



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación


R user meeting

An-Chi Ho and Eva Rifà

contributor: 🎄

01/12/2022

Agenda

1. Ice-breaker:
2. News
 - General R
 - s2dv
 - startR
 - ClimProjDiags
 - multiApply
 - CStools
 - CSIndicators
3. Q&A
4. User presentation: XMAS SPECIAL 

Ice-breaker

Two truths and one lie

- A. Only CTools and CSIndicators current packages work with s2dv_cube objects
 - B. Only s2dv and CTools have plotting functions
 - C. Subset() function is part of s2dv
-
- A. Python is older than R
 - B. R is named R is partly based on the names of the first two R authors
 - C. CRAN package repository features 15000-16000 available packages
-
- A. startR last version number is 2.2.1
 - B. startR was first published on CRAN in April 2017
 - C. startR was firstly created for the ERA4CS project MEDSCOPE
-
- A. s2dverification first release date 2013-08-02
 - B. s2dverification has a published paper on Environmental Modelling & Software journal
 - C. s2dv first release date on 2021-06-05
-
- A. We have 11 users on the shiny server
 - B. s2dverification is not used by anyone now
 - C. According to the survey, the most used R packages in our department is startR

Two truths and one lie

A. Only CStools and CSIndicators current packages work with s2dv_cube objects

B. Only s2dv and CStools have plotting functions

C. Subset() function is part of s2dv → Subset() function is part of ClimProjDiags

A. Python is older than R

B. R is named R is partly based on the names of the first two R authors

C. CRAN package repository features 15000-16000 available packages → 18908 available packages

A. startR last version number is 2.2.1

B. startR was first published on CRAN in April 2017

C. startR was firstly created for the ERA4CS project MEDSCOPE → startR was created purely for our department's need

A. s2dverification first release date 2013-08-02

B. s2dverification has a published paper on Environmental Modelling & Software journal

C. s2dv first release date on 2021-06-05 → s2dv first release date on 2020-02-07

A. We have 11 users on the shiny server

B. s2dverification is not used by anyone now

C. According to the survey, the most used R packages in our department is startR →

s2dv ≥ Start > multiApply >
CStools > ClimProjDiags >
s2dverifiacion > easyNCDF >
CSIndicators

General R

Last meeting follow-up: print summary() with log4r

To print nice summary in [log4r](#): use ``capture.output()``, which save the output as one string per line.

```
library(log4r)

logger <- logger(threshold = 'INFO', appenders = file_appender("test.log"))
info(logger, "I'm printing to a log file!")

a <- 1:10
summary <- summary(a)
summary_string <- capture.output(summary)
for (i in summary_string) {
  info(logger, i)
}
```

```
INFO [2022-11-18 13:38:10] I'm printing to a log file!
INFO [2022-11-18 13:39:41]      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
INFO [2022-11-18 13:39:41]      1.00    3.25    5.50    5.50    7.75   10.00
```

R TOOLS ROADMAP 2023

Reasons to plan a release

User needs

Tool needs

Project needs

Reverse dependency break

CRAN status

issues
(21/11/2022)

Winter 2023

Spring 2023

Long term

s2dv

v1.3.0
Oct. 2022

19

- Plotting function integration
- Singificance test functionality consistency

- Efficiency test

- Plot different projections

startR

v2.2.1
Nov 2022

28

- Better conversion to s2dv_cube
- Bugfixes & refine error messages

- ecFlow usage improvement
- Efficiency test

- Read and interpolate irregular grid
- Multiple steps

CSTools

v4.1.1
Nov. 2022

20

- Development of new s2dv_cube structure
- Work with multiple datasets

- Downscaling integration
- Plotting function improvement

- CST_Load() internal use of Start()
- Convert s2dv_cube to xarray

CSIndicators

v0.0.2
Oct. 2022

10

- New s2dv_cube structure change for CST functions

- New vignettes

- Health indicators integration

ClimProjDiags

v0.2.0
Nov 2022

1

multiApply

v2.1.3
Feb 2021

3

- Bugfixes and sanity checks

easyNCDF

v0.1.0
Mar 2020

2

General

—

—

- shape file function - license change - RStudio-server version upgrade - Individual user meeting

Guidelines for R-related questions in Earth Sciences

Please read the document about how to open an issue for R-related questions:

<https://docs.google.com/document/d/1zRlmsRwFDJctDB94x6HGf6ezu3HFHhEjaBu0hVcrwTI/edit?usp=sharing>.

Table of contents

1. When to open an issue?
2. Where to open the issue?
3. What to put in the issue?
 - a. Who to tag?
 - b. What is the suitable title?
 - c. What is the necessary information?
 - d. Which environment do you use?
4. How to build a Minimal Reproducible Example
 - a. Minimal necessary dataset
 - b. Minimal necessary code
 - c. Reproducibility

Guidelines for R-related questions in Earth Sciences

Especially pay attention to [section 4. How To Build A Minimal Reproducible Example](#)

a. Minimal necessary dataset

- i. NetCDF files under /esarchive,
- ii. A .Rds or .Rdata file
- iii. Sample data in a package, e.g., CTools::lonlat_temp
- iv. Random synthetic data

b. Minimal necessary code

- i. DO NOT just copy-paste or link your original full script.
- ii. DO NOT post screenshots of code.
- iii. If you create any files in your example, include code to delete them, (e.g. use unlink())
- iv. Include short comments to make the code easier to understand

c. Reproducibility

- i. Do others have the access to the files provided in the issue?
- ii. Open a brand new R console to run the script.
- iii. Put the script on your personal GitLab repository (recommended for long script)

s2dv

CDORemap() bugfix

Error shows if the input array reaches all the following conditions:

- (1) dimension number is more than 4
- (2) more than 3 dimensions are not of length 1
- (3) the last dimension is not lat or lon.

```
Error in R_nc4_def_var_double: NetCDF: NC_UNLIMITED in the wrong index
Name of variable that the error occurred on: "var"
[1] "-----"
[1] "Var:  var"
[1] "Ndims:  4"
[1] "Dimids:  "
[1] 3 2 1 0
Error in ncvar_add(nc, vars[[ivar]], verbose = verbose, undefine = TRUE) :
  Error in ncvar_add, defining var var
```

status: Fixed, in master

More details: <https://earth.bsc.es/gitlab/es/s2dv/-/issues/83>

Corr() output `sign`

- Argument `conf.lev = 0.95` has been changed to `alpha = 0.05`
- New argument `sign = FALSE`
- New output `sign`: A logical array of the statistical significance

status: Done, in branch [develop-Corr-significance](#)

```
p.val <- pt(t, eno_expand - 2, lower.tail = FALSE)
if (sign) signif <- !is.na(p.val) & p.val <= alpha
```

RMSSS() development (1/2)

- New argument `alpha = 0.05`
- New argument `sign = FALSE`
- New output `sign`: A logical array of the statistical significance

*Anyone wants confidence level?

status: Done, in branch [develop-RMSSS-significance](#)

```
F.stat <- (eno2 * rms2^2 / (eno2 - 1)) / ((eno1 * rms1^2 / (eno1 - 1)))
tmp <- !is.na(eno1) & !is.na(eno2) & eno1 > 2 & eno2 > 2
p_val <- 1 - pf(F.stat, eno1 - 1, eno2 - 1)
if (sign) signif <- p_val <= alpha
# If there isn't enough valid data, return NA
p_val[which(!tmp)] <- NA
if (sign) signif[which(!tmp)] <- NA
```

RMSSS() development (2/2)

1. Include:

- Parameter “memb_dim” to compute ensemble mean
- Parameter “ref” to allow the computation of the RMSSS of the forecast with respect to a reference forecast (other than the climatological forecast)

2. Use .RMS() inside .RMSSS():

Since $\text{RMSSS} = 1 - \text{RMS_exp} / \text{RMS_ref}$, .RMSSS() can use .RMS() instead of its own code.

3. Different significance test method:

Current one is one-sided Fisher test; add two-sided Random Walk test as another method?

status: Nothing done yet, under discussion

Check issue: <https://earth.bsc.es/gitlab/es/s2dv/-/issues/84>

Trend() p.val bug

Trend() return wrong/misplaced p-values if there is time vector full of NAs (hence the function returns NAs for all the outputs)

status: Fixed, in branch [develop-Trend_pval_NA](#)

Check issue: <https://earth.bsc.es/gitlab/es/s2dv/-/issues/85>

startR

New release 2.2.1

Check NEWS.md <https://earth.bsc.es/gitlab/es/startR/-/blob/master/NEWS.md>

ClimProjDiags

New release 0.2.1

Check NEWS.md <https://earth.bsc.es/gitlab/es/ClimProjDiags/-/blob/master/NEWS.md>

Note that 0.2.0 and 0.2.1 are the same. There was a mistake in the 0.2.0 submission to CRAN and one new function `WeightedCells()` was missing. Version bump in order to submit to CRAN again.

But nothing impacted on users' side.

multiApply

Mind the warnings

When using `Apply()` or the functions that use `Apply()`, if you encounter the warning like:

```
In arrays_of_results[[component]][(1:prod(component_dims)) + ... :  
  number of items to replace is not a multiple of replacement length
```

or other warnings not intendedly produced by `Apply()`, **it probably has problems**.

→ Check the function used in `Apply()`. *Does the output has the same dimensions all the time?*

```
data <- array(1:12, dim = c(time = 4, member = 3))  
res <- Apply(data, fun = mean, target_dims = 'time')
```

What does it mean? → `mean()`'s input is a 1-dim array [time = 4] and it is run 3 times (margin dim [member = 3]). So, the outputs of the 3 times should have the same dimensions.

status: Fixed, in branch [develop-inconsistent_output_length](#)

Check issue: <https://earth.bsc.es/gitlab/ces/multiApply/-/issues/12>

CSTools

New release 4.1.1

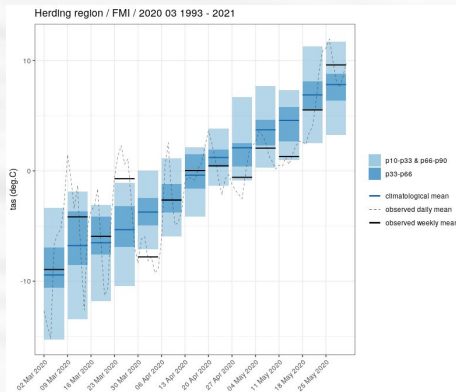
Short release on due to ClimProjDiags dependency of Subset() function.

Fixes:

- CST_Analogs corrected input of ClimProjDiags::Subset()
- PlotCombinedMap corrected use of 'cex_bar_titles' parameter
- CST_Anomaly added 'memb_dim', 'dat_dim' and 'ftime_dim' and improved use for 'dim_anom' parameters

Next steps:

- New plotting function for case studies temporal visualization
- Change s2dv_cube object structure



Fixes in CSTools 4.1.1

- **CST_Analogs()**

- corrected: input of `ClimProjDiags::Subset()`

```
obslocal <- SelBox(obsVar, lon = lonVar, lat = latVar, region = region)$data
Analog_fields <- Subset(obslocal, ...)
```

- **PlotCombinedMap()**

- corrected use of 'cex_bar_titles' parameter

```
GradientCatsColorBar([...], title_scale = 1)
GradientCatsColorBar([...], title_scale = cex_bar_titles)
```

- **CST_Anomaly()**

- Added 'memb_dim', 'dat_dim' and 'ftime_dim' and improved 'dim_anom' parameters

```
dim_anom = 'sdate': Input parameter time_dim in Ano_CrossValid() and Clim()
memb_dim = 'member': Input parameter memb_dim in Ano_CrossValid() and Clim()
dat_dim = c('dataset', 'member'): Input parameter dat_dim in Ano_CrossValid() and Clim()
ftime_dim = 'ftime': Target dimension for Apply() with .Loess when filter_span is not NULL.
```

dat_dim in CStools

- `dat_dim` parameter has been added into **(CST_)BiasCorrection** and **(CST_)Calibration** functions. It allow to use the function between multiple datasets.

```
> dim(exp$data)
member sdate ftime lat lon dataset
3      4      5      6      7      2
```

```
> dim(obs$data)
dataset member sdate ftime lat lon
1      1      4      5      6      7
```

```
CST_Calibration(exp = exp, obs = obs, dat_dim = 'dataset')
CST_BiasCorrection(exp, obs, dat_dim = 'dataset')
```

```
> dim(a$data)
member sdate ftime lat lon nexp nobs
3      4      5      6      7      2      1
```

status: under test (develop-dat_dim branch)

CSIndicators

Two thresholds

- Functions **TotalSpellTimeExceedingThreshold()**, **TotalTimeExceedingThreshold()** and **AccumulationExceedingThreshold()** now accept two thresholds instead of one.

How it works?

1. First the function checks the number of operators:

1 threshold

a) If op (operator) parameter is:
'>', '<', '>=' or '<='

2 thresholds

b) If op is a vector pair of operators:
`c('<', '>')`, `c('<', '>=')`,
`c('<=', '>')`, `c('<=', '>=')`,
`c('>', '<')`, `c('>', '<=')`,
`c('>=', '<')`, `c('>=', '<=')`

2. Then, the threshold can be:

A MD array with named dimensions

A scalar

A vector of length `time_dim`

A pair of scalars `c(thres1, thres2)`

A list of 2 MD array with named dimensions

A list of 2 vectors of length `time_dim`

Example: `CST_AccumulationExceedingThreshold(data = data, threshold = list(thres1, thres2), op = c('>', '<'), time_dim = 'ftime', start = list(03, 03), end = list(03, 25))`

status: under test (develop-Threshold)

Q & A

User presentation XMAS SPECIAL

Create your own Christmas tree!

Plot your own Christmas tree by **R base functions** and spice up with an **emoji** twist 🌲 😊 💕

✨ Inspired by

<https://github.com/R-CoderDotCom/christmas-tree>

🎭 Package “emo”: <https://github.com/hadley/emo>

🎨 Colors reference:

https://www.rapidtables.com/web/color/RGB_Color.html

<https://www.datanovia.com/en/blog/awesome-list-of-65>

[7-r-color-names/](https://www.datanovia.com/en/blog/awesome-list-of-65)



Create your own Christmas tree!

Copy functions and template to your workstation:

```
cp /home/Earth/aho/aho-testtest/ChristmasTree/ChristmasTree.R
cp /home/Earth/aho/aho-testtest/ChristmasTree/plot_ChristmasTree.R
cp /home/Earth/aho/aho-testtest/ChristmasTree/recipe_ChristmasTree_template.yml
```

Or get them on GitLab:

<https://earth.bsc.es/gitlab/aho/aho-testtest/-/tree/master/ChristmasTree>

1. Modify the recipe
2. Run `plot_ChristmasTree.R`.
3. Share your tree with us 🎄

Note that saving into a file is faster than plotting in a pop-up window.

Thanks for joining