About the new ‘s2dv_cube’

This document contains a description of the new ‘s2dv_cube’ object and the motivation of the development.

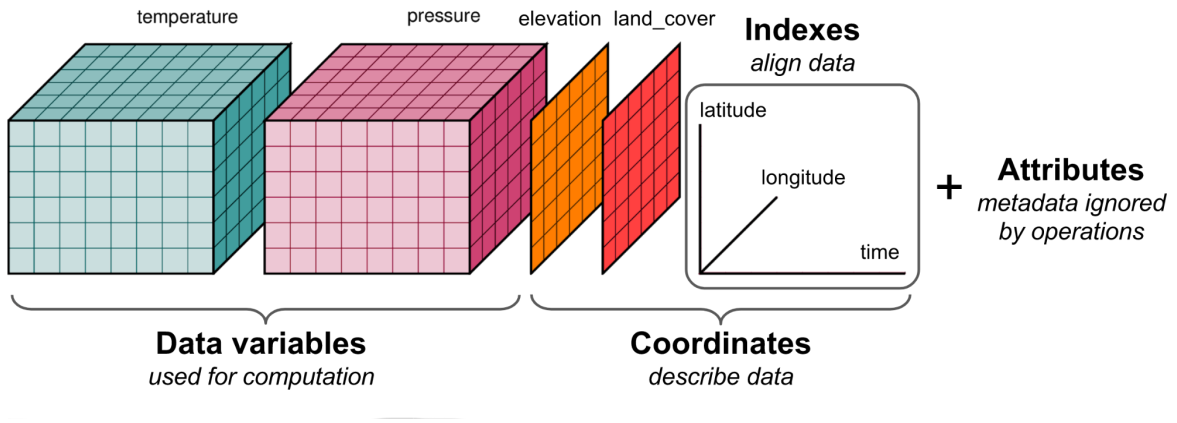
Barcelona, March 2023

Overview

An ‘s2dv_cube’ is an object class that is mainly a list of named elements to keep data and metadata in a single object. One of its elements, “data”, is a multi-dimensional array with named dimensions containing the values of a variable. The rest of the elements are metadata such as spatial coordinates.[1]

Until July 2022, there was a discussion about ‘s2dv_cube’ object structure due to different reasons. As a result of the discussion some requirements of the object were set:

1) To standardise its structure to make it more consistent with dimensions and attributes.
2) To be compatible with the startR package.[2]
3) To have a similar structure as a DataArray from the xarray python package.



[3]

[1] https://gmd.copernicus.org/articles/15/6115/2022/gmd-15-6115-2022.html
[2] https://earth.bsc.es/gitlab/external/cstools/-/issues/94
[3] Example image of a multidimensional data array from python package xarray https://docs.xarray.dev/en/stable/index.html

New structure

```
$data: [array with named dimensions]
$dims: [named vector with dimensions]
$coords
  # (same elements as dim(data))
  $dataset: [named indexes vector of dataset]
  $member: [named indexes vector of members]
  $sdate: [named vector of character string with sdates]
  $ftime: [named indexes vector of ftime]
  $lat: [named vector of latitudes]
  $lon: [named vector of longitudes]
$attrs
  $Variable
    $metadata [list of all the variables information]
    $varName [character string of the variable name]
  $Datasets [information of Datasets (name, initialization dates, ...)]
  $source_files [data path]
  $when [date when data has been loaded]
  $Dates: [array with named time dimensions with Dates of POSIXct class]
  $load_parameters: [input parameters of Load/Start function]

* In CST_* functions, if they need to use the values of lat/lon/Dates/etc.,
find the name in a pre-defined list. E.g.,
s2dv:::.KnownLonNames() has several possible options for longitude
** When `$Dates` is modified by functions SelectPeriodOnData/Dates() the
original dates will be stored as attributes in element `$Dates`.
```

Old structure

```
List of 10
 $ data: num [1, 1:6, 1:3, 1:31, 1:4, 1:4] 1.20e-08 ...
 $ lon : num [1:4(1d)] 6 7 8 9
 $ lat : num [1:4(1d)] 47 46 45 44
 $ Variable
      ..$ varName : chr "prlr"
      ..$ level
 $ Datasets
 $ Dates
      ..$ start : POSIXct[1:93], format: "2011-03-01" "2011-03-02" ...
      ..$ end   : POSIXct[1:93], format: "2011-03-02" "2011-03-03" ...
 $ when         : POSIXct[1:1], format: "2022-10-24 16:53:30"
 $ source_files : "/exp/ecmwf/system5c3s/prlr_s0-24h/prlr_20101101.nc"
 $ not_found_files
 $ load_parameters:
      ..$ var   : chr "prlr"
      [...]
 - attr(*, "class")= chr "s2dv_cube"
```