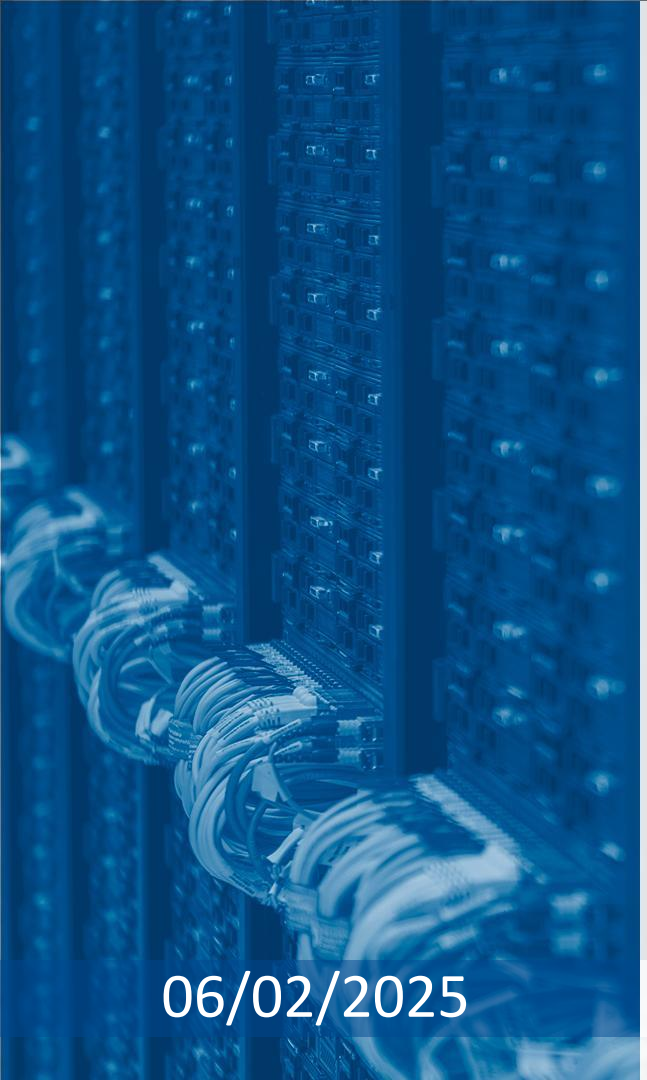


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

R user meeting

06/02/2025

Victòria A., Ariadna B., Theertha K., Raúl C.

Agenda

1. Ice-breaker: Variable assignment in R
2. News
    - General
    - CSTools
    - esviz
    - SUNSET
    - GHR Packages
3. User presentation: Unit testing (Raúl Capellán)
4. Q&A

Ice-breaker: Assignment operators in R

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Assignment operators in R

One of the unusual features of R compared to other commonly used scripting languages is the use of “<-” instead of “=” for variable assignment:

```
# We use:
a <- 10
# Instead of:
a = 10
```

But… **why?**

Is there a difference between “=” and “<-”?

Are there any other assignment operators in R?

Why <-?

1. **Historical reasons (mostly):**

In S (the precursor of R) and other older programming languages, the character "<-" was used for variable assignment before "=" was popularized. "=" was used as a comparison operator, because "==" did not exist yet.

"<-" was chosen as the assignment operator with R for backwards-compatibility with S code.

2. **Readability:**

<- is recommended in R style guides for readability, because it shows the direction of the assignment:

```
# a <- b is more explicit than a = b
a <- 12
b <- 10
b = a # b <- a or b -> a?
```

3. **They are not exactly the same.**

Is there any difference between “=” and “<-”?

There are some subtle differences between “=” and “<-” that are related to **environments**. “=” Should be used when passing function arguments, whereas “<-” will do assignment on the top level:

```
# ‘x =’ is used to define the value of the ‘x’ parameter in the mean() function:
> mean(x = c(1, 2, 3))
[1] 2
> x
Error: object 'x' not found

# Here we are actually defining a new variable ‘x’
> mean(x <- c(1, 2, 3)) # Note: don’t do this
[1] 2
> x
[1] 1 2 3
```

Is there any difference between “=” and “<-”?

The difference is clearer in this example where we use ‘y’ instead of ‘x’:

```
# This:
> mean(y <- c(1, 2, 3))
# Is the same as these two steps:
> y <- c(1, 2, 3)
> mean(x = y)

# Whereas doing the same with ‘=’ results in an error:
> mean(y = c(1, 2, 3))
Error in mean.default(y = c(1, 2, 3)) :
  argument "x" is missing, with no default
```

Are there any other assignment operators in R?

★ **The "<<-" operator** is used for global assignment in functions. A variable defined with "<<-" within a function will also be defined in the global environment:

```
> my_fun <- function(x) {
    y <<- x*25
    x <- x + y
    return(x)
  }
> z <- my_fun(x = 10)
> z
[1] 260
> y
[1] 250
```

★ It is also possible (but usually not recommended) to use "->" and "->>":

```
> rnorm(5) -> my_array
> my_array
[1] -0.4179281 -0.5568826 -0.7010551 -0.3259824  0.526703
```

Are there any other assignment operators in R?

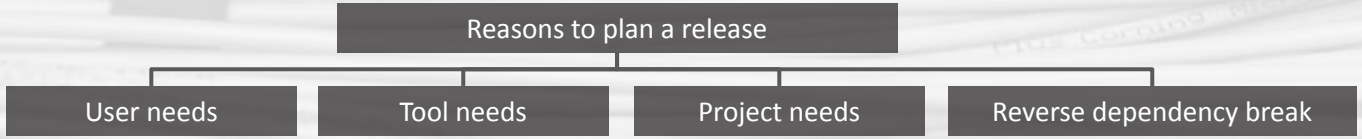
To know more:

- Assignment Operators (R Documentation)
- Why do we use arrow as an assignment operator? - Colin Fay
- The R Inferno - Burns Statistics

General

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

R TOOLS ROADMAP 2025



	General Direction	Q1 2025	Q2 2025	Q3 2025	Long term
s2dv (v2.1.0)	Enhance statistical methods	Include new statistical methods and improve the existing ones			Efficiency improvement
startR (v2.4.0)	- Maintain the flexibility while improving user experience - Focus on using Autosubmit as workflow manager	- Ensure that Compute() works on the Hub and with the new HPC machines - Update use cases and documentation	- Investigate limitations in data size - Deprecate ecFlow as workflow manager		- Load and interpolate irregular grid - Load GRIB - Multiple steps Code refactoring
CSTools (v5.2.0)	Increase the general methods and enhance the relevance of the s2dv_cube	- Include new methods for s2dv_cube - Enhance unit tests		Add new use cases	Convert s2dv_cube to xarray/other objects
CSIndicators (v1.1.1)	Facilitate the calculation of new indicators through existing functionalities	- Minor release with function improvements	- Health Indicators integration - Add use cases		Add new vignettes for bioclimatic and drought indicators
esviz (WIP)	Establish a common function structure for various needs	- Standardize function parameters - Time series and scorecard functions	First release		- Reduce device dependency - Improve different projections
ClimProjDiags **multiApply** **easyNCDF**	No planned development. Only maintenance.				
General	User meeting, user support, shiny server maintenance While developing the packages for SUNSET or any specific project, it's important to keep the packages as general and individual as possible.				

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Made by Theertha Kariyathan, Ariadna Batalla, Victoria Agudetse (Dec. 2024)

CONTRIBUTING.md files in packages

We have added a CONTRIBUTING.md file to the R packages, which cointains a summary of contribution rules, including:

- ★ Questions and bug reports (who to ask and how)
- ★ Steps to contribute
- ★ How to add new functions
- ★ Style guide with examples

Example: https://earth.bsc.es/gitlab/es/startR/-/blob/master/CONTRIBUTING.md

CSTools

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

New function CST_Summary()

Develop function to provide a summary of the data in the s2dv_cube

Features :

- Months that have been loaded
- Range of the dates that have been loaded
- Object dimensions
- Basic statistical summary of the data
- Variables that have been loaded, along with their units
- Missing files
- Number of NAs per time dimension and latitude/longitude dimensions

New function CST_Summary()

CST_summary(data_cube, loaded_files = TRUE, na_dim = TRUE)

```
[1] "2025-01-30 15:34:43 DATA SUMMARY:"
[1] "2025-01-30 15:34:43 data_cube months: February"
[1] "2025-01-30 15:34:43 data_cube range: Feb 01 2017 to Feb 01 2017"
[1] "2025-01-30 15:34:43 data_cube dimensions:"
[1] "2025-01-30 15:34:43      dat      var    sdate ensemble     time      lat      lon "
[1] "2025-01-30 15:34:43        2        2        1        1        1        5        5 "
[1] "2025-01-30 15:34:44 Statistical summary of the data in  data_cube :"
[1] "2025-01-30 15:34:44 Variable:  tas  (units:  K )"
[1] "2025-01-30 15:34:44    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's "
[1] "2025-01-30 15:34:44   247.2   247.3   247.3   248.1   248.8   249.7       4 "
[1] "2025-01-30 15:34:44 Variable:  sfcWind  (units:  m s**-1 )"
[1] "2025-01-30 15:34:44    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. "
[1] "2025-01-30 15:34:44   5.769   5.811   6.111   6.142   6.504   6.572 "
[1] "2025-01-30 15:34:44 Dimensions with NA values"
[1] "2025-01-30 15:34:44 dat: 2 var: 1 sdate: 1 ensemble: 1 time: 1 lat: 3,5 lon: 2,3"
[1] "2025-01-30 15:34:44 Number of NAs per dimension"
[1] "2025-01-30 15:34:44 dat: 4 var: 4 sdate: 4 ensemble: 4 time: 4 lat: 2,2 lon: 2,2"
[1] "2025-01-30 15:34:44 Loaded files"
[1] "2025-01-30 15:34:44 /esarchive/exp/ecmwf/system4_m1/monthly_mean/tas_f6h/tas_20170101.nc"
[2] "2025-01-30 15:34:44 /esarchive/exp/ecmwf/system5_m1/monthly_mean/tas_f6h/tas_20170101.nc"
[3] "2025-01-30 15:34:44 /esarchive/exp/ecmwf/system4_m1/monthly_mean/sfcWind_f6h/sfcWind_20170101.nc"
[4] "2025-01-30 15:34:44 /esarchive/exp/ecmwf/system5_m1/monthly_mean/sfcWind_f6h/sfcWind_20170101.nc"
[1] "2025-01-30 15:34:44 ------------------------------------------"
```

issue: https://earth.bsc.es/gitlab/external/cstools/-/issues/156

status: in branch dev-cst_summary

esviz

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

New caption parameters in VizEquiMap()

Two new parameters in **`VizEquiMap()`** to add a caption at the bottom left of the plot.

➔ **`caption`**

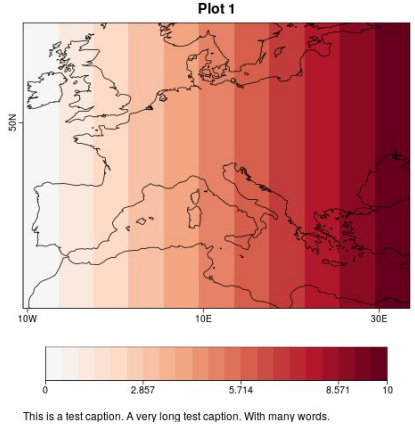
- Character string with the caption text.
- Captions with multiple lines can be constructed using string manipulation functions like **`paste0()`** and **`"\n"`** to indicate line breaks:

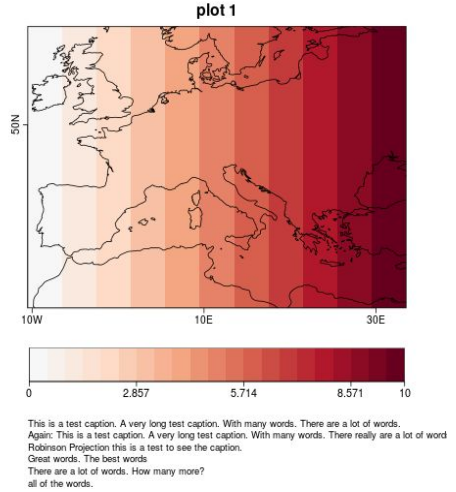
```
caption = paste0("This is a test caption.", "\n",
                 "This is a new line.")
```

- Default is `NULL`.

➔ **`caption_size`**

- Scale factor for the figure caption.
- Default is 0.8 (1 if **`vertical = TRUE`**).





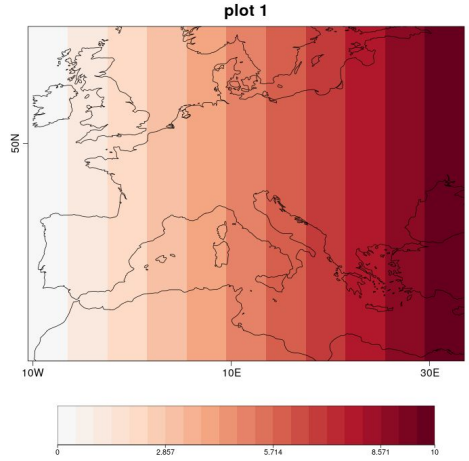
New vertical colorbar parameter in VizEquiMap()

A new parameter in **`VizEquiMap()`** to change the orientation of the color bar.

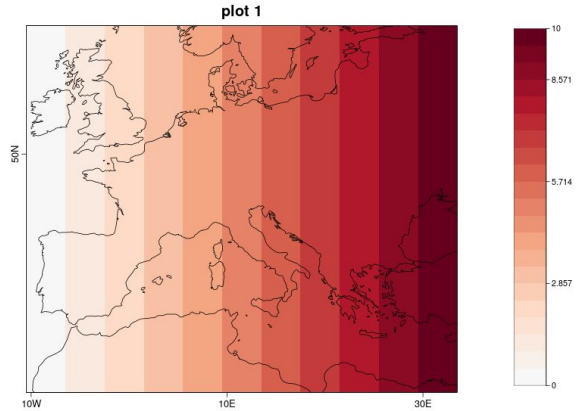
➔ **`vertical`**

- `TRUE/FALSE` for vertical/horizontal color bar.
- Default is `FALSE` (horizontal).
- Parameters 'width' and 'height' might need to be modified to accommodate the vertical colour bar.

`vertical = FALSE`

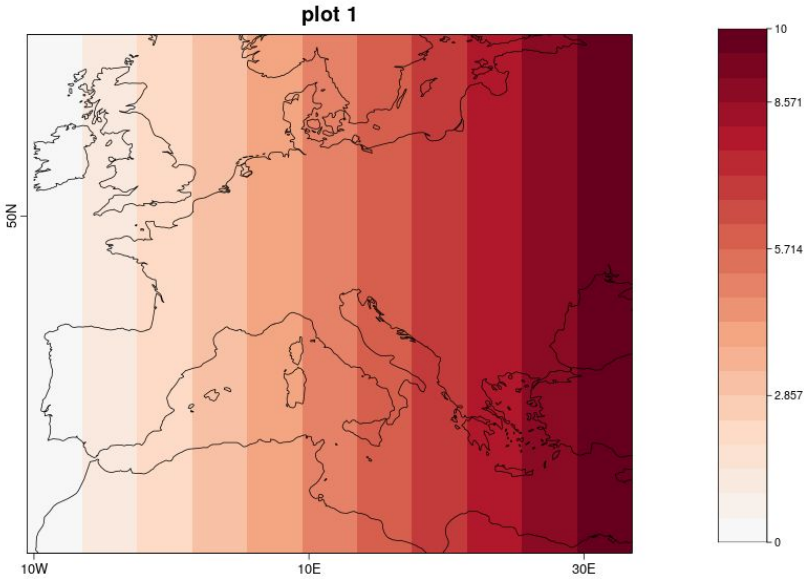


`vertical = TRUE`



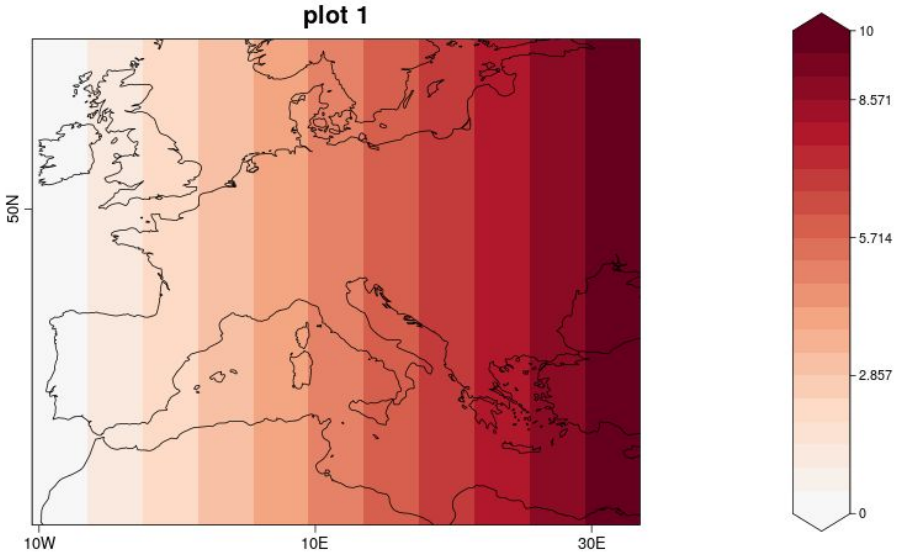
Caption + Vertical color bar

One-line caption



This is a test caption. A one line test caption. With many words. There are a lot of words.

Multiple lines caption (+ triangle_ends)



Robinson Projection this is a test to see the caption.
A super long caption. A very long caption with a lot of words that are wording all together.
Great words. The best words
There are a lot of words. How many more?
Who knows, words are the best.
there needs to be more words
sorry there needs to be even more words
all of the words.

SUNSET

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Fair vs. not-Fair Brier Skill Score

The **Brier Skill Scores** (bss10 and bss90) were being computed as 'fair' (assuming an infinite ensemble size) by default in the Skill module.

This has now been fixed, and requesting `'bss10'` and `'bss90'` will return the "unfair" version of the metrics. The new metrics `'fbss10'` and `'fbss90'` are now available as the fair versions of the Brier Skill Score.

For more details, see the documentation of s2dv::RPSS().

issue: https://earth.bsc.es/gitlab/es/sunset/-/issues/165
status: in master

Visualization module improvements

New parameters for Visualization:

1. `forecast_method: mean, median, IQR`

   New methods for the forecast map: **ensemble mean, median and inter-quartile range**. One or more of these statistics can be specified and the Visualization() will iterate through them.

2. `shapefile: /esarchive/scratch/<user>/<my_shapefile>.shp`

   Path to a shapefile to be incorporated into the maps.

3. `logo: tools/BSC_logo_95.png`

   The path to a PNG image, such as an institutional logo, to add to the bottom right of plots.

issue: https://earth.bsc.es/gitlab/es/sunset/-/issues/148

status: in master

Visualization module improvements

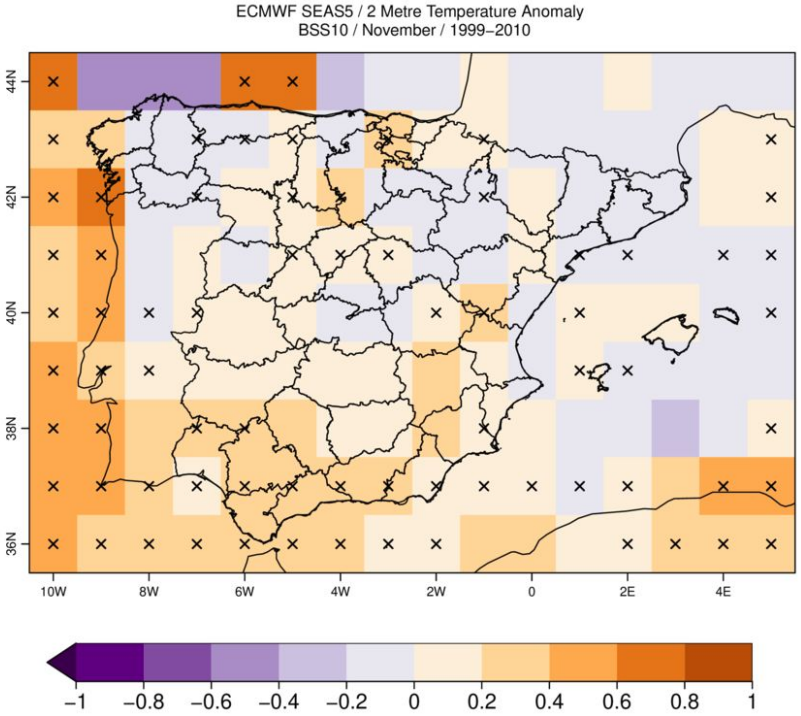
Other improvements include:

- Changes to function names.
- plot_most_likely_terciles() accepts probabilities in array form
- Captions now available for all projections
- Code refactoring
- Removal of unneeded data in captions

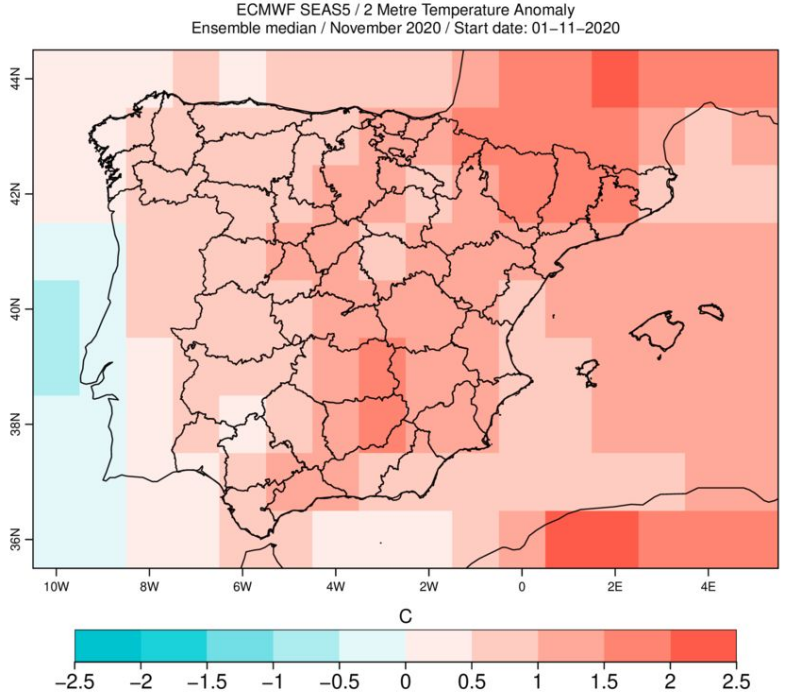
You can see the full list of changes in the merge request:
https://earth.bsc.es/gitlab/es/sunset/-/issues/148

issue: https://earth.bsc.es/gitlab/es/sunset/-/issues/148
status: in master

Visualization module improvements



Nominal start date: 1st of November
Forecast month: 01
Reference: ERA5
alpha = 0.05

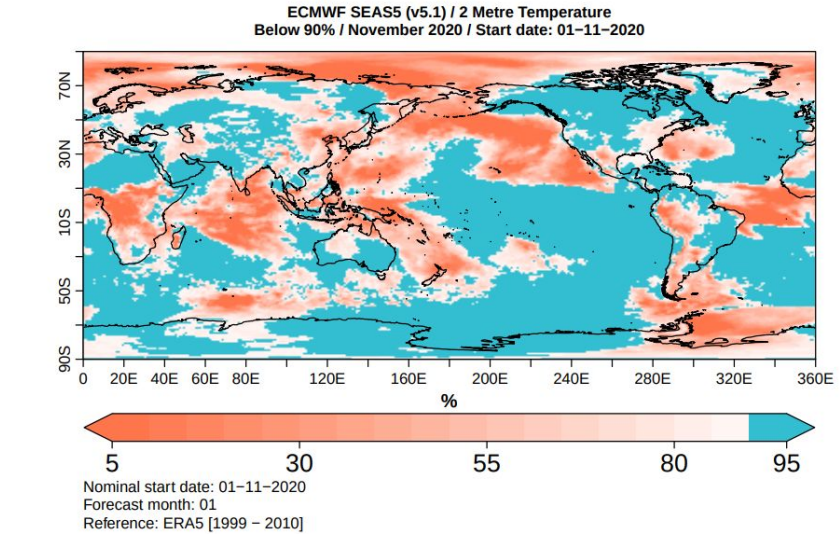


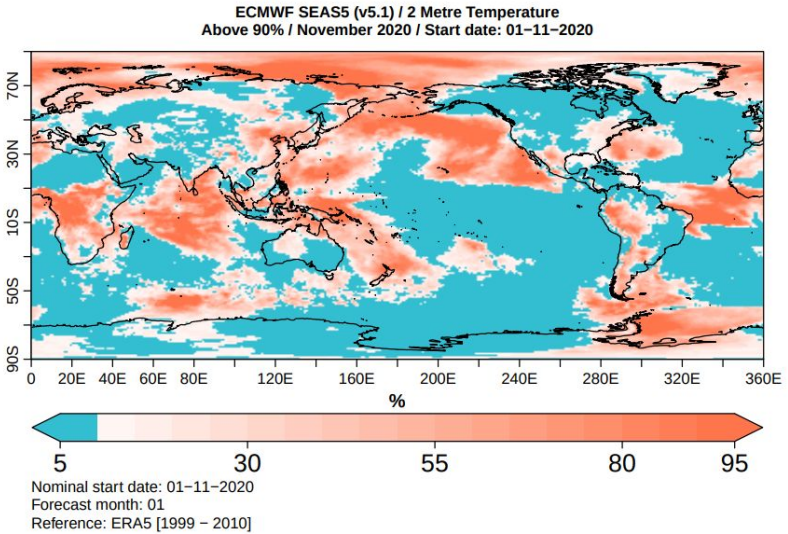
Nominal start date: 01−11−2020
Forecast month: 01
Reference: ERA5
Units: C

New visualization function: plot_extreme_probs()

New function in the Visualization module to plot probabilities below and above a percentile:

```
Recipe$Analysis$Workflow$Probabilities$percentiles: [[9/10]]
```

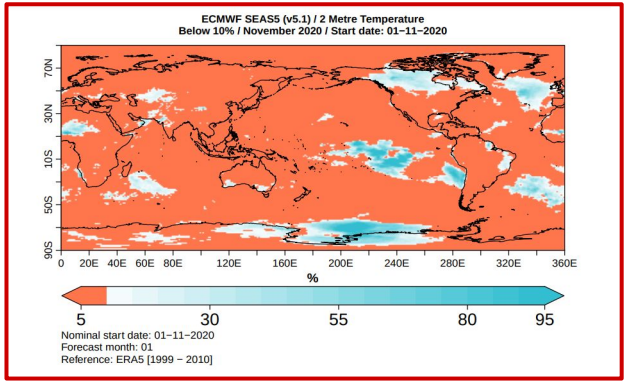


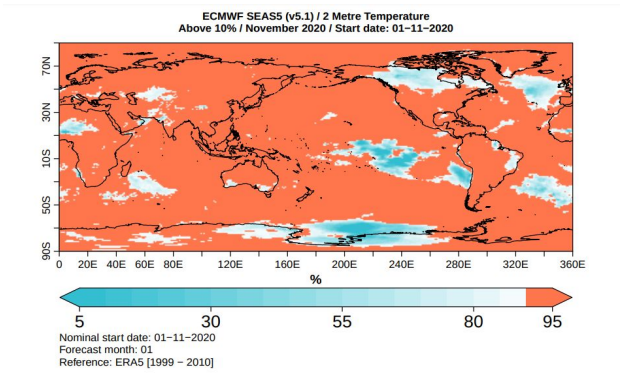


New visualization function: plot_extreme_probs()

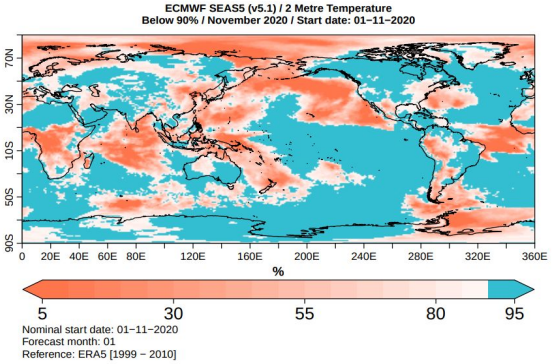
`Recipe$Analysis$Workflow$Probabilities$percentiles: [[1/3, 2/3], [1/10], [9/10]]`

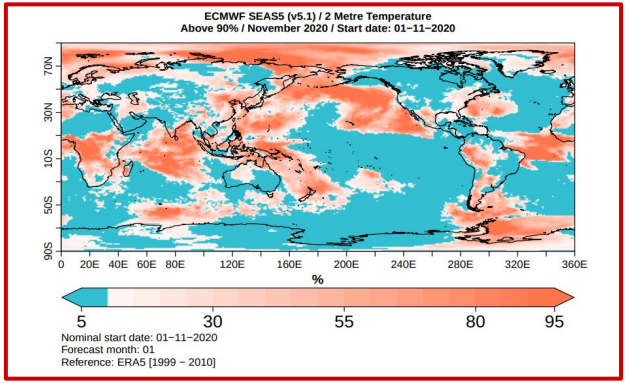
[1/10]:





[9/10]:





MR: https://earth.bsc.es/gitlab/es/sunset/-/merge_requests/184

status: in branch dev-plot_extreme_probs

GHR packages

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

ghr_packages

Objectives

01 **Existing packages**

- clim4health
- GHRexplore
- GHRmodel

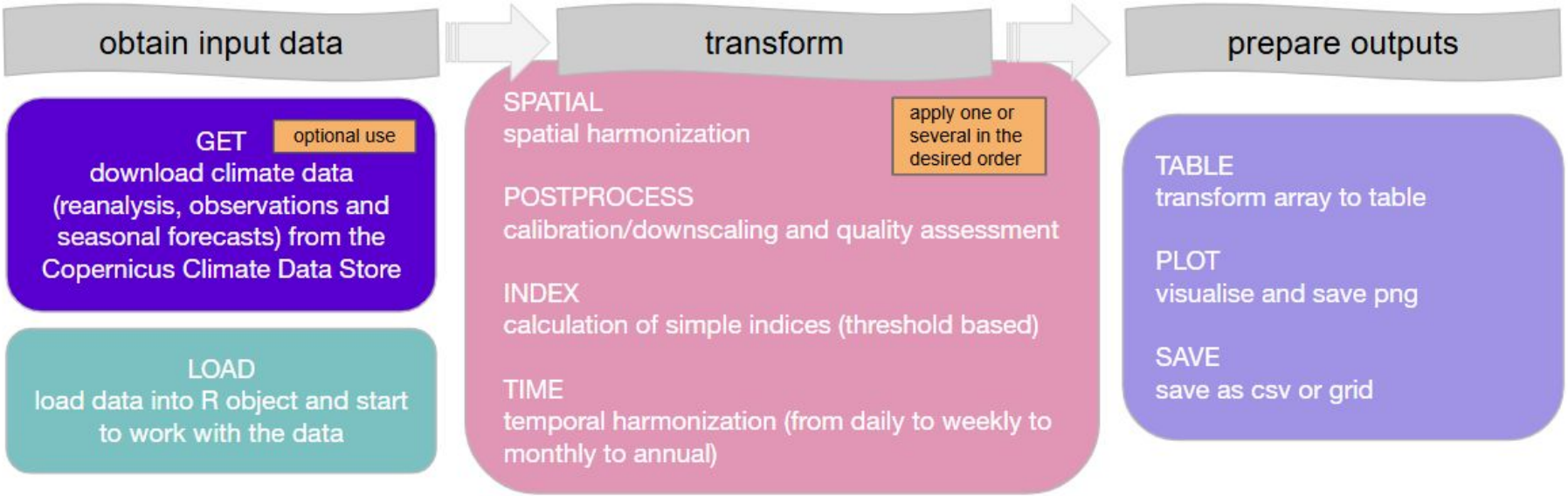
02 **New packages**

- GHRdashboard
- GHRpredict

clim4health

clim4health

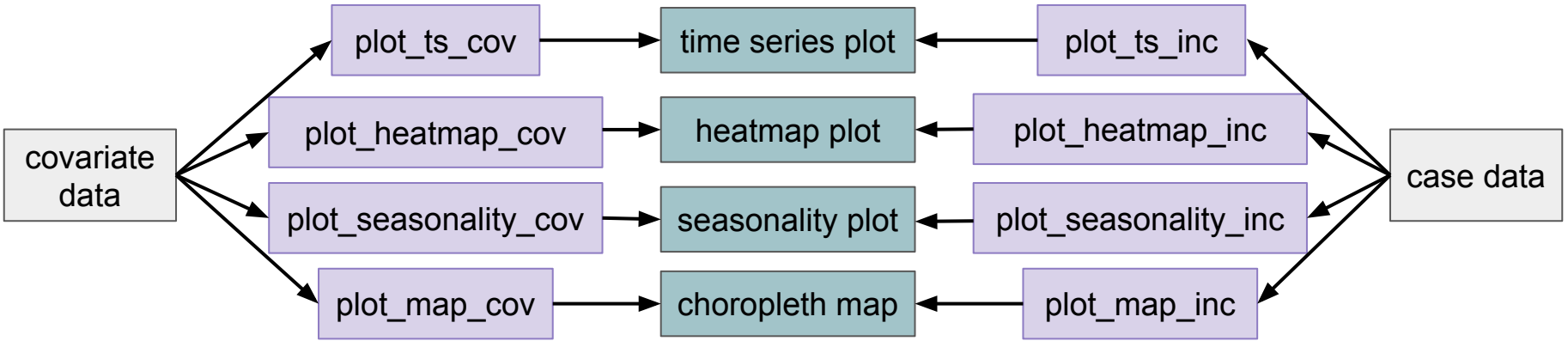
HARMONIZE



Repo: https://earth.bsc.es/gitlab/ghr/harmonize-clim4health

ghr_explorer

This package is designed to explore covariate and case data to identify the possible drivers and predictors of disease risk, by visualizing these variables over time and space.



Repo: https://earth.bsc.es/gitlab/ghr/ghrexplore

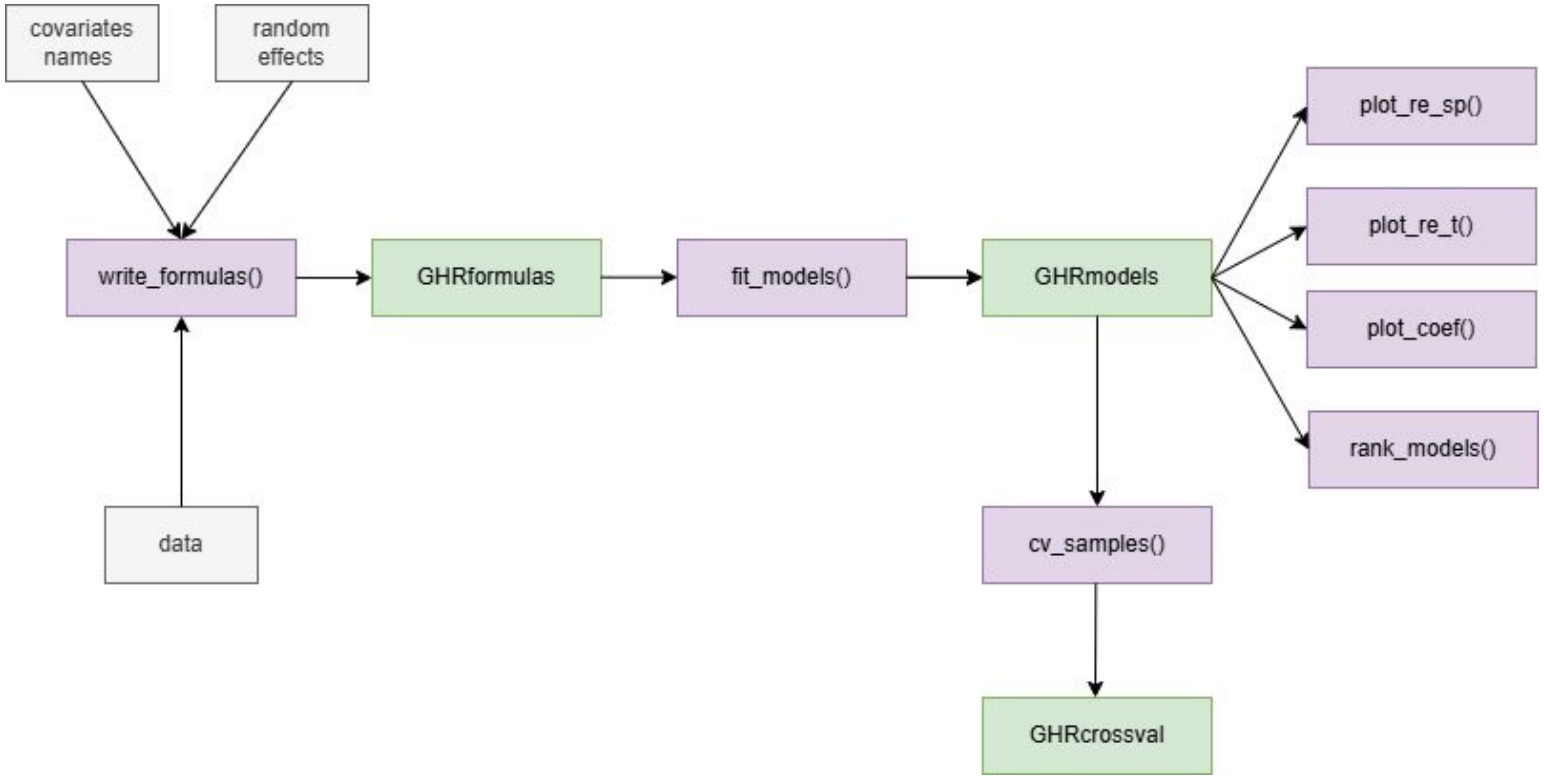
ghr_model

This package is designed to identify key drivers and predictors of disease risk, by offering functions for Bayesian hierarchical spatio-temporal models of varying complexity.

The package depends on R-INLA package for model fitting. Model outputs are evaluated through metrics such as DIC, WAIC, CRPS, and MAE. Additional functionalities include tools for extracting model outputs, simplifying interpretation, plotting both fixed and random effects and identifying optimal lagged associations between covariates and outcomes. Users can also perform cross-validation to determine the best-performing models.

Repo: https://earth.bsc.es/gitlab/ghr/ghrmodel

ghr_model


covariates names
random effects
write_formulas()
GHRformulas
fit_models()
GHRmodels
plot_re_sp()
plot_re_t()
plot_coef()
rank_models()
data
cv_samples()
GHRcrossval

User presentation

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Unit testing

URL presentation: https://drive.google.com/drive/folders/1d7EDnatONU2gMmP4VOr2qn7_oDboGQ61?usp=drive_link

Q&A


BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Thanks for joining

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Next meeting: April 3rd