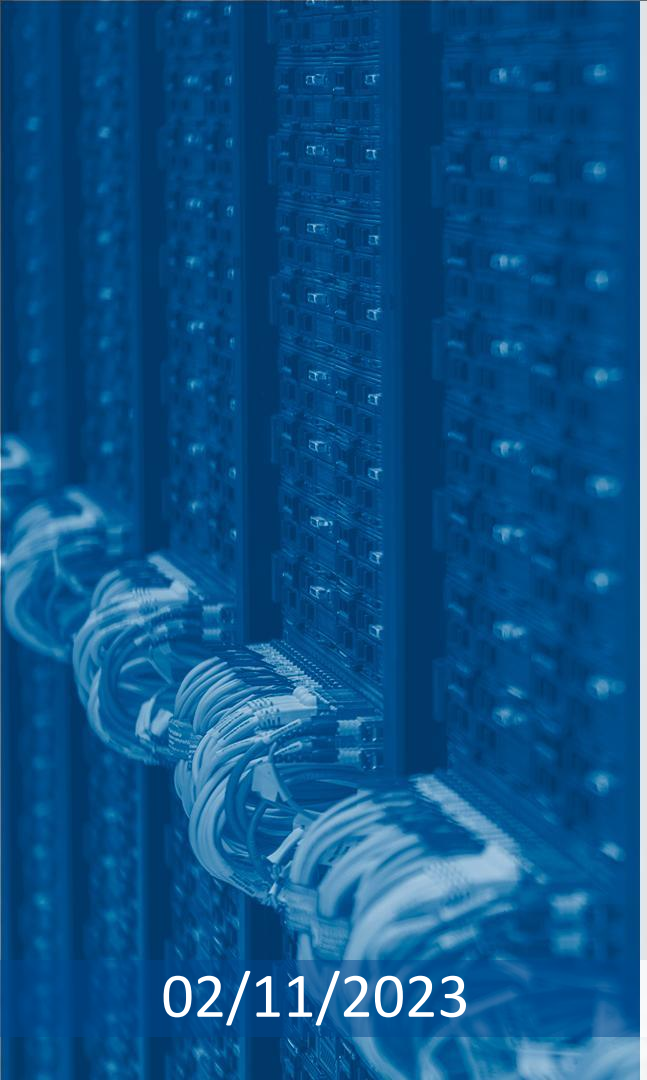


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

R user meeting

An-Chi Ho, Eva Rifà, Victòria Agudetse

contributor: Giovenale

02/11/2023

Agenda

1. Ice-breaker: lintr
2. News
    - General
    - s2dv
    - CSTools
    - CSIndicators
    - esviz
    - SUNSET
3. Presentation: vdiffr [Giovenale]
4. Q&A

Ice-breaker

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

lintr: Are your scripts linted?

lintr

“Checks adherence to a given style, syntax errors and possible semantic issues. Supports on the fly checking of R code edited with 'RStudio IDE', 'Emacs', 'Vim', 'Sublime Text', 'Atom' and 'Visual Studio Code'.”

Introduction:
https://cran.r-project.org/web/packages/lintr/vignettes/lintr.html

General

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

BSC Training (former PATC)

In the BSC Training, we had an R tools hands-on session. You can find the slides on earth wiki, and the hands-on materials on GitLab.

(1) Use **startR** to load and pre-process data: https://earth.bsc.es/gitlab/es/startR/-/blob/master/inst/doc/tutorial/PATC2023/

(2) Data assessment with **CSTools** and **s2dv**: https://earth.bsc.es/gitlab/external/cstools/-/tree/master/inst/doc/tutorial/PATC2023

(3) **SUNSET**: https://earth.bsc.es/gitlab/vagudets/bsc-trainings-r/-/tree/main/earth_sciences_simulation_environments/2023/sunset

Be aware of the impact of different R versions

R version 4.1.2 (2021-11-01) -- "Bird Hippie"

```
>  c(F,T) || F
[1] FALSE
```

R version 4.2.1 (2022-06-23) -- "Funny-Looking Kid"

```
> c(F,T) || F
[1] FALSE
Warning message:
In c(F, T) || F : 'length(x) = 2 > 1' in coercion to 'logical(1)'
```

R version 4.3.1 (2023-06-16) -- "Beagle Scouts"

```
>  c(F,T) || F
Error in c(F, T) || F : 'length = 2' in coercion to 'logical(1)'
```

s2dv

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

GetProbs() abs_thresholds

New parameter ‘abs_thresholds’ in GetProbs() to take the pre-calculated absolute thresholds (quantiles). So, instead of using relative thresholds to calculate quantiles, the function skips this step and directly uses the abs_thresholds input to calculate forecast probabilities.

```
res1 <- GetProbs(data = data, time_dim = 'sdate', memb_dim = 'ensemble',
                 prob_thresholds = NULL, abs_thresholds = c(-0.2, 0.3))

abs_thr <- array(c(-0.2, 0.3) + rnorm(40) * 0.1, dim = c(cat = 2, sdate = 20))
res2 <- GetProbs(data = data, time_dim = 'sdate', memb_dim = 'ensemble',
                 prob_thresholds = NULL, abs_thresholds = abs_thr, bin_dim_abs =
'cat')
```

CSTools

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

CSTools new release 5.1.0 (17-10-2023)

Fixes

- Calibration() show warnings in atomic function when multiple cores are used #130
- PlotForecastPDF fix background color in different machines. #131
- Correct CST_Subset indices for metadata. #185
- CST_Analogs: Added sdate_dim parameter and improve initial checks. #187
- Remove CST_Anomaly repeated checks in order to accept no member dimensions in obs. #188
- CST_SaveExp developments: improve warning, save metadata correctly. #132

Development

- PlotWeeklyClim to allow years outside the reference period and allow setting y limits. #118 #124
- PlotCombinedMap() has upper triangle_end; the color bars can have different breaks. #125
- New print method #128
- CST_MultiEOF development treat spatial NAs. #127
- New function CST_Start(). #126

CSTools new release 5.1.1 (19-10-2023)

Fixes

- Added **startR namespace in all CST_Start** calls of the vignettes and sample data #139

```
repos_exp <- '/esarchive/exp/ecmwf/system5c3s/monthly_mean/$var$_f6h/$var$_$sdate$.nc'
exp <- CST_Start(dataset = repos_exp, var = clim_var,
                 member = startR::indices(1:10), sdate = dateseq, ftime = startR::indices(2:4),
                 lat = startR::values(list(lat_min, lat_max)),
                 lat_reorder = startR::Sort(decreasing = TRUE),
                 lon = startR::values(list(lon_min, lon_max)),
                 lon_reorder = startR::CircularSort(0, 360),
                 synonims = list(lon = c('lon', 'longitude'), lat = c('lat', 'latitude'),
                                 member = c('member', 'ensemble'), ftime = c('ftime', 'time')),
                 transform = startR::CDORemapper,
                 transform_params = list(grid = 'r256x128', method = 'bilinear'),
                 transform_vars = c('lat', 'lon'),
                 return_vars = list(lat = NULL, lon = NULL, ftime = 'sdate'), retrieve = TRUE)
```

CSTools next developments

Fixes

- CST_SaveExp doesn't save well ‘time_bounds’ (`object$attrs$time_bounds`) #137
- CST_Start use startR functions without the explicit namespace #140

Developments

- Add time frequency attribute in as.s2dv_cube #133
- Improve s2dv_cube() function #123

…

CSIndicators

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Change time_dim default value to ‘time’

- Changed parameter to **identify the leadtime dimension name** default value to ‘time’

```
CST_PeriodAccumulation <- function(data, start = NULL, end = NULL,
                                   time_dim = 'time', na.rm = FALSE,
                                   ncores = NULL) {
PeriodAccumulation <- function(data, dates = NULL, start = NULL, end = NULL,
                               time_dim = 'time', na.rm = FALSE,
                               ncores = NULL)
```

status: In branch master

Check issues: https://earth.bsc.es/gitlab/es/csindicators/-/issues/35

Corrected subset of s2dv_cube coordinates

- Added subset of coordinates to make ‘s2dv_cube’ elements consistent.

Where is the consistency?

➢ The s2dv_cube elements: `$data`, `$coords`, `$dims` and `$attrs$Dates` need to match always.

➢ If in the result the dimensions of `$data` are different than the original object, we need to update manually inside the functions also the dimensions of other elements.

status: In branch master

Check issues: https://earth.bsc.es/gitlab/es/csindicators/-/issues/33

Added ClimProjDiags dependency

➢ Readded **ClimProjDiags** dependency with **Subset** → Removed auxiliary functions added due to the dependency issue of package 'climdex.pcic'

ClimProjDiags dependencies

- **ClimProjDiags::Subset()**: AccumulationExceedingThreshold, PeriodAccumulation, PeriodMax, PeriodMin, PeriodVariance, PeriodMean, QThreshold. SelectPeriodnData, TotalSpellTimeExceedingThreshold, TotalTimeExceedingThreshold.

Other dependencies

- Packages multiApply, stats and utils

status: In branch master

Check issues: https://earth.bsc.es/gitlab/es/csindicators/-/issues/34

Substitute CST_Load by CST_Start

- Substituted CST_Load by CST_Start in **Agricultural Indicators** vignette.

```
CST_Load(var = 'prlr',
         exp = list(S5path_prlr),
         obs = list(path_ERA5prlr_CDS),
         sdates = sdates,
         lonmax = 353, lonmin = 352.25,
         latmax = 41.75, latmin = 41,
         storefreq = 'daily',
         leadtimemin = 1,
         leadtimemax = 214,
         nmember = 3,
         output = "lonlat",
         grid = "r1440x721",
         method = 'bicubic')
```

```
CST_Start(dataset = S5path_prlr, var = "prlr",
          member = startR::indices(1:3),
          sdate = sdates,
          ftime = startR::indices(1:214),
          lat = startR::values(list(lat_min, lat_max)),
          lat_reorder = startR::Sort(decreasing = TRUE),
          lon = startR::values(list(lon_min, lon_max)),
          lon_reorder = startR::CircularSort(0, 360),
          synonims = list(lon = c('lon', 'longitude'),
                          lat = c('lat', 'latitude'),
                          member = c('member', 'ensemble'),
                          ftime = c('ftime', 'time')),
          transform = startR::CDORemapper,
          ...)
```


status: In branch master

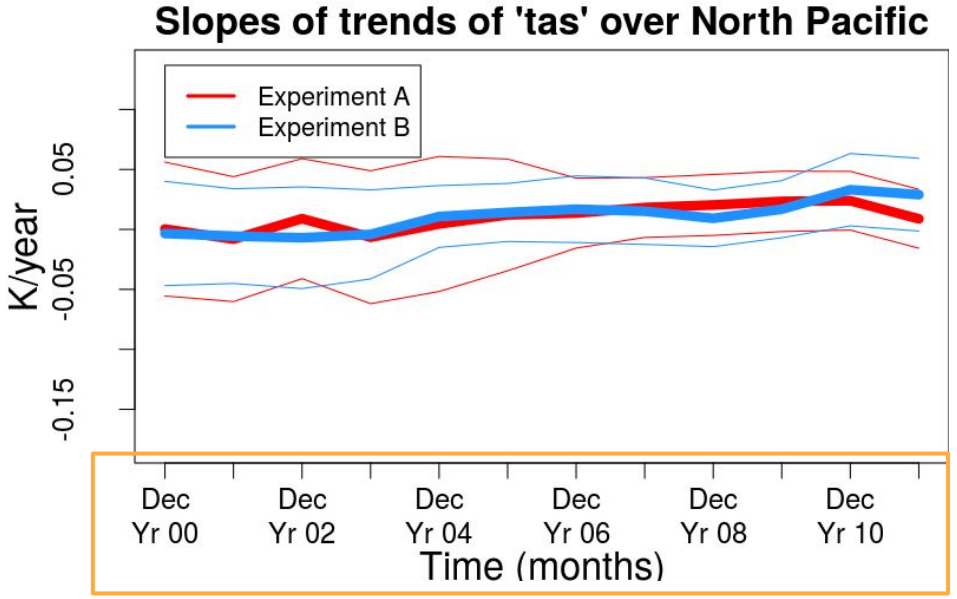
Check issues: https://earth.bsc.es/gitlab/es/csindicators/-/issues/37

esviz

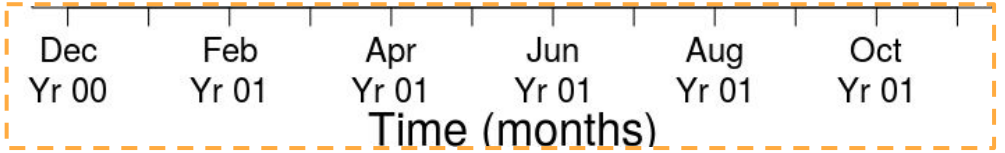
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Does PlotVsLTime() create wrong x-axis tick label?

PlotVsLTime() says "Plot a score along the forecast time with its confidence interval"
So, the x-axis should be forecast time, which is month, as the X label says. But the tick labels say "Yr"



Should be this?



New function names

See the function list in https://earth.bsc.es/gitlab/es/esviz/-/issues/1

PlotEquiMap → **Viz**EquiMap
PlotRobinson → **Viz**Robinson
PlotForecastPDF → **Viz**ForecastPDF
PlotMostLikelyQuantileMap → **Viz**MostLikelyQuantileMap
…

SUNSET

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Anomaly computation for the Downscaling use case

`CSTools::CST_Anomaly()`, which is used within the Anomalies module, uses the per-pair method to compute the climatology, and therefore it works **only for the case where the hindcast and the reference share the same grid**.

But sometimes the anomalies need to be computed before regridding, like in the downscaling use case. Now `Anomalies()` includes an option for simple anomaly computation (no per-pair, no cross-validation) when hcst and obs have different grids.

status: in master

Custom directives in Autosubmit

Custom slurm directives for jobs sent to the cluster can now be included in the Autosubmit section of the recipe. This allows the user to specify options such as the type of node, whether is should be requested exclusively, the job queue, etc.

```
auto_conf:
    script: example_scripts/test_scorecards_workflow.R
    expid: a6ae
    hpc_user: bsc32762
    wallclock: 03:00
    processors_per_job: 14
    platform: nord3v2
    custom_directives: ['#SBATCH --exclusive', '#SBATCH --constraint=medmem']
    email_notifications: yes
    email_address: victoria.agudetse@bsc.es
    notify_completed: yes
    notify_failed: no
```

status: in master

‘Corr’ metric name change

In the Skill module, the metric ‘corr’ has been renamed to ‘corr_individual_members’.

This metric returns the correlation coefficient between each individual member of the hindcast and its corresponding observations, which is the result of `s2dv::Corr()` when the parameter memb is set to `FALSE`.

The correlation coefficient for the ensemble mean is still named ‘enscorr’.

status: in master

New detailed use case: Hands-on tutorial

The "BSC Training Course: Earth Sciences Simulation Environments" (formerly known as PATC) took place on the 31st of October and included a session on R tools for Climate Forecast analysis, with hands-on tutorials for startR, CSTools, and a simple SUNSET use case.

You can find the slides here: Climate Forecast Analysis hands-on tutorial: R tools

And the tutorial here: MN4 SUNSET Hands-on Tutorial

We hope to add an updated version for Nord3v2 and workstation/hub to the SUNSET repository soon.

[User presentation] by Giovenale

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Testing R functions that generate graphics

The “vdiffr” package

is a “testthat” extension that makes it easy to automatically check code that generates R graphics.

1. Suggest “vdiffr”

In your R package folder, you need to suggest:

```
usethis::use_package(package = "vdiffr", type = "Suggests")
usethis::use_package(package = “testthat", type = "Suggests")
```

2.Program your own plotting function:

```
---------------------------------------------------------------------------------
plot_ts <- function(data,
                    var = "",
                    time = "",
                    area = NULL,
                    aggregate_space = NULL,
                    aggregate_time = NULL,
                    aggregate_fun = "mean",
                    main = NULL,
                    ylab = NULL, xlab = NULL,
                    panel = FALSE,
                    highlight = NULL,
                    ...) { ... }
---------------------------------------------------------------------------------
```

3.Test R functions that generate graphics

```
usethis::use_test("plot_ts")
-------------------------------------------------------------------------------------------------------
test_that("plot_ts", {

  vdiffr::expect_doppelganger(“plot_ts_test1",plot_ts(data=data_brazil,
                                  var="dengue_cases",
                                  aggregate_space="state_code",
                                  time="date",
                                  aggregate_time="month",
                                  panel=TRUE,
                                  aggregate_fun=mean))
})
-------------------------------------------------------------------------------------------------------
```

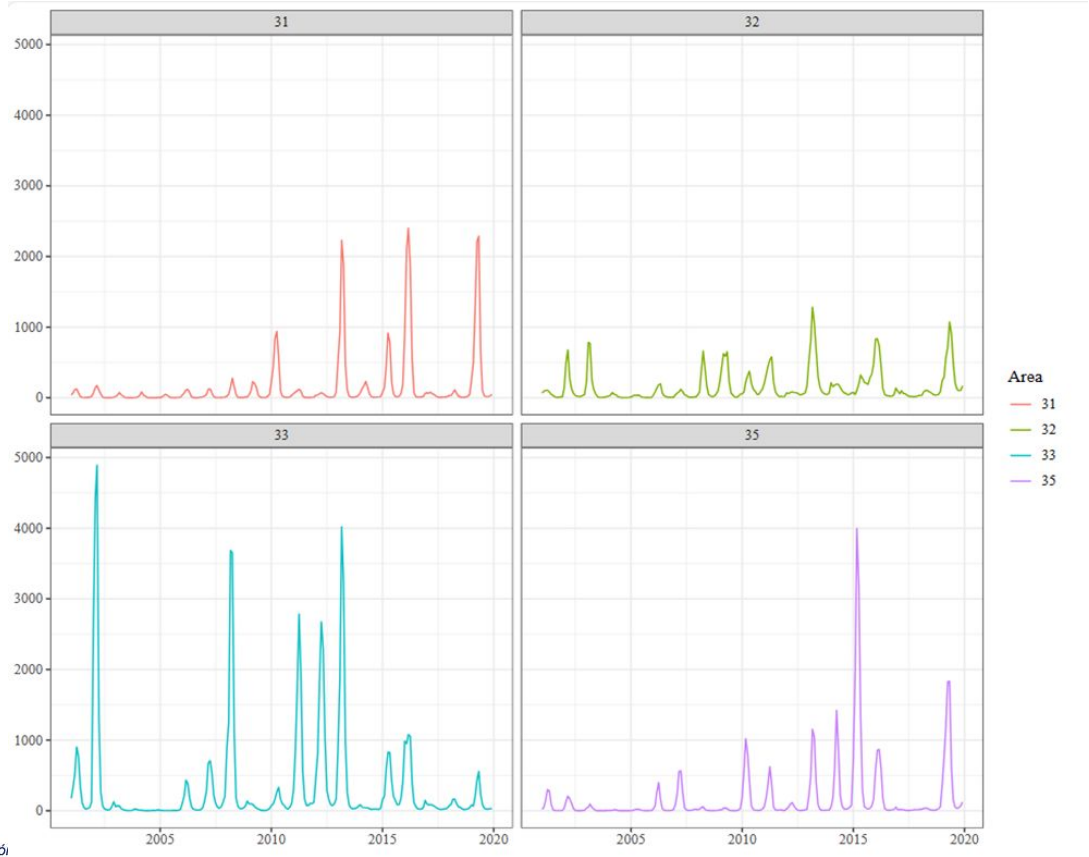
4. Test

---

run devtools::test()

The first time it is run, expect_doppelganger() generates a reproducible SVG file that represents the expected appearance of your plot and saves it inside your test folder. After that, the generated SVG is compared to the saved version and any mismatch is reported as a failure in the testthat output.

5. The SVG generated by “plot_ts”



EXTRA:1

We test if “expect_doppelganger” detects differences, by modifying the unit test we wrote before

```
-----------------------------------------------------------------------------------------------------------
test_that("plot_ts", {

  vdiffr::expect_doppelganger(“plot_ts_test1",plot_ts(data=data_brazil,
                                var="dengue_cases",
                                aggregate_space="state_code",
                                time="date",
                                aggregate_time="month",
                                panel=FALSE,
                                aggregate_fun=mean))
})
-----------------------------------------------------------------------------------------------------------
```

EXTRA:2

We re-run the test

---

run devtools::test()

```
> devtools::test()
i Testing idextreme.test
✓ | F W S  OK | Context
✖ | 1       0 | plot_ts [0.6s]
──────────────────────────────────────────────────────────────────────────────
Failure (test-plot_ts.R:3): plot_ts
Snapshot of `testcase` to 'plot_ts/ts-agg-test1.svg' has changed
Run `testthat::snapshot_review('plot_ts/')` to review changes
Backtrace:
 1. vdiffr::expect_doppelganger(...)
      at test-plot_ts.R:3:2
 3. testthat::expect_snapshot_file(...)
──────────────────────────────────────────────────────────────────────────────
```

EXTRA:3

Checking the differences…

Thanks for joining

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Next meeting: 7th December, 12h