Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Climate Forecast Analysis hands-on tutorial: R tools

An-Chi Ho and Eva Rifà

PATC 2022
Earth Sciences Simulation Environments

10th Nov, online

Earth Science Department

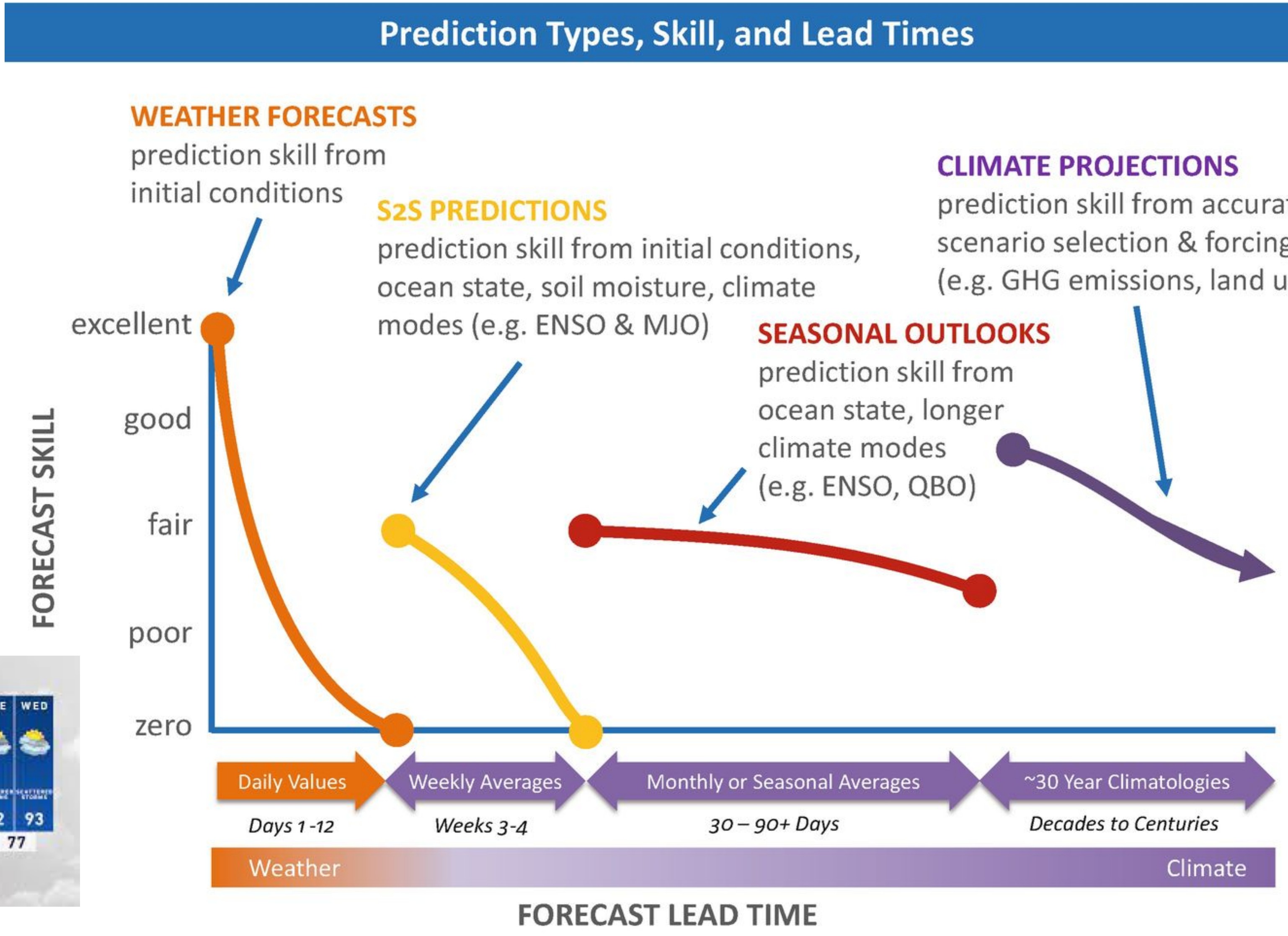
Outline

1. Introduction to Climate Forecasts
2. Introduction to the Climate Forecast Analysis Tools
3. Case studies and Hands-On I
4. startR overview
5. Hands-On II

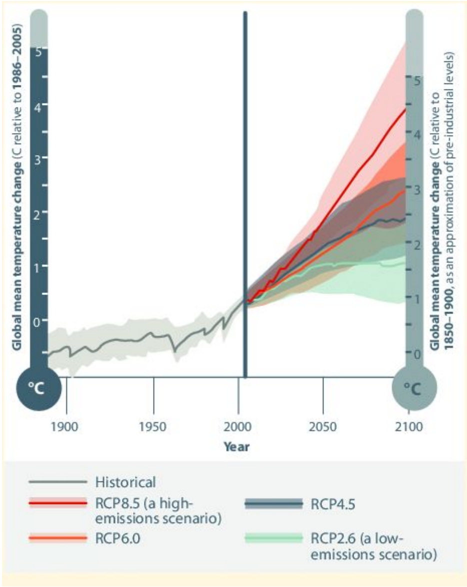
1. Introduction to Climate Forecasts

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Climate Forecast: Forecast horizon



Adapted from iri.columbia.edu/news/qa-subseasonal-prediction-project

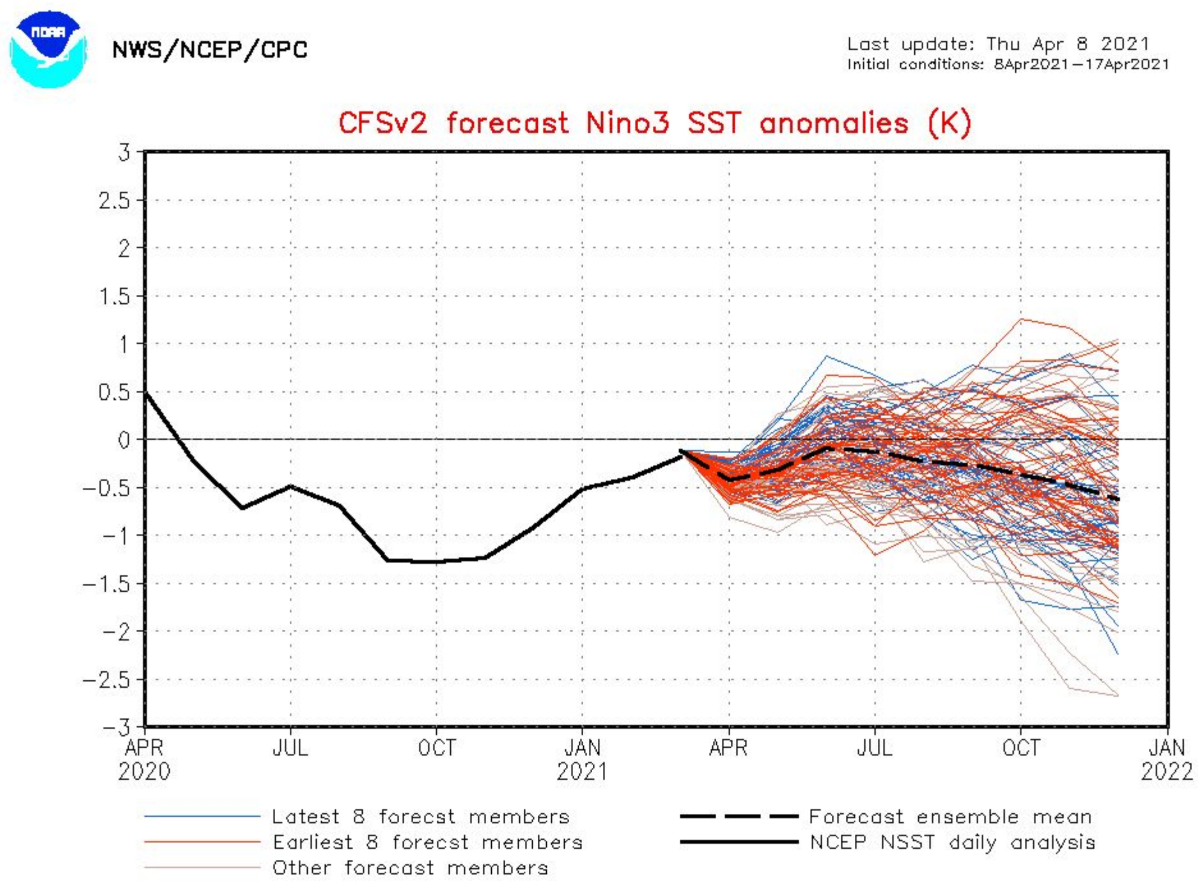


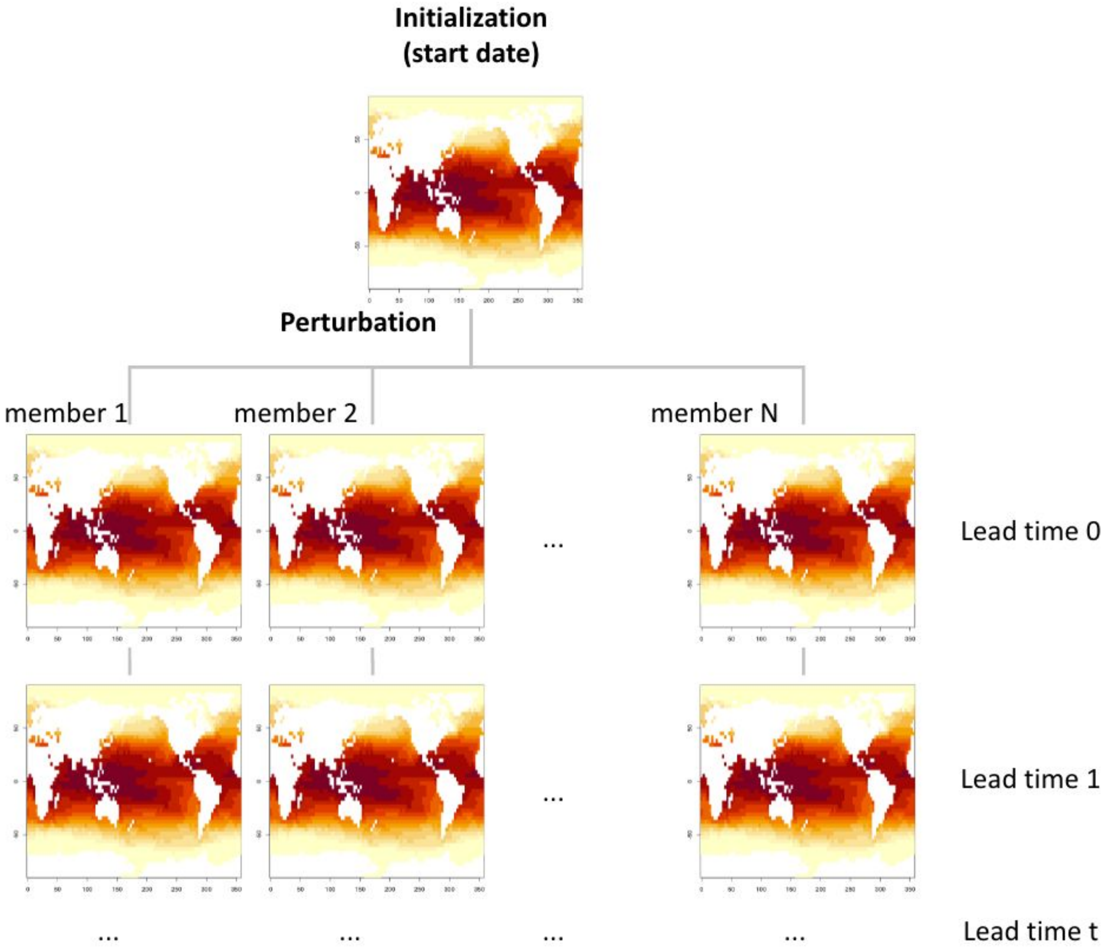
Source: DOI:10.13140/RG.2.2.21145.62564

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Climate Forecast: Ensemble generation

- Different models represent the equations using different parameterizations
- Perturbations on initial conditions are included to generate a ensemble of simulations





Multi-dimensional array with named dimension:

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

Climate Forecast: From Climate data to Climate product

How to turn climate data into useful result or products?

RAW CLIMATE PREDICTIONS

Predictions obtained directly from different climate prediction systems.

BIAS ADJUSTMENT

These adjustments have been applied to improve as much as possible reliability of the climate predictions.

FORECAST QUALITY ASSESSMENT

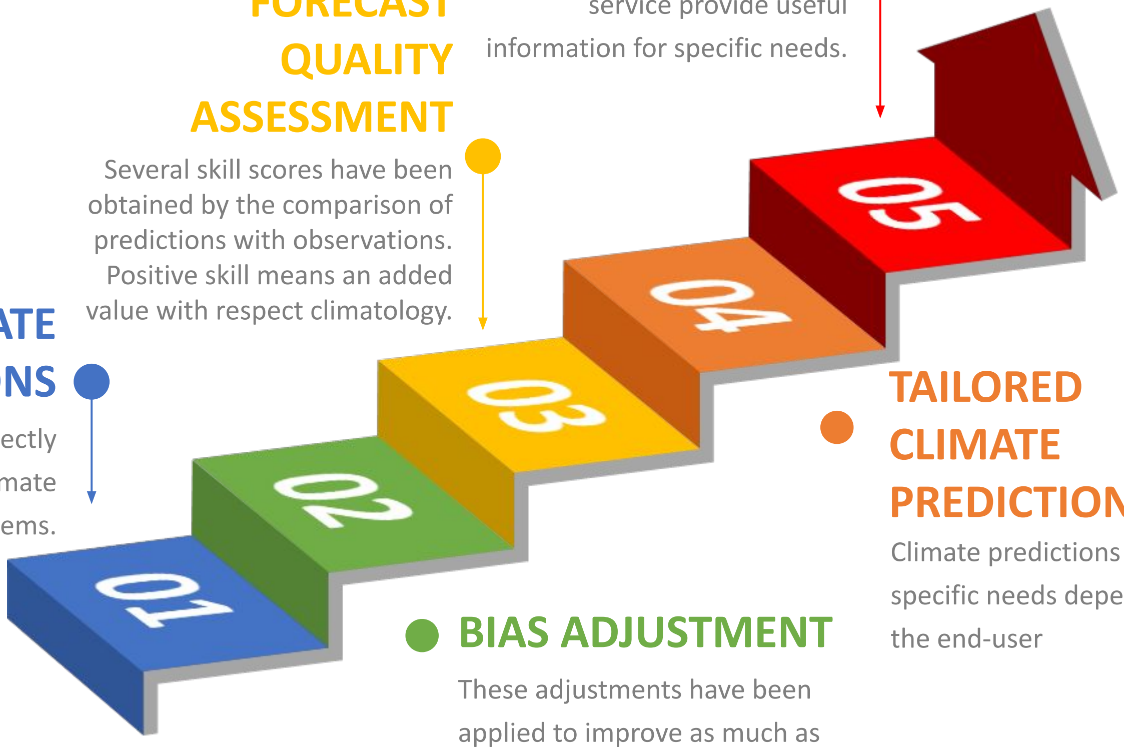
Several skill scores have been obtained by the comparison of predictions with observations. Positive skill means an added value with respect climatology.

TAILORED CLIMATE PREDICTIONS

Climate predictions tailored to specific needs depending on the end-user

CLIMATE SERVICE PRODUCT

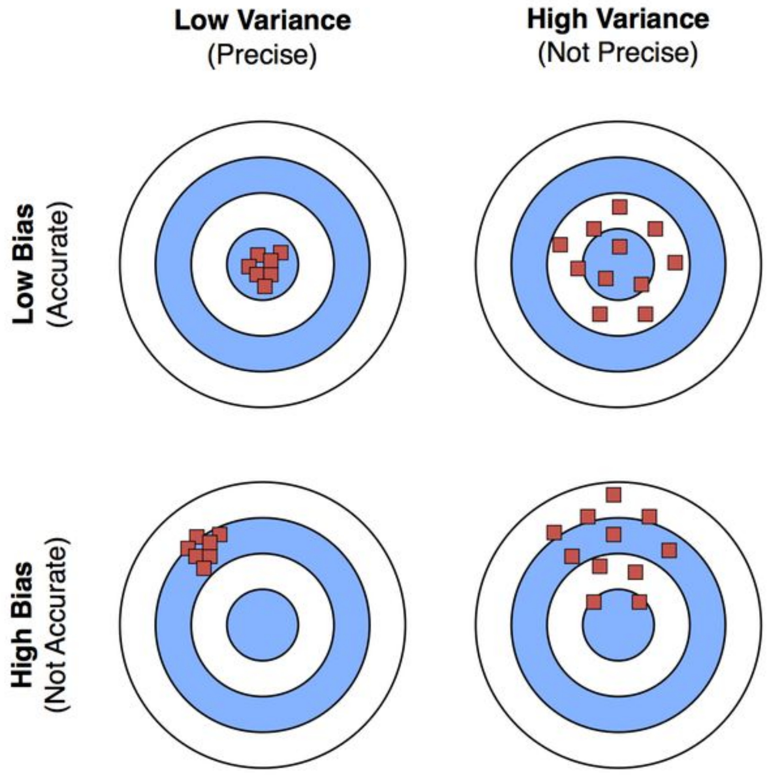
The final products from the service provide useful information for specific needs.



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Climate Forecast: Processing

- The raw experimental output may have room to be improved by reference data (e.g., observation)
- Many methods can improve the quality of forecast, e.g., bias correction, variance inflation, minimized mean-squared error, etc.



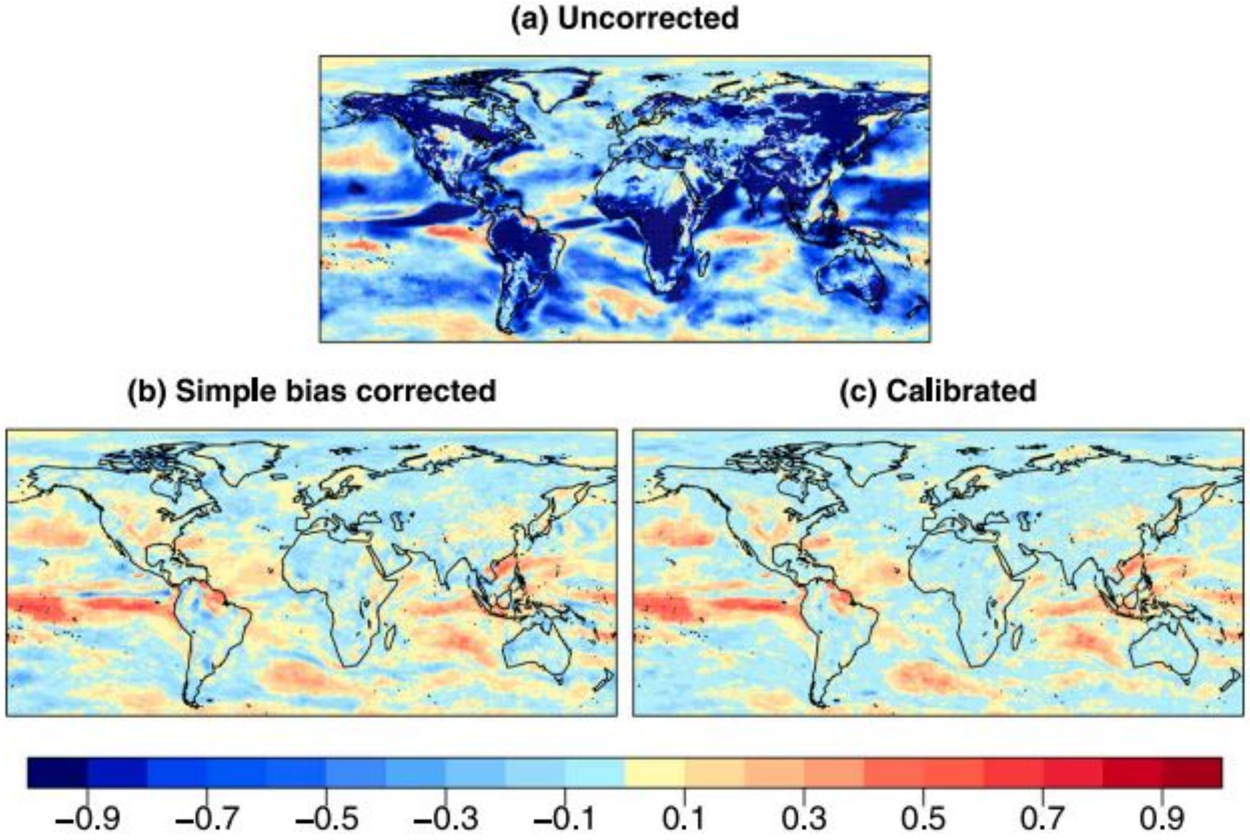


FIG. 4. Fair RPSS for tercile events of 10-m wind speed forecasts from ECMWF System 4 and ERA-Interim reanalysis in winter (DJF). These predictions have been initialized on 1 Nov for the period of 1981–2012.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Torralba, V., Doblas-Reyes, F. J., MacLeod, D., Christel, I., & Davis, M. (2017). Seasonal Climate Prediction: A New Source of Information for the Management of Wind Energy Resources, *Journal of Applied Meteorology and Climatology*, *56*(5), 1231-1247. Retrieved Apr 9, 2021, from https://journals.ametsoc.org/view/journals/apme/56/5/jamc-d-16-0204.1.xml

Climate Forecast: From Climate data to Climate product

How to turn climate data into useful result or products?

CLIMATE SERVICE PRODUCT

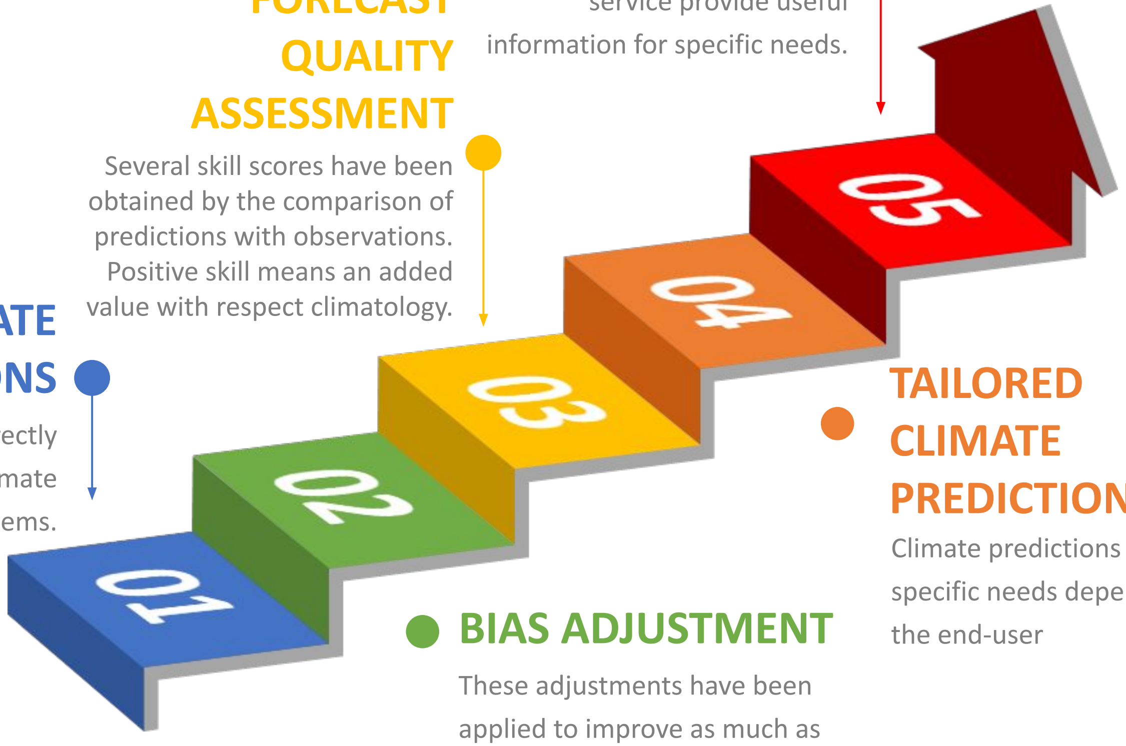
The final products from the service provide useful information for specific needs.

FORECAST QUALITY ASSESSMENT

Several skill scores have been obtained by the comparison of predictions with observations. Positive skill means an added value with respect climatology.

RAW CLIMATE PREDICTIONS

Predictions obtained directly from different climate prediction systems.



TAILORED CLIMATE PREDICTIONS

Climate predictions tailored to specific needs depending on the end-user

BIAS ADJUSTMENT

These adjustments have been applied to improve as much as possible reliability of the climate predictions.

We need TOOLs for data processing.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

2. Introduction to Climate Forecast Analysis Tools

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

R programming language

R is a generic programming language, especially features in a strong framework for statistical computing and graphics.

- Free software (under the GNU GPL license.) You can install R from https://cran.r-project.org/
- Provides a wide variety of statistical techniques (linear and non-linear modelling, classical statistical tests, classification and simulation...)
- Well developed plotting tools (e.g., ggplot2)
- Once R is installed, the base R packages are installed along. But you can further install other packages. See the list of all the available packages on CRAN https://cran.r-project.org/web/packages/available_packages_by_name.html

Climate Forecast Analysis Tools: Table of packages

★ Functions are split into packages depending on their objective or project requirement

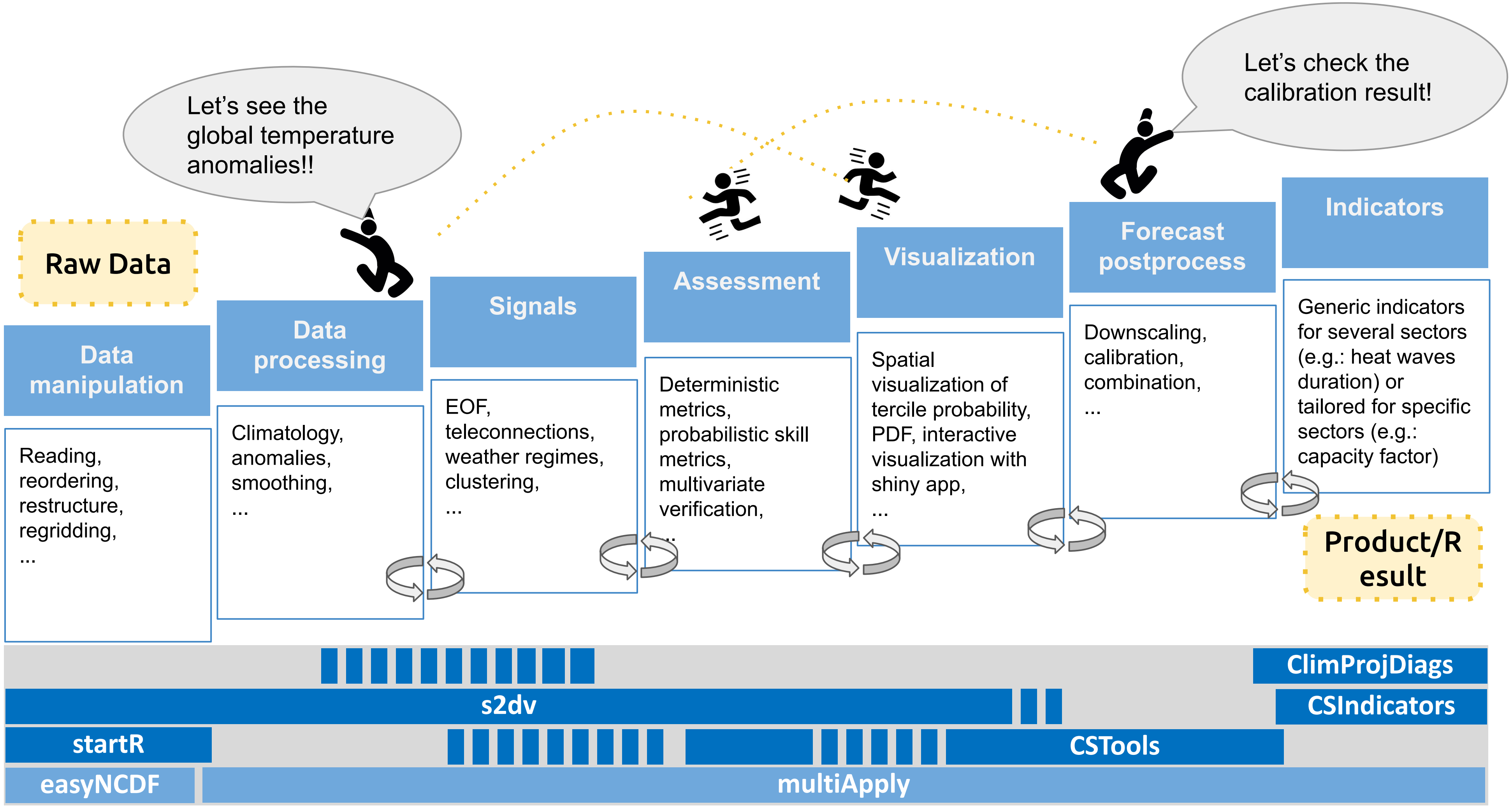
★ Functions from different packages (including external packages) can be used together to perform analyses or obtain climate service products

	Package name	Short description	Link to CRAN and GitLab
Data loading and manipulation	**easyNCDF**	Read/write netCDF files into/from multidimensional R array.	https://CRAN.R-project.org/package=easyNCDF https://earth.bsc.es/gitlab/es/easyNCDF
	startR	Data retrieval and processing tools	https://CRAN.R-project.org/package=startR https://earth.bsc.es/gitlab/es/startR
	multiApply	Apply functions to multiple multidimensional arrays or vectors allowing parallel computation	https://CRAN.R-project.org/package=multiApply https://earth.bsc.es/gitlab/ces/multiApply
Analysis and processing	**s2dv**	Functions for Forecast Verification and visualization	https://CRAN.R-project.org/package=s2dv https://earth.bsc.es/gitlab/es/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 https://earth.bsc.es/gitlab/external/cstools
Climate indicators	**CSIndicators**	Sectorial Indicators for Climate Service	https://CRAN.R-project.org/package=CSIndicators https://earth.bsc.es/gitlab/es/csindicators
	ClimProjDiags	Climate extreme indices, evaluation of the agreement between models, weight and combination functions.	https://CRAN.R-project.org/package=ClimProjDiags https://earth.bsc.es/gitlab/es/ClimProjDiags

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

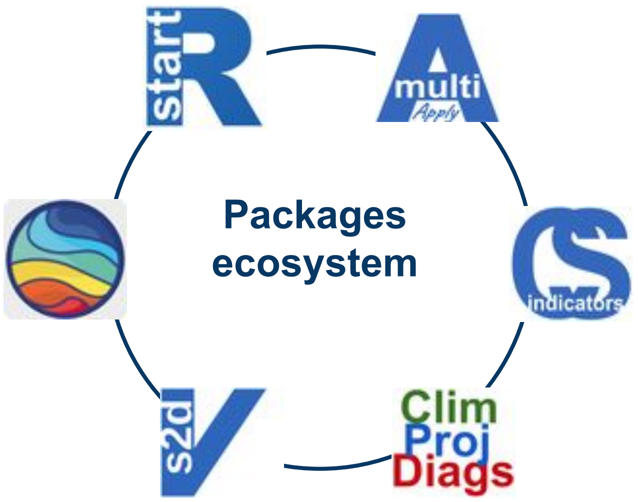
Climate Forecast Analysis Tools: Interoperability & Methods

- The tools can be used interchangeably, depending on the needs
- The package ecosystem aims to cover the whole data analysis cycle.



Climate Forecast Analysis Tools: Academic researches

★ **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; Delgado-Torres et al. 2022
Climate Projections	C3S MAGIC	ESMValTool papers: python and R synergy

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Climate Forecast Analysis Tools: Climate services

Data providers

Climate Change Service
climate.copernicus.eu

Subseasonal-to-Seasonal S2S Prediction Project

NOAA

Tools

Services provided

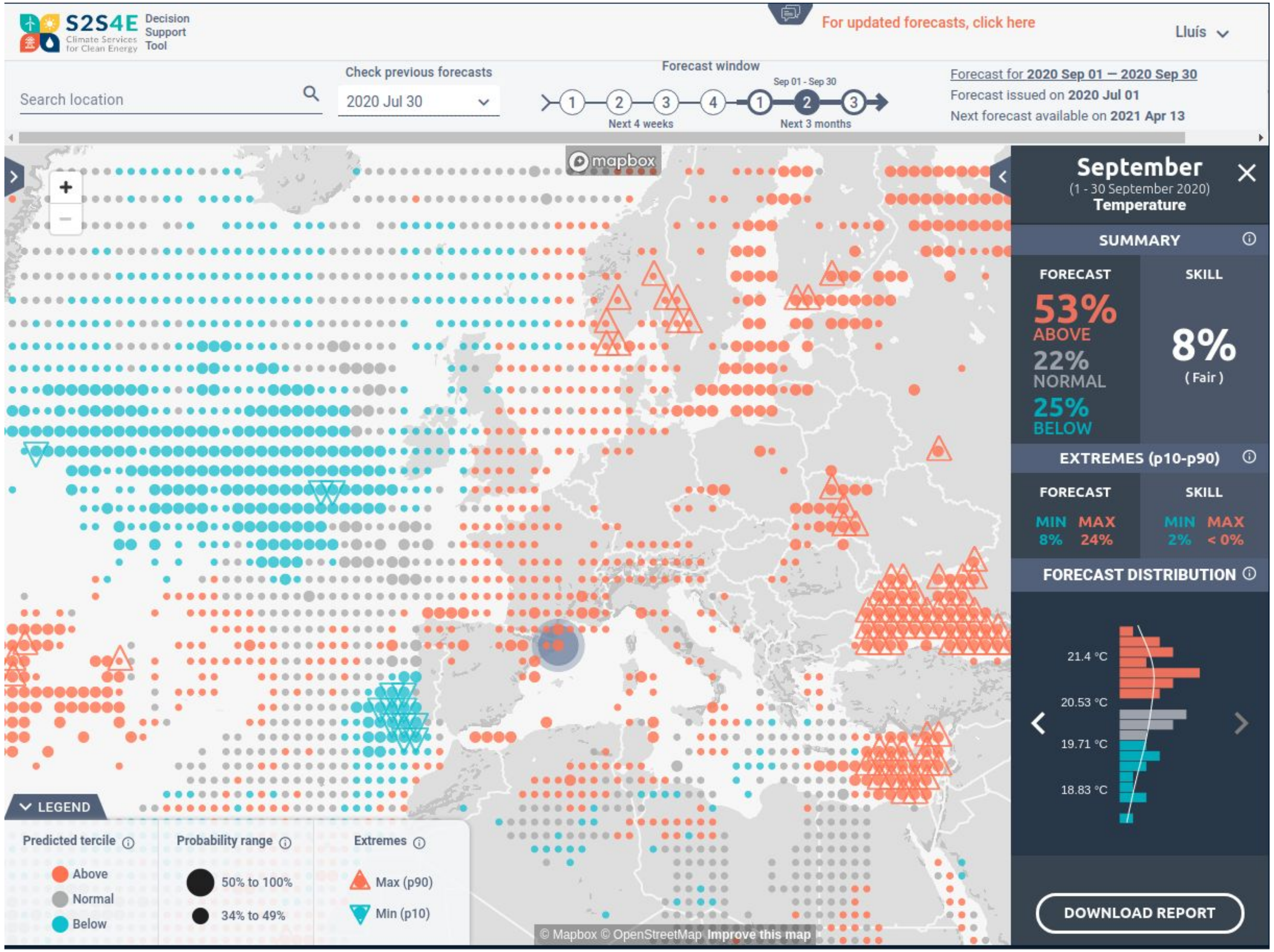
S2S4E Climate Services for Clean Energy

VISCA Vineyards Integrated Smart Climate Application

VITIGEOSS

MED-GOLD

Barcelona Supercomputing Center
Centro Nacional de Supercomputación



NOTE: This tool is no longer updated as the S2S4E project came to an end in 2020.

3. Case studies and Hands-On I

CSTools package

- Used to exploit dynamical seasonal forecasts in order to provide information relevant to stakeholders at the seasonal timescale.
- We develop the package using GitLab: https://earth.bsc.es/gitlab/external/cstools
- The package is published on CRAN: https://cran.r-project.org/web/packages/CSTools/index.html

Basic functions	Correction	Downscaling	Evaluation
CST_Load	CST_BiasCorrection	CST_Analogs	CST_MultivarRMSE
CST_Anomaly	CST_Calibration	CST_RFTemp	CST_MultiMetric
CST_SaveExp	CST_QuantileMapping	CST_RainFARM	
CST_SplitDim	CST_BEI_Weighting	CST_RFSlope	
CST_MergeDims	BEI_PDFBest	CST_RFWeights	
s2dv_cube	CST_DynamicalBC	CST_ADAMONT	
as.s2dv_cube		CST_AnalogsPredictors	

Plotting functions	Classification
PlotMostLikelyQuantileMap	CST_WeatherRegimes CST_RegimeAssign
PlotForecastPDF PlotPDFsOLE	CST_CategoricalForecast
PlotCombinedMap PlotTriangles4Categories	CST_EnsClustering CST_MultiEOF

CST functions and s2dv_cube objects

- An **s2dv_cube** is an object class created to work with multidimensional arrays with named dimensions, specific coordinates and stored metadata.
- **CST_*** functions from **CSTools** and **CSIndicators** packages work with these objects:

```
CST_RainFARM <- function(data, weights = 1.,
                         slope = 0, nf,
                         ...){ })

RainFARM <- function(data, lon, lat, nf,
                     weights = 1., nens = 1,
                     ...) { })
```

```
s2dv_cube object elements
 $ data        : An array with named dimensions
 $ lon         : 1D array of longitudes
 $ lat         : 1D array of latitudes
 $ Variable    : list of 2 elements: "varName" and "level"
 $ Datasets    : list of 2 : "InitiatlizationDates" and "Members"
 $ Dates       : list of 2 with "start" and "end" vectors of dates
 $ when        : A time stamp of data creation time
 $ source_files: Vector with paths to all the found files
 - attr(*, "class")= chr "s2dv_cube"
```

CST_Functions

Both types of functions, with the prefix 'CST_' and without this prefix, can be used. The CST_ functions will produce a more simplified code since these functions can be easily chained up.

CST_ functions work on s2dv_cube object: it is a multi-dimensional array with named dimensions containing the data + metadata

~~CST_~~Functions

The data and/or the analysis must be divided into pieces (chunks) to complete the analysis. This can be done looping or using startR chunking capabilities. In any case, the code would be easier to be managed using functions without prefix.

Functions without prefix work on multi-dimensional array with named dimensions.

Case study: 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: hydropower generation, agriculture, industry, and meltwater shortage can cause heavy economic loss.

→ Reliable seasonal forecasts of snow resources that estimate the snow accumulation at the end of spring.

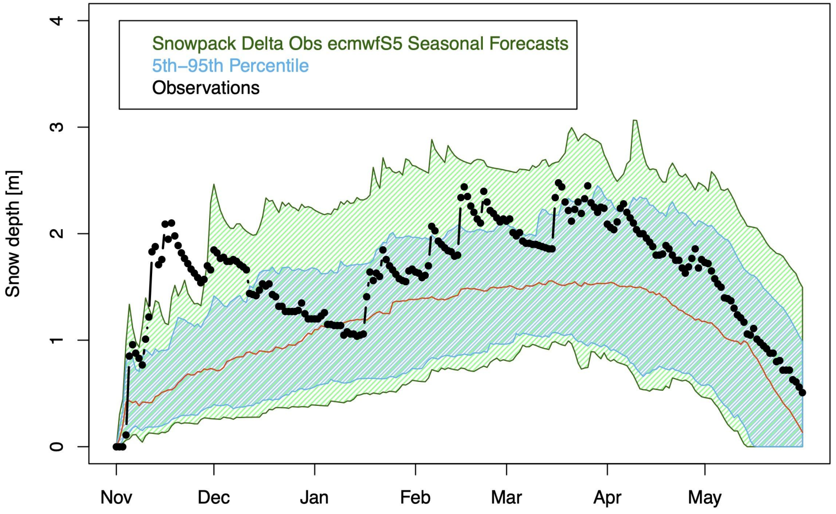
→ The RainFARM downscaling method incorporated within CSTools is employed to downscale precipitation and then used as input for the SNOWPACK model.

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

Case study: SNOWPACK

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

Bocchetta delle Pisse (2410 m) – 2014

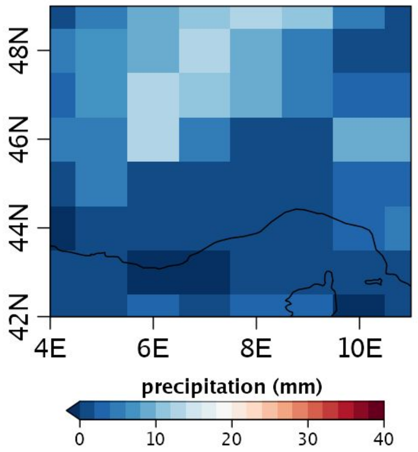


★ **Using CSTools package**, the climate forecast data can be postprocessed to obtain **relevant information for the end-users.**

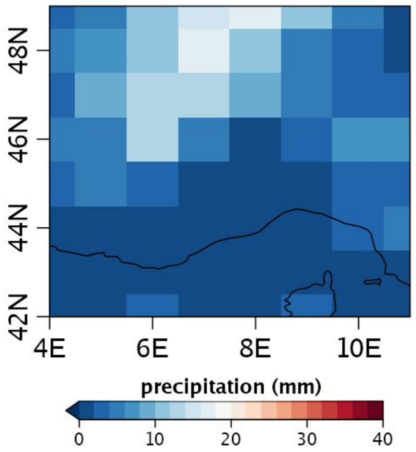
Step 1

Load the data

ECMWF-S5C3S



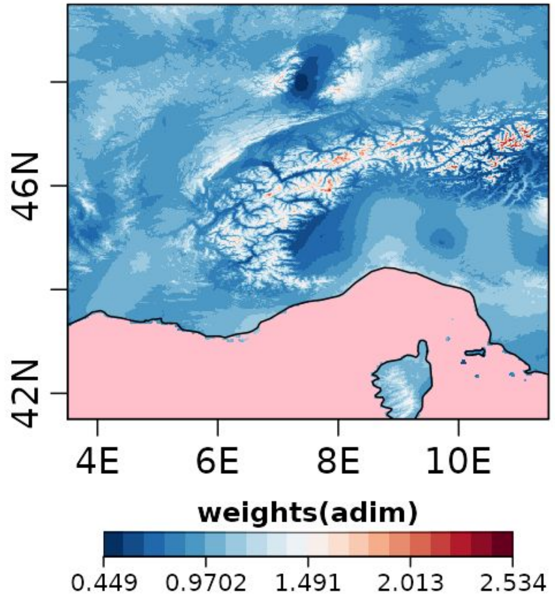
Bias Corrected



Step 2

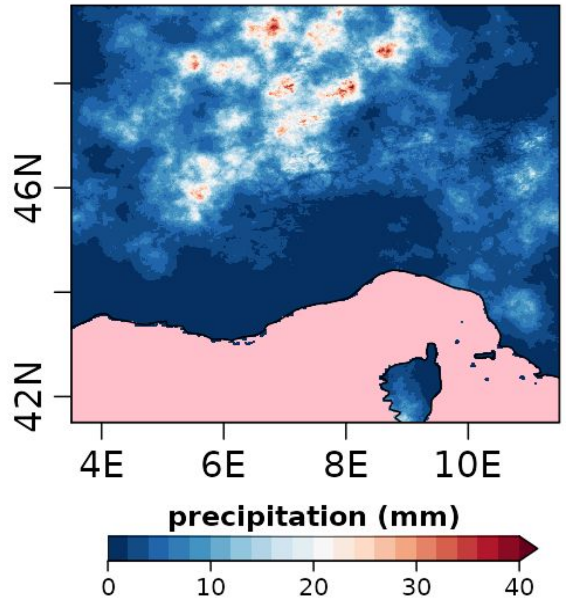
Step 3

Weights



Step 4

Downscaled



Result

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Prepare the environment on MN4

1. If you use Windows, remember to open Xming first for plotting later. (download here)
2. log in VM: `ssh -XY patc{xx}@bsceshandson01.bsc.es`
3. ssh to mn4: `ssh -XY mn4` (passwordless)
4. Require resources:

   ```
   salloc -t 02:00:00 -n 1 -c 16 -J patc_test --x11 --qos=training
   ```
5. Load the required modules:

   ```
   module load gcc/7.2.0 pcre2 intel R/4.1.2 CDO/1.8.2
   ```
6. Open R: `R`

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Hands-on 1: Running CSTools and RainFARM

- Link to the CRAN vignette page:

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

- Link to the GitLab vignette page:

https://earth.bsc.es/gitlab/external/cstools/-/blob/master/vignettes/RainFARM_vignette.Rmd

STEP I: explore a sample data

```
library(CSTools)
exp <- lonlat_prec

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

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Hands-on 1: Running CSTools and RainFARM

STEP II: Downscale using CST_RainFARM function

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

STEP III: 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()
```

STEP IV: 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()
```

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Hands-on 2: Wind Power Generation

The **energy sector** is affected by the atmospheric circulation in many ways.

- Energy supply from renewable sources like wind, solar or hydropower relies on availability of wind, sunshine or water.
- Energy demand is affected by changes in near-surface temperature. A number of indicators derived from atmospheric variables can be useful as proxies of energy production/demand.

We can compute two indicators for wind power generation:

- **WindPowerDensity** - Wind power that is available for extraction from the wind flow, per square meter of swept area.
- **WindCapacityFactor** - Wind power generation of a wind turbine, normalized by the maximum power that the turbine can deliver (rated capacity).

Hands-on 2: Running CSIndicators - Energy Indicators

- Link to CRAN page of CSIndicators package: https://cran.r-project.org/web/packages/CSIndicators/index.html
- Link CRAN: https://cran.r-project.org/web/packages/CSIndicators/vignettes/EnergyIndicators.html
- Link GitLab: https://earth.bsc.es/gitlab/es/csindicators/-/blob/master/vignettes/EnergyIndicators.Rmd

i) WindPowerDensity

STEP I: create sample data

```
set.seed(1)
wind <- rweibull(n = 1000, shape = 2, scale = 6)
WPD <- WindPowerDensity(wind)
mean(WPD)
```

STEP II: visualize histogram

```
par(mfrow=c(1, 2))
hist(wind, breaks = seq(0, 20))
hist(WPD, breaks = seq(0, 4000, 200))
WPD <- WindPowerDensity(wind, ro = 1.15)
```

ii) WindCapacityFactor

STEP I

```
WCFI <- WindCapacityFactor(wind, IEC_class = "I")
WCFIII <- WindCapacityFactor(wind, IEC_class = "III")
```

STEP II: visualize histogram

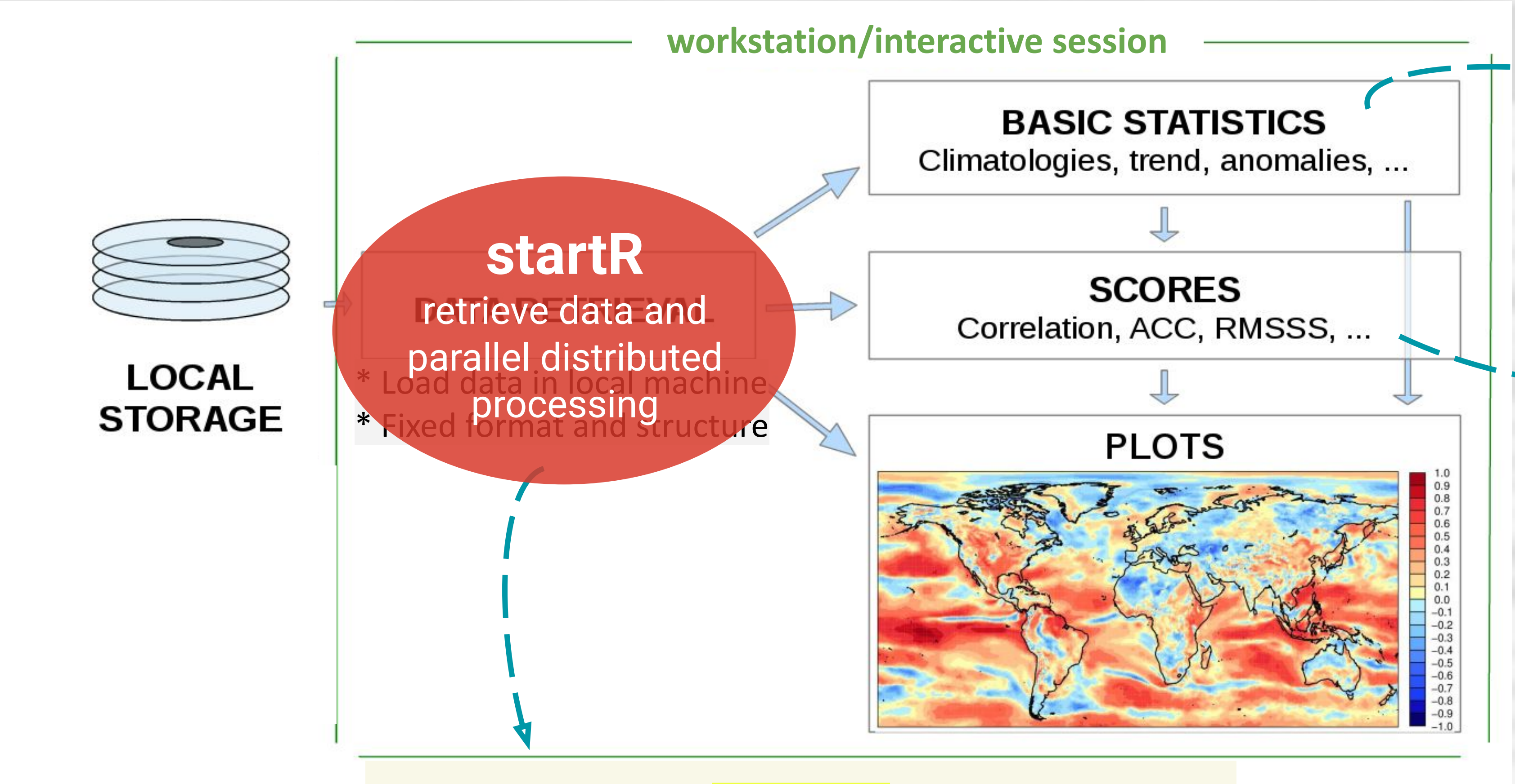
```
par(mfrow=c(1, 3))
hist(wind, breaks = seq(0, 20))
hist(WCFI, breaks = seq(0, 1, 0.05), ylim = c(0, 500))
hist(WCFIII, breaks = seq(0, 1, 0.05),
                      ylim = c(0, 500))
```

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

4. startR overview

How was startR born?

[Data analysis procedure]



Complex analysis

- bootstrapping
- machine learning model
- etc.

Issues

- Longer data loading and processing time
- Limited memory space in local machine

Big data

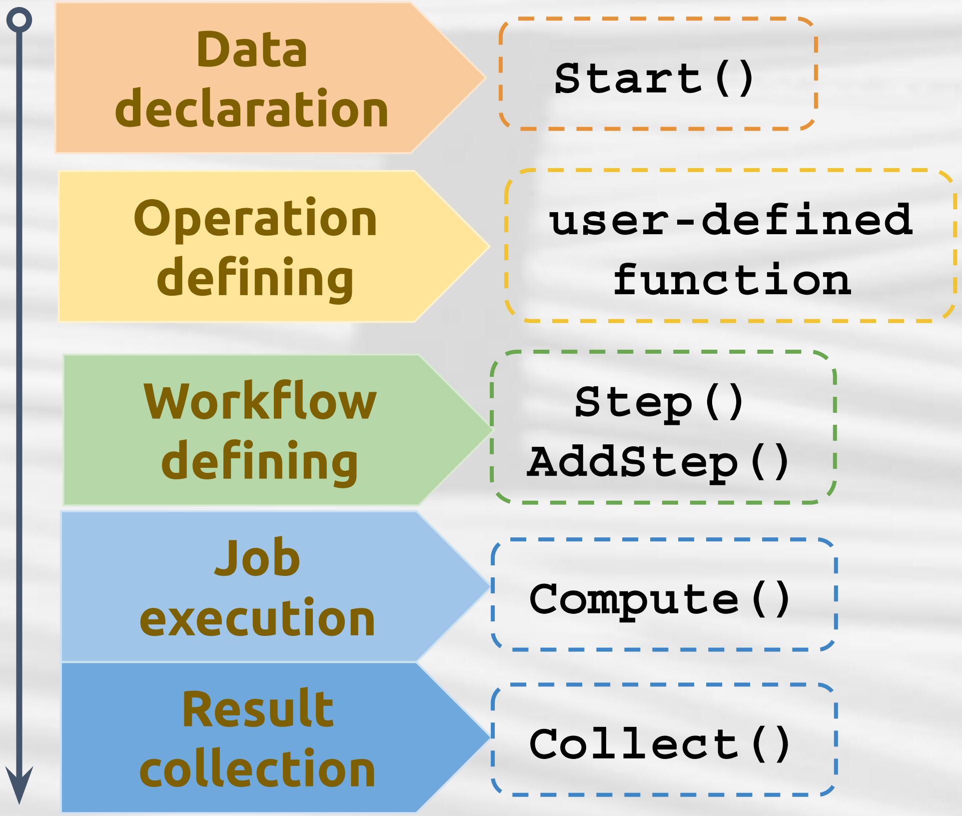
- Higher resolution in all dimensions
- Various types of data required

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

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

Preparation:

```
module load ecFlow
```

Check the whole script on GitLab:
https://earth.bsc.es/gitlab/es/startR/-/blob/develop-PATC2022/inst/doc/tutorial/PATC2022/nord3_demo.R

Demo: Compute on HPCs

Data declaration

```
repos <- "/esarchive/exp/ecmwf/system5_m1/monthly_mean/$var$_f6h/$var$_$sdate$.nc"
lon.min <- 0
lon.max <- 359.9
lat.min <- -90
lat.max <- 90

data <- Start(dat = repos,
              var = 'tas',
              sdate =  c('20170101', '20170201'),
              ensemble = 'all',
              time = 'all',
              latitude = values(list(lat.min, lat.max)),
              latitude_reorder = Sort(),
              longitude = values(list(lon.min, lon.max)),
              longitude_reorder = CircularSort(0, 360),
              synonims = list(latitude = c('lat', 'latitude'),
                              longitude = c('lon', 'longitude')),
             return_vars = list(time = 'sdate',
                                longitude = NULL, latitude = NULL),
                                retrieve = FALSE)
```

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Demo: Compute on HPCs

Data declaration

Operation defining

```
func <- function(x, conf, pval) {
  # x: [ensemble, time]
  # ensemble mean
  ens_mean <- apply(x, 2, mean)
  # temporal trend
  trend <- s2dv::Trend(ens_mean, conf = conf, pval = pval)$trend

  return(trend)
 }
```

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Demo: Compute on HPCs

Data declaration

Operation defining

Workflow defining

```
func <- function(x, conf, pval) {
  # x: [ensemble, time]
  # ensemble mean
  ens_mean <- apply(x, 2, mean)
  # temporal trend
  trend <- s2dv::Trend(ens_mean, conf = conf, pval = pval)$trend

  return(trend)
}
```

```
step <- Step(func, target_dims = c('ensemble', 'time'),
             output_dims = list(trend = 'stats'),
             use_libraries = c('s2dv'))
wf <- AddStep(data, step, conf = FALSE, pval = FALSE)
```

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Demo: Compute on HPCs

Data declaration

Operation defining

Workflow defining

Job execution

```
#------------------user-defined--------------------
  queue_host <- 'nord4' # short name in .ssh/config
  temp_dir <- '/gpfs/scratch/bsc32/bsc32734/startR_hpc/'
  ecflow_suite_dir <- '/home/Earth/aho/startR_local/'
#--------------------------------------------------

  # Nord3-v2
  res <- Compute(wf,
                 chunks = list(latitude = 2, longitude = 2),
                 threads_load = 2, threads_compute = 4,
                 cluster = list(queue_host = 'nord4',
                                queue_type = 'slurm',
                                temp_dir = temp_dir,
                                cores_per_job = 16,
                                job_wallclock = '01:00:00',
                               max_jobs = 4,
#                               extra_queue_params = list('#SBATCH --constraint=medmem'),
                               bidirectional = FALSE,
                               polling_period = 10),
                 ecflow_suite_dir = ecflow_suite_dir,
                 wait = TRUE)
```

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

5. Hands-on II

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Useful commands to check objects

# Find class **class()**	# summary of object **summary()**
# object: list **str()** **names()**	To show the figures on MN4: **display** xxx.png
# object: array **dim()**	
# attributes **attributes()** **attr(...)**	

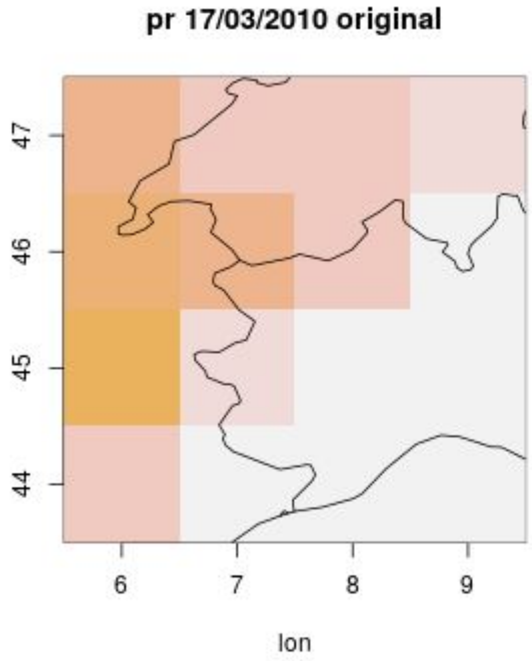
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Hands-on 3: Use startR to load the data

Use package “startR” to load the data used in CSTools [RainFARM vignette]

Data description

This sample data set contains a small cutout of gridded seasonal precipitation forecast data from the Copernicus Climate Change ECMWF-System 5 forecast system. Specifically, for the 'prlr' (precipitation) variable, for the first 6 forecast ensemble members, daily values, for all 31 days in March following the forecast starting dates in November of years 2010 to 2012, for a small 4x4 pixel cutout in a region in the North-Western Italian Alps (44N-47N, 6E-9E). The data resolution is 1 degree.



[Question]
https://earth.bsc.es/gitlab/es/startR/-/blob/develop-PATC2022/inst/doc/tutorial/PATC2022/handson_3-rainfarm.md

[Answer]
https://earth.bsc.es/gitlab/es/startR/-/blob/develop-PATC2022/inst/doc/tutorial/PATC2022/handson_3-rainfarm_ans.md

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Hands-on 4: Skill score in startR workflow

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 function, building the workflow, and executing the operation.

The Ranked Probability Skill Score (RPSS) and the root mean square error skill score (RMSSS) are used to verify the seasonal forecast.

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

[Question]
https://earth.bsc.es/gitlab/es/startR/-/blob/develop-PATC2022/inst/doc/tutorial/PATC2022/handson_4-skill_workflow.md
[Answer]
https://earth.bsc.es/gitlab/es/startR/-/blob/develop-PATC2022/inst/doc/tutorial/PATC2022/handson_4-skill_workflow_ans.md

*To learn more, check: Delgado-Torres et al. DOI: https://doi.org/10.1175/JCLI-D-21-0811.1

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Hands-on 5: 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-PATC2022/inst/doc/tutorial/PATC2022/handson_5-interpolation.md
[Answer]
https://earth.bsc.es/gitlab/es/startR/-/blob/develop-PATC2022/inst/doc/tutorial/PATC2022/handson_5-interpolation_ans.md

Thank you

If you have any question, feel free to contact us!

- **An-Chi Ho** (an.ho@bsc.es)
- **Eva Rifà** (eva.rifarovira@bsc.es)