



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación

Climate Forecast Analysis Tools Framework

Núria Pérez-Zanón and An-Chi Ho

PATC 2021

12 Nov, online

Earth Science Department



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



Climate Forecast Analysis Tools Framework

Núria Pérez-Zanón, An-Chi Ho, Francesco Benincasa, Pierre-Antoine Bretonnière, Louis-Philippe Caron, Chihchung Chou, Carlos Delgado-Torres, Llorenç Lledó, Nicolau Manubens, Lluís Palma

useR! 2021

5-10 July, online

Earth Science Department



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Climate Forecast Analysis Tools Framework

Núria Pérez-Zanón, An-Chi Ho, Francesco Benincasa, Pierre-Antoine Bretonnière, Louis-Philippe Caron, Chihchung Chou, Carlos Delgado-Torres, Llorenç Lledó, Nicolau Manubens, Lluís Palma

Outline

1. Introduction to Climate Forecasts
2. Introduction to the Climate Forecast Analysis Tools
3. Case Study (hands - on)
4. Successful research
5. startR (hands-on)

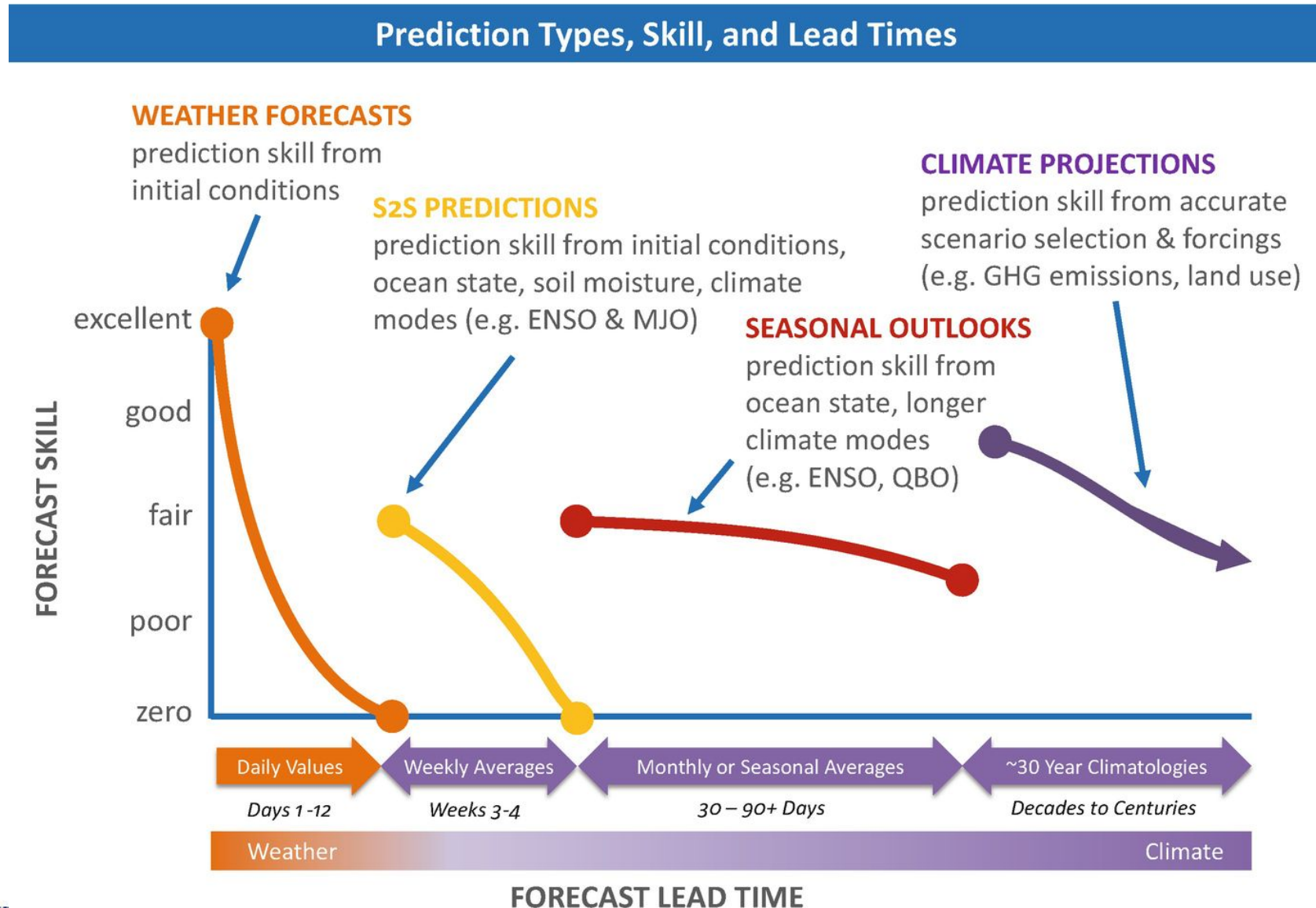
Climate Forecasts



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Climate Forecast: Forecast horizon

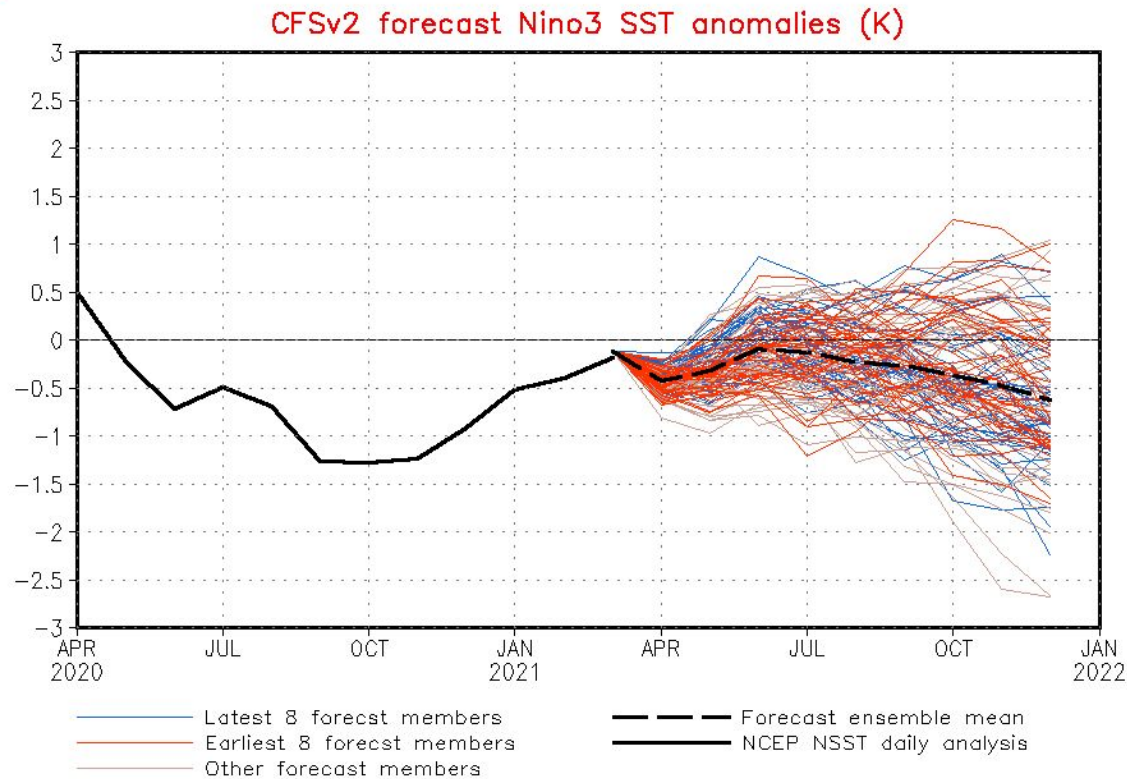


Climate Forecast: Ensemble generation

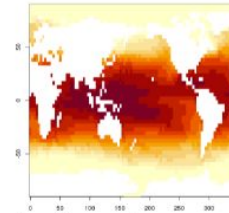


NWS/NCEP/CPC

Last update: Thu Apr 8 2021
Initial conditions: 8Apr2021–17Apr2021



Initialization
(start date)

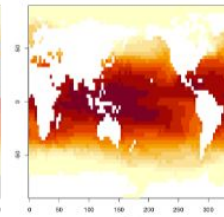
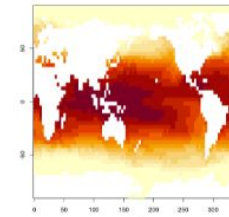


Perturbation

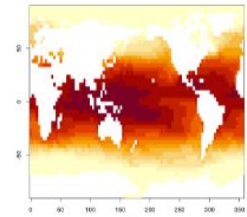
member 1

member 2

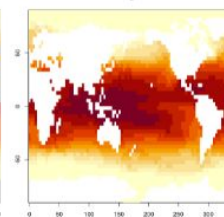
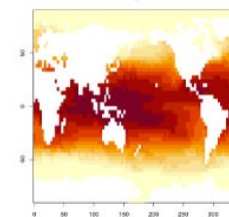
member N



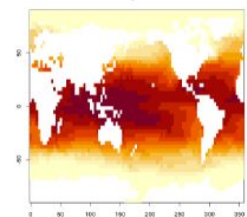
...



Lead time 0



...



Lead time 1

...

...

...

...

Lead time t



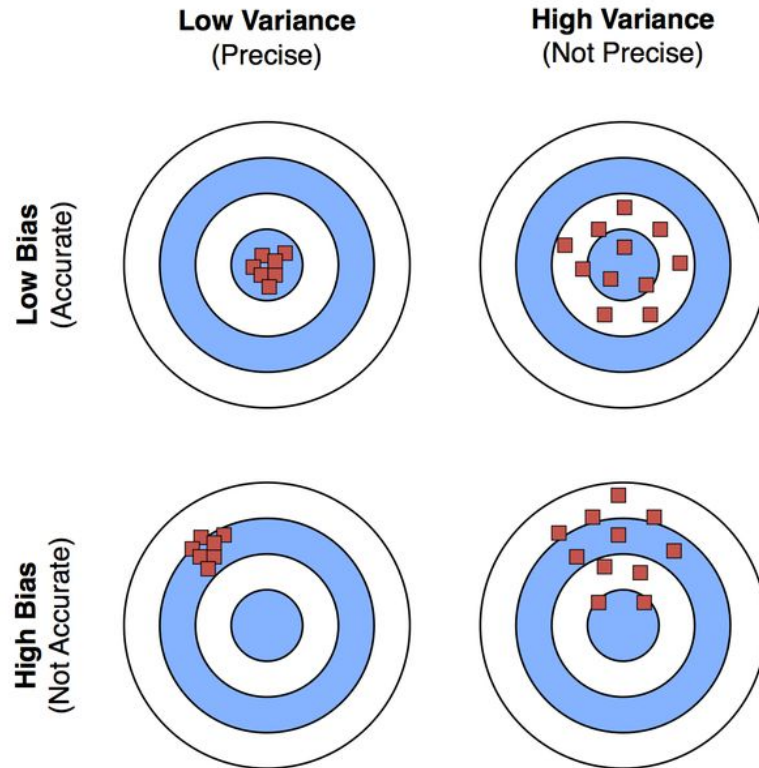
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación


Multi-dimensional array with named dimension:

e.g.: {model = 2, sdates = 30, members = 25, ltime = 7, lat = 90, lon = 360, nlevels = 10}

Climate Forecast: Processing

Bias Correction is an example of one method applied to raw forecast to improve their quality



 This work by Sebastian Raschka is licensed under a Creative Commons Attribution 4.0 International License.

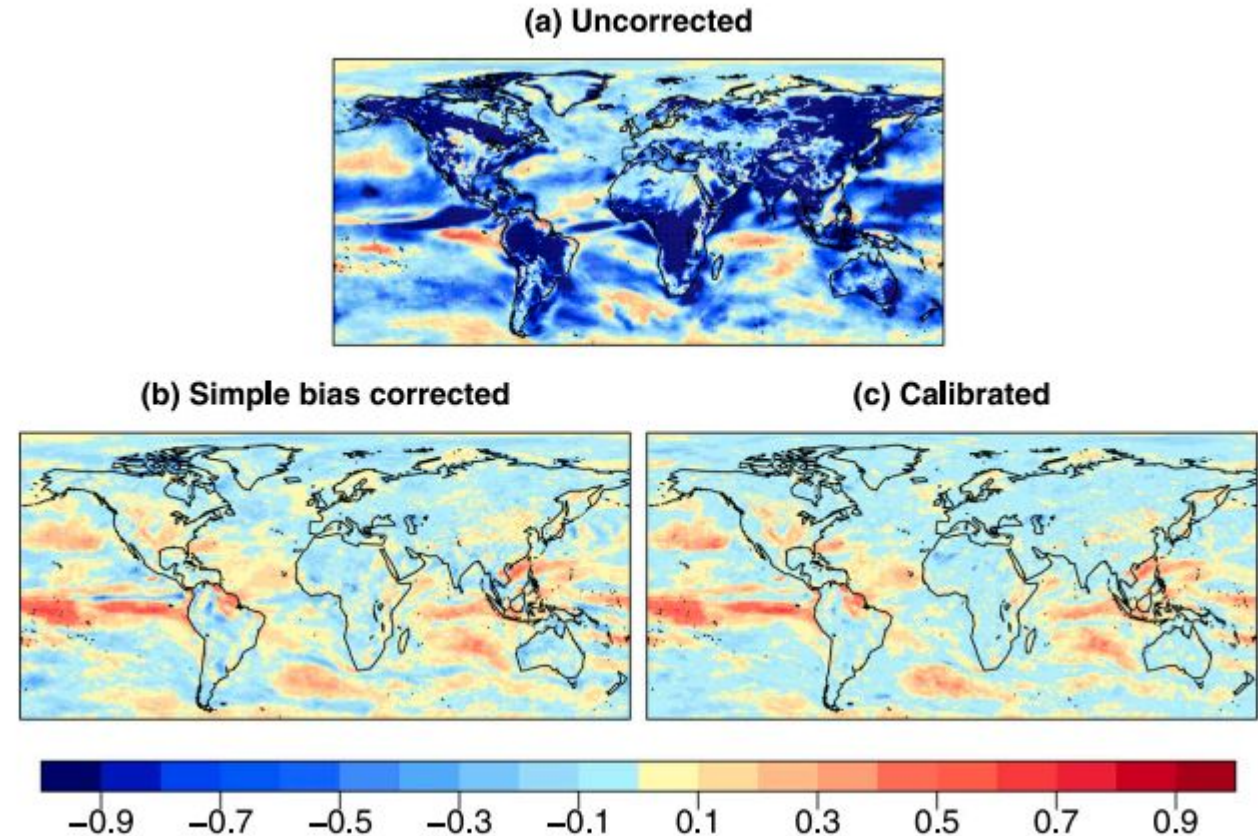
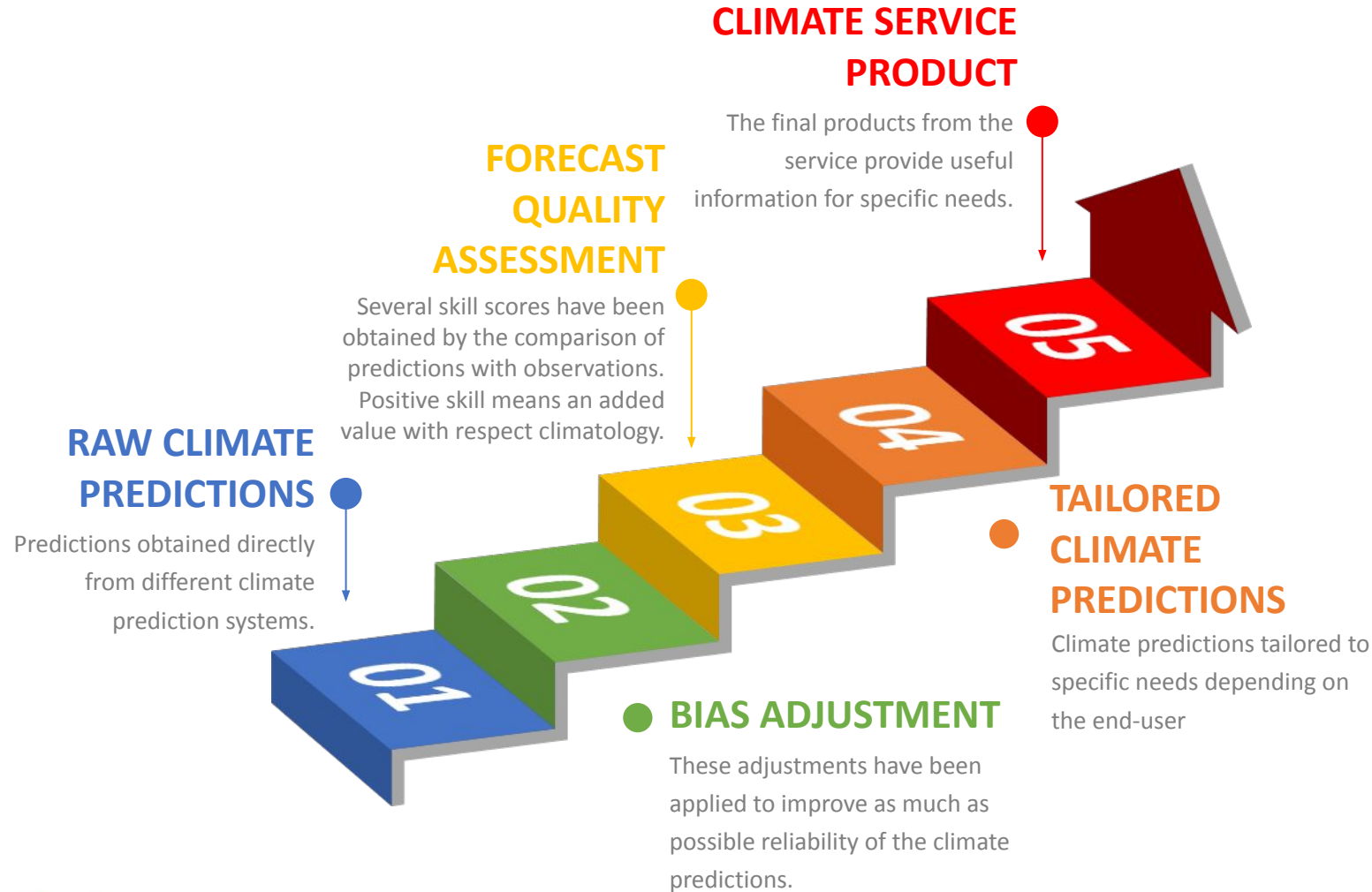


FIG. 4. Fair RPSS for tercile events of 10-m wind speed forecasts from ECMWF System 4 and ERA-Interim reanalysis in winter (DJF). These predictions have been initialized on 1 Nov for the period of 1981–2012.

Climate Forecast: From Climate data to Climate product



- ★ Availability of climate data does not automatically imply to have access to useful climate information
- ★ Tools required to process climate forecast

**We need a programming
language to process
climate data**



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

R programming language

R is a programming language and framework for statistical computing and graphics.

- Available as free software (under the GNU GPL license)
- Provides a wide variety of statistical techniques (linear and non-linear modelling, classical statistical tests, classification and simulation...)
- Well developed plotting tools

You can install R from <https://cran.r-project.org/>

- You will have the base R packages installed in your machine
- You can look for more packages https://cran.r-project.org/web/packages/available_packages_by_name.html

E.g.: The ggplot2 package for visualization: <https://www.r-graph-gallery.com/ggplot2-package.html>

Climate Forecasts Analysis Tools Introduction

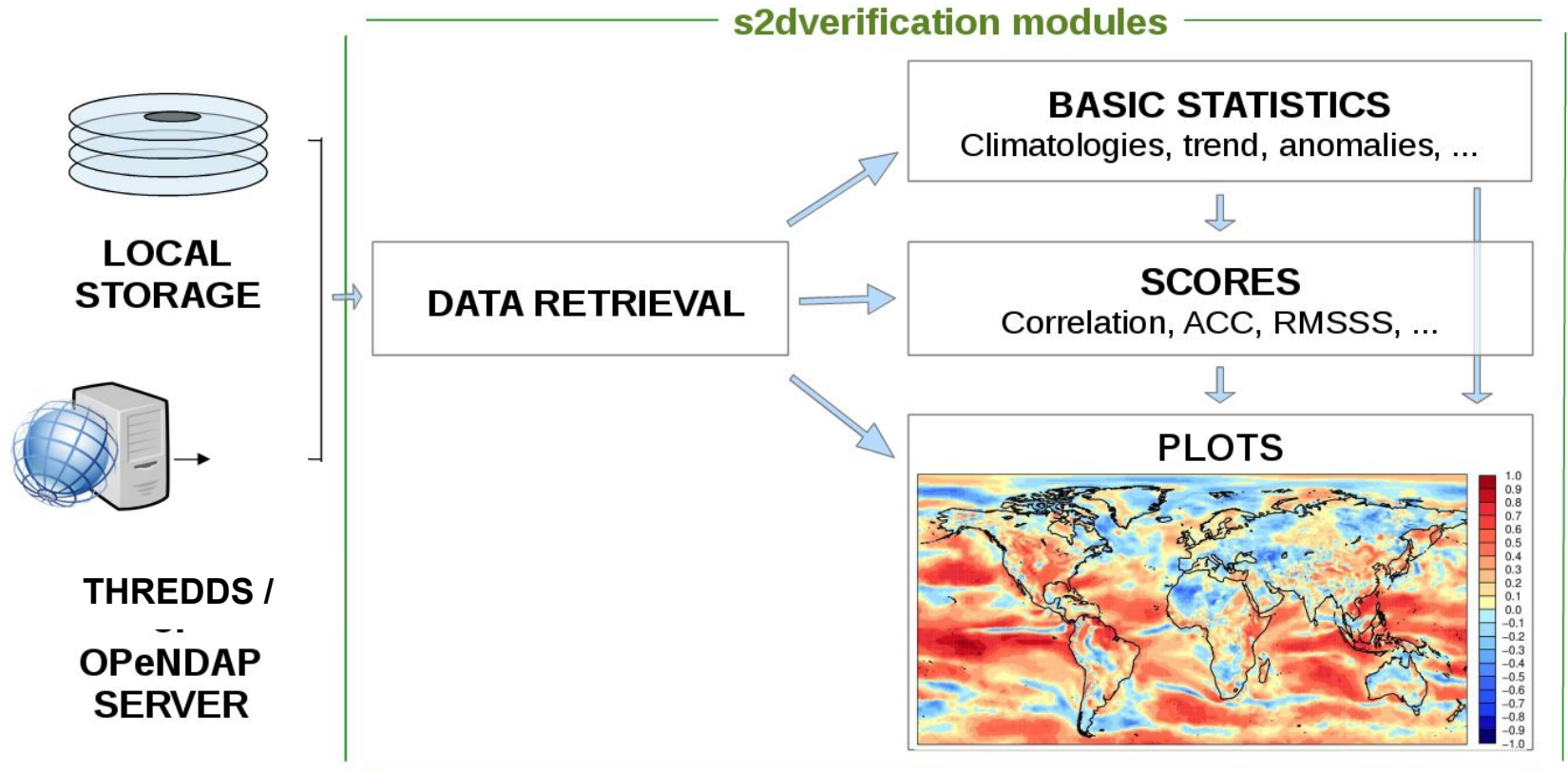


**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Climate Forecast Analysis Tools: Early version

- ★ Methods developed by the department to assess the quality of the forecast were gathered in s2dverification R package
- ★ Researchers could easily share their methods and replicate colleagues analysis on their own data
- ★ Common needs detected (e.g.: reading data from files and visualization tools)



Origin (V.Guemas)

2009

On CRAN

2013

2015

2017

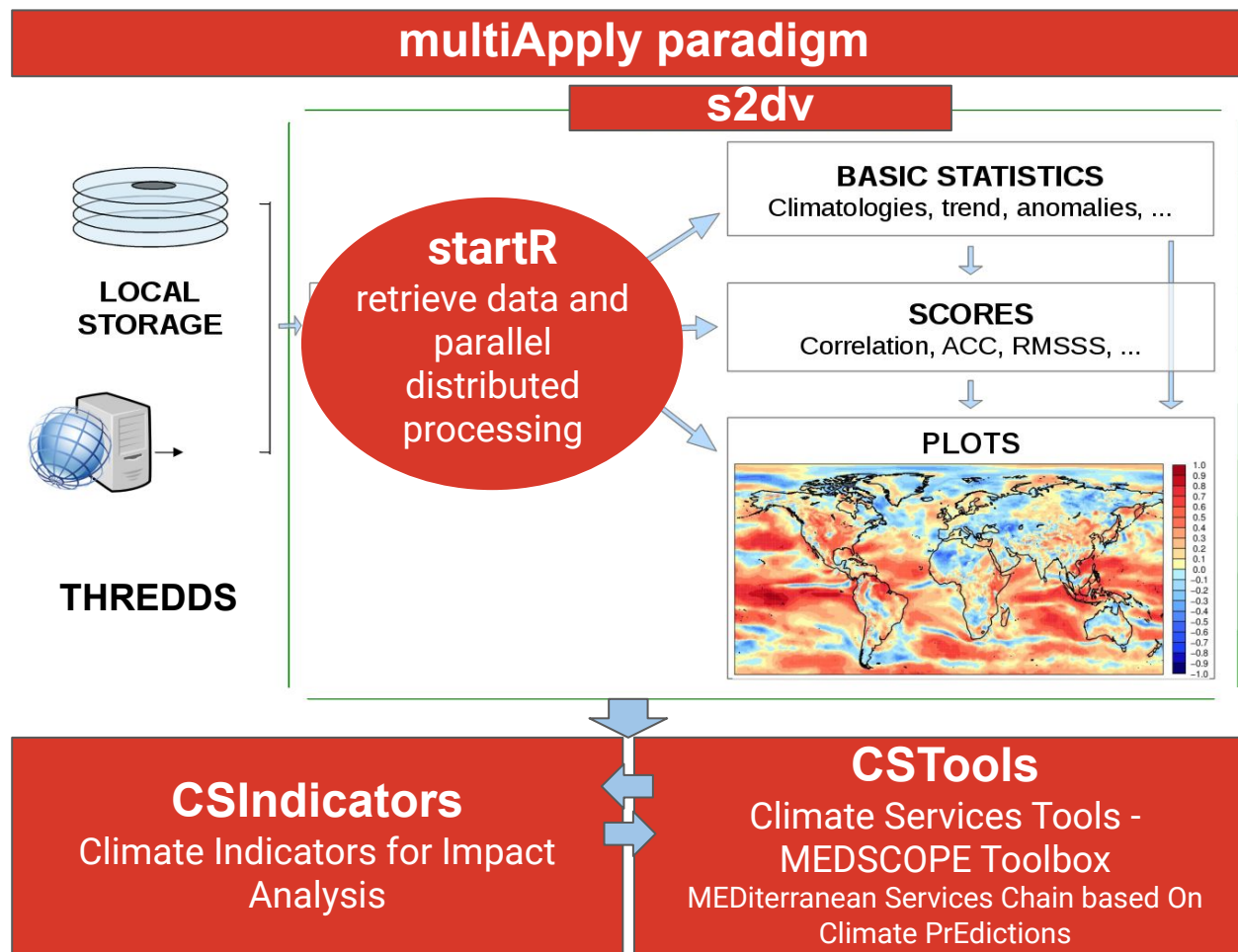
2019

v2.9.0

2021

Climate Forecast Analysis Tools: Current version

- New forecast horizons to be analyzed
 - tools flexibility required
- Increase of data size (more members, finer grid resolution, multi-model analysis)
 - multi-core and multi-node analysis on HPC
- New methods to postprocess the forecast are being developed



Climate Forecast Analysis Tools: Table of packages

- ★ Functions are split on packages depending on their objective
- ★ Functions from different packages (even external packages) can be used to perform an analysis or obtain a product

	Package name	Short description	Link to CRAN
Data manipulation	easyNCDF	Read/write netCDF files into/from multidimensional R array.	https://CRAN.R-project.org/package=easyNCDF
	startR	Data retrieval and processing tools	https://CRAN.R-project.org/package=startR
	multiApply	Apply functions to multiple multidimensional arrays or vectors allowing parallel computation	https://CRAN.R-project.org/package=multiApply
Analysis and processing	s2dverification	Functions for Forecast Verification and visualization	https://CRAN.R-project.org/package=s2dverification
	s2dv	Adaptation of s2dverification to multiApply	https://CRAN.R-project.org/package=s2dv
	CSTools	Methods for forecast calibration, statistical and stochastic downscaling, optimal forecast combination and tools to obtain tailored products.	https://CRAN.R-project.org/package=CSTools
Climate indicators	CSIndicators	Sectorial Indicators for Climate Service	https://CRAN.R-project.org/package=CSIndicators
	ClimProjDiags	Climate extreme indices, evaluation of the agreement between models, weight and combination functions.	https://CRAN.R-project.org/package=ClimProjDiags

Case studies



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Climate Forecast Analysis Tools: SNOWPACK

Snowpack is an essential water reservoir that is fed by snowfall during the **cold season** and then released in late spring and summer when the precipitation contribution is low and the water request has a peak.

Mountain meltwater is essential for several economic activities including **hydropower generation**, **agriculture**, **industry**, and **meltwater shortage** can cause heavy economic loss.

SNOWPACK model estimates snow depth and snow water equivalent at selected high-elevation sites.

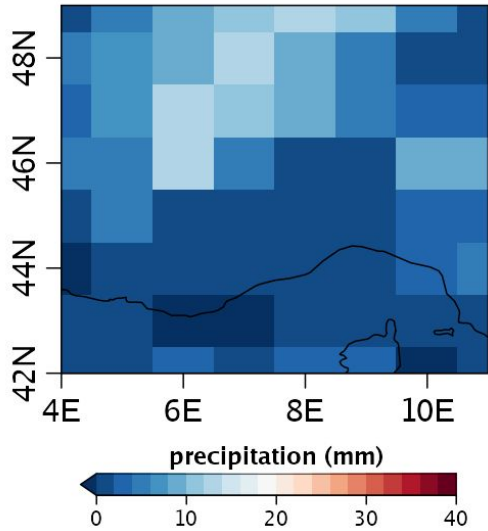
REQUIREMENT: Postprocessed precipitation forecast



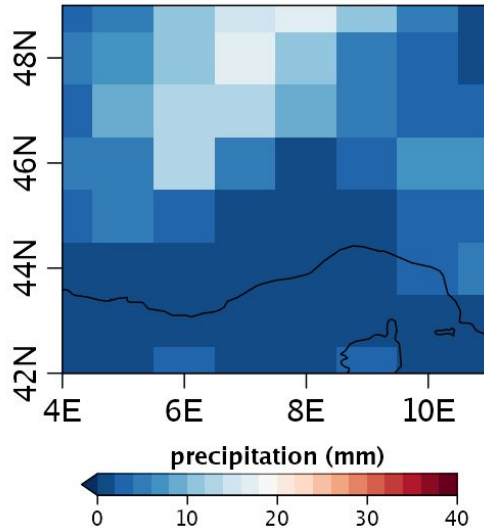
Digging a snowpit on Taku Glacier, in Alaska to measure snowpack depth and density (wikipedia)

Climate Forecast Analysis Tools: SNOWPACK

ECMWF-S5C3S



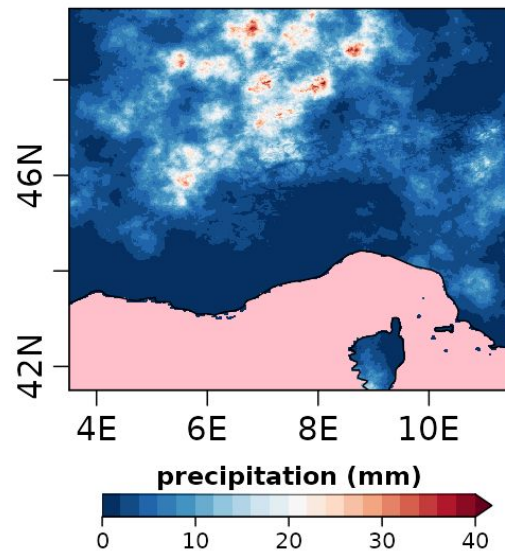
Bias Corrected



Step 2

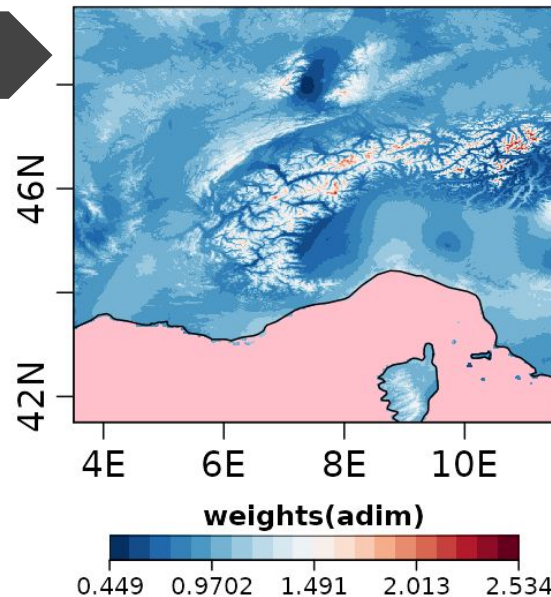
Step 4

Downscaled



Result

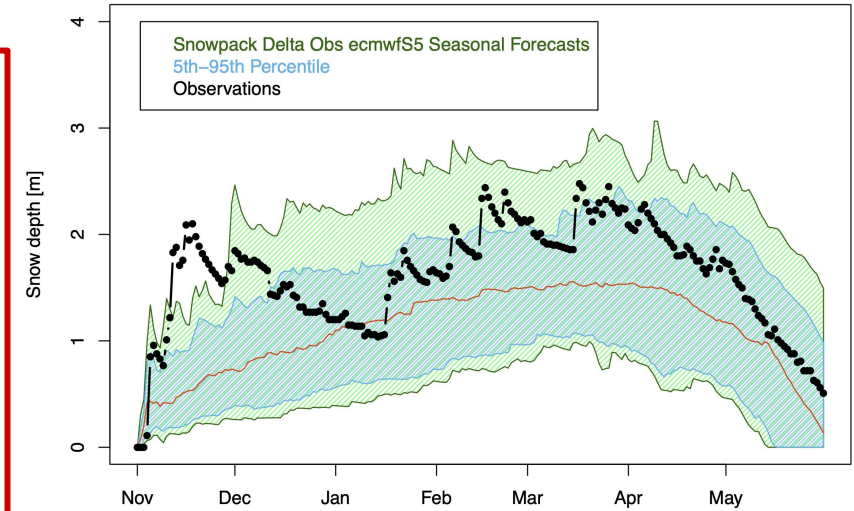
Weights



Step 3

SNOWPACK model estimation of snow depth in a specific site on the Alps

Bocchetta delle Pisse (2410 m) – 2014



★ Using CTools package, the climate forecast data can be postprocessed to obtain relevant information for the end-users.

Hands-on: Running CStools and RainFARM

- 1- Open a terminal
- 2- connect to MN4
- 3- Interactive session `salloc -p interactive`
- 4- Open an R session

Objective: explore a sample data

```
library(CStools)
exp <- lonlat_prec

dim(exp$data)
names(exp)
str(exp)
exp$lat
```



Hands-on: Running CStools and RainFARM

Objective Downscale

```
exp_down <- CST RainFARM(exp, nf=20, kmin = 1, nens = 3,  
                          time_dim = c("member", "ftime"))  
dim(exp_down$data)
```

Objective Visualize original data

```
a <- exp$data[1, 1, 1, 17, , ] * 86400 * 1000  
a[a > 60] <- 60  
  
png("original_data.png", width = 10, height = 10,  
    units = 'cm', res = 150)  
image(exp$lon, rev(exp$lat), t(apply(a, 2, rev)),  
      xlab = "lon", ylab = "lat",  
      col = rev(terrain.colors(20)), zlim =  
      c(0, 60))  
map("world", add = TRUE)  
title(main = "pr 17/03/2010 original")  
dev.off()
```

Objective: Visualize downscaled data

```
a <- exp_down$data[1, 1, 1, 1, 17, , ] * 86400 * 1000  
a[a > 60] <- 60  
  
png("downscaled_data.png", width = 10, height = 10, units  
    = 'cm', res = 150)  
image(exp_down$lon, rev(exp_down$lat), t(apply(a, 2,  
rev)), xlab = "lon", ylab = "lat",  
      col = rev(terrain.colors(20)), zlim = c(0, 60))  
map("world", add = TRUE)  
title(main = "pr 17/03/2010 downscaled")  
dev.off()
```

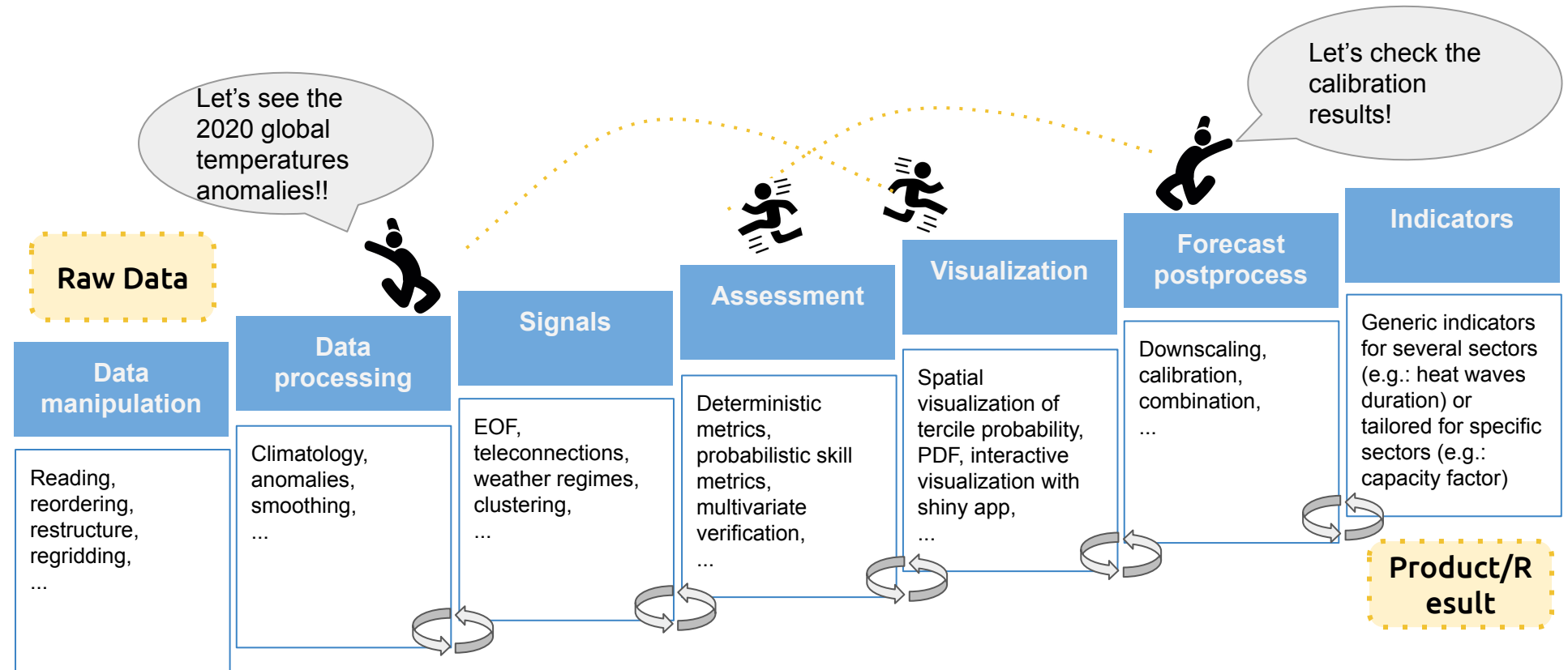
https://cran.r-project.org/web/packages/CStools/vignettes/RainFARM_vignette.html

Development Strategy



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación

Climate Forecast Analysis Tools: Interoperability & Methods



Climate Forecast Analysis Tools: Development Guidelines

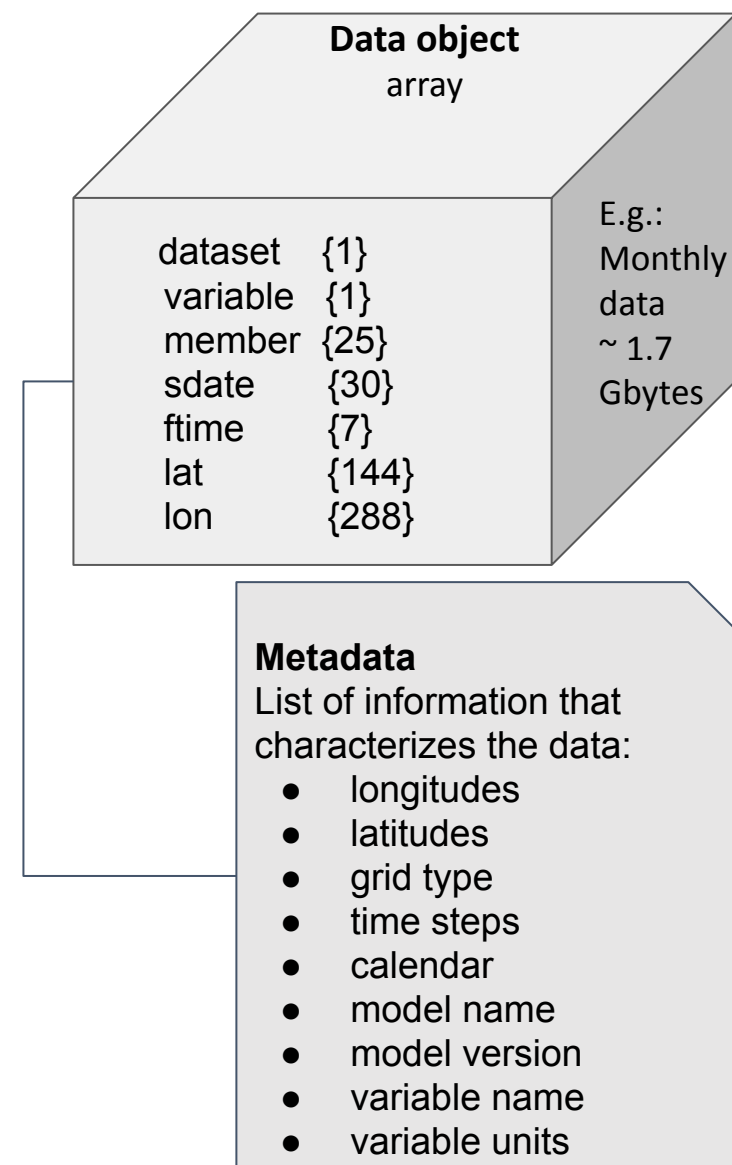
- ★ **Data** is read from (NetCDF) files on **multi-dimensional arrays with named dimensions**
- ★ The attributes of the data (**metadata**) can be also retrieved
- ★ An **atomic function** is defined working on the minimum number of dimensions required (e.g.: spatial mean requires 'lat' and 'lon' dimensions)

```
sp_mean <- function(data) {  
  mean(data)  
}
```

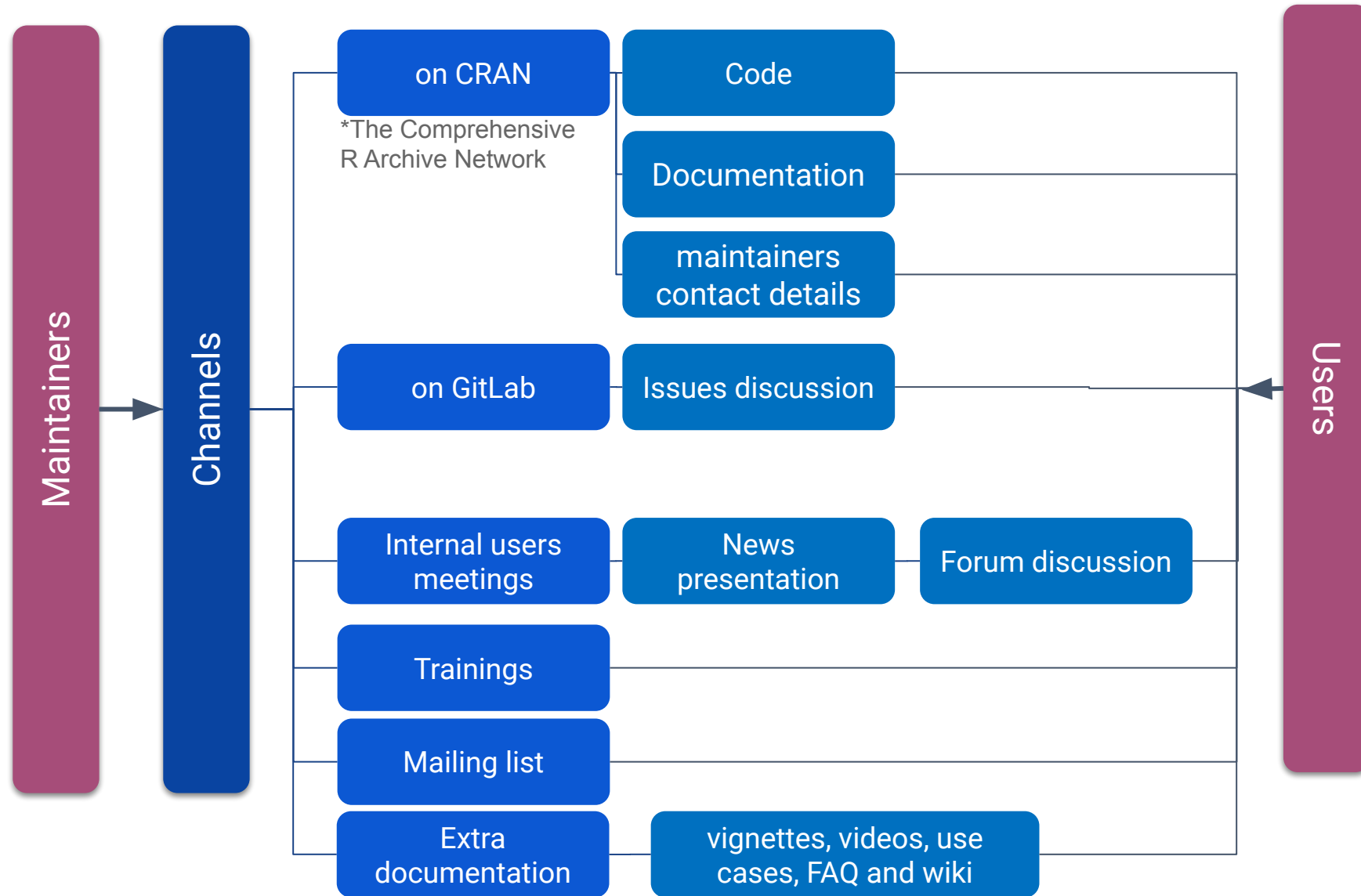
- ★ Using **Apply function from multiApply package**, the function will be applied to the res of the dimensions:

```
SPMean <- function(data, ncores = 1) {  
  Apply(list(data), target_dims = c('lat', 'lon'), fun = sp_mean,  
    ncores = ncores)  
}
```

- ★ Parameter 'ncores' allows **parallel computing**
- ★ Extra guidance (e.g.: **checks, formatting, automatic tests, etc.**) is provided in the development guidelines



Climate Forecast Analysis Tools: Users support



Climate Forecast Analysis Tools: Vignettes

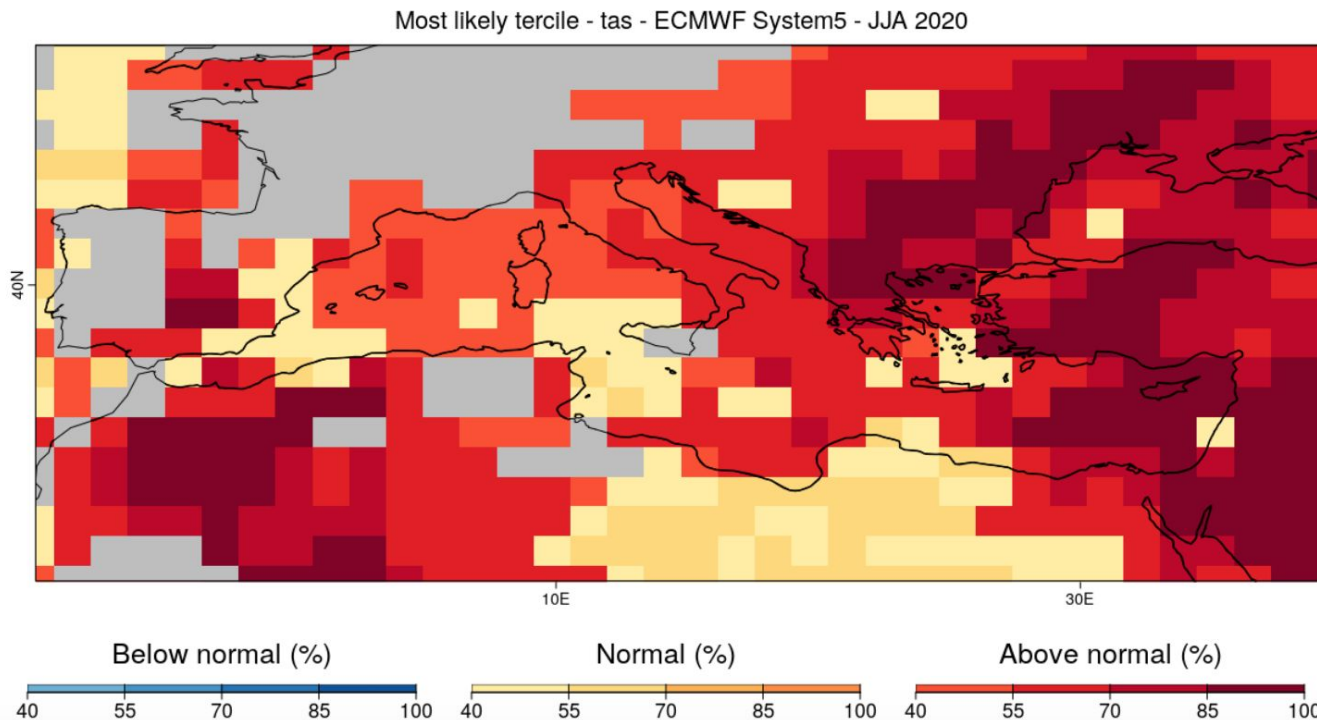
6. Simultaneous visualization of probabilities and skill scores

From the RPSS, we create a mask: regions with RPSS ≤ 0 will be masked.

```
mask_rpss <- RPSS[[1]]
mask_rpss[RPSS[[1]] <= 0] <- 1
mask_rpss[is.na(RPSS[[1]])] <- 1
mask_rpss[RPSS[[1]] > 0] <- 0
```

Finally, we plot the latest forecast, as in the previous step, but add the mask we just created.

```
PlotMostLikelyQuantileMap(probs = prob_map, lon = Lon, lat = Lat, coast_width = 1.5,
                          legend_scale = 0.8, mask = t(mask_rpss),
                          toptitle = paste('Most likely tercile -', clim_var,
                                           '- ECMWF System5 - JJA 2020'))
```



- ★ R packages allow to share a document including text, code and figures called **vignettes**.
- ★ Users and researchers can follow them to obtain learn on how to use the tools or reproduce an example.

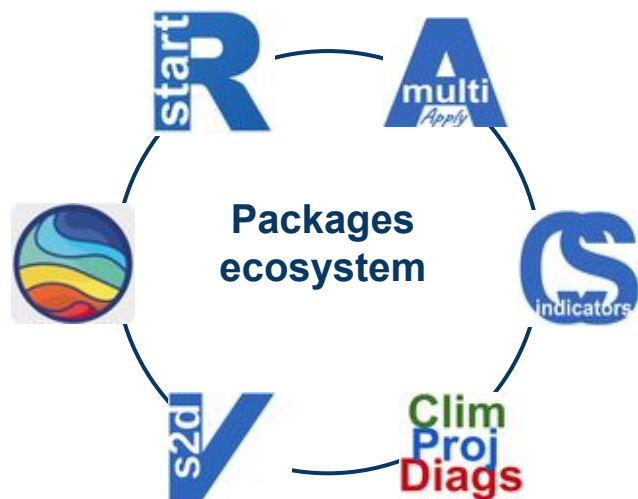
E.g.: Computing and displaying the most likely tercile of a seasonal forecast

See

https://cran.r-project.org/web/packages/CS-Tools/vignettes/MostLikelyTercile_vignette.html

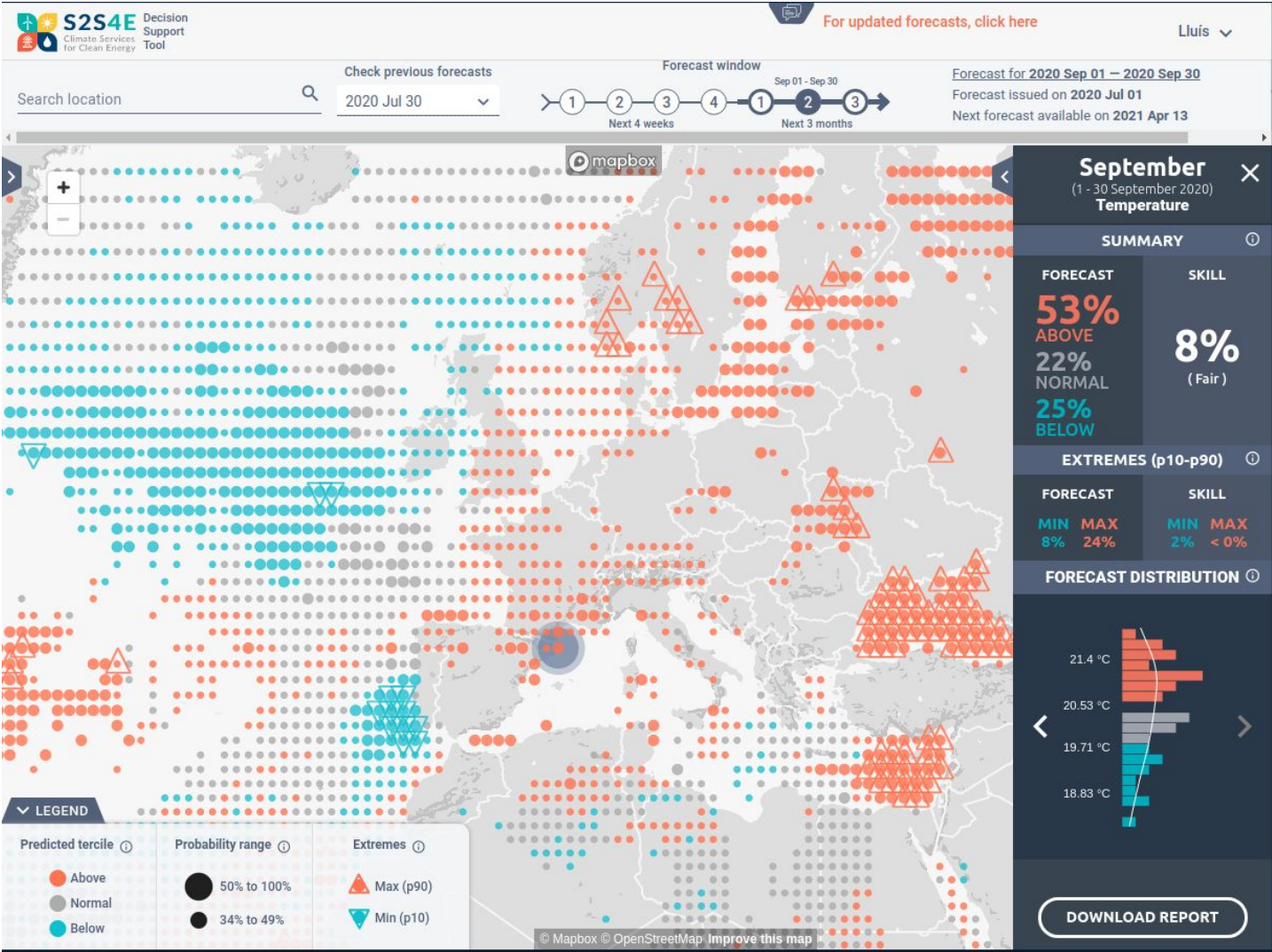
Climate Forecast Analysis Tools: Successful research

- ★ R tools are being used in several research lines and operationals



Research line	Projects	Publication e.g.
In-situ observations	Indecís, S2S4E	Tall towers and reanalysis Ramon et al. 2019
Atmospheric Composition	Ongoing collaboration in CALIOPE-Urban	
Sub-seasonal Forecast	S2S4E VITIGEOSS	Verification Manrique et al. 2020
Seasonal Forecast	S2S4E, Visca, Medscope, INTAROS, Medgold QA4Seas	Wind power generation Lledó et al., 2019
Decadal Predictions	EUCP, C3S 34c	CMIP6 Assessment Bilbao et al. 2021
Climate Projections	C3S MAGIC	ESMValTool papers: python and R synergy

Climate Forecast Analysis Tools: Successful research



Climate Forecast Analysis Tools: Last Remarks

The tools ...

- ★ **Don't need to be use all at once**
 - E.g.: Users can choose only to load the data but not use the methods
- ★ **Common needs are a priority**
 - E.g.: Calculating climatologies is a common need but exploring new methods to calculate them can be done by the users

Future work

- ★ **Special focus on the retrieving step which is a common need**
 - E.g.: Increase the flexibility when retrieving multi-dimensional datasets
- ★ **Keep increasing the quality of the support**
 - E.g.: Exploring to create a readthedocs to allow searching for in-house functions
- ★ **Keep collaborating with users to understand their needs**



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Thanks for your attention and thanks to all contributors and dependencies developers

ncdf4, plyr, abind, data.table, reshape2, ggplot2, RColorBrewer,
abind, bigmemory, GEOmap, geomapdata, NbClust,
SpecsVerification, future, ...

Contact

- **Núria Pérez-Zanón** (nuria.perez@bsc.es)
- **An-Chi Ho** (an.ho@bsc.es)

Questions?

startR overview

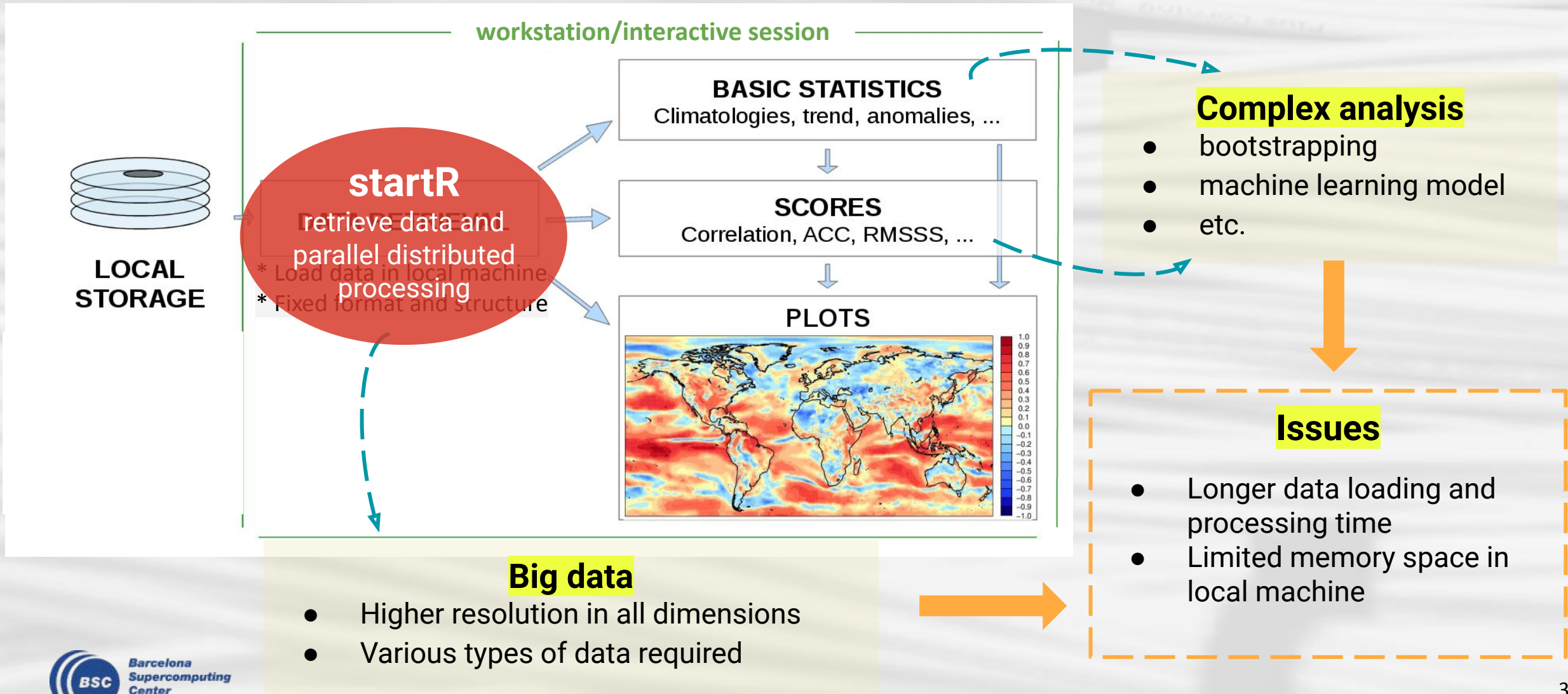


**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

How was startR born?

[Data analysis procedure]

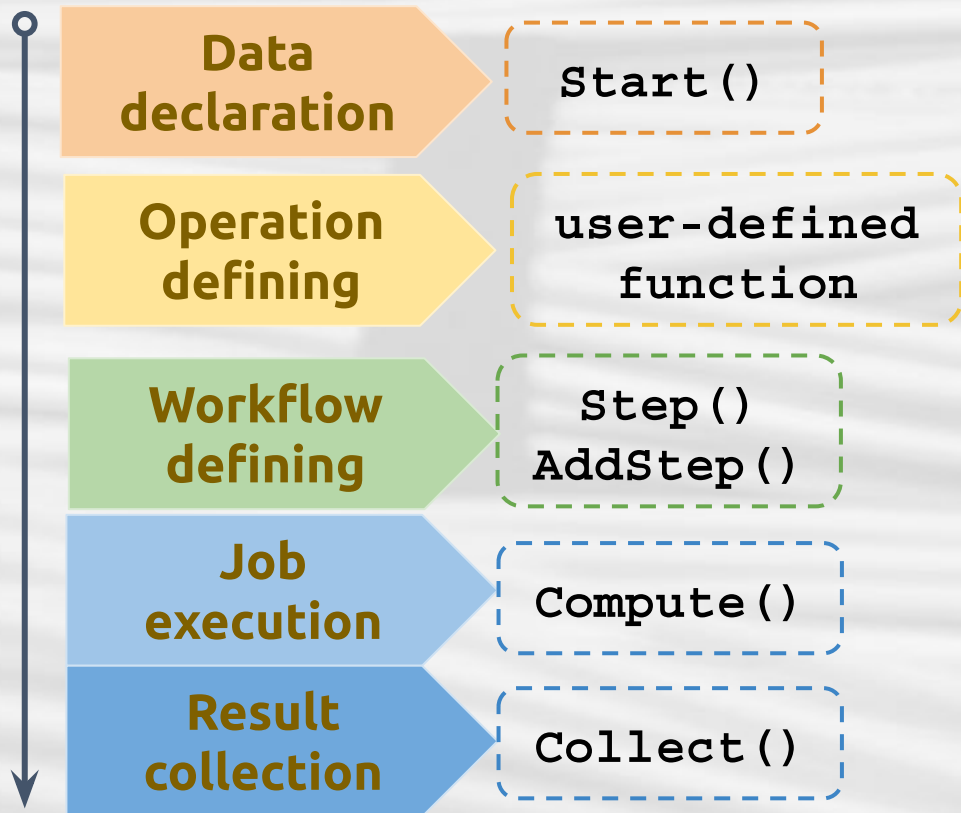


startR features

- ★ An R package tailored for **big multi-dimensional data** retrieval and processing
- ★ Apply **multiApply** paradigm, which provides flexibility in multi-dimensional data processing
- ★ Implement the MapReduce paradigm (i.e., chunking) on HPCs for **parallel distributed data-processing**
- ★ Pre-processing: data **transformation** or **reordering/reshaping/renaming** dimensions before performing analysis
- ★ Well-preserved metadata during the whole process
- ★ Use **ecFlow** workflow manager for job distribution and monitoring on HPCs
- ★ Acceptable data format: **netCDF** for now, but may be available for other formats.

startR functions and workflow

With startR, users can create a concise script for data analysis with all the information needed.



1. Declare the data sources and the required file/inner dimensions.
2. Define the operations to be applied.
3. Combine the elements from the previous steps to build up the workflow.
4. Set the configuration for the chosen machine and trigger job execution.
5. Collect the results when the execution is finished.

Demo: Compute on HPCs



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Demo: Compute on HPCs

Explain the startR workflow and demonstrate how to submit the job to HPCs for resource-consuming computation.

https://earth.bsc.es/gitlab/es/startR/-/blob/develop-tutorial_PATC2021/inst/doc/tutorial/PATC2021/nord3_demo.R

Prepare the environment on MN4

(log in VM)

- ssh mn4
- module load intel impi mkl R/3.6.1 CDO/1.8.2
- R
- library(startR)

Use case 1: Seasonal forecast verification



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Use case 1: Seasonal forecast verification

In this use case, we learn how to use the startR workflow to finish a piece of analysis, including defining and pre-processing the desired data, defining the operation, building the workflow, and executing the operation.

The ensemble-adjusted Continuous Ranked Probability Score (CRPS) and the root mean square error skill score (RMSSS) are computed to verify the forecast.

To make the process faster, the required data size is small here so we can run the execution on workstation.

[Question]

https://earth.bsc.es/gitlab/es/startR/-/blob/develop-tutorial_PATC2021/inst/doc/tutorial/PATC2021/hands-on_1_workflow.md

[Answer]

https://earth.bsc.es/gitlab/es/startR/-/blob/develop-tutorial_PATC2021/inst/doc/tutorial/PATC2021/hands-on_1_workflow_ans.md

Use case 2: Spatial grid interpolation



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

Use case 2: Spatial grid interpolation

In this use case, we will learn how to use `Start()` to load the data to the local memory space and do the spatial interpolation. The default transformation function is `startR::CDORemapper`, a wrapper function of `s2dv::CDORemap` that uses `cdo` inside.

[Question]

https://earth.bsc.es/gitlab/es/startR/-/blob/develop-tutorial_PATC2021/inst/doc/tutorial/PATC2021/hands-on_2_interpolation.md

[Answer]

https://earth.bsc.es/gitlab/es/startR/-/blob/develop-tutorial_PATC2021/inst/doc/tutorial/PATC2021/hands-on_2_interpolation_ans.md