Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

R tools user meeting

07/11/2019

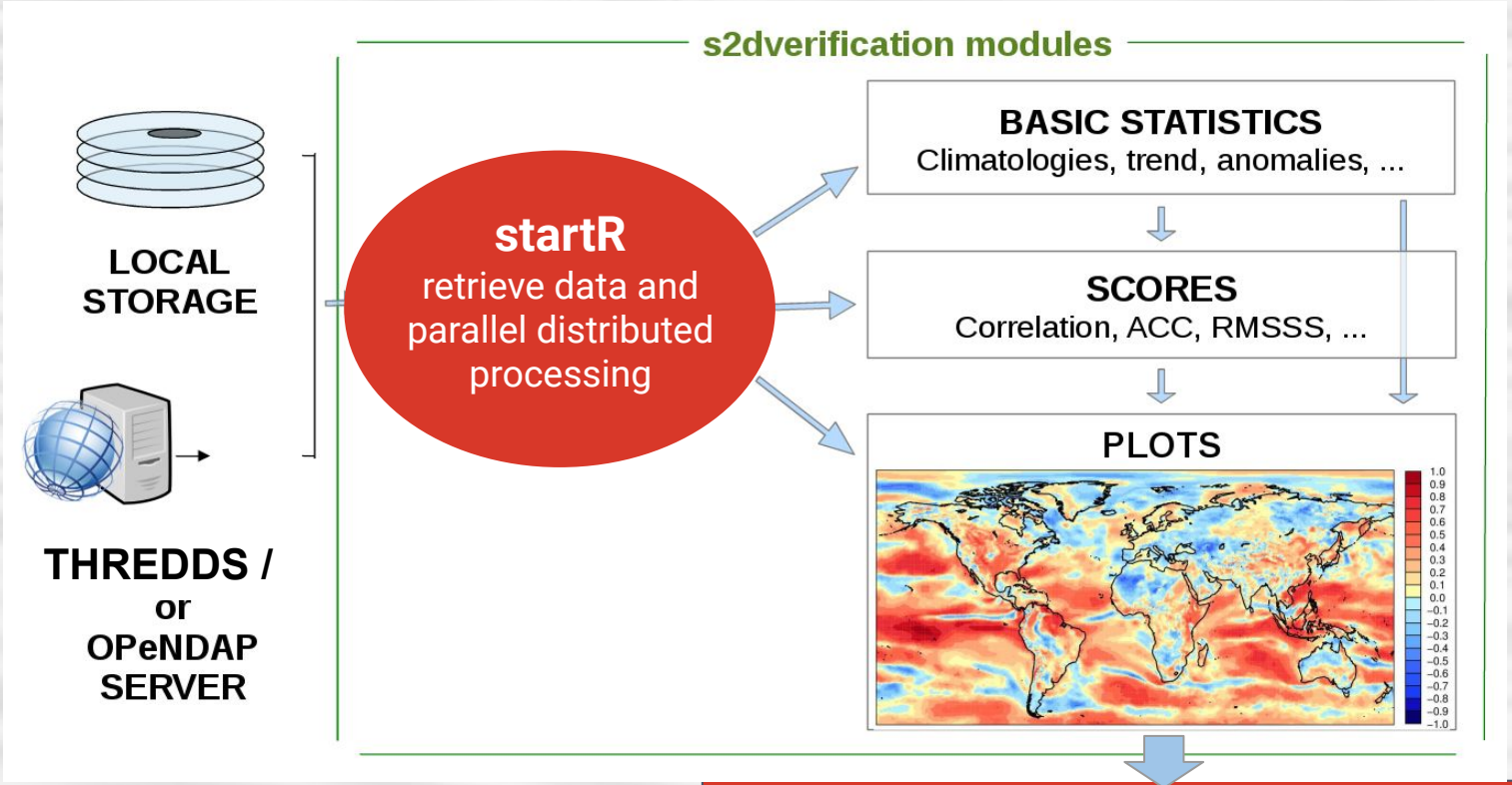
An-Chi Ho and Núria Pérez-Zanón

Package feature

- **s2dverification**: Functions to load, analyze, and verify seasonal to decadal forecast
- **CSTools**: Functions to be used in Climate Services (from postprocessing to plotting tools)
- **ClimProjDiags**: A set of tools to compute indices (including formatting)
- **startR**: Load large dataset and do parallel data processing on HPCs
- Other helper packages: **easyNCDF**, **multiApply**

Find vignettes, code, documentation, issue tracker and much more! on GitLab!

R tool scheme



CSTools

Climate Services Tools - MEDSCOPE Toolbox

MEDiterranean Services Chain based On Climate PrEdictions

ClimProjDiags

Introduction to ClimProjDiags

https://earth.bsc.es/gitlab/es/ClimProjDiags

- It was created during the **MAGIC C3S contract**
- It is used in **ESMValTool recipes**
- It was **named** during a department **R user meeting**
- It is following the **multiApply paradigm** for compatibility with startR:
  - basic function working in the minimum number of dimensions
  - a wrapper function using Apply* on the basic function to work in N-dimensional arrays with named dimensions

* or other functions of the apply family (e.g. apply, lapply, sapply, mapply, aaply)

Introduction to ClimProjDiags

- It contains **13 functions**: 8 to compute indices and 5 for formatting
- It has documentation in **roxygen2** format
- It contains **4 vignettes** showing the computation of indicators:
  - Anomaly agreement
  - Diurnal temperature range indicator
  - Extreme indices t90p, t10n, rx5days, cdd, wx
  - Heat and coldwaves duration
- It is published on CRAN

https://cran.r-project.org/package=ClimProjDiags

Introduction to ClimProjDiags

Formatting functions

DailyAno(): Daily anomalies
SeasonSelect(): Selects a season from daily data for multidimensional arrays
SelBox(): Selects spatial region
Subset(): Subsets an N-dimensional array (along, indices, drop parameters)
WeightedMean(): Calculate spatial area-weighted average

Computing Indicators

AnoAgree(): Percentage of anomalies which agrees with the sign of the mean anomaly
Climdex(): Wrapper to compute ETCCDI* climate change indices
CombineIndices(): Combine weighted indices
DTRIndicator(): Diurnal temperature range indicator
DTRRef(): Diurnal temperature range
Extremes(): Sum of spell lengths exceeding daily threshold
Threshold(): Daily thresholds based on quantiles
WaveDuration(): Heat and cold waves duration

*Expert Team on Climate Change Detection and Indices

CSTools

Introduction to CSTools

https://earth.bsc.es/gitlab/es/CSTools

- It is being developed in **MEDSCOPE** ERA4CS project
- It is integrating functions from different institutions
- Its name emerged in a department **R user meeting**
- It is following the **multiApply paradigm** for compatibility with startR:
  - *CST_function()* working in s2dv_cube objects (OOP)
  - *function()* working in N-dimensional array with named dimensions

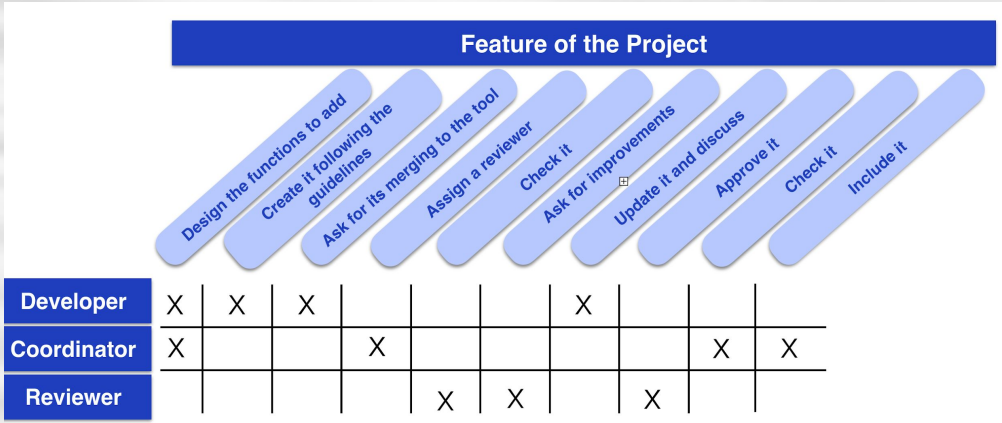
* or other functions of the apply family (e.g. apply, lapply, sapply, mapply, aaply)
*Expert Team on Climate Change Detection and Indices

Introduction to CSTools

- Currently on CRAN **v1.0.1** with **14 functions, 3 data samples and 3 vignettes**:
  - Multi-model Skill Assessment
  - Multivariate RMSE
  - RainFARM

https://cran.r-project.org/package=CSTools

- Working in the next release (November 2019) with 25 functions
- Development Guidelines:
  - Roles
  - Workflow
  - Branching strategy



Introduction to CSTools

Basic functions

CST_Load
CST_Anomaly
CST_SaveExp
CST_SaveNC
s2dv_cube
as.s2dv_cube

Correction

CST_BiasCorrection
CST_Calibration
CST_QuantileMapping
CST_BEI_Weighting
BEI_Weights
BEI_PDFBest*
CST_CatgFc

Downscaling

CST_Analogs
CST_RainFARM
CST_RFSlope
CST_RFWeights
RainFARM
RFSlope

Evaluation

CST_MultivarRMSE
CST_MultiMetric
CST_MultiEOF

Ploting functions

PlotMostLikelyQuantileMap
PlotForecastPDF
PlotCombinedMap

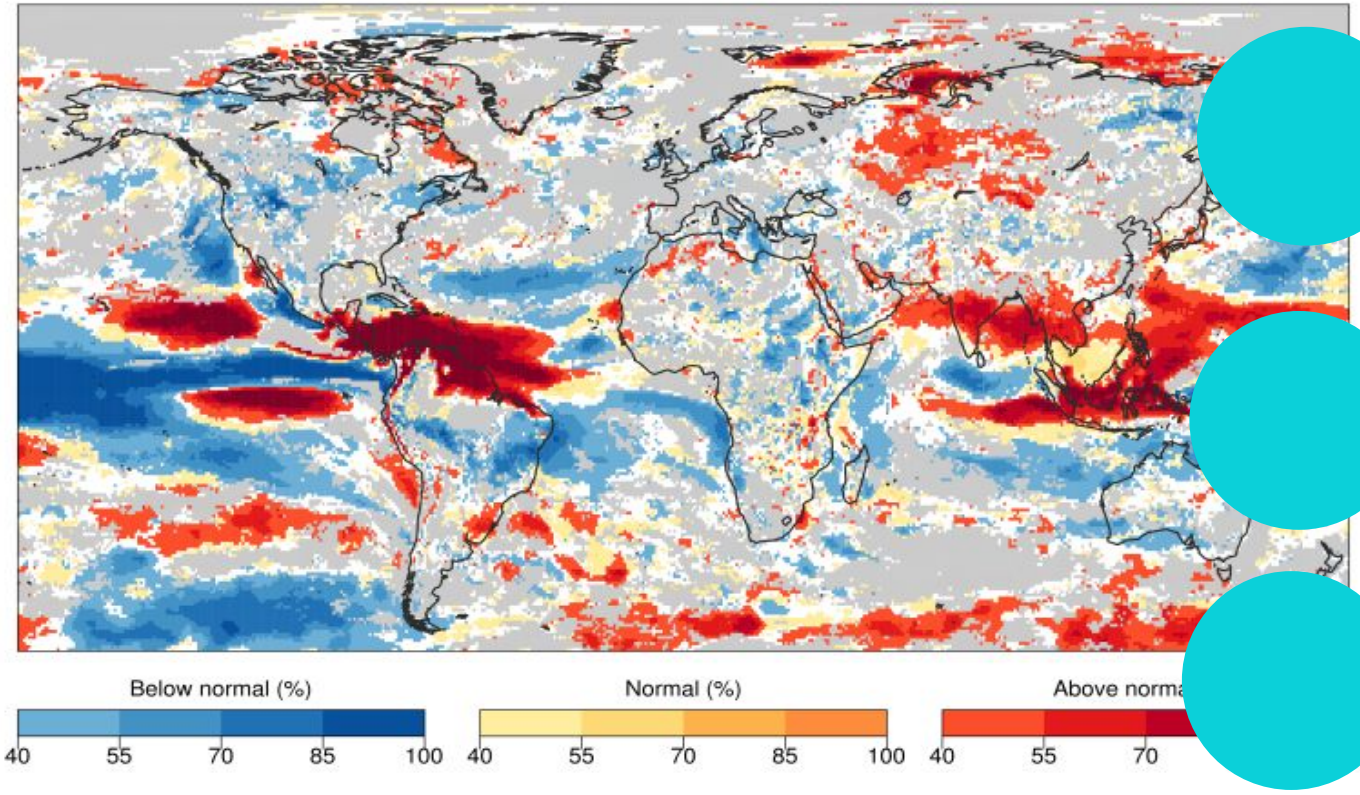
Future future functionalities: ADAMONT, SMOP and Dynamical Bias Correction

*BEI: Best Estimated Index

*RainFARM: Rainfall Filtered Autoregressive Model

Introduction to CSTools

PlotMostLikelyQuantileMap



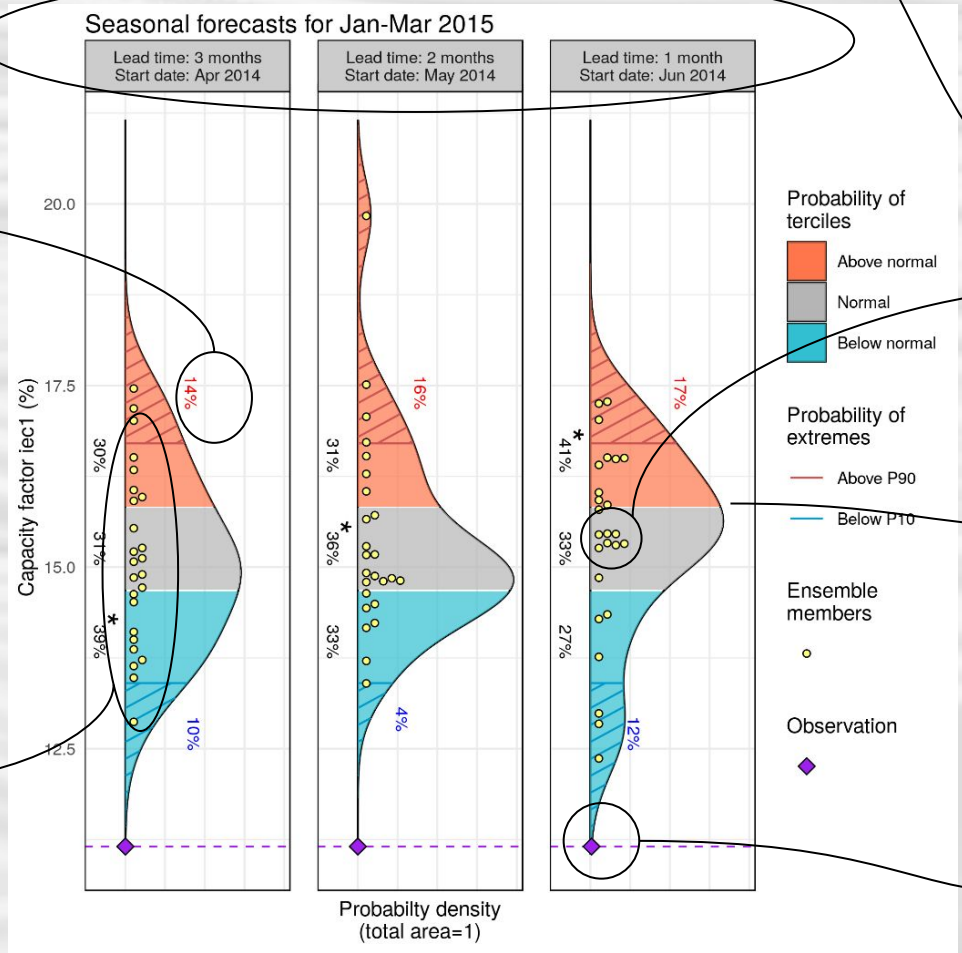
How will be the next month/season temperature?

Only areas with positive skill (RPSS) are shown to the users

Summarising information in one map only

Introduction to CSTools

PlotForcastPDF



Three forecasts with different lead time

Extremes: prob. above P90

Ensemble members

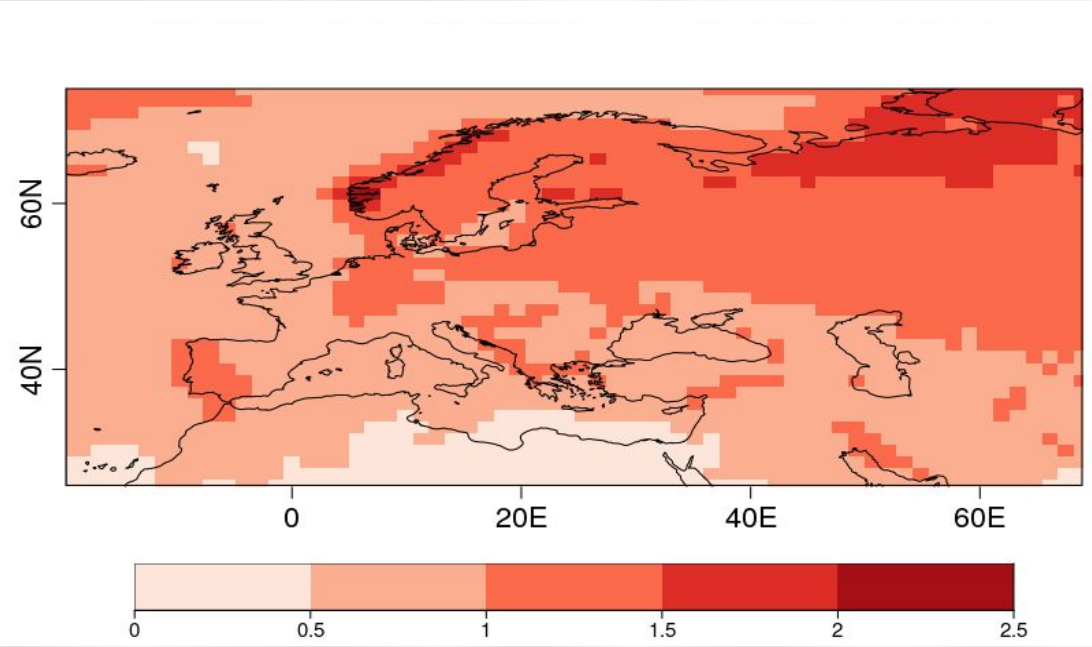
PDF from ensemble dressing

Tercile probabilities and Most Likely Tercile

Observed value

Introduction to CSTools

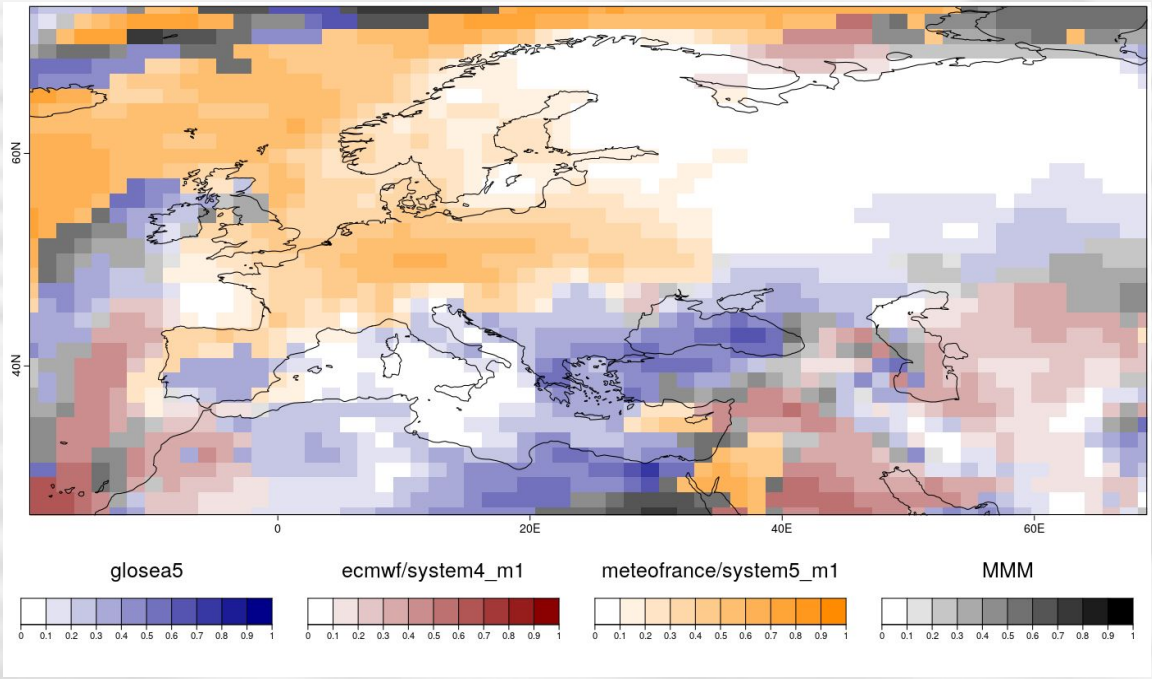
MultivarRMSE



This function calculates the RMSE from multiple variables at once. The multivariate RMSE is computed as the mean of each variable's RMSE scaled by its observed standard deviation. The variables can also be weighted based on their relative importance, as defined by the user.

Introduction to CSTools

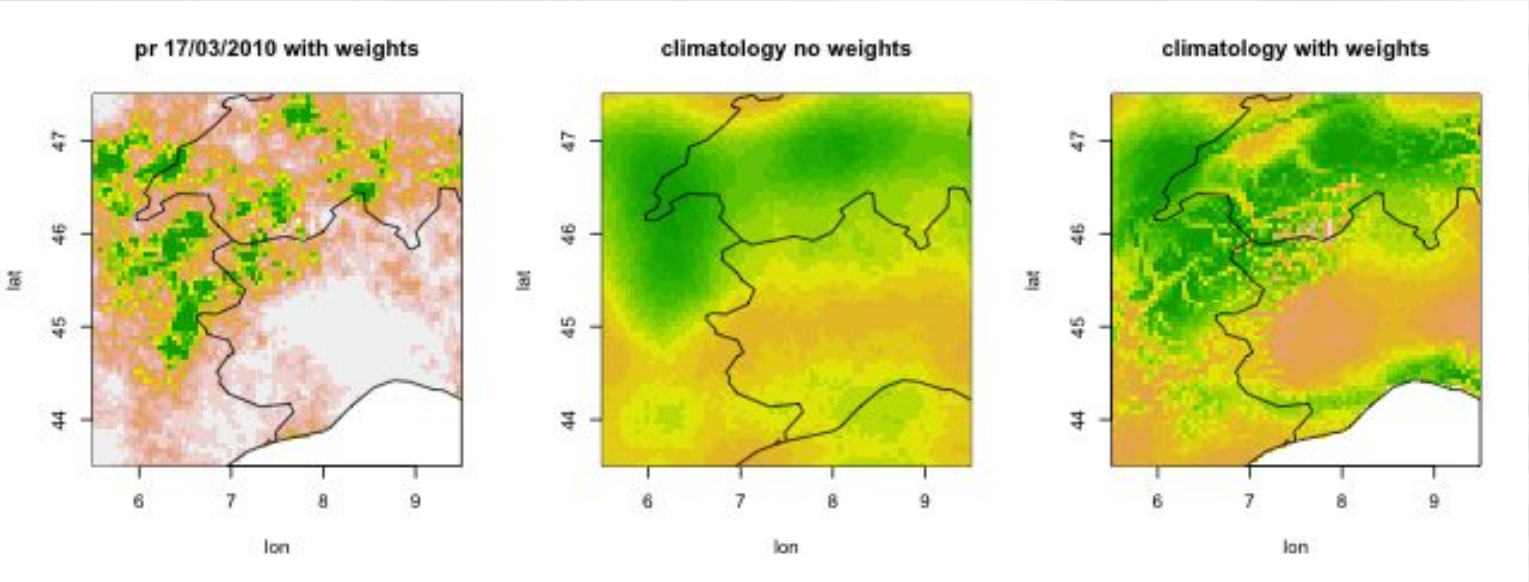
MultiMetric + PlotCombinedMap



ACC (anomaly correlation coefficient),
RMS (root mean square error) and **RMSSS** (root mean square error skill score) of individual models and multi-model ensemble forecasts

Introduction to CSTools

RainFARM: Rainfall Filtered Autoregressive Model



The RainFARM climatology downscaled without weights presents on average a very coarse structure, comparable with that of the original fields, while when using the weights a much more realistic distribution is achieved.

See the vignette!!
https://cran.r-project.org/web/packages/CSTools/vignettes/RainFARM_vignette.html

Tools Update

General announcement

→ **Have you joined the R user email?**

earth-rtools@bsc.es

→ **List of R modules and package version**

https://docs.google.com/spreadsheets/d/1gA-O1WkgH50VnyxOz9gbec_DsyvWMkyF-9SIFANXLqI/edit#gid=0

s2dverification announcement

- NEWS.md on GitLab

s2dverification 2.8.6

- Apply Roxygen2 format to all the files.
- Bug fix in Composite().
- Bug fix in Ano(). Recommend to assign the dimensions by name to avoid confusion when the dimensions have same length.
- Trend() documentation error fix.
- Introduce new function PlotMatrix().

s2dverification announcement

- **PlotMatrix()**

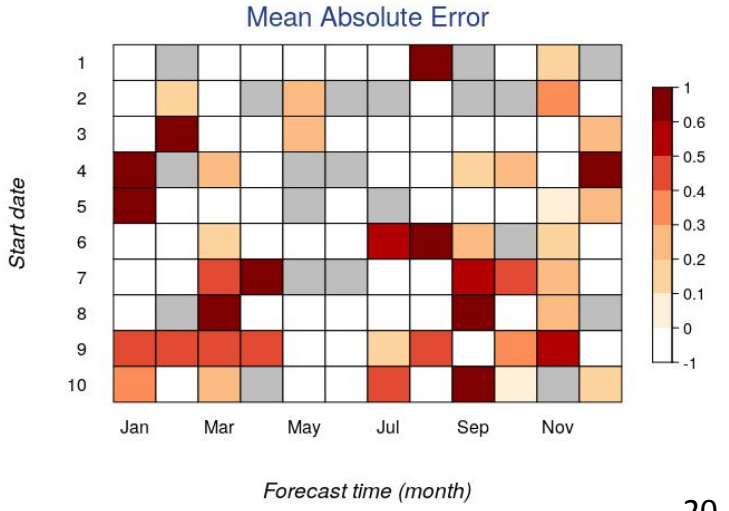
PlotMatrix *Function to convert any numerical table to a grid of coloured squares.*

Description

This function converts a numerical data matrix into a coloured grid. It is useful for a slide or article to present tabular results as colors instead of numbers.

Usage

```
PlotMatrix(var, brks = NULL, cols = NULL, toptitle = NULL,
  title.color = "royalblue4", xtitle = NULL, ytitle = NULL,
  xlabels = NULL, xvert = FALSE, ylabels = NULL, line = 3,
  figure.width = 1, legend = TRUE, legend.width = 0.15,
  xlab_dist = NULL, ylab_dist = NULL, fileout = NULL, size_units = "px",
  res = 100, ...)
```



s2dverification announcement

- **Roxygen2 format** Please remember to follow the header format for your development!!

header

```
#'Function to convert any numerical table to a grid of coloured squares.
#'
#'This function converts a numerical data matrix into a coloured
#'grid. It is useful for a slide or article to present tabular results as
#'colors instead of numbers.
#'
#'@param var A numerical matrix containing the values to be displayed in a
#'  colored image.
.
.
#'@return A figure in popup window by default, or saved to the specified path.
#'
#'@examples
.
.
#'@importFrom grDevices dev.new dev.off dev.cur
#'@export
PlotMatrix <- function(var, brks = NULL, cols = NULL,
                       toptitle = NULL, title.color = "royalblue4",
                       xtitle = NULL, ytitle = NULL, xlabels = NULL, xvert = FALSE,
                       ylabels = NULL, line = 3, figure.width = 1, legend = TRUE,
                       legend.width = 0.15, xlab_dist = NULL, ylab_dist = NULL,
                       fileout = NULL, size_units = 'px', res = 100, ...){
```

@title

@description

startR announcement

- **Documentation revision** (in branch 'develop-doc')

  README.md, pracitcal_guide.md… etc.

- **FAQs**

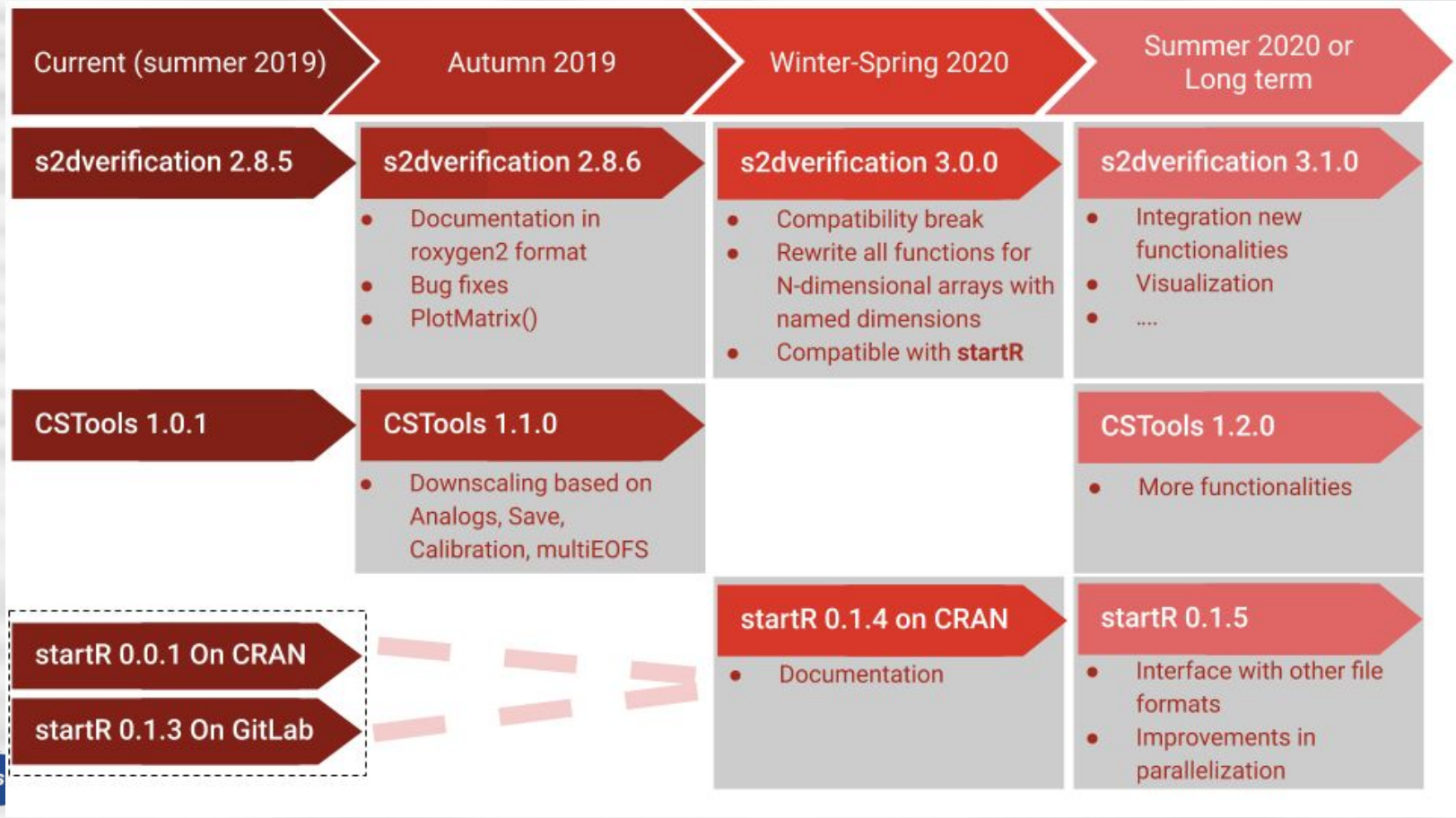
  Under inst/doc/, faq.md

- **Example scripts (usecase)**

  Under inst/doc/, usecase.md

Any feedback is welcomed!!

R package future plan



Result of s2dverification function usage poll

Most frequently used functions:

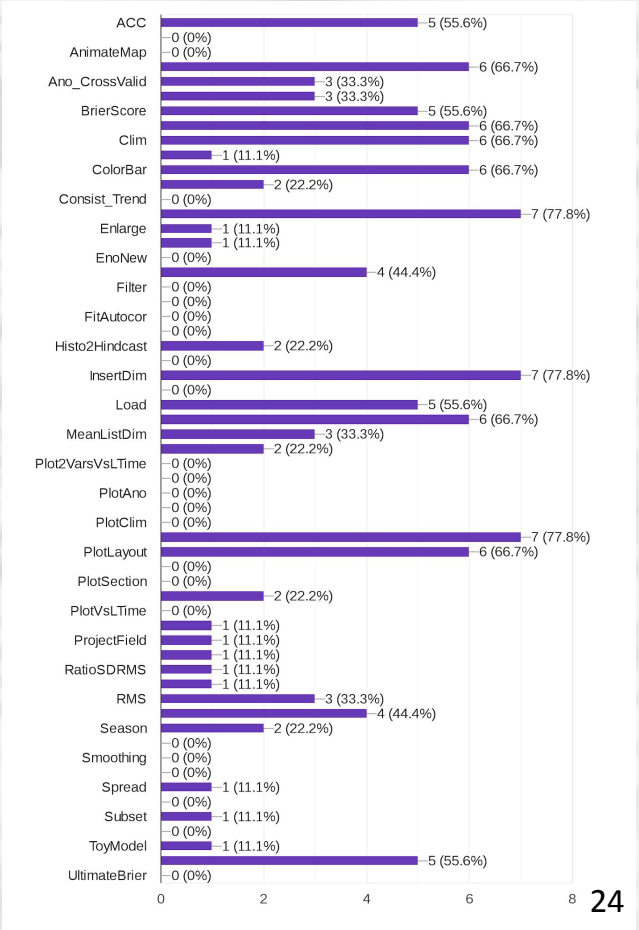
(7) Corr, InsertDim, PlotEquiMap

(6) Ano, Clim, CDORemap, ColorBar, Mean1Dim, PlotLayout

(5) ACC, BrierScore, Load, Trend

(4) EOF, RMSSS

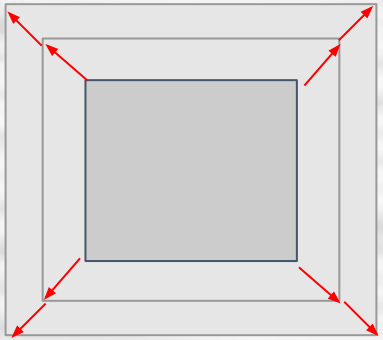
(3) ...



Issue Discussion

Discussion: transformation within Start()

```
obs <- Start(...,
             transform = CDORemapper,
             transform_extra_cells = 2,
             transform_params = list(grid = 'r360x181',
                                     method = 'conservative',
                                     crop = c(lons.min, lons.max, lats.min, lats.max)),
             transform_vars = c('latitude', 'longitude'),
              ...)
```



When lat/lon is subset of the whole region...

cdo	Start(..., transform_extra_cell = 0)	Start(..., transform_extra_cell = 2)	Load(grid = ...)	CDORemap()
1	2	1	1	2

Which number should be the default of transform_extra_cells?

Discussion: Histo2Hindcast() improvement

- Add parameter 'leap'?
- Add more time frequency options (annual, daily, hourly...etc.)?

`Histo2Hindcast` *Chunks Long Simulations For Comparison With Hindcasts*

Description

This function reorganizes a long run (historical typically) with only one start date into chunks corresponding to a set of start dates. The expected input structure is the one output from `Load()` with 4 to 7 dimensions.

Usage

```
Histo2Hindcast(varin, sdatesin, sdatesout, nleadtimesout)
```