

BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación

# **R** user meeting

07/11/2024

Victòria Agudetse, Ariadna Batalla, Theertha Kariyathan

### Agenda

- 1. Ice-breaker: package 'styler'
- 2. News
  - startR
  - o s2dv
  - $\circ$  CSTools
  - CSIndicators
  - $\circ$  esviz
  - $\circ$  SUNSET
- 3. Q&A

# Ice-breaker: R package 'styler'



### R Package 'styler'

styler is a package that automatically formats your code according to the tidyverse style guide. It has functions to format:

- A character string...
- An entire file...
- An entire directory...
- ... Even an entire R package!

The package also allows you to define your own custom style guide and even publish it.

Let's take a look at some simple examples.

#### R Package 'styler': function style\_text()

```
library(styler)
```

```
# style_text() formats a string
style_text("my_vector = c(1,2,3,4)")
\# my_vector <- c(1, 2, 3, 4)
# It can also be a multi-line string
style_text(
  "if(!is.character(dat_dim)){
       error()
     }"
# if (!is.character(dat_dim)) {
   error()
#
# }
```

### R Package 'styler': function style\_file()

```
library(styler)
```

```
# style_file() formats an entire file (.R, .Rmd, and other related extensions)
style_file("R/CST_Subset.R")
Styling 1 files:
    R/CST_Subset.R i
```

- Status Count Legend
- Ø File unchanged.
- 1 File changed.
- Styling threw an error.

Please review the changes carefully!

**WARNING:** style\_file() rewrites your file! Before using it, it is strongly recommended to have it tracked in git to avoid permanent unwanted changes

### startR



★ Improvement in error message for Compute.R

★ Compute(), returns the following error message when dat dimension is chunked : Error in Compute(wf, chunks = list(sdate = 2, dat = 2)) : "Chunking along the pattern dimension dat is not allowed for now."

```
if (any(names(chunks) == attr(workflow$inputs$input1, "PatternDim"))) {
    stop(paste0("Chunking along the pattern dimension ",
        attr(workflow$inputs$input1, "PatternDim"),
        " is not allowed for now."))
```

MR: https://earth.bsc.es/gitlab/es/startR/-/merge\_requests/242 status: in master

# s2dv



The new version of startR, v2.1.0, was installed on all machines in October. It includes several bugfixes and the following developments:

- ★ NAO(): new parameter "exp\_cor" to calculate forecast
- New parameter "abs\_threshold" in GetProbs()
- ★ New parameter "return\_mean" in RPS() and CRPS()
- ★ New parameter "print\_sys\_msg" in CDORemap()
- ★ New function SprErr()
- ★ New parameter "alpha" in Bias()
- ★ New parameter "N.eff" in RandomWalkTest()

Previous maintainer: An-Chi Ho. New maintainer: Ariadna Batalla If you encounter any issues, please report them!



# **CSTools**



#### New function CST\_Bind()

This is a wrapper of the abind::abind() function. The function has two levels:

- 1. Bind(): Works with **arrays with named dimensions**. Binds two arrays along the specified dimension
- 2. CST\_Bind(): Works with **s2dv\_cube objects**. Binds the to arrays in \$data, as well as the corresponding metadata in \$dims, \$coords and \$attrs.

MR: <u>https://earth.bsc.es/gitlab/external/cstools/-/merge\_requests/215</u> status: in development (in branch dev-CST\_abind) Parameters:

- ★ x: A list of two or more objects of class s2dv\_cube to be bound together.
- ★ along: A character string indicating the name of the binding dimension.
- ★ dat\_dim: A character string indicating the name of dataset dimension. The default value is NULL. Specifying this dimension allows the function to ensure the dataset metadata is correctly preserved.
- ★ var\_dim: A character string indicating the name of the variable dimension. The default value is NULL. Specifying this dimension allows the function to ensure the variable metadata is correctly preserved.

MR: <u>https://earth.bsc.es/gitlab/external/cstools/-/merge\_requests/215</u> status: in development (in branch dev-CST\_abind)

# **CSIndicators**



# CST\_PeriodPET and CST\_PeriodStandardization(): update for metadata parameters

CST\_PeriodPET() : modified to return a PET item in the attrs\$Variable\$metadata list with the longname ("Potential evapotranspiration") and the units ("mm"). CST\_PeriodStandardization() : longname of the variable is modified to append the word "standardized".

```
CST_PeriodPET(...)
metadata :
    PET
    long name : Potential evapotranspiration
    units : mm
    other :
```

```
CST_PeriodStandardisation(...)
```

```
metadata :
    prlr
    units : m s-1
        long name : Total precipitation
standardized
```

CST\_PeriodPET() : MR: <u>https://earth.bsc.es/gitlab/es/csindicators/-/merge\_requests/69</u> status: in master CST\_PeriodStandardization() : MR: <u>https://earth.bsc.es/gitlab/es/csindicators/-/merge\_requests/68</u>

status: to be merged

### esviz



# VizPlotForecastPDF(): include multiple observations and improve plot styling

Improved **VizPlotForecastPDF(),** it is now possible to:

- include several observations for one forecast panel, include different observations for different forecast panels.
- change the text and size of the legends text, legends and titles.
- include/exclude the horizontal line for the observations.
- adjust the height, width and resolution of the figure.

issue: <u>https://earth.bsc.es/gitlab/es/esviz/-/issues/17</u>
status: in branch imprv\_VizforecastPDF (ready to test)

Spain - Variable: TAS - Season: Annual - Multi-model - Reference dataset: ERA5 Calibration: False; Climatological period: False CRPSS = -0.62\*: RPSS = 0.4\*



### New mask parameters in VizEquiMap()

VizEquiMap () now includes 2 parameters to control data masking on the map.

- mask
  - Controls which areas on the map are masked
  - Input: array with the same dimensions as var, containing binary values of [1, 0] or logical values.
     Default is NULL
    - 0 or FALSE: points that should be masked
    - 1 or TRUE: points that should be plotted normally
- mask\_color
  - Sets the color of the masked areas
  - O Input: any valid color name or hex code ("e.g., "gray", "#FF5733"). Default is "white"





# **SUNSET**



### Launching jobs on CTE-AMD and MN5

The launch\_SUNSET.sh script can now be used to launch jobs on CTE-AMD and MN5.

- 1. Without Autosubmit: Your output directory 'Output\_dir' should be in **/gpfs/.** You can launch the script directly from the HPC machine.
- 2. With Autosubmit: Your output directory 'Output\_dir' should be in **/esarchive/**, and you should launch the script from the Hub or a workstation. A temporary directory will be created in your /gpfs/scratch/ automatically. When all the jobs finish, the outputs will be transferred back to /esarchive/ and the temporary files will be cleaned.

We are working on a hands-on tutorial to show the step-by-step process.

issue: https://earth.bsc.es/gitlab/es/sunset/-/issues/149
status: in branch dev-launch\_on\_MN5 (ready to test)

### New plotting option: masking significance

For the metric plots, the options 'dots' and 'mask' (and 'both') are now available to choose how to indicate statistical significance of grid points. This <u>example script</u> shows all the possibilities. This option can be specified from <u>the recipe</u> or directly when calling the Visualization() function: Visualization(..., significance = "mask")

For now, the 'mask' option (shown here) is only available for the 'robinson' and 'lambert\_europe' projections.





MR: <u>https://earth.bsc.es/gitlab/es/sunset/-/merge\_requests/163</u> status: in master

### **Bugfixes in plotting functions**

Several bugs spotted in the plotting functions have been fixed:

- The **captions of forecast plots** now show the correct time unit depending on the horizon ("weeks" for weekly subseasonal data; "months" for seasonal and decadal data).
- plot\_metrics(): Now display correct time bounds for time-aggregated data when the months span different years (e.g.: "December to February").
- A bug in the conditional statements that was causing errors for decadal plot titles.

MR: <u>https://earth.bsc.es/gitlab/es/sunset/-/merge\_requests/171</u>

### **Bugfix in Aggregation module**

The time aggregation module failed when Time\_aggregation:execute was set to 'no' in the recipe but the Aggregation() function was called anyway.

Now, in this case the function returns the original data with a warning message: "The Aggregation module has been called but parameter Time\_aggregation:execute is set to 'no' in the recipe. The data is returned unchanged."

issue: <u>https://earth.bsc.es/gitlab/es/sunset/-/issues/154</u> status: in master



## Thanks for joining



Next meeting: December 12th