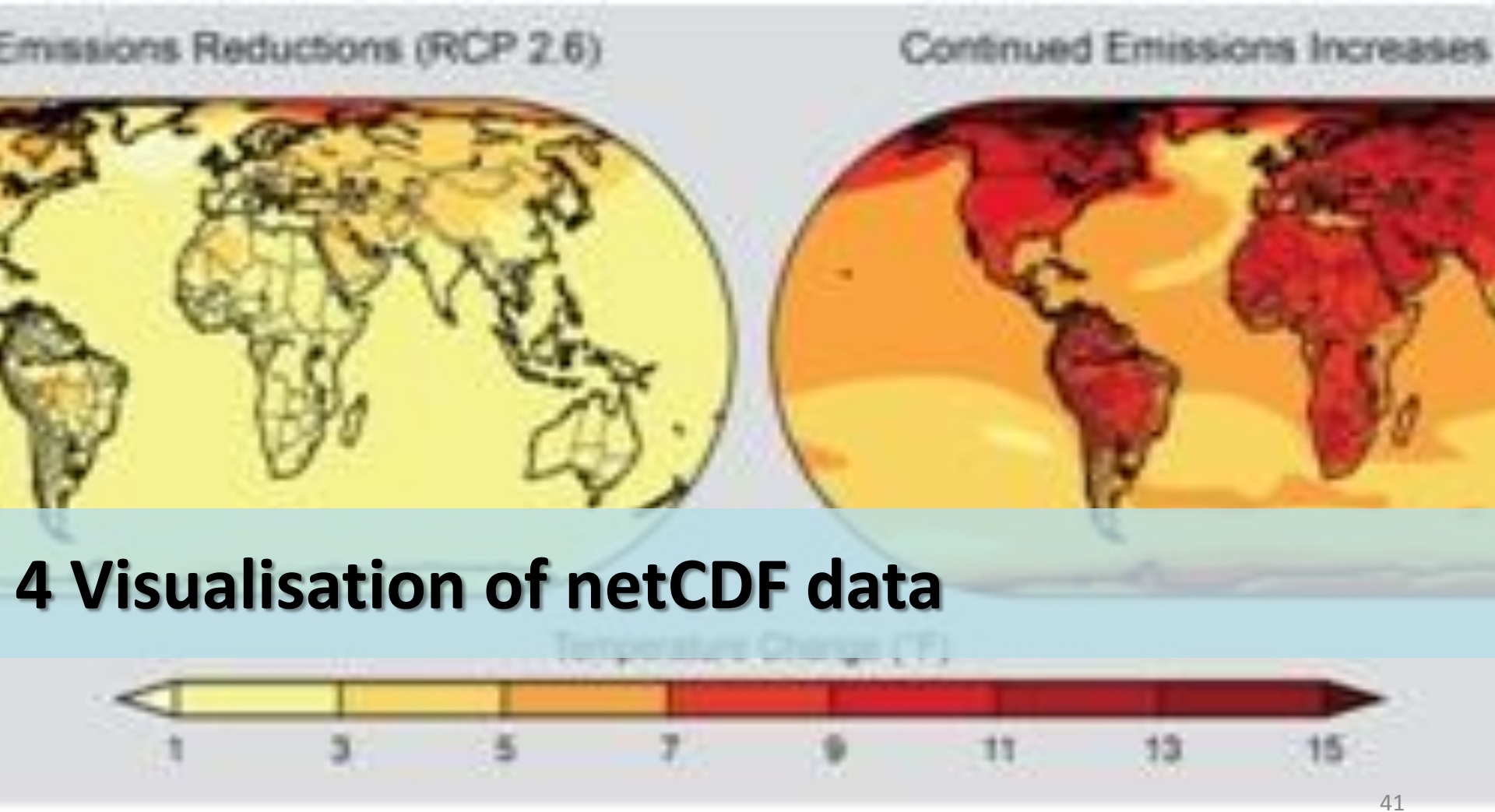


3 Analysis and Processing of netCDF data



4 Visualisation of netCDF data

“We do not inherit the Earth from our Ancestors, we borrow it from our Children”

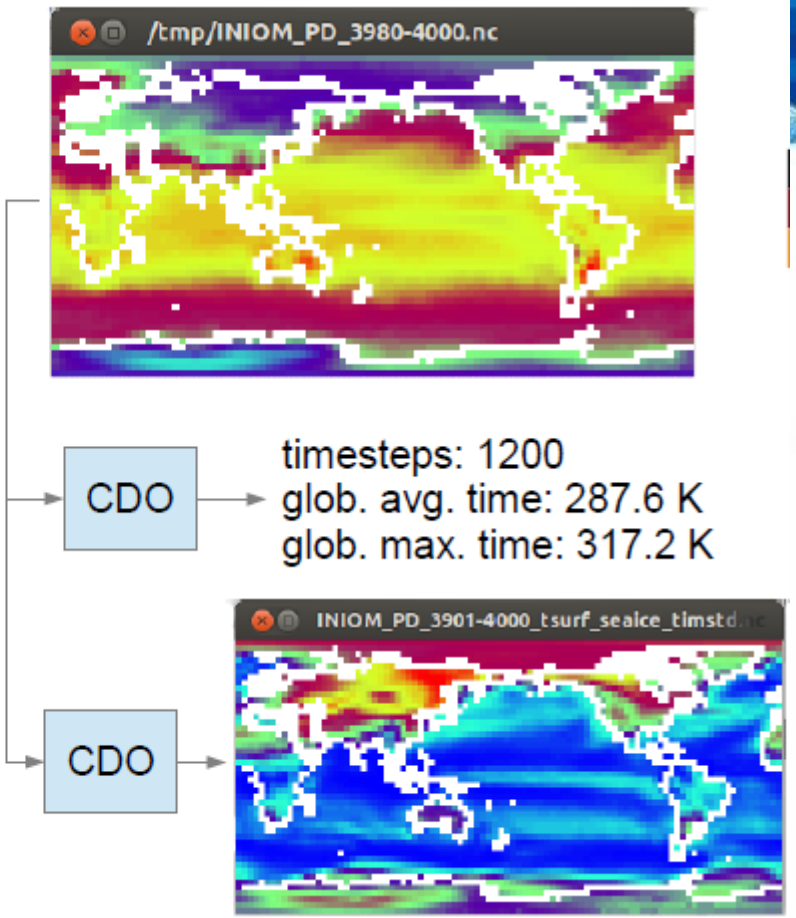


4 Visualisation of netCDF data

4- Visualisation of netCDF data

Special tools necessary for data analysis, plotting, inspection, ...

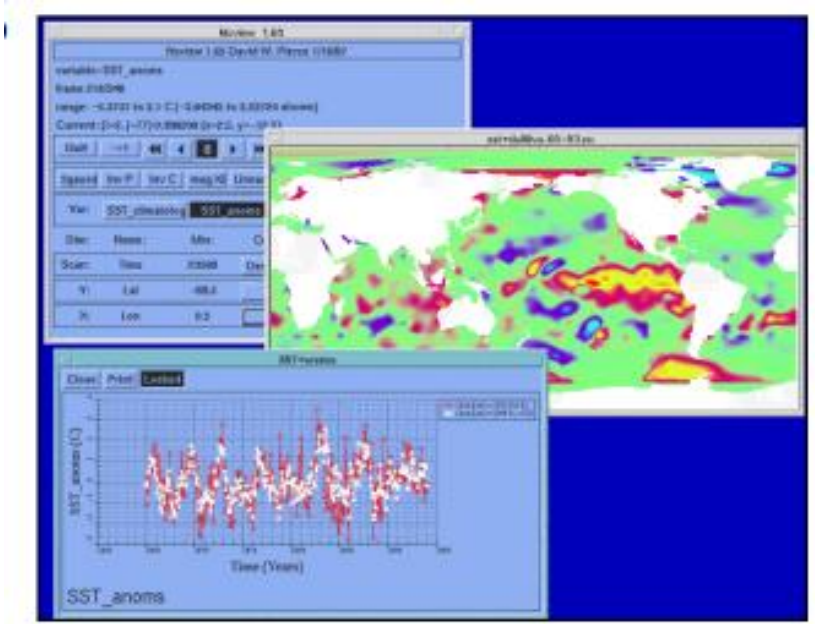
- UNIX has many tools available
  - viewers: ncview/panoply
  - translators: ncdump/ncgen
  - analysis tools
- Windows
  - Some tools available, e.g. cygwin
  - But: reduced functionality



4- Visualisation of netCDF data

ncview:

- Visual browser for netCDF
- View simple animations along various dimensions
- Change look of graphs/maps



4- Visualisation of netCDF data

Ncview – getting started

- To run ncview type:

```
$ ncview <data_file>
```

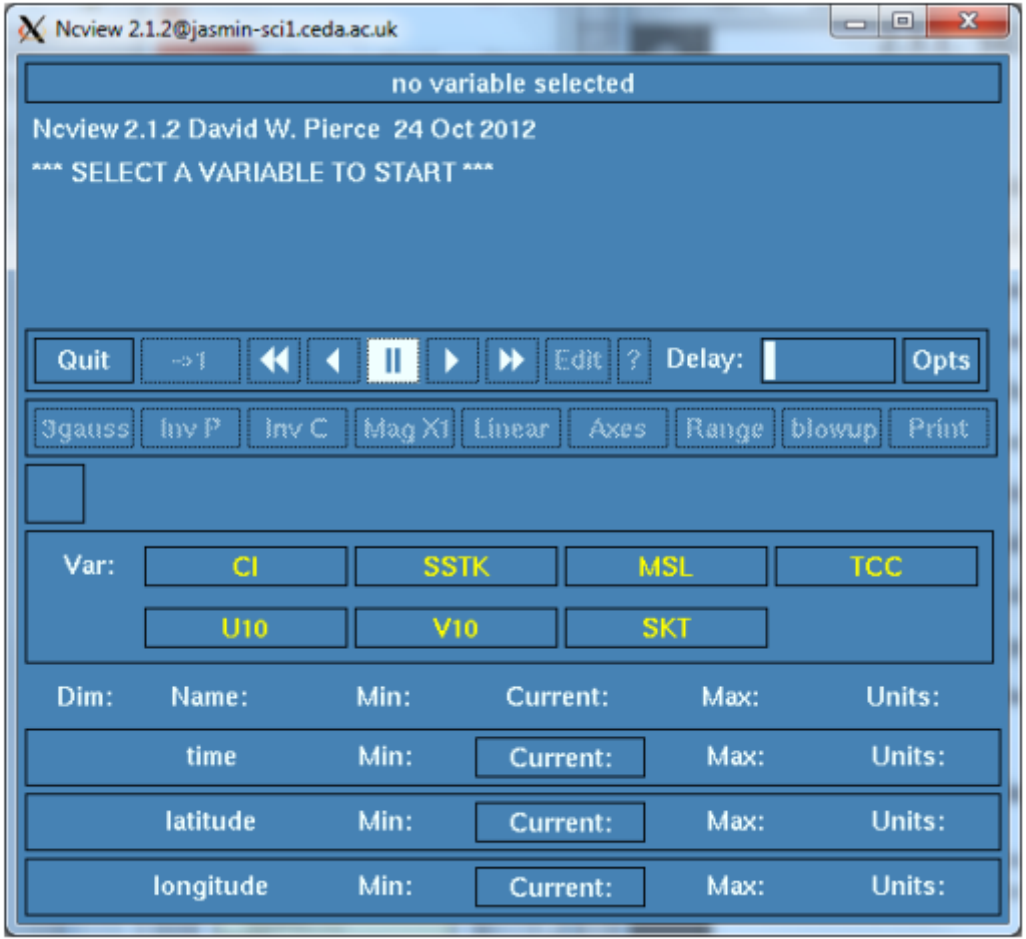
- E.g.:

```
$ ncview example_data/ggas2014121200_00-18.nc &
```

4- Visualisation of netCDF data

You are presented with:

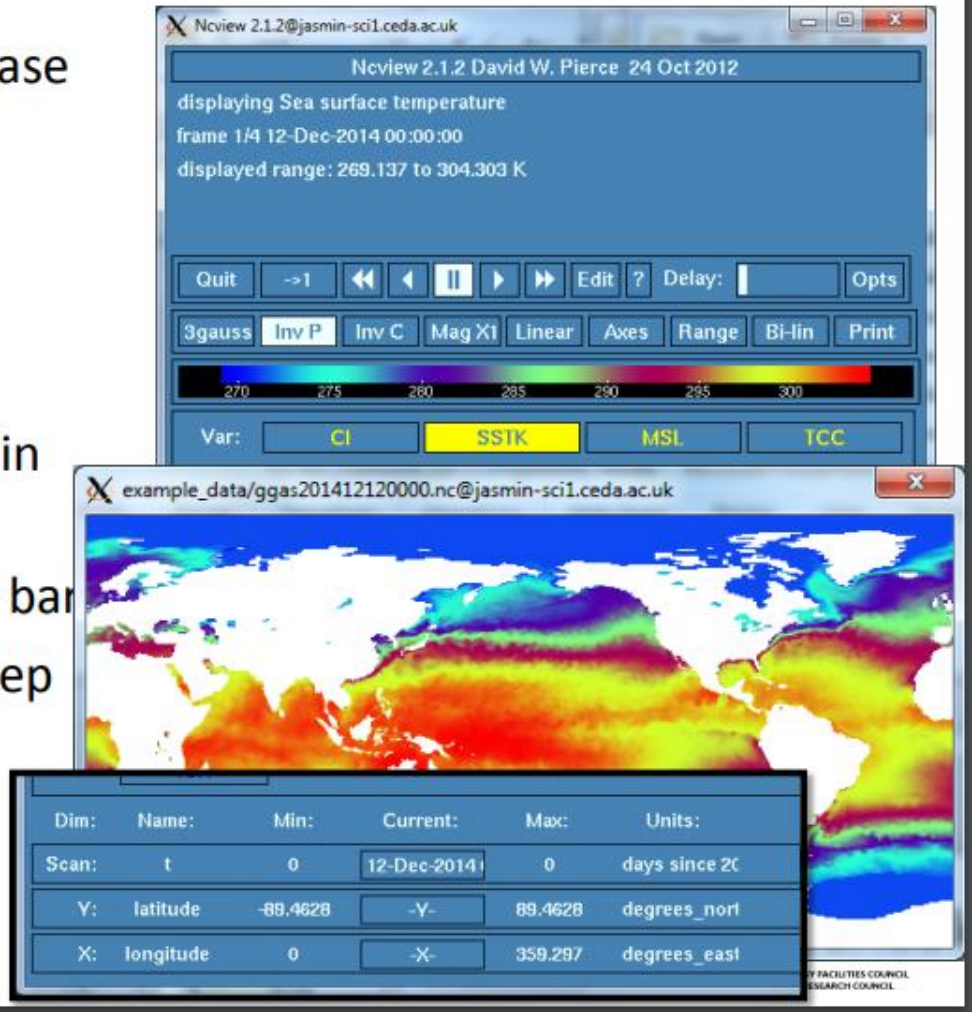
- An information panel
- A control panel for visualization
- A variable selection panel
- Details of dimensions for the selected variable



Select variable: in this case we have selected Sea Surface Temperature ("SSTK").

Ncview:

- displays information in the top panel
- Populates the colour bar
- Plots the first time step of the variable
- Displays information about dimensions



4- Visualisation of netCDF data

ncview – animate

Use the animation panel to click through time steps or animate the plot.

Press the  button to run the animation. Adjust the "Delay" to slow it down.

4- Visualisation of netCDF data

ncview – other options

The plotting panel has a range of useful functions that are best investigated by manually testing each.

E.g.: **3gauss**: click to select a different colour scale

Inv P: Invert the plot.

Inv C: Invert the colour scale

Mag X1: right/left click to zoom in/out

Axes: modify axes of plot

Print: print to a file or printer

4- Visualisation of netCDF data

Further Reading

Cdo,
Ncview
ArcGIS
Ferret
Excel
NCO (netcdf operator)
IDL
ncBrowse
Matlab
R
CDAT
GrADS Graphic Analysis and Display System
IDV (Integrated data Viewer)

CDO Home page:
https://code.zmaw.de/projects/cdo

ncview:
http://meteora.ucsd.edu/~pierce/ncview_home_page.html

Acknowledgments

The slides were prepared with information mainly retrieved from :

-ESRI website: https://pro.arcgis.com

-Unidata website https://www.unidata.ucar.edu/software/netcdf/

-Some online resources

- Slides of Dr. rer. nat. Christian Stepanek (AWI- Germany)
- Slides of Graziano Giuliani - ICTP ESP


DANKE FÜR IHRE AUFMERKSAMKEIT
THANK YOU FOR YOUR ATTENTION
MERCI POUR VOTRE ATTENTION
GRACIAS POR SU ATENCIÓN