



**Barcelona  
Supercomputing  
Center**

*Centro Nacional de Supercomputación*

# R user meeting

An-Chi Ho, Eva Rifà, Victòria Agudetse

contributor: Jaume

07/09/2023

# Agenda

1. Ice-breaker: Brainstorming for the name of the new package
2. News
  - General R
  - s2dv
  - startR
  - CSTools
  - CSIndicators
  - SUNSET
3. Presentation: package “config” [Jaume]
4. Q&A

# Ice-breaker



**Barcelona  
Supercomputing  
Center**

Centro Nacional de Supercomputación

# Name for new package?

Analysis of Spatial Data: <http://lib.stat.cmu.edu/R/CRAN/web/views/Spatial.html>

List of all the relevant spatial analysis/plotting R packages

## Combination of...

s2d	plot
clim	viz
CS	vis
	map

## Acronyms...

**ASAP**: Atmospheric Science Array Plotting

**PACS**: Plotting Array of Climate Science

**PACT**: Plotting Array Ciencia Tierra

**PACTS**: Plotting dAta ClimaTe Science

**PICS/picsR**: Plotting In Climate Science (in R)

**MESA**: Mapping Earth Science Array

**CLIP**: Climate Plotting (cliplot and clipr are taken)

And new function names? E.g., PlotEquiMap → ???

# General R



**Barcelona  
Supercomputing  
Center**

Centro Nacional de Supercomputación

# Issue template

We created the template for startR, s2dv, CStools, CSIndicators, ClimProjDiags, SUNSET GitLab. It reminds you of the essential things for an issue. Feel free to write the issue in your style but please be aware of the important things mentioned in the template.

s2dv new issue: <https://earth.bsc.es/gitlab/es/s2dv/-/issues/new>

# New plotting package

- Move all the plotting-related functions there, but with new names to avoid confusion in use.
- Include shapefile function as well.
- Scorecard (?)
- New colorbar function for discrete values.

GitLab project: <https://earth.bsc.es/gitlab/es/climplot2> (Name TBD)

Function list: <https://earth.bsc.es/gitlab/es/climplot2/-/issues/1>

*(Note that the URLs will be changed when we decide the official name!)*

# New ColorBar function for discrete values

```
ColorBarDiscrete(  
  brks = 0:4,  
  cols = c("green1", "green2", "green3", "green4"),  
  vertical = F, labels = paste0('lev ', 1:4),  
  label_scale = 1.5, extra_margin = c(0.5,2,0.5,2)  
)
```

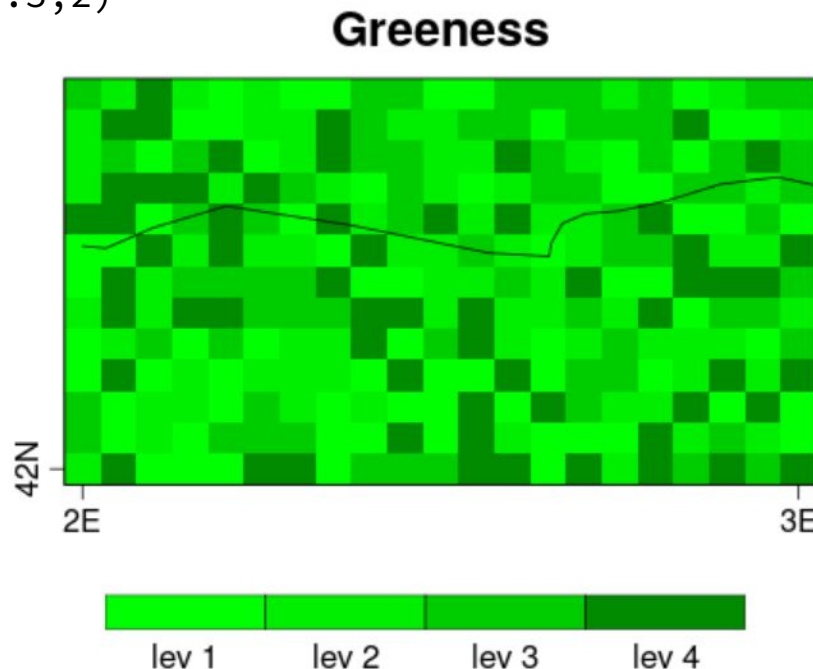
TODO:

- Integrate it with plotting functions
- Add space between boxes

status: In main, need further improvement

function:

<https://earth.bsc.es/gitlab/es/climplot2/-/blob/main/R/ColorBarDiscrete.R>





# shp\_mask developments

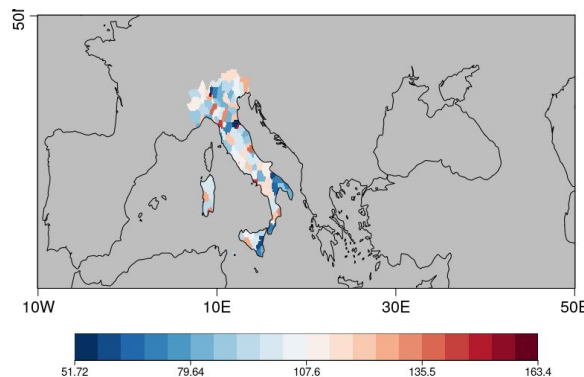
- **New parameters:**

- `shp.system.name`: Any specified column name where to subset `reg.ids`.
- `region`: Logical to add a new dimension for the region in mask.
- `target_crs`: Target Coordinate Reference System.
- `check_valid`: Whether to apply the function `st_make_valid` into the shapefile and coordinates.
- `max_dist`: Maximum distance allowed for no intersection between shapefile and reference grid.
- `...`: Additional parameters to be added in the `st_intersection` call.

- **Developments:**

- Included Alba's development to find the closest distance between the shapefile and the reference grid when there is no intersection:

```
dist <- sqrt((xy.sfg[,1] - x.centroid.shpi)**2 +  
             (xy.sfg[,2] - y.centroid.shpi)**2)  
tmp_coords <- array(xy.sfg[which(dist == min(dist)),],  
                    dim = c(1,2))
```



**Figure 1.** Mask array applied to random data over Italian provinces. Source: Alba Llabrés code.

**status:** In branch [develop-shp\\_mask\\_params](#) → as **ShapeToMask** in new plotting package  
Check issue: <https://earth.bsc.es/gitlab/es/s2dv/-/issues/75>

s2dv



**Barcelona  
Supercomputing  
Center**

Centro Nacional de Supercomputación

# RPS and RPSS NA criterion

# CRPSS ref = NULL no cross-validation

Possibility of building the climatological forecast (when ref=NULL) without using cross-validation with new parameter `clim.cross.val`.

# PlotEquiMap() plot all-NA array

PlotEquiMap() used to fail when all the data are NAs. Now, it can plot the map and the map will be filled with the NA color, and the warnings will be shown.

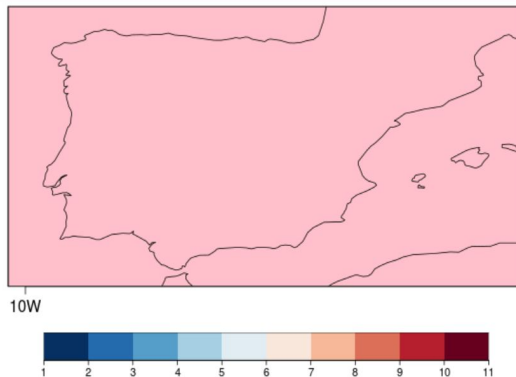
```
> PlotEquiMap(data, lon, lat, filled.continents = F)
```

Warning messages:

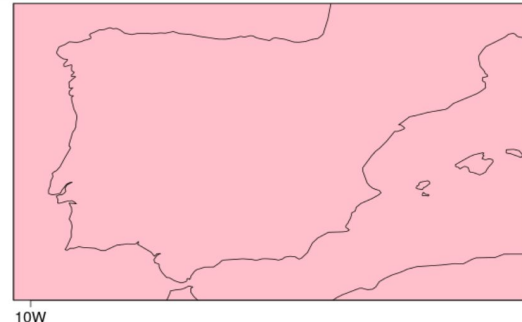
```
1: ! Warning: All the data are NAs. The map will be filled with colNA.
```

```
2: ! Warning: All data are NAs. Color bar won't be drawn. If you want to have  
!color bar still, define parameter 'brks' or 'bar_limits'.
```

(brks/bar\_limits is specified)



(brks/bar\_limits is not provided)



status: in master

issue:

<https://earth.bsc.es/gitlab/es/s2dv/-/issues/96>

# startR



**Barcelona  
Supercomputing  
Center**

*Centro Nacional de Supercomputación*

# New release 2.3.0

- Load variable metadata when retrieve = F
- Change Compute() "threads\_load" to 1 to be consistent with documentation
- Add Autosubmit as workflow manager
- SelectorChecker() to recognize class integer

Will be submitted to CRAN after internal test.

# Compute() with Autosubmit

To use the whole startR workflow (check [these slides](#) if you have no idea what it is) and submit jobs to HPCs, we need to use workflow manager. Aside from ecFlow, now we can use **Autosubmit** ( $\geq 4.0.0$ )!

The advantages of Autosubmit include:

- ★ Internal support
- ★ Clear GUI and logs
- ★ Simple usage, less set-up required

**Q:** Can I keep using the ecFlow scripts? → Yes you can, we will support startR-ecFlow usage until critical bugs occur.



# Compute() with Autosubmit

How to do it?

1. Use the same startR script of Start(), self-defined function, Step(), and AddStep()
2. Change some parameters in Compute():
  - a. `workflow_manager = 'autosubmit'`
  - b. `autosubmit_suite_dir = <local_path_for_temp_files>`
  - c. `autosubmit_server = 'bscesautosubmit01'`
  - d. `cluster`: Some elements are different. Check [practical\\_guide.md](#)
3. You can find logs and Autosubmit-related files under `/esarchive/autosubmit/<expid>`

# Bugfix for integer selector

The selectors in Start() can be integer now (it used to return error with class integer; only accepted numeric)

```
data  <- Start(dat = ...,  
               lon = values(list(lon.min, lon.max)),  
               lat = values(list(lat.min, lat.max)),  
               ...)
```

*usually with class “numeric”, but  
can be “integer” as well*

# CSTools



**Barcelona  
Supercomputing  
Center**

*Centro Nacional de Supercomputación*

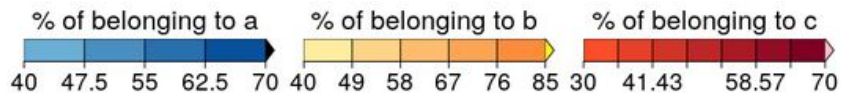
# Add 'print' method for 's2dv\_cube'

```
's2dv_cube'
Data          [ 1.2009e-08, 5.6514e-09, 0, 8.4771e-09, 2.6138e-08 ... ]
Dimensions    ( dataset = 1, member = 6, sdate = 3, ftime = 31, lat = 4, lon = 4 )
Coordinates
  dataset : expl
  member  : 1, 2, 3, 4, 5, 6
  sdate   : 20101101, 20111101, 20121101
  ftime   : 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, [...]
Attributes
  varName  : prlr
  metadata :
    prlr
      units : m s-1
      long name : Total precipitation
      other info : use_dictionary, description, daily_agg_cellfun, monthly_agg_cellfun, verification_time
    lon
      cdo_grid_name : r360x181
      other info : dim, data_across_gw, array_across_gw, first_lon, last_lon, projection
  Datasets : expl
  Dates    : 2011-03-01 2011-03-02 2011-03-03 2011-03-04 2011-03-05 ...
  when     : 2023-04-11 17:05:57
  source_files : /esarchive/exp/ecmwf/system5c3s/daily_mean/prlr_s0-24h/prlr_20101101.nc ...
  load_parameters :
    ( var ) : = prlr
  ...
```

# PlotMostLikelyQuantileMap legend

The changes are in PlotCombineMap().

- The color bars can have upper limit < 100% and upper triangle.
- The color bars can have different breaks.
- One new function ColorBar\_onebox() especially for plotting only one color box with the same break width as others.

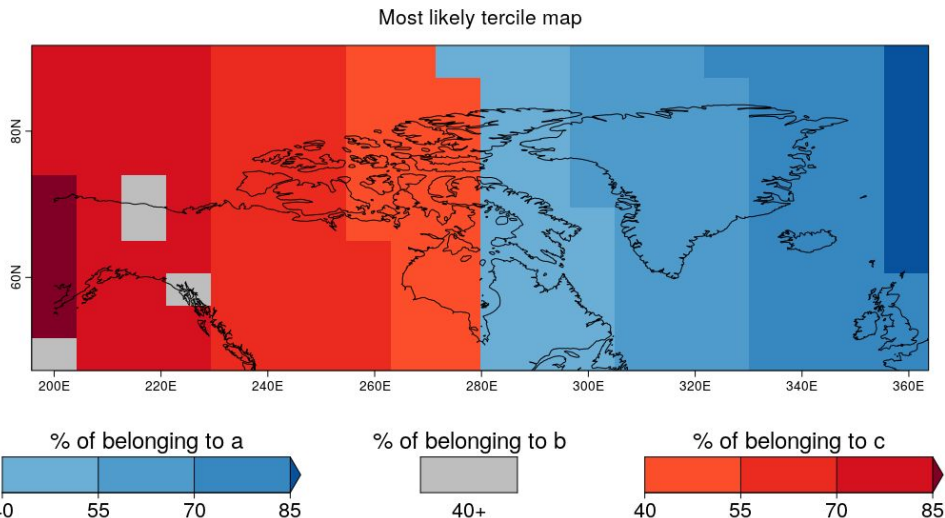


status:

In master. Source s2dv::ColorBar from master as well for bugfix.

issue:

<https://earth.bsc.es/gitlab/external/cstools/-/issues/125>



# CSIndicators



**Barcelona  
Supercomputing  
Center**

*Centro Nacional de Supercomputación*

# Added memb\_dim to MergeRefToExp

- This function can merge observations and forecasts, known as the ‘blending’ strategy through **time\_dim**.
- Only **memb\_dim** is allowed to be of different length (and **time\_dim**). Two cases:
  - (a) One array doesn’t have **memb\_dim** or is equal to 1 and the other has multiple **memb\_dim**:
    - Repeated values of array 1 up to the length of member dimension of array 2
  - (b) Both arrays have **memb\_dim** dimension and is greater than 1:
    - All combinations of member dimension will be returned

```
ref <- array(1001:1009, c(ftime = 3, sdate = 1, member = 3))
data <- array(1:4, c(ftime = 2, sdate = 1, member = 2))
res <- MergeRefToExp(data1 = ref, data2 = data)
> dim(res)
  ftime sdate member
    5     1     6
```

```
> res[,1,] # 1st sdate [ftime x member]
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 1001 1001 1004 1004 1007 1007
[2,] 1002 1002 1005 1005 1008 1008
[3,] 1003 1003 1006 1006 1009 1009
[4,]    1    3    1    3    1    3
[5,]    2    4    2    4    2    4
```

# SUNSET



**Barcelona  
Supercomputing  
Center**

*Centro Nacional de Supercomputación*



# New repository name: From 'Auto-S2S' to 'SUNSET'

The name and the URL of the SUNSET GitLab repository have been changed **from Auto-S2S to SUNSET**.

The URL of the repository has also changed. When pushing or pulling, you may see this message:

```
remote: Project 'es/auto-s2s' was moved to 'es/sunset'.  
remote:  
remote: Please update your Git remote:  
remote:  
remote: git remote set-url origin https://earth.bsc.es/gitlab/es/sunset.git
```

Just run the command as directed:

```
git remote set-url origin https://earth.bsc.es/gitlab/es/sunset.git
```

# New module: Units

New module for unit conversion is in development. Users will be able to specify the target units for each variable in the recipe:

Variables:

- {name: tas, units: C, freq: monthly\_mean}
- {name: prlr, units: mm, freq: monthly\_mean}

The `Units()` function checks if the units are consistent among all the `s2dv_cube` objects in the `data` object, and whether the target units are the same as the original units. If they are not, it performs unit conversion, changing the data and metadata.

See the full discussion in: <https://earth.bsc.es/gitlab/es/sunset/-/issues/57>

**status:** in branch dev-units

# New module: Scorecards

New module that can be used to produce the Scorecards from the in-house [CSScorecards package](#), for the verification metrics computed with SUNSET. The specifications for the Scorecards can be defined in the recipe:

```
Scorecards:
  execute: yes
  regions:
    Extra-tropical NH: {lon.min: 0, lon.max: 360, lat.min: 30, lat.max: 90}
    Tropics: {lon.min: 0, lon.max: 360, lat.min: -30, lat.max: 30}
    Extra-tropical SH : {lon.min: 0, lon.max: 360, lat.min: -30, lat.max: -90}
  start_months: NULL
  metric: mean_bias enscorr rpss crpss enssprerr
  metric_aggregation: 'score'
  table_label: NULL
  fileout_label: NULL
  col1_width: NULL
  col2_width: NULL
  calculate_diff: no
```

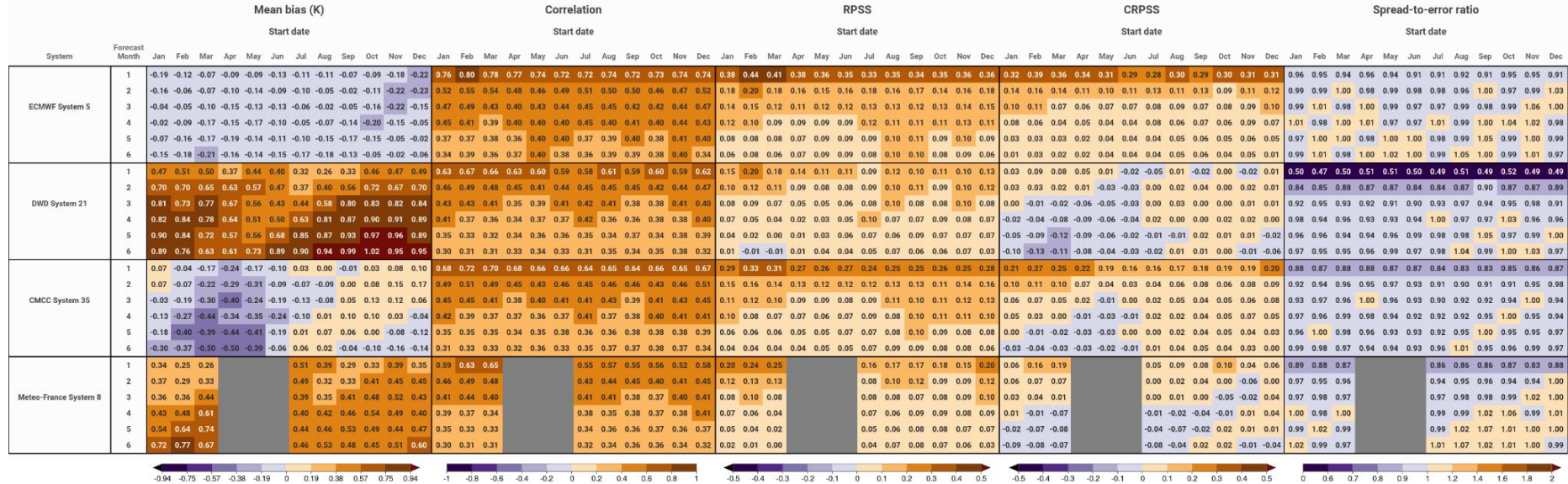
**status:** in branch dev-scorecards

# New module: Scorecards

## Scorecard example (from the [CSScorecards vignette](#)):

Near-Surface Air Temperature. Region: Extra-tropical SH (Interpolation = to system, Aggregation level = skill, Cross-validation = none)

(Ref: ERA5 1993-2016)



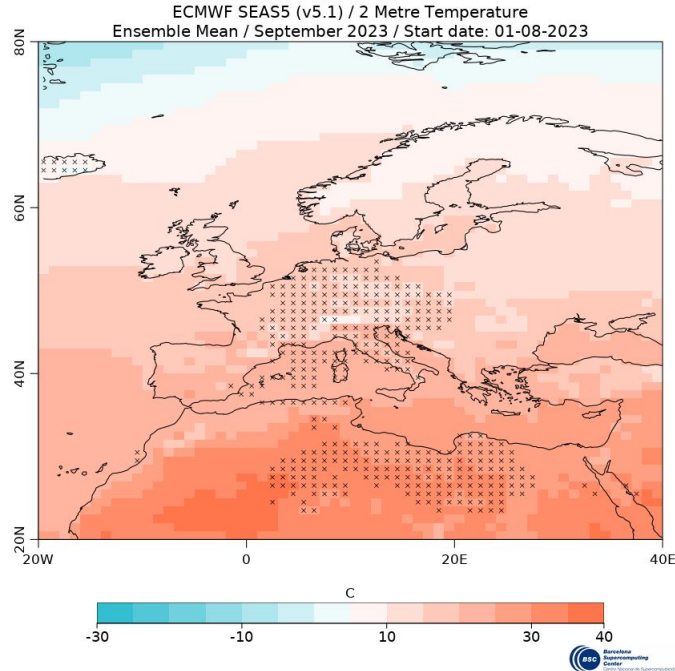
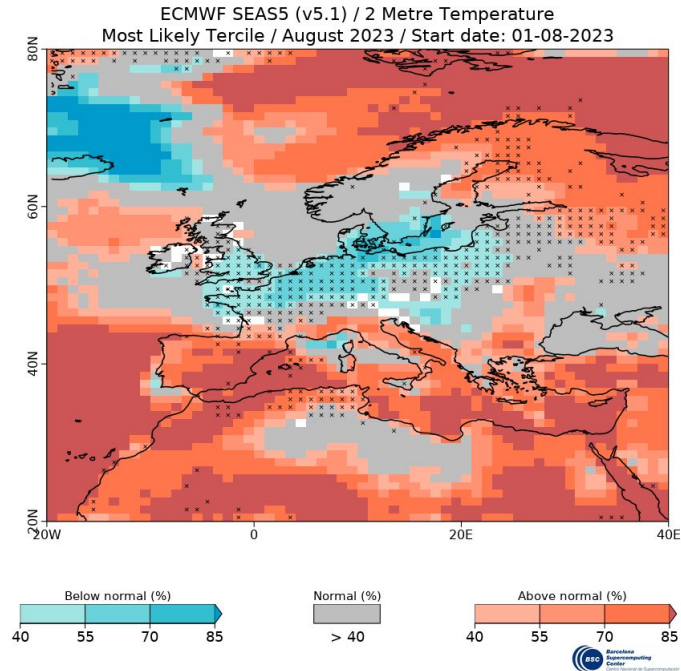
# Plot enhancements

The following enhancements are planned and in development for the plots:

- Adjustment of the colors and categories in the **Most Likely Terciles** plot, following the discussion in CStools: [external/cstools#125 \(closed\)](#)
- Adjustments for the **Forecast Ensemble Mean** plots: units, new colors, new title, and allowing masking/dots by ensemble correlation.
- A yaml configuration file will be used to **adjust the plot size and layout for pre-defined regions**. New regions can be added any time.
- Added units in the legend of the mean bias skill metric plot.

**status:** in branch dev-csoper ([https://earth.bsc.es/gitlab/es/auto-s2s/-/merge\\_requests/83](https://earth.bsc.es/gitlab/es/auto-s2s/-/merge_requests/83))

# Plot enhancements



status: in branch dev-csoper ([https://earth.bsc.es/gitlab/es/auto-s2s/-/merge\\_requests/83](https://earth.bsc.es/gitlab/es/auto-s2s/-/merge_requests/83))

# Module function name changes

In the next release, the names of the module functions will be changed to be the same as the module name.

For example:

- `load_datasets()` will become `Loading()`
- `calibrate_datasets()` will become `Calibration()`

The old names will be maintained, but using them will raise a warning asking you to change your scripts. They will be fully deprecated in a future release.

**status:** in branch `dev-rename_functions_adapt_recipe`

# User presentation



**Barcelona  
Supercomputing  
Center**

*Centro Nacional de Supercomputación*



# 'config' package

To use `config`, create a file `config.yml` with default as well as other arbitrary configurations.

```
default:
  origin.data.path: "/esarchive/obs/tall_towers"
  output.data.path: "/esarchive/obs/tall_towers"
  read.towers.path: "/esarchive/scratch/jramon/GitLab_jramon/TallTowers/read_tower_data/"
  plot.path: "/esarchive/scratch/jramon/INDECIS/plotting"
  tower.metadata: "/esarchive/scratch/jramon/GitLab_jramon/TallTowers/generic/read_tower_info.R"
  write.nc: "/esarchive/scratch/jramon/GitLab_jramon/TallTowers/generic/write_talltower_nc.R"

publication:
  origin.data.path: "/esarchive/obs/tall_towers"
  output.data.path: "/esarchive/scratch/glope3/The_Tall_Tower_Dataset_v4"
```

To read configuration values you call the `config::get` function, which returns a list containing all of the values for the currently active configuration:

```
library(config)
config <- config::get()
config[["output.data.path"]]
#> [1] /esarchive/obs/tall_towers
```

# an alternative to 'config' package

Create a file `config.R` with a list containing all the information

```
config <- list(origin.data.path = "/esarchive/obs/tall_towers",  
              output.data.path = "/esarchive/obs/tall_towers",  
              read.towers.path = "/esarchive/scratch/jramon/GitLab_jramon/TallTowers/read_tower_data/",  
              plot.path       = "/esarchive/scratch/jramon/INDECIS/plotting",  
              tower.metadata  = "/esarchive/scratch/jramon/GitLab_jramon/TallTowers/generic/read_tower_info.R",  
              write.nc        = "/esarchive/scratch/jramon/GitLab_jramon/TallTowers/generic/write_talltower_nc.R")
```

Source `config.R` to read the configuration information

```
source("./config.R")  
config[["output.data.path"]]  
#> [1] /esarchive/obs/tall_towers
```

# Thanks for joining