National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

# (CF-)NetCDF: conceptual summary

## Including `netcdf4python` summary & CF Data Tools recap

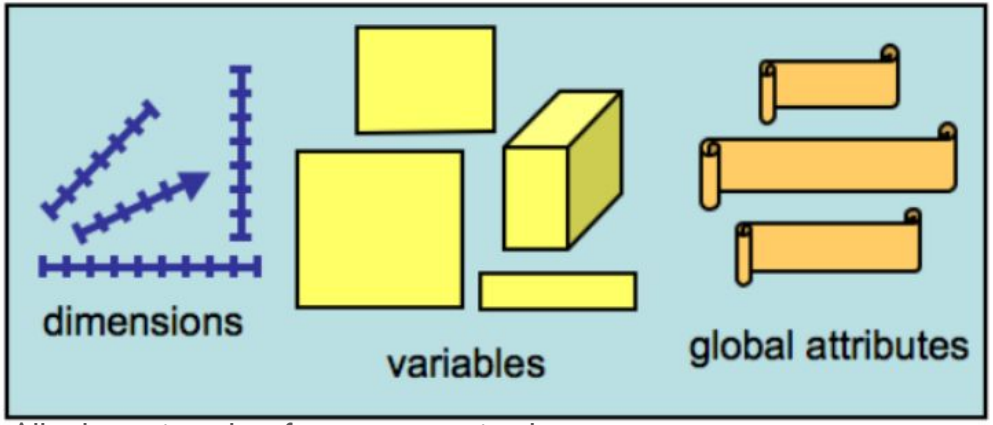
Sadie Bartholomew
Computational Scientist / Senior RSE
NCAS-CMS & Reading Met. Dept.
NCAS Intro. to Scientific Computing, 2025-11-20

# netCDF: a file (data) format and a data model

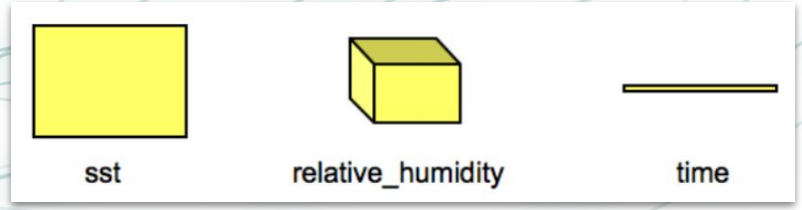
- **Binary file format (.nc extension)** adopted currently as very popular standard for exchange & storage of earth science data and data in some other domains
  - developed by UCAR's Unidata project: UCAR netCDF homepage, including documentation, release & support details, a tutorial, FAQs & more: www.unidata.ucar.edu/software/netcdf/
- **open(-source), so free to use** and to install tools to work with it
- **self-describing** (metadata categorises each data array) **and flexible, but that means it requires interpretation on its own**, hence a metadata standard to control use of metadata is, in practice, required: step in the CF (Metadata) Conventions
- **NetCDF files + CF Metadata Conventions = CF-netCDF**
  - flexible self-describing storage for array-based geoscientific data
  - plus standardised metadata to facilitate comparison & processing

# netCDF: centers around 3 core components

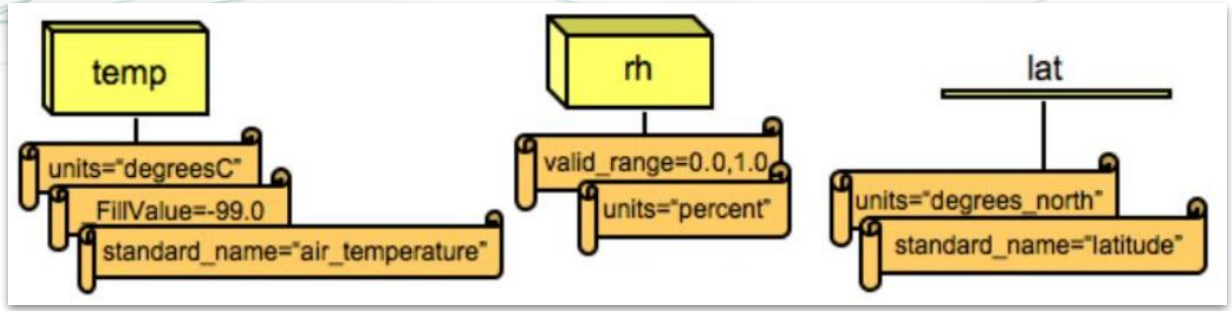
- netCDF dataset contains **dimensions, variables, and attributes**, which all have both a name & an ID




Examples could be:

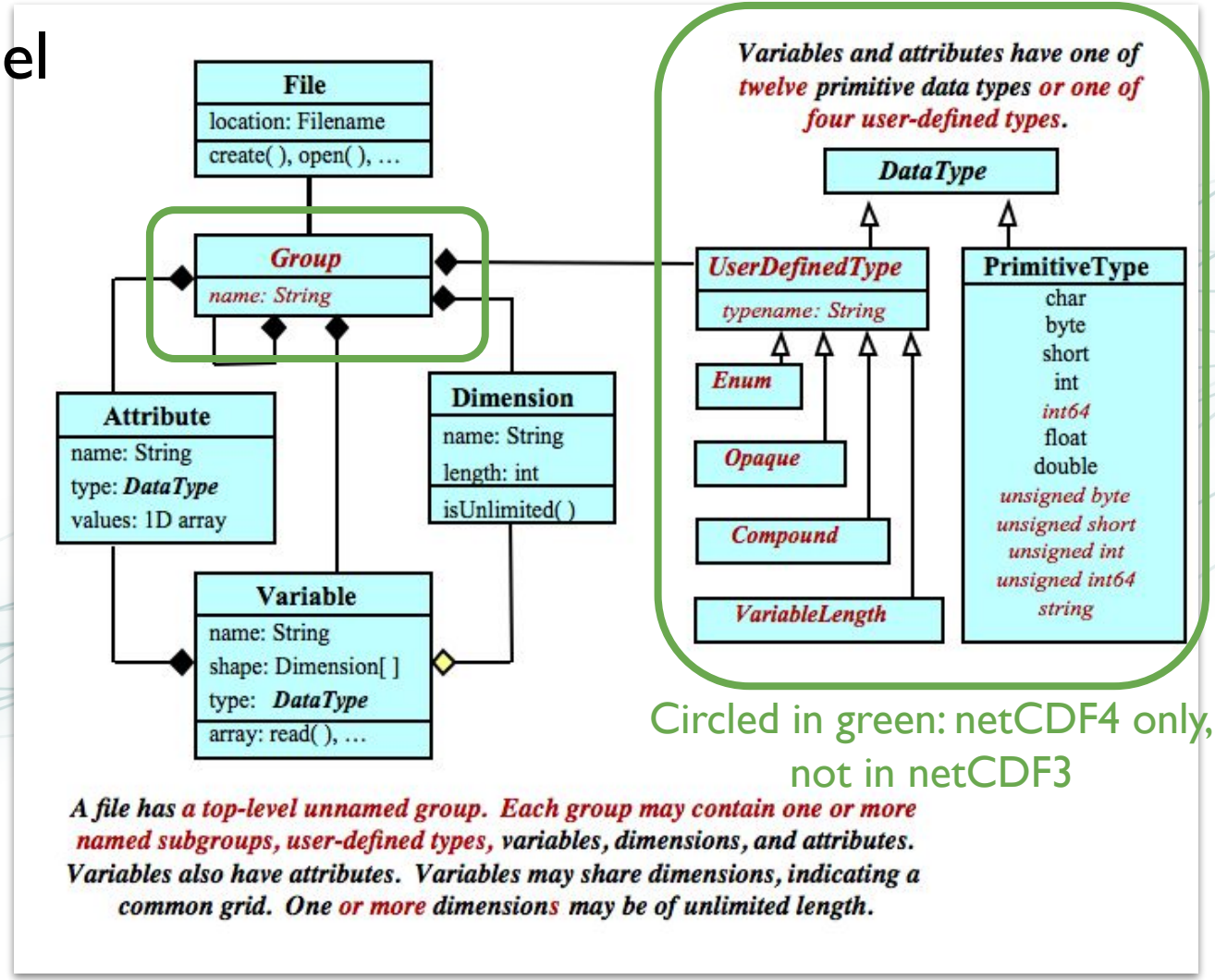

All schematics taken from: presentation by Dr. Charlène Gava (see: https://indico.ictp.it/event/8847/session/71/contribution/246/material/slides/0.pdf) via (I believe, from memory!) older UNIDATA netCDF documentation




University of Reading

# netCDF data model & netCDF 3 vs 4

- netCDF3 dataset contains **dimensions, variables, and attributes**, which all have both a name and an ID
- netCDF4 also adds groups and user-defined types
- these components can be used together to capture the meaning of data and relations among data fields in an array-oriented dataset




Circled in green: netCDF4 only, not in netCDF3

University of Reading

# `netcdf4python`: lowest level Python interface

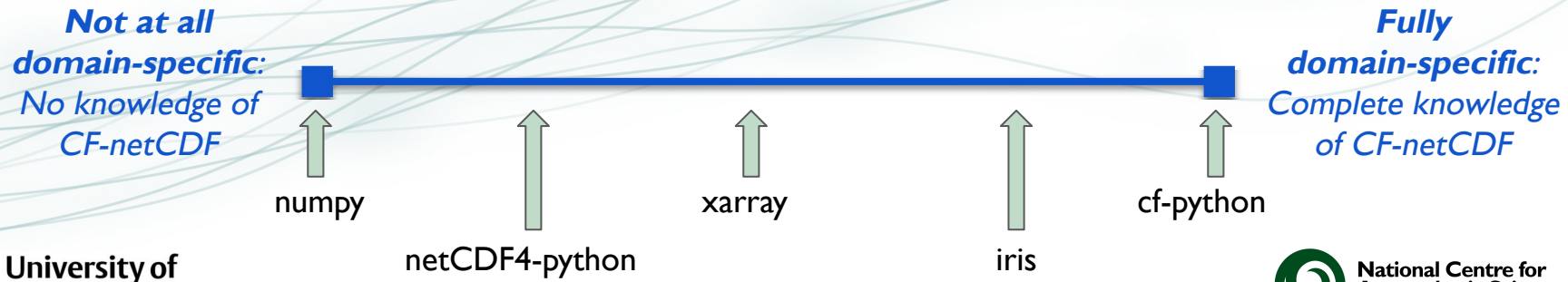
- If you want to work with netCDF data directly, without a higher-level interface such as xarray or cf-python, you should use `netcdf4python` (or another Unidata netCDF interface in another language e.g. C, also available). You can work with attributes, variables and dimensions directly, and create new ones etc.
- See: https://unidata.github.io/netcdf4-python/ (documentation) and/or https://github.com/Unidata/netcdf4-python (repo/codebase)
- Example code:

```
>>> # Import the library and read in a netCDF datasat
>>> from netCDF4 import Dataset
>>> rootgrp = Dataset("test.nc", "w", format="NETCDF4")
>>> print(rootgrp.data_model)
NETCDF4
>>> # Create some new dimensions!
>>> level = rootgrp.createDimension("level", None)
>>> time = rootgrp.createDimension("time", None)
>>> lat = rootgrp.createDimension("lat", 73)
```

# Python tools for netCDF: from general to domain-specific

How the various Python tools you can use (not an exhaustive list!) to work with netCDF compare:

- **NumPy**: data arrays and operations on these only
- **netcdf4-python**: a tool for working with netCDF without CF Conventions awareness
- **xarray (and cf_xarray)**: a general tool which extends numpy (the fundamental package for scientific computing in Python) to notably introduce labelling for dimensions, coordinates and attributes. The cf_xrray tool adds some CF Conventions awareness, but not much!
- **Iris**: developed by the Met Office, similar functionality to CF Data Tools but less CF support
- **CF Data Tools (cf-python and cf-plot):**



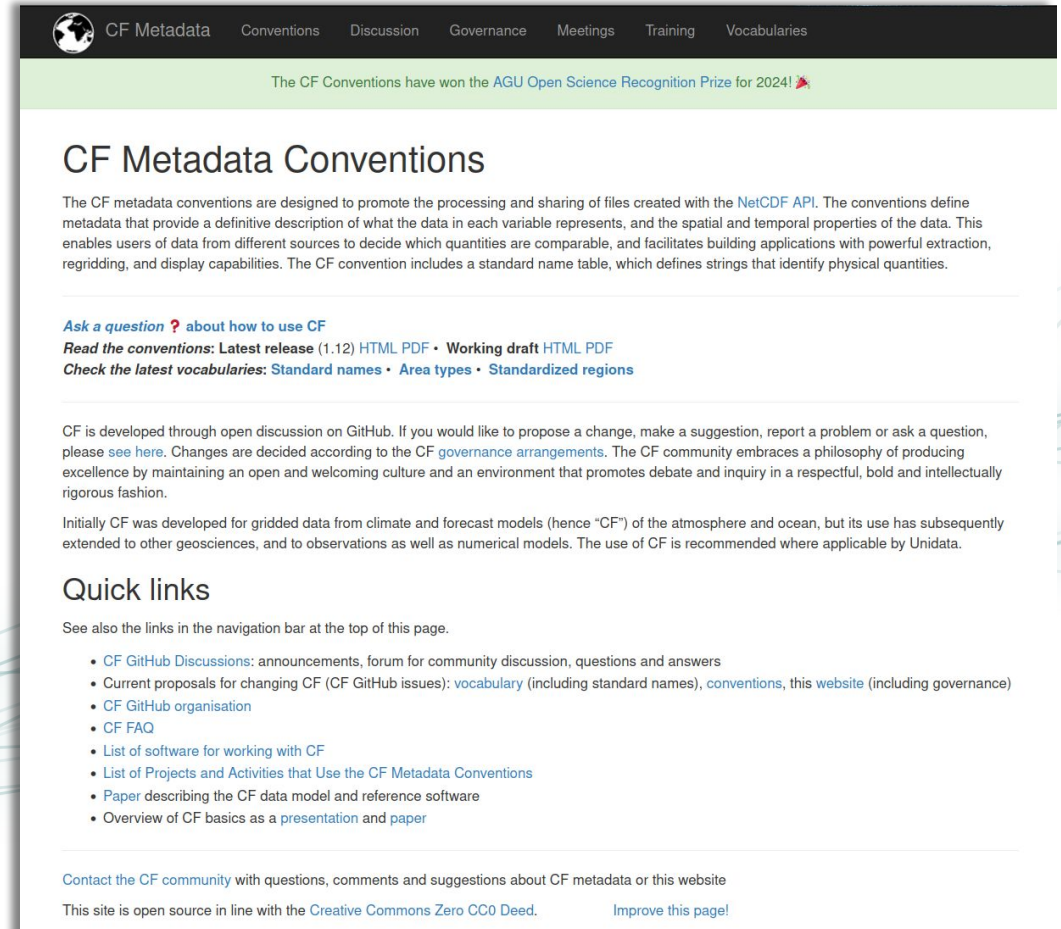
# ‘CF’ refers to the CF Conventions

‘CF’ is in the name for all of the tools because they center on supporting the **CF Conventions**, a metadata standard used throughout the geoscience e.g. Meteorology and atmospheric science communities that makes it easier to share, use and understand our datasets.

- The ‘C’ and ‘F’ from the CF Conventions stands for **Climate and Forecast**
- Some terminology: **NetCDF files + CF Metadata Conventions = CF-netCDF**
- Though cf-python and cf-plot etc. can work with datasets that aren’t (very, even at all) CF-compliant, they are designed to work best for CF-compliant data (not just netCDF but other formats e.g. Met Office PP and fields files)

- The idea is that they can use this standardised metadata intelligently so that they can tell what you want to do. Simple example: if you square your data array values, you probably want the units to become the squared version - so cf-python changes them as such automatically for you

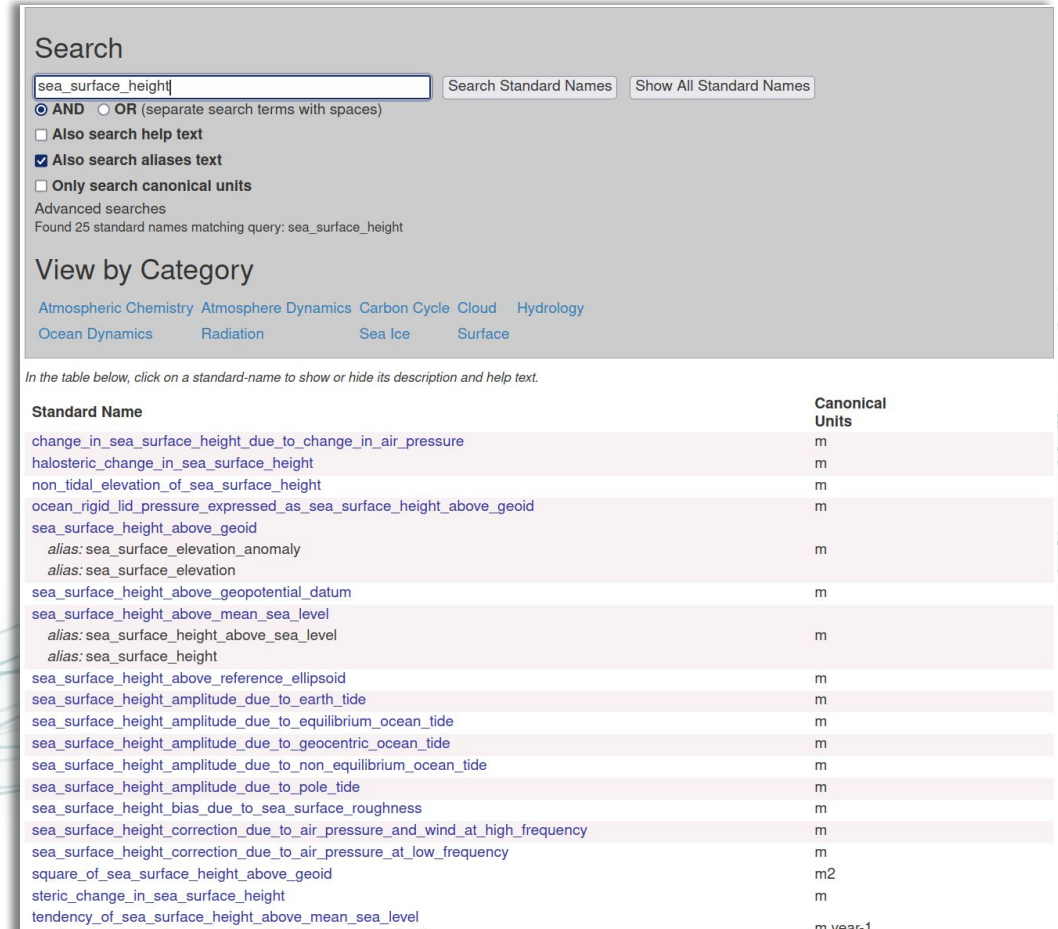
# The CF Conventions: a community metadata standard




- The CF Conventions provide a standardised way to describe and encode multi-dimensional geospatial data, which is crucial for sharing, comparing, and analysing data across different platforms, models, and research groups.
- netCDF datasets are flexibly self-describing so, without a standard, concepts such as variable names and geometry encoding would need interpretation as to precise meaning
- See https://cfconventions.org/ for e.g. the canonical latest document and list of standard names

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

# The CF Conventions: a community metadata standard



Search

sea_surface_height | Search Standard Names | Show All Standard Names

AND OR (separate search terms with spaces)

Also search help text

Also search aliases text

Only search canonical units

Advanced searches

Found 25 standard names matching query: sea_surface_height

View by Category

Atmospheric Chemistry | Atmosphere Dynamics | Carbon Cycle | Cloud | Hydrology | Ocean Dynamics | Radiation | Sea Ice | Surface

*In the table below, click on a standard-name to show or hide its description and help text.*

| Standard Name | Canonical Units |
|---|---|
| change_in_sea_surface_height_due_to_change_in_air_pressure | m |
| halosteric_change_in_sea_surface_height | m |
| non_tidal_elevation_of_sea_surface_height | m |
| ocean_rigid_lid_pressure_expressed_as_sea_surface_height_above_geoid | m |
| sea_surface_height_above_geoid<br>*alias:* sea_surface_elevation_anomaly<br>*alias:* sea_surface_elevation | m |
| sea_surface_height_above_geopotential_datum | m |
| sea_surface_height_above_mean_sea_level<br>*alias:* sea_surface_height_above_sea_level<br>*alias:* sea_surface_height | m |
| sea_surface_height_above_reference_ellipsoid | m |
| sea_surface_height_amplitude_due_to_earth_tide | m |
| sea_surface_height_amplitude_due_to_equilibrium_ocean_tide | m |
| sea_surface_height_amplitude_due_to_geocentric_ocean_tide | m |
| sea_surface_height_amplitude_due_to_non_equilibrium_ocean_tide | m |
| sea_surface_height_amplitude_due_to_pole_tide | m |
| sea_surface_height_bias_due_to_sea_surface_roughness | m |
| sea_surface_height_correction_due_to_air_pressure_and_wind_at_high_frequency | m |
| sea_surface_height_correction_due_to_air_pressure_at_low_frequency | m |
| square_of_sea_surface_height_above_geoid | m2 |
| steric_change_in_sea_surface_height | m |
| tendency_of_sea_surface_height_above_mean_sea_level | m year-1 |

- A key case of how the CF Conventions help us all to work together (and alone) is the **standard names**.
- Example: if you created or edited data describing sea surface height, there is ambiguity in what that exactly means - e.g. is it a change, amplitude or correction, and what constitutes the starting point for 'height', what effect is it encoding, etc.?
- *Note the CF Conventions are community-driven* and anyone can request changes or updates e.g. new standard names - please contribute/get involved if useful!

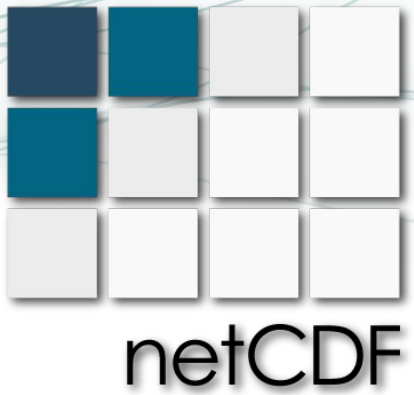
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

# NCAS CF Data Tools: aimed at netCDF & MO data users

Ultimately, cf-python and cf-plot (and friends) are **targeted/designed for use by anyone who wants to work with data in the format of netCDF or its human-readable CDL form, or Met Office file formats i.e. PP files & fields/UM files, however CF Compliant they may be.**

Given the focus on the CF Conventions, they are especially aimed at those using datasets related to Climate and Forecasting data (model, satellite, observational, etc.) for atmosphere, surface & ocean, etc. Notable clients:

- researchers (at any level, including students)
- Research Software Engineers
- data center workers
- data producers/managers
- and other technical/research professionals





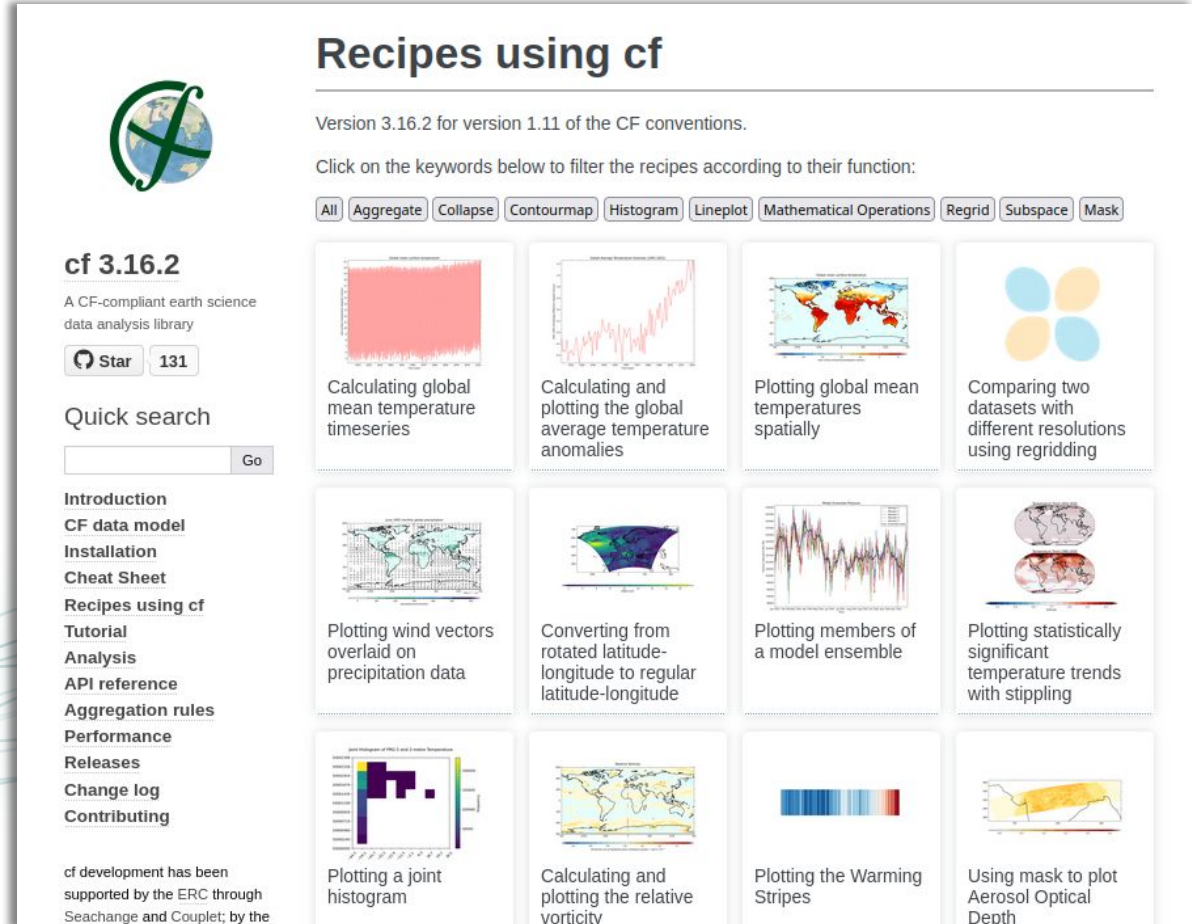

# CF Data Tools (cf-python & cf-plot): what we covered

In the CF Data Tools components of the course, we looked at 6 sections, each with a teaching and a practical/exercise Notebook, covering the following topics, working with example netCDF (and PP) data:

1. **Reading and inspecting datasets**
2. **Writing and editing datasets**
3. **Reducing down datasets by subspacing and collapsing**
4. **Visualising datasets**
5. **Example data analysis of datasets**
6. **Regridding (gridded) datasets**

There is also a cheat sheet which you may find useful:
https://ncas-cms.github.io/cf-python/cheat_sheet.html or in the course repo at
https://github.com/ncasuk/ncas-isc/blob/main/python-data/cheat_sheets/cfpython_cheatsheet.pdf

# cf-python & cf-plot: ‘recipes’ illustrating capability




In the cf-python documentation, we have a page listing created code recipes which demonstrate some possible applications of cf-python and cf-plot, see:

https://ncas-cms.github.io/cf-python/recipes/index.html

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL