R tools user meeting

An-Chi Ho and Núria Pérez-Zanón

06/11/2020

Agenda

1. Package review & update
   - s2dverification
   - s2dv
   - startR
   - CSTools

2. Development Roadmap

3. User engagement and expectations

4. Shiny app introduction

5. Profiling in R

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

In-house packages status

PACKAGE	LAST RELEASE	NEXT RELEASE	#ISSUES
s2dverification	October 2020	One more release & deprecated	84
s2dv	February 2020	November 2020	7
CSTools	July 2020	December 2020	24
ClimProjDiags	January 2020	-	2
startR	October 2020	Winter 2020	17
easyNCDF	March 2020	-	2
multiApply	September 2019	-	6

ACTION Review your issues and branches!

s2dverification

New version **v2.9.0**

- Improved functions: Ano(), Corr(), Clim(), Composite()
- Correct the time retrieval in Load():

    When start date and the first lead time in netCDF file do not match, Load() retrieved the wrong time metadata as well as unaligned data among data sets.

See details in:

- NEWS.md
- CRAN

```
start date: 200403

system5c3s
time = "2004-03-02 12", "2004-03-03 12", "2004-03-04 12", …

era5
time = "2004-03-01 11:30", "2004-03-02 11:30", "2004-03-03 11:30", ...
```

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

startR

New version **v2.1.0**

- /dev/shm automatic cleaning on Compute(). Solve the error 'No space left on device' which happened when the jobs are aborted.
- New parameter 'largest_dims_length' in Start()

    It can examine all the files to find the largest inner dimension length. It is useful when certain inner dimension among the files does not have consistent length (e.g., different ensemble number).
- Metadata retrieval bugfixes

See details in:
- NEWS.md
- CRAN

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

CSTools

Basic functions
CST_Load
CST_Anomaly
CST_SaveExp
CST_SplitDim
CST_MergeDims
s2dv_cube
as.s2dv_cube

Correction
CST_BiasCorrection
CST_Calibration
CST_QuantileMapping
CST_BEI_Weighting
BEI_PDFBest
CST_CategoricalForecast
CST_DynamicalBC

Downscaling
CST_Analogs
CST_RFTemp
CST_RainFARM
CST_RFSlope
CST_RFWeights
CST_ADAMONT
CST_AnalogsPredictors

Evaluation
CST_MultivarRMSE
CST_MultiMetric

Plotting functions
PlotMostLikelyQuantileMap
PlotForecastPDF PlotPDFsOLE
PlotCombinedMap PlotTriangles4Categories

Classification
CST_WeatherRegimes WeatherRegimes
CST_RegimeAssign RegimeAssign
CST_EnsClustering CST_MultiEOF

■ To be include in December   ■ Enhancements for December + vignettes

*RainFARM: Rainfall Filtered Autoregressive Mode
*BEI: Best Estimate Index

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

see vignettes and documentation: https://cran.r-project.org/web/packages/CSTools/index.html

Current (autumn 2020)	Autum-Winter 2020	Winter-Spring 2021	Long term
s2dverification 2.8.6	**s2dverification 2.9.0**		**s2dverification 2.9.1**
	• Fixes: Load, Clim,…		• To be deprecated
s2dv 0.0.1	**s2dv 0.0.2**	**s2dv 1.0.0**	**s2dv 1.0.3**
	• Fixes: Clim,.. • New functions from s2dverification • New 6 functions	• Remaining transformation from s2dverification	• Refine plotting function (use MapGenerator)
startR 2.0.1	**startR 2.1.0**	**startR 2.1.1**	**startR 3.0.0**
	• Fixes shm/dev, metadata …	• New developments: regriding, …	• R workflow manager, multiple steps in AddStep
CSTools 3.0.1	**CSTools 4.0.0**	**CSTools 4.0.1**	
	• All functions promised • Fixes for the manuscript	• Fixes	

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Users enagement / expectations

Given the latest news from HHRR that **working from home** is the primary option ….

… we propose you to have **R users monthly meetings**

We could **add new content** or redefine the format of the meetings to fulfill your **expectations and needs**

- ❏ These meetings can be set for a specific date like **the 1st Friday of the month at 3.00 pm** to avoid extra doodles and emails
- ❏ Usual content:
    - ❏ NEWS: fixes, developments and releases
    - ❏ Open issues to discuss
    - ❏ Users questions
    - ❏ Specific topics proposed by An-Chi or Núria
- ❏ To help with point d)
    - Would you like presenting a piece of code?
    - Example: There are new functions in s2dv RandomWalkTest() how to use it?
    - There is a place on gitlab or other shared place where an example is shown?
- ❏ Ideas and suggestion?

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

shiny app

cp_shiny-figure

This Shiny application is for visualizing and monitoring the BSC-ES experimental/diagnostic data. It demonstrates the pictures of the diagnostics.

First release v1.0.0
http://bscesshiny.bsc.es:3838/es/cp_shiny-figure_v1.0.0/

Figure storage structure:

/esarchive/exp/ecearth/a32p/plots/

atmos
land
ocean
ocnBgchem
sealce

heatc0-300m
heatcsum0-300m
maxmoc26N
maxmoc45N
mlotst

msftbarot
msftyz
sos
tos
zos

climantarctic_tos_xxx.png
climarctic_tos_xxx.png
climglobal_tos_xxx.png

*Figures are produced by ESMValTool and mapgenerator.

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Profiling in R

Profiling in R

Rprof and Profvis

Profvis (https://rstudio.github.io/profvis/) is a tool for helping you to understand how R spends time and memory. It provides a <u>interactive graphical interface</u> for visualizing data from **Rprof**, R's built-in tool for collecting profiling data.

Alternative to Profvis: *profr* package [to be explored]

Which to use?
- are you in an interactive session with access to RStudio? Use **Profvis**
- are you submitting a job with an R script? Use **Rprof**

General usage:

```
Rprof()
  What <- Ever()
  Code <- You()
  Want <- ToRun()
Rprof(NULL)
profile.info <- summaryRprof()
```

```
library(profvis)

profvis({
  What <- Ever()
  Code <- You()
  Want <- ToRun()
})
```

Profiling in R

Profvis in **RStudio**,

➜ the profile visualization is **automatically generated** and you can navigate on it;

➜ you can **save the output** in *.html* format to share it;

➜ some of the info you can see:

◆ total time

◆ time and memory by function (if no info it is considered instantaneous)

◆ *flame graph:* the horizontal direction represents time in milliseconds, and the vertical direction represents the call stack.

Profile of *Urban background sheme* step in CALIOPE-urban:
https://earth.bsc.es/gitlab/ac/caliope-urban/uploads/df563c53d82d35dccd5c131e7606dfef/profile.html

After detecting the bottleneck of this code, the code was improved with using in-house R package **multiApply** achieving **53 % of speed increase**.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Thanks for your attention
Any questions?

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación