

Current status of MEDSCOPE CS-Tools R package

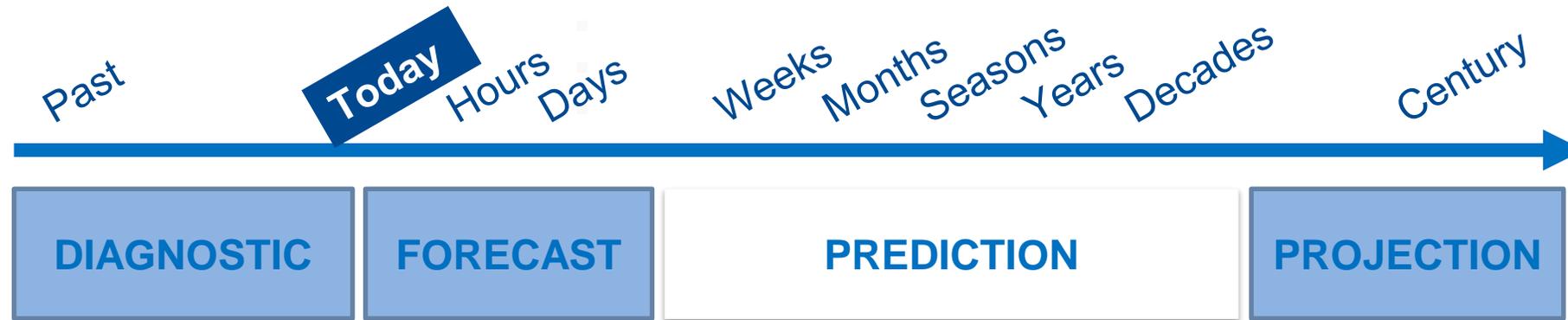
Louis-Philippe Caron, BSC

with contributions from

N Pérez-Zanón, C Alvarez-Castro, L Batté, S Corti, M Dominguez, F Fabiano, S Gualdi,
J von Hardenberg, L Lledó, N Manubens, P Marson, S Materia, E Sánchez, B Van
Schaeybroeck, V Torralba, S Terzago, D Verfaillie, D Volpi and others



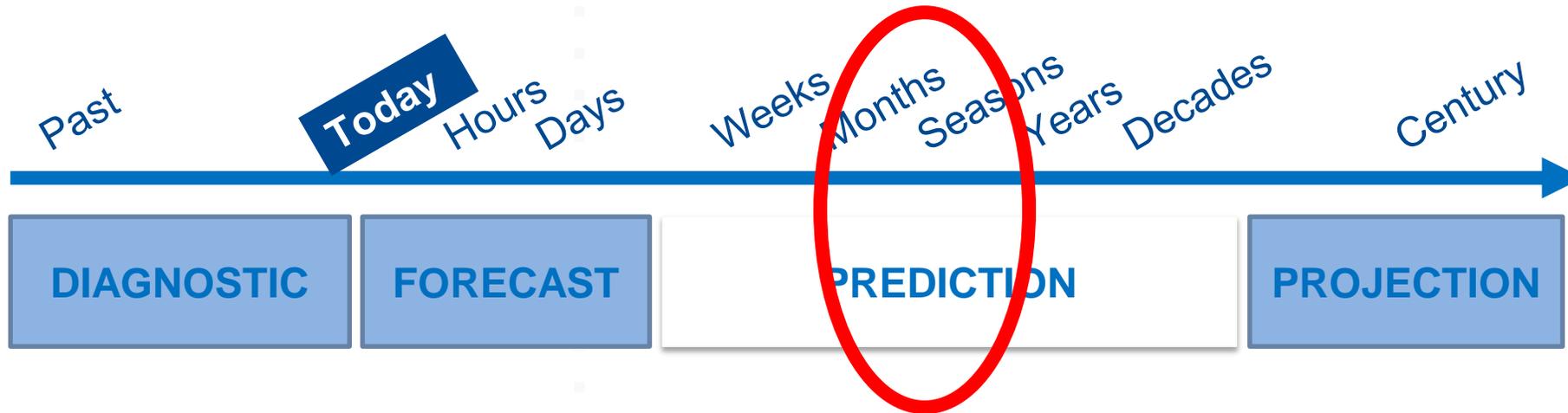
A toolkit for climate predictions



Progression from initial-value problems with weather forecasting at one end and multi-decadal to century projections as a forced boundary condition problem at the other, with climate prediction (**sub-seasonal, seasonal and decadal**) in the middle. Prediction involves initialization and systematic comparison with a **simultaneous** reference.



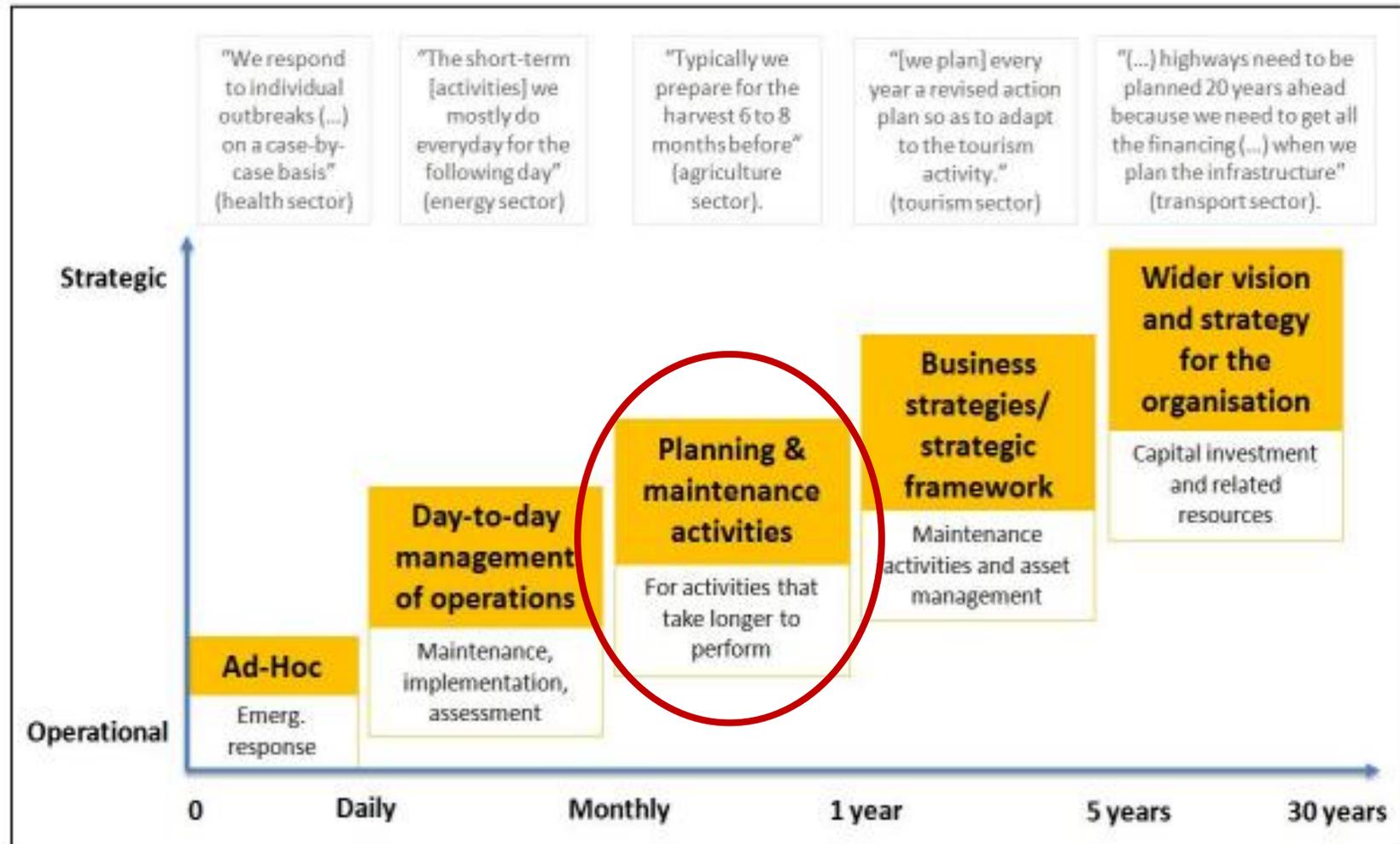
A toolkit for climate predictions



The predictions are based in the initial conditions of the **sea surface temperature, snow cover, soil condition or sea ice**, which have a slow evolution that can range from few months to years.



Near term: A user's view



Dessai and Bruno-Soares (2013)



An example: Users in the energy sector

Long-term user engagement has allowed identifying a number of decisions that could benefit from climate predictions

Energy producers: Resource management strategies

Energy traders: Resource effects on markets, Anticipate energy prices

Plant operators: Planning for maintenance works, especially offshore wind O&M

Plant investors: Anticipate cash flow, optimize return on investment

Grid operators: Anticipate hotter/colder seasons to schedule power plants to reinforce supply.



Near term: Sector readiness

In all sectors there are potential applications but in some sectors the decision making processes that would benefit from predictions are better defined.



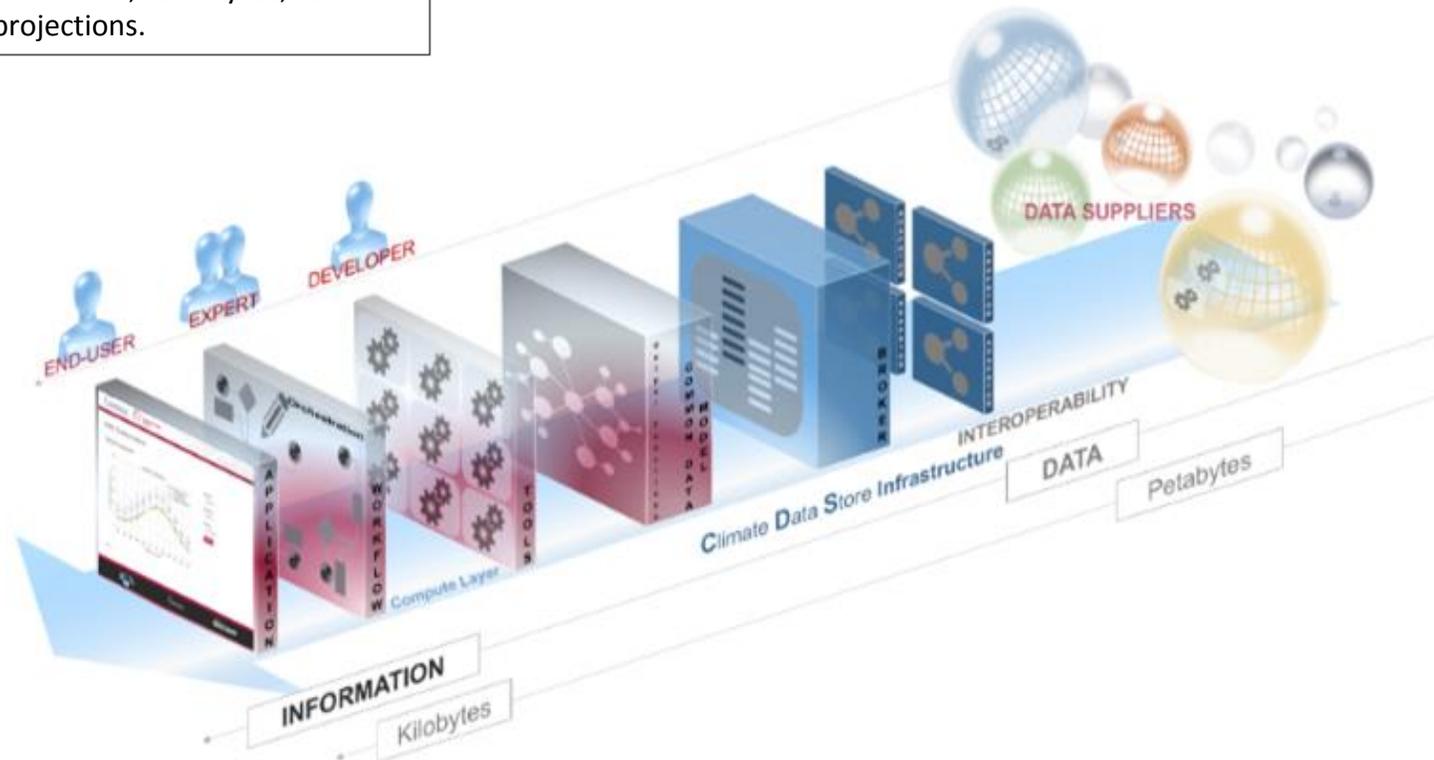
Provided the added value of predictions is illustrated to the users.

From project deliverables of EUCP (D 6.4), PRIMAVERA (D11.6) and EUPORIAS (D12.3). Additional sectoral comments in user engagement by S2S4E, APPLICATE, MED-GOLD, HIATUS and VISCA.



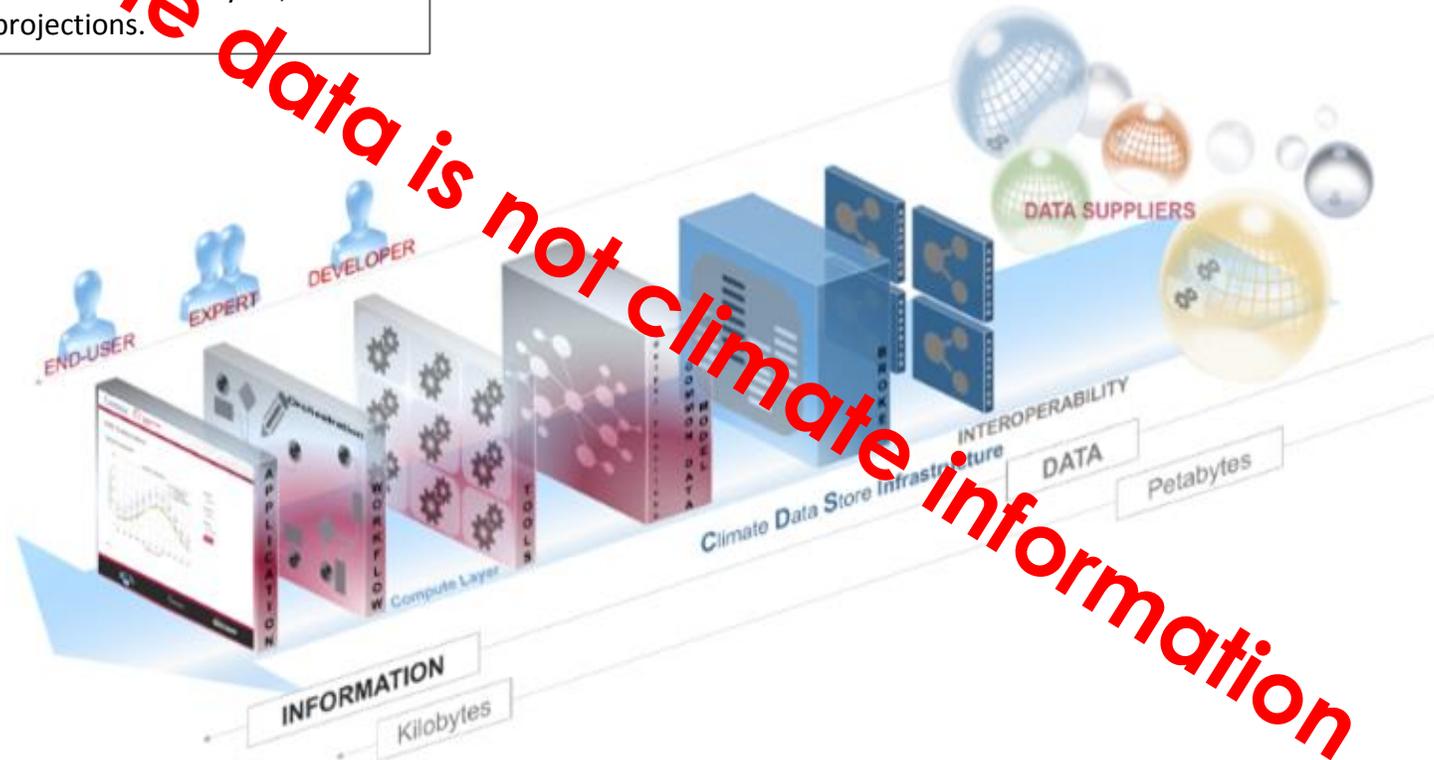
Copernicus Climate Change Service (C3S)

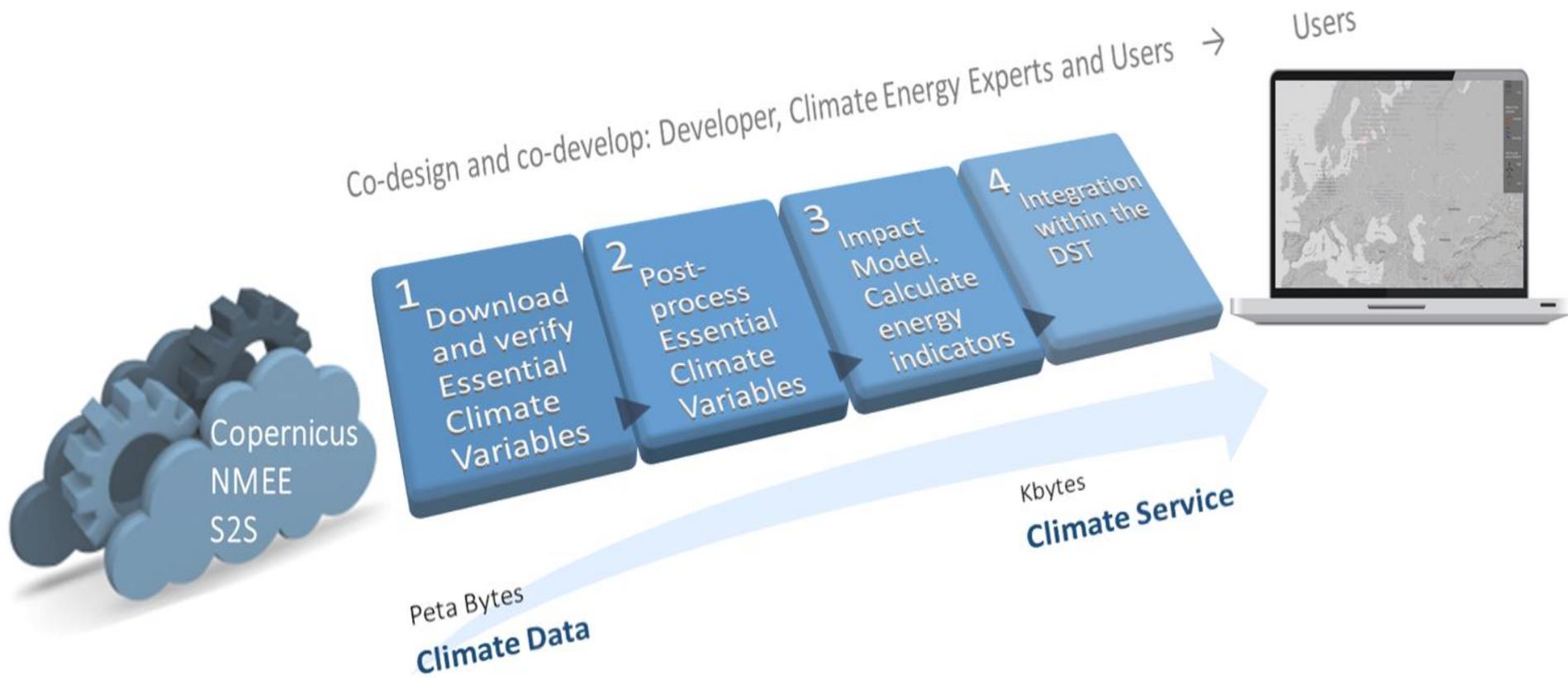
The *Climate Data Store (CDS)* provides a single point of access to a wide range of climate datasets, namely satellite and in-situ observations, reanalyses, seasonal forecasts and climate projections.

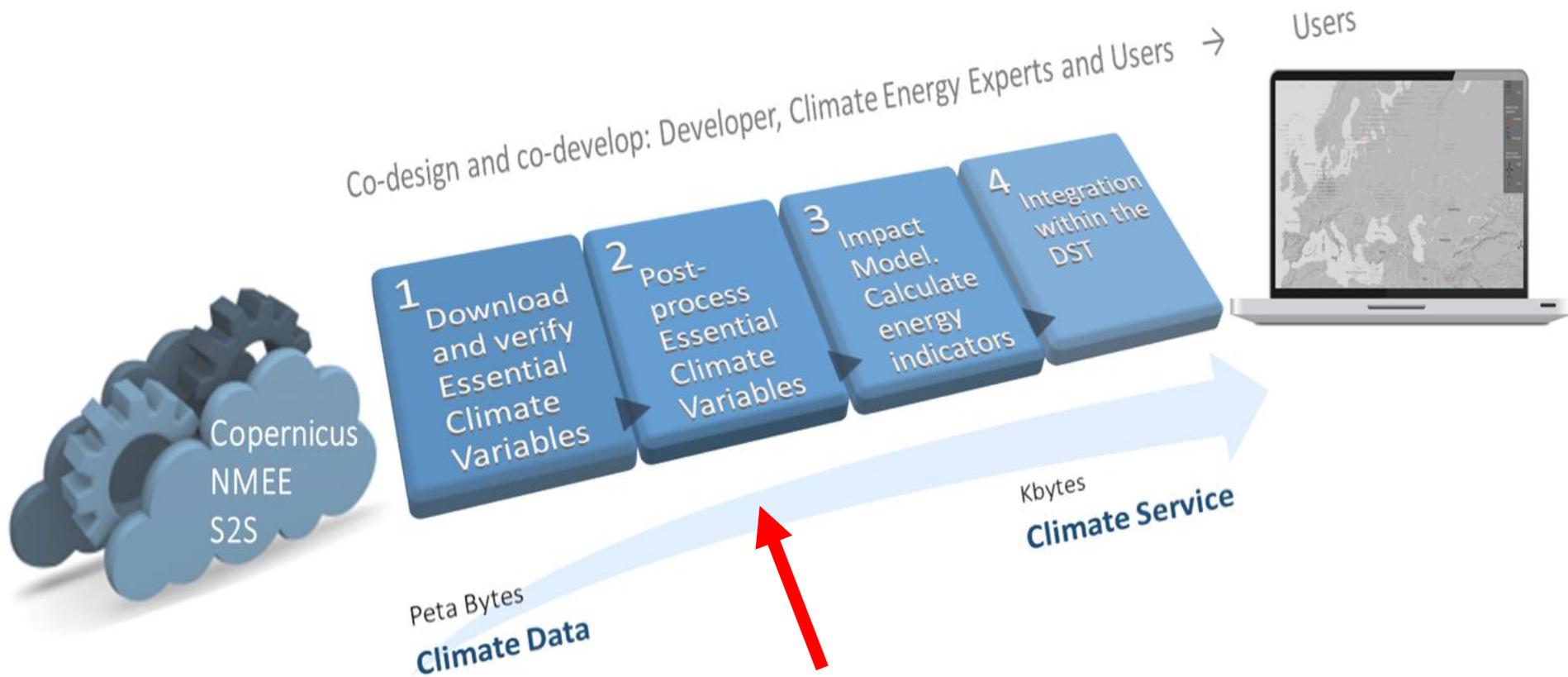


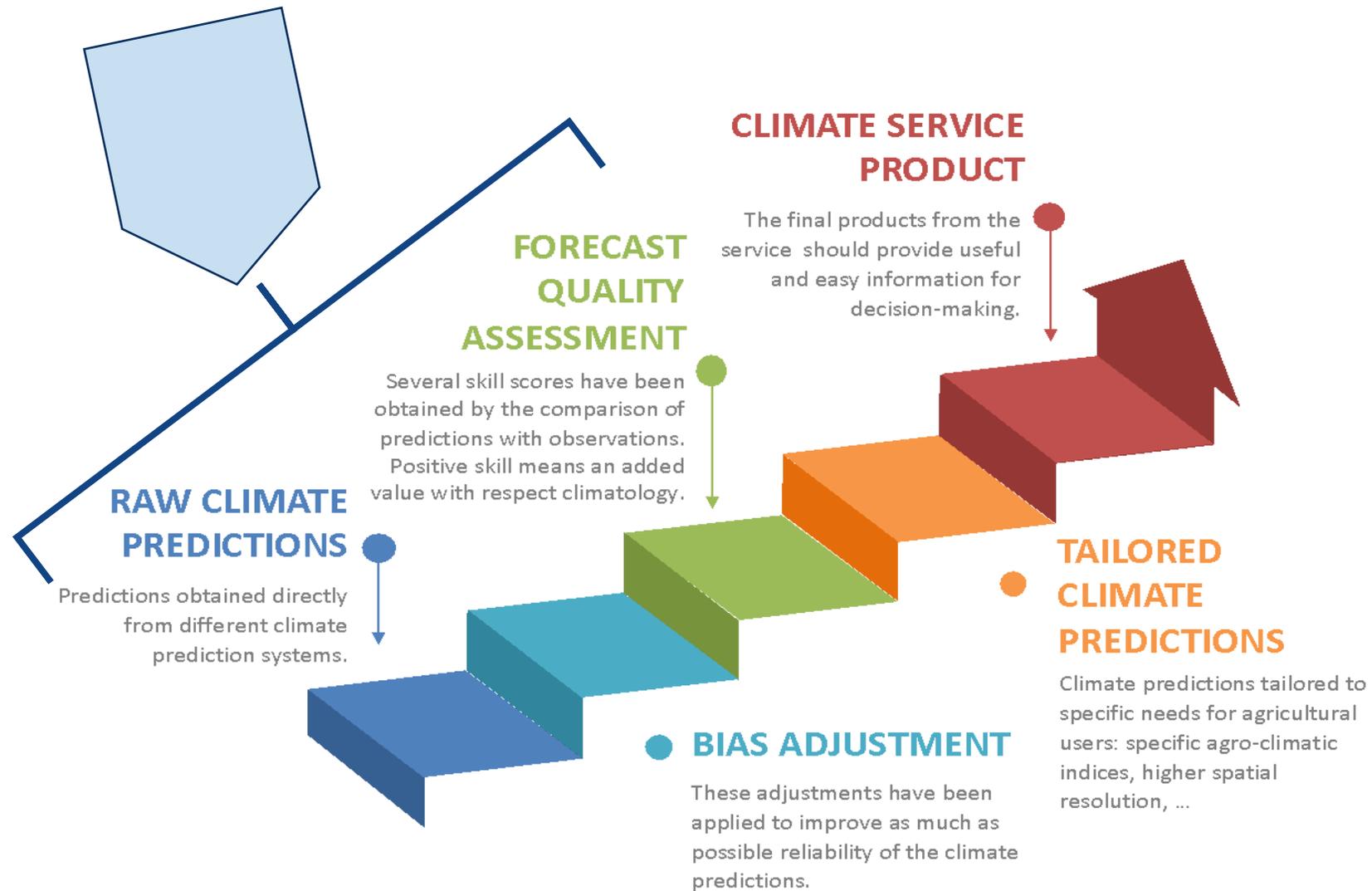
Copernicus Climate Change Service (C3S)

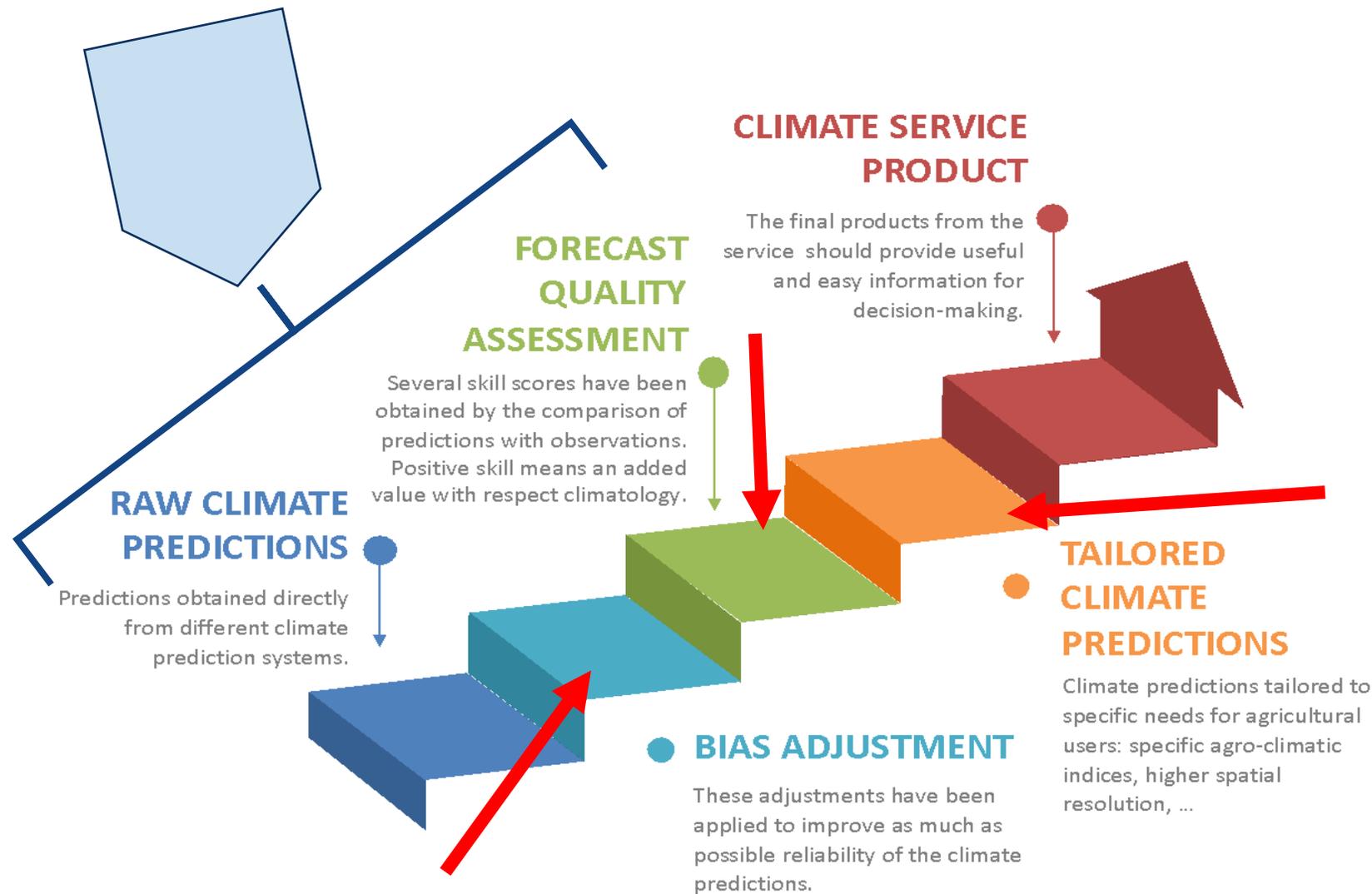
The *Climate Data Store (CDS)* provides a single point of access to a wide range of climate datasets, namely satellite and in-situ observations, reanalyses, seasonal forecasts and climate projections.











Climate Service Tools

- MEDSCOPE *R* package is called **CSTools**
- Currently contains ~25 functions
- Should contain >30 functions upon completion
- Available on CRAN (R repository) since April 24th
- Second release: November 2019
- Further releases: Spring 2020 / Fall 2020
- License: Apache License 2.0
 - A permissive license whose main conditions require preservation of copyright and license notices. Contributors provide an express grant of patent rights. Licensed works, modifications, and larger works may be distributed under different terms and without source code.

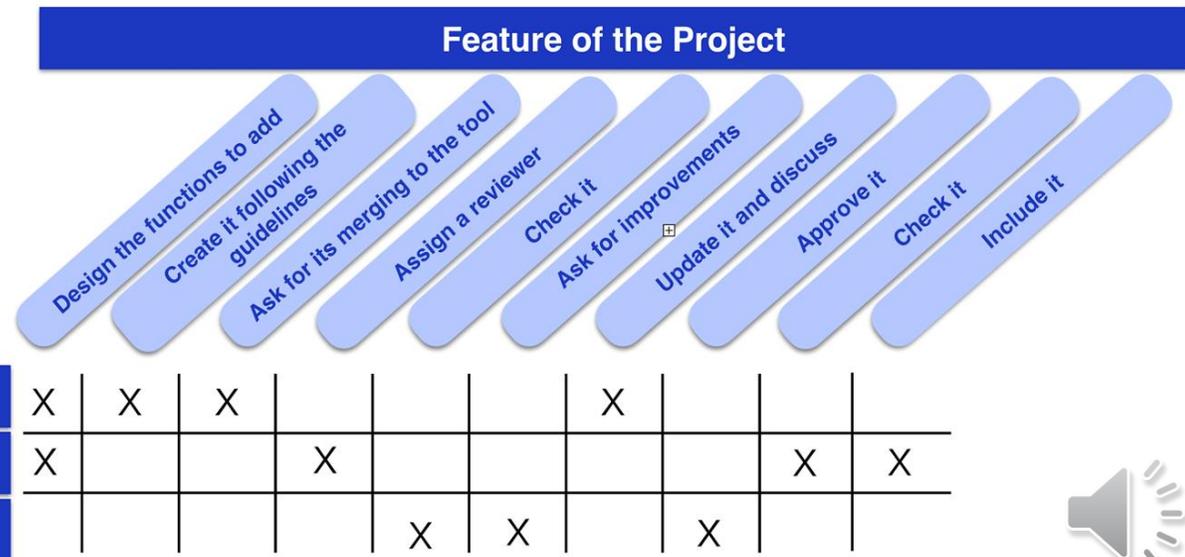


Climate Service Tools

- Use VCS for development: MEDSCOPE Gitlab [page](#)
- [Common procedure](#) and conventions for adding new functions to the Medscope prototype.
- [Documentation](#) and development policy for the MEDSCOPE

prototype

- 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_CategoricalEnsCombina
tion

Downscaling

CST_Analogs
CST_RainFARM
CST_RFSlope
CST_RFWeights
RainFARM
RFSlope

Evaluation

CST_MultivarRMSE
CST_MultiMetric
CST_MultiEOF

Plotting functions

PlotMostLikelyQuantileMap
PlotForecastPDF
PlotCombinedMap

*BEI: Best Estimated Index



Bias adjustment methodologies

Method	Equation	Description	Reference
Simple bias correction	$y'_{i,m} = (y_{i,m} - \bar{y}) \frac{\sigma_x}{\sigma_y} + \bar{x}$	Based on the assumption that both the reference and forecasted distribution are well approximated by a Gaussian distribution.	<i>Leung et al. (1999)</i>
Calibration method	$y'_{i,m} = \alpha \hat{y}_i + \beta (y_{i,m} - \hat{y}_i) + \bar{x}$	Variance inflation modifies the predictions to have the same inter-annual variance as the reference dataset and corrects the ensemble spread to improve the reliability.	<i>Doblas-Reyes et al. (2005)</i>

$y'_{i,m}$ Bias adjusted forecast

$y_{i,m}$ Original forecast

\hat{y}_i Ensemble mean

\bar{y} Forecast climatology

σ_y Forecast standard deviation

\bar{x} Reference climatology

σ_x Reference standard deviation

ρ correlation between the ensemble mean forecast and the observations

σ_e standard deviation of the difference between the ensemble members and the ensemble mean

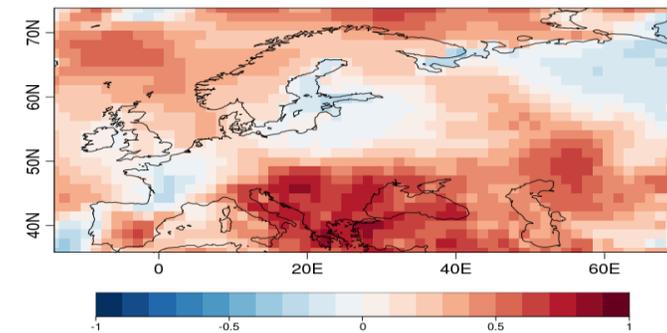
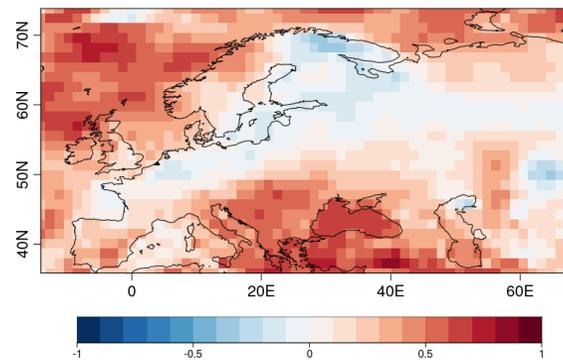
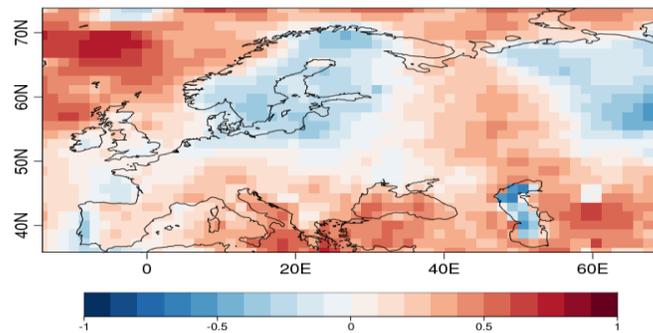
$$\alpha = \text{abs}(\rho) \frac{\sigma_x}{\sigma_y}$$

$$\beta = \sqrt{1 - \rho^2} \frac{\sigma_x}{\sigma_e}$$



MultiMetric

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

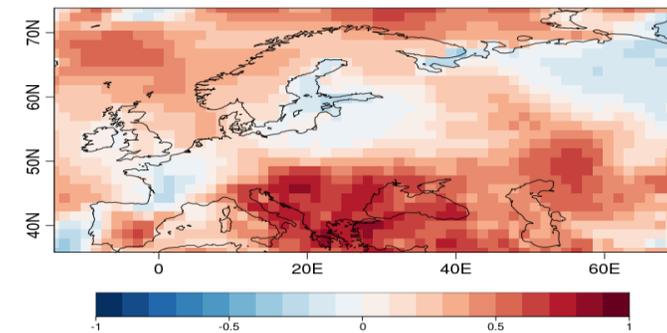
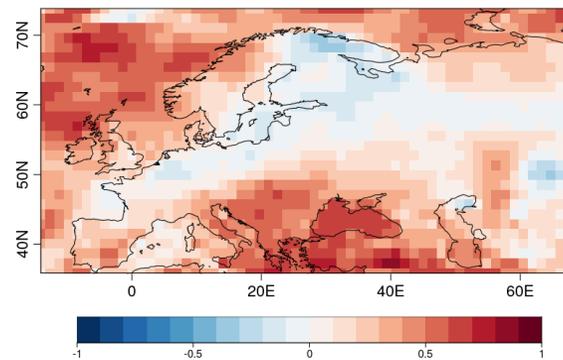
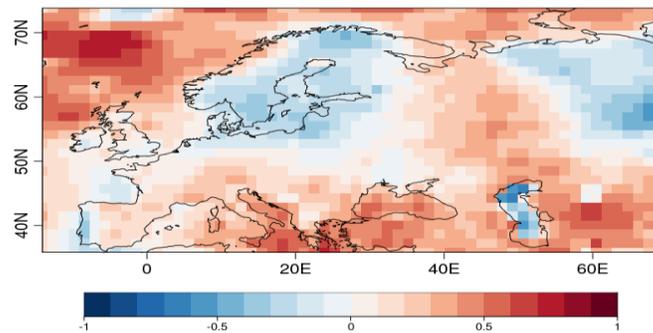


ACC of 2m temperature for 3 forecast system for JJA.



MultiMetric

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



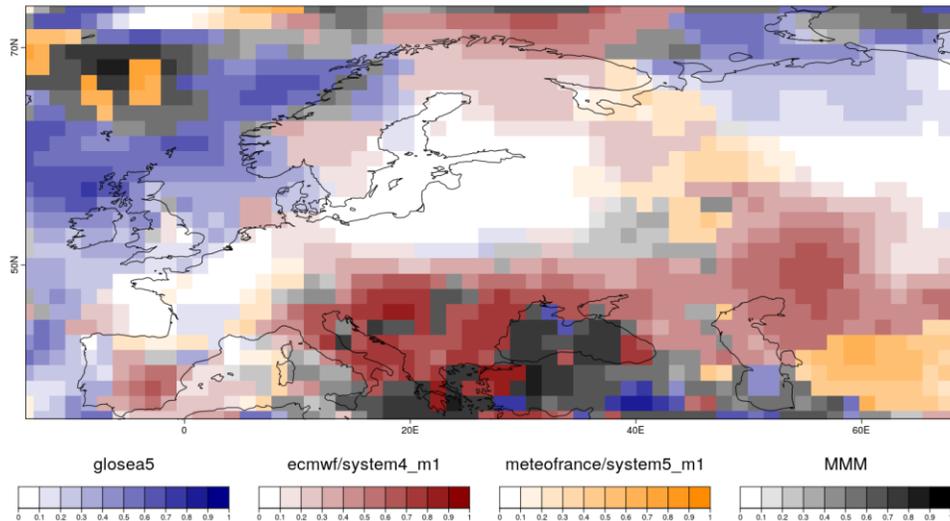
ACC of 2m temperature for 3 forecast system for JJA.

Which one is more skilful for my region?
Better to use the ensemble mean?



PlotCombinedMap

It can be used to identify the best model/forecast over a particular region, as well as the particular level of skill over that region.

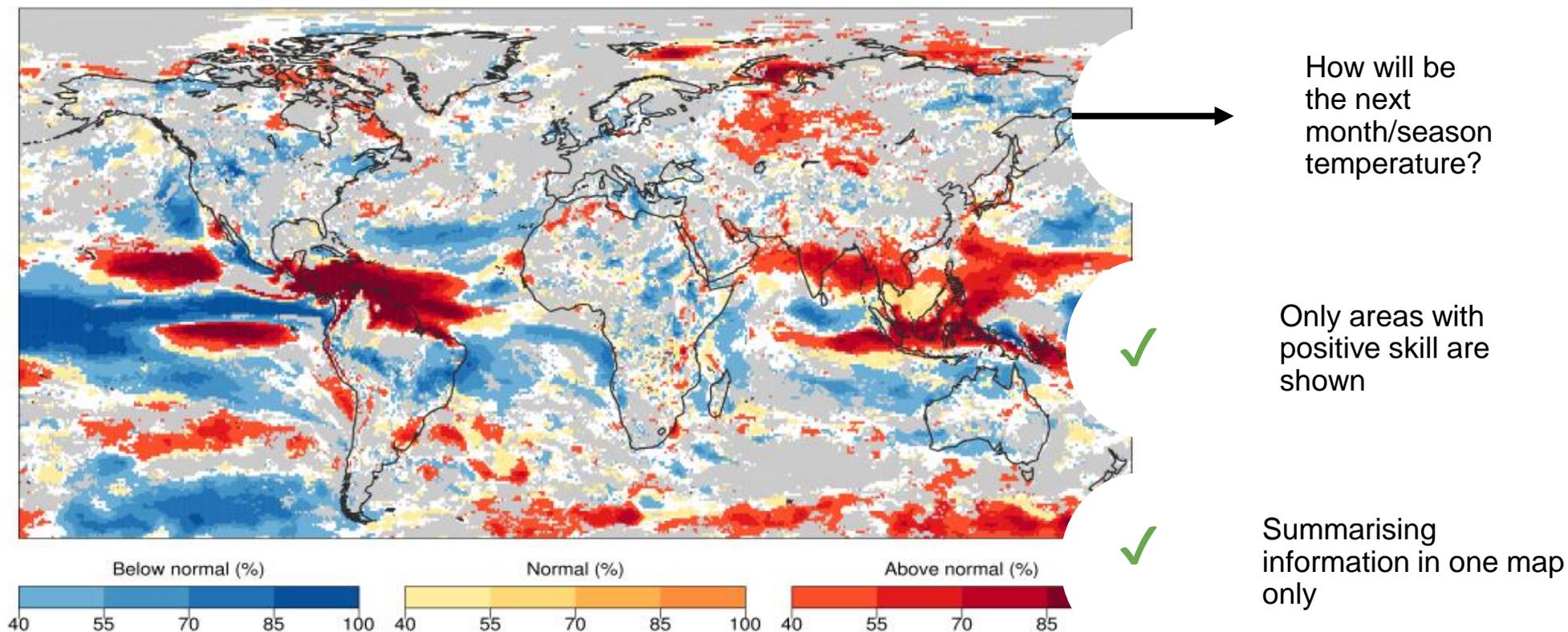


Spatial representation of the highest correlation values for each grid point obtained for three different models - **GloSea5** (blue), **ECMWF System 4** (red) and **Météo-France System 5** (yellow) seasonal forecasting systems as well as the **ensemble mean** (MMM - grey) versus a reference dataset.

Vignette available: https://cran.r-project.org/web/packages/CSTools/vignettes/MultiModelSkill_vignette.html



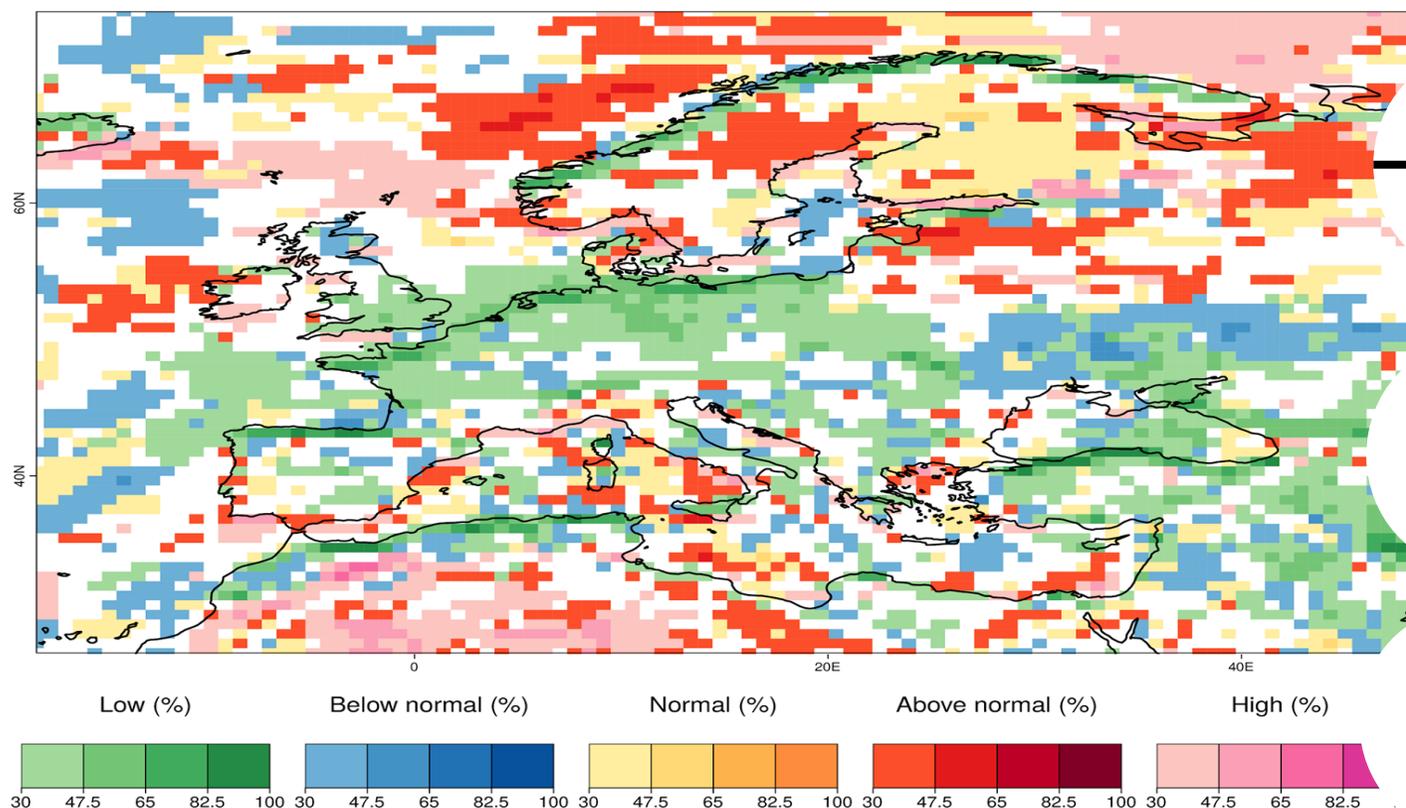
PlotMostLikelyQuantileMap



ECMWF S4 10-m wind speed seasonal forecast for DJF 2015 initialized the 1st of November. The most likely wind speed category (below-normal, normal or above normal) and its percentage probability to occur is shown. White areas show where the probability is less than 40 % and approximately equal for all three categories. Grey areas show where the seasonal forecasts don't improve the climatology.



PlotMostLikelyQuantileMap



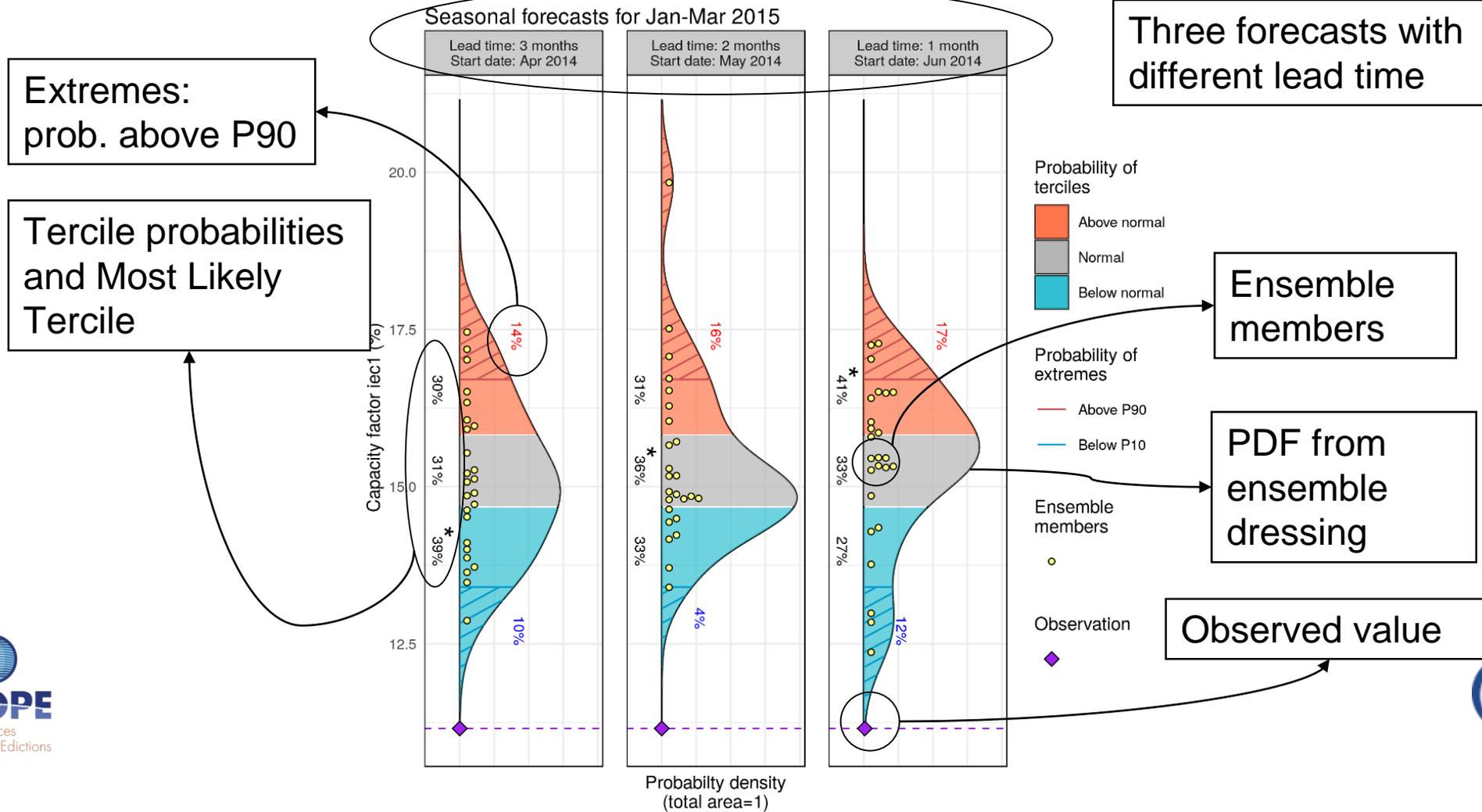
How will be the next month/season temperature?

Only areas with positive skill are shown

Summarising information in one map only

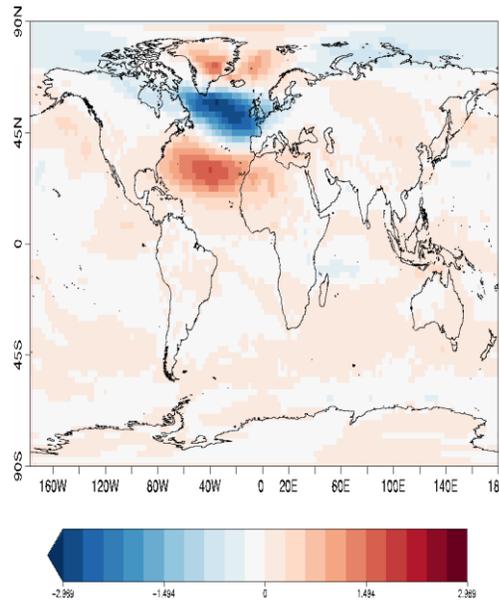


PlotForecastPDF

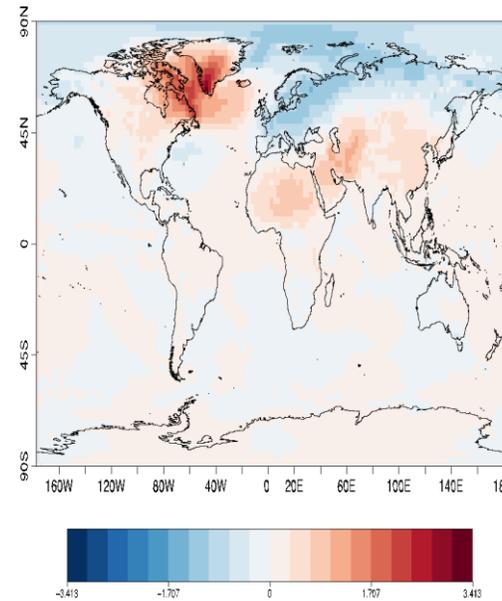


Multi-variable EOFs

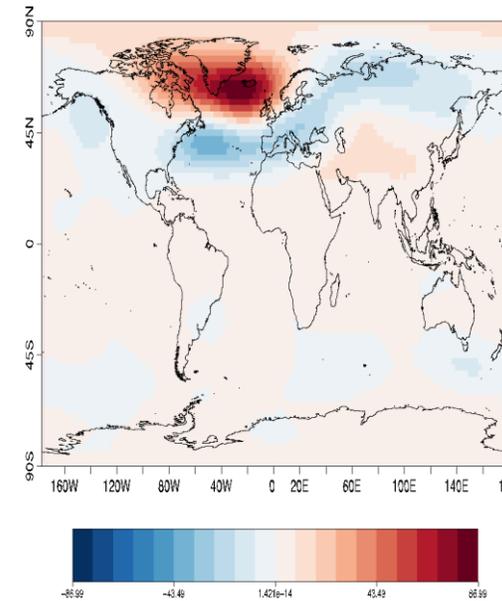
EOF1 for u10



EOF1 for tas



EOF1 for z500

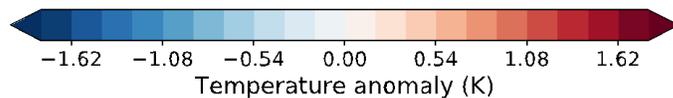
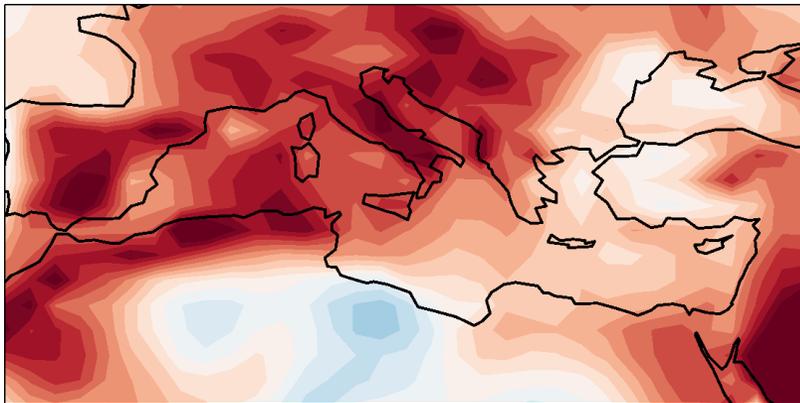


EnsClustering

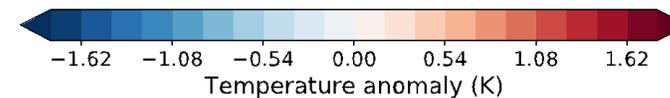
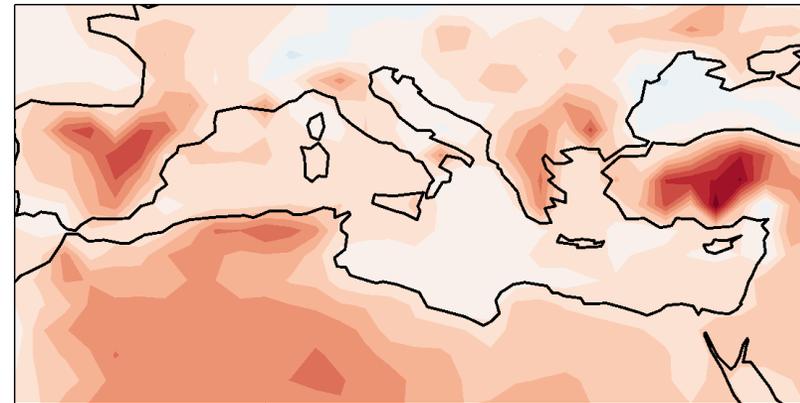
The method has been applied here to the new ECMWF seasonal forecasts (System 5) of 2m temperature in the Mediterranean area.

Summer 2017.

Observed anomaly

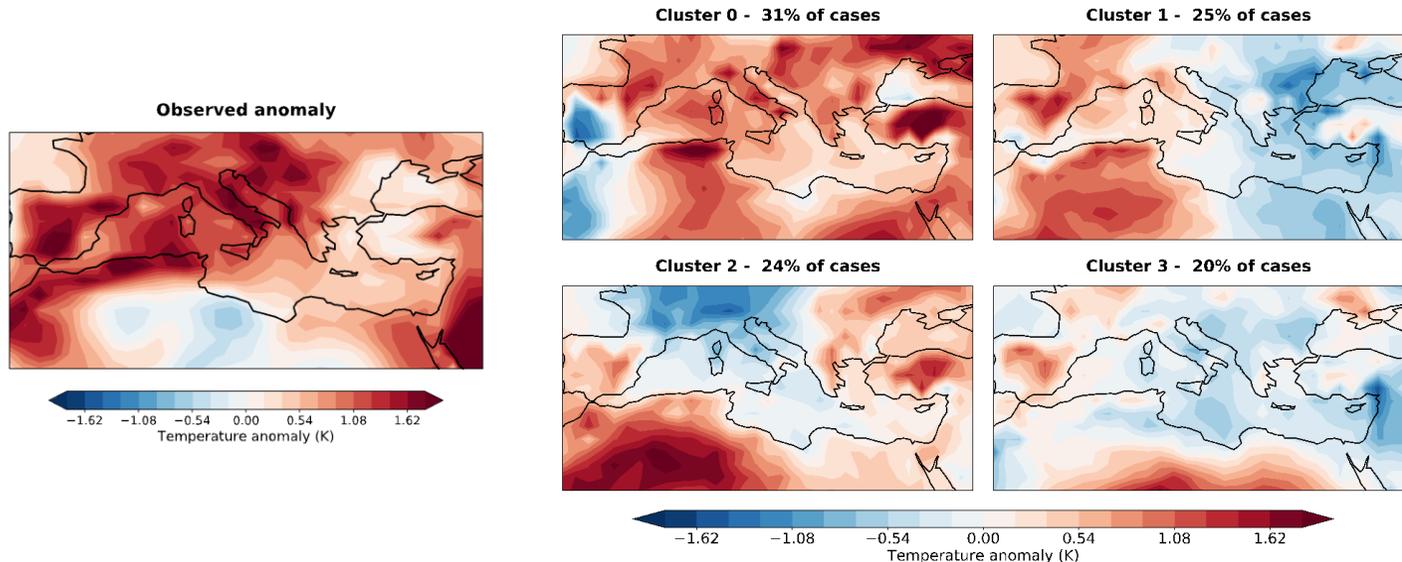
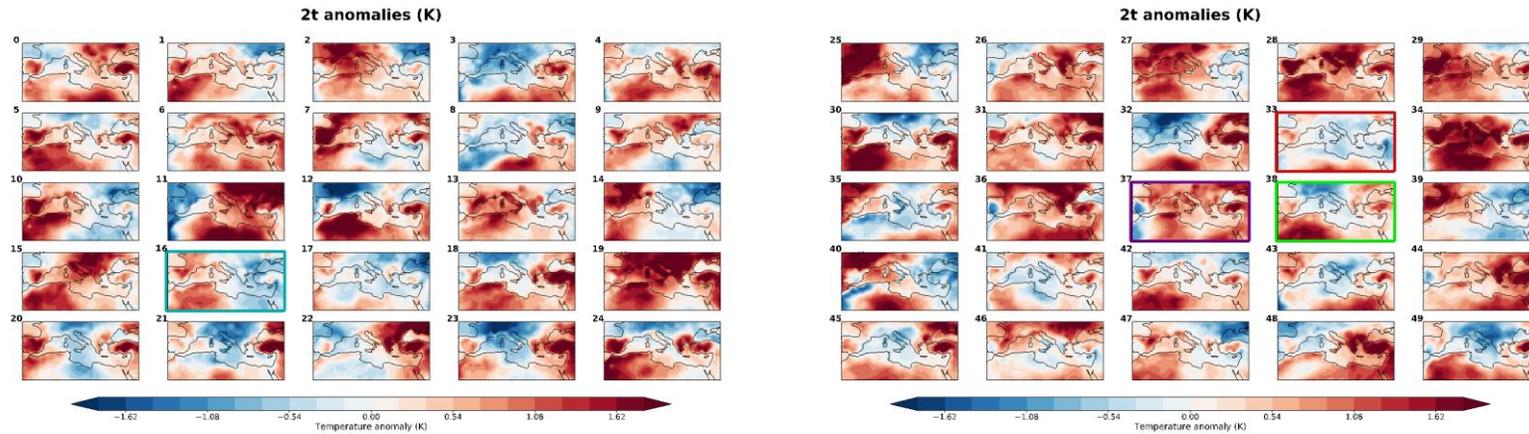


Ensemble mean



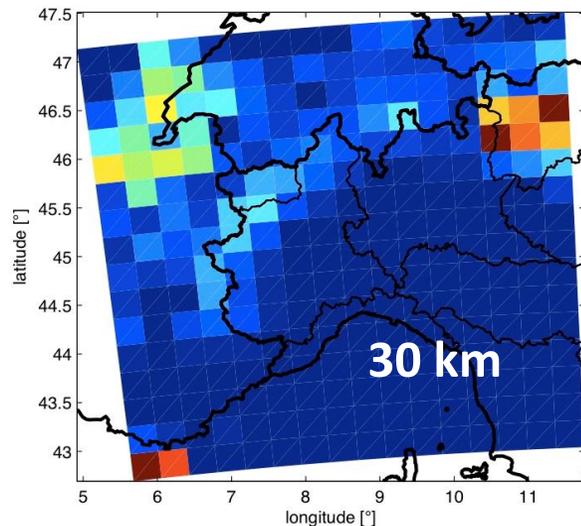
EnsClustering

The 50 ensemble forecasts anomalies



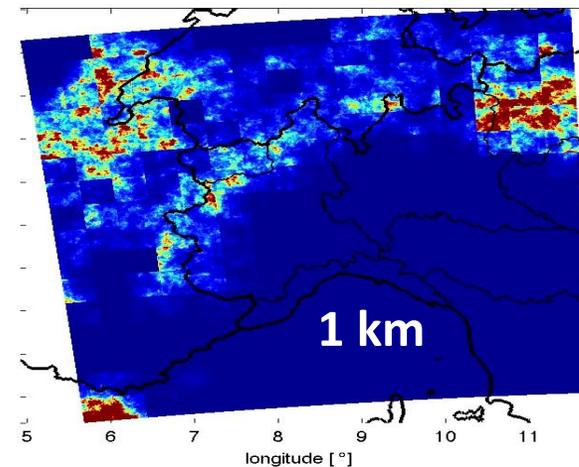
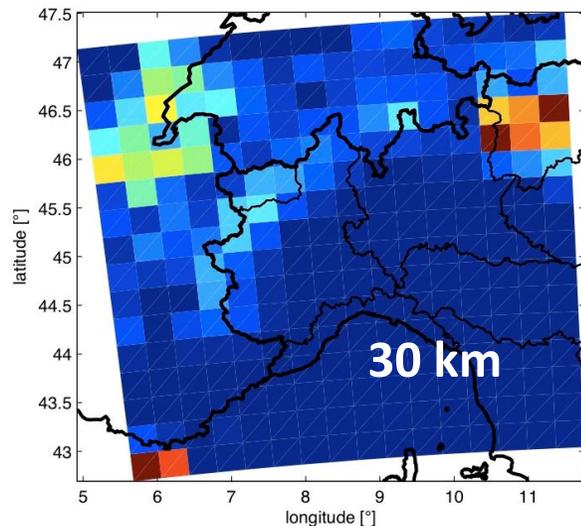
RainFARM

Impact studies at a local scale, particularly in the water sector, may require a representation the unresolved small-scale structure of precipitation, in particular subgrid variability and precipitation extremes. Examples include studies on the impact of extreme rainfall, flood impacts in small basins and ecosystem studies.



RainFARM

Impact studies at a local scale, particularly in the water sector, may require a representation the unresolved small-scale structure of precipitation, in particular subgrid variability and precipitation extremes. Examples include studies on the impact of extreme rainfall, flood impacts in small basins and ecosystem studies.

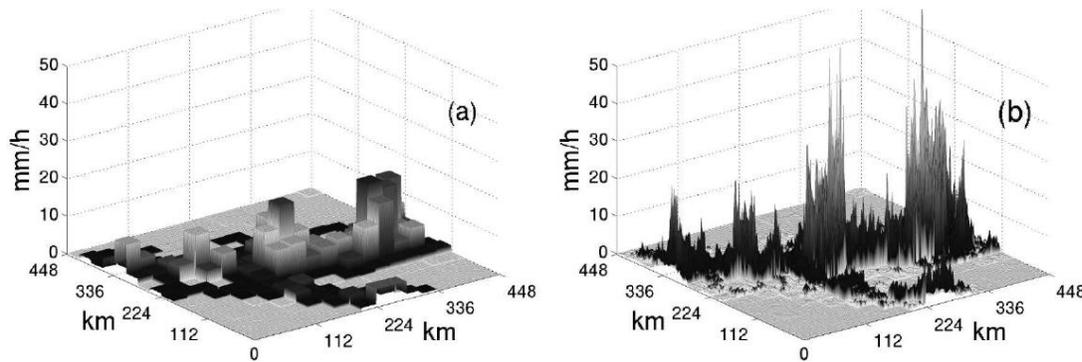


Example of a daily precipitation field
Downscaled with
RainFARM



RainFARM stochastic precipitation downscaling

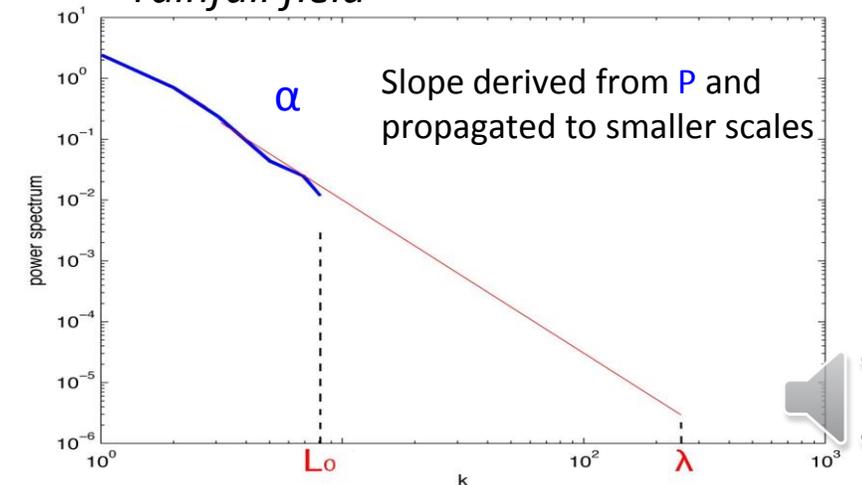
The **RainFARM** method belongs to the class of stochastic precipitation downscaling methods: it derives its only parameter from the original large-scale precipitation field and does not require further variables/information. It conserves precipitation at the scale of the original model.



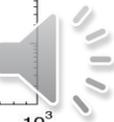
RainFARM uses simple statistical properties of large-scale field (shape of the power spectrum) and generates small-scale rainfall fields propagating this information to smaller scales in Fourier space.

The precipitation fields generated by stochastic procedures are consistent with the large-scale features imposed by the original precipitation fields, such as the total rainfall volume, and with the known statistical properties of precipitation at multiple scales.

Spatial Power spectrum of the rainfall field



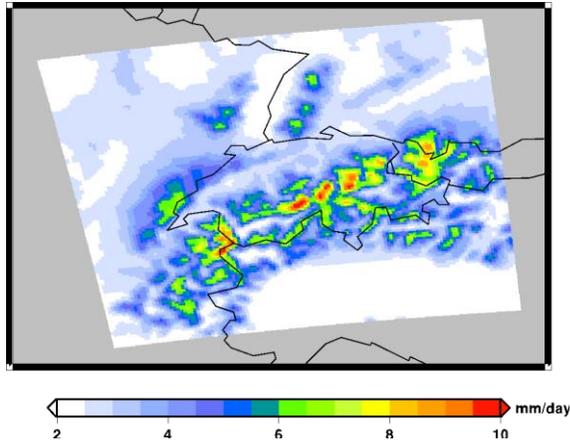
- N. Rebora, L. Ferraris, J. von Hardenberg, A. Provenzale, 2006; *RainFARM: Rainfall Downscaling by a Filtered Autoregressive Model*. *J. Hydrometeorology*, 7, 724-738
- D'Onofrio, D.; Palazzi, E.; von Hardenberg, J., Provenzale A., Calmanti S.; *Stochastic Rainfall Downscaling of Climate Models*. *J of Hydrometeorology* 15 (2), 830-843 (2014)



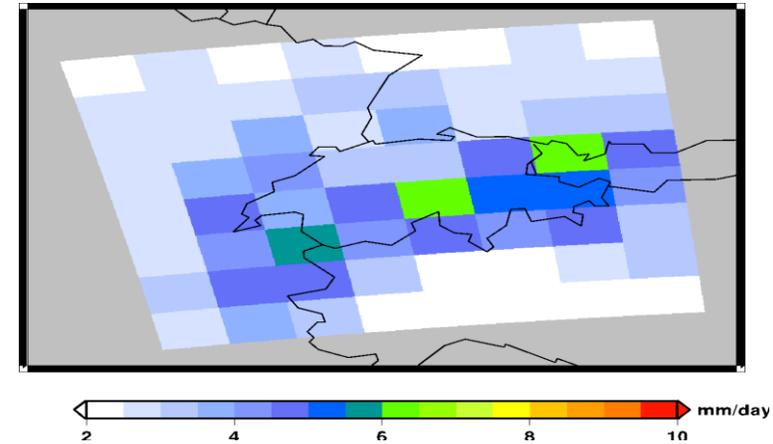
A “perfect model” experiment

WRF precip, 0.04° res. (1979-2008, forced by ERA Interim)

WRF at 0.04° (1979-2008)



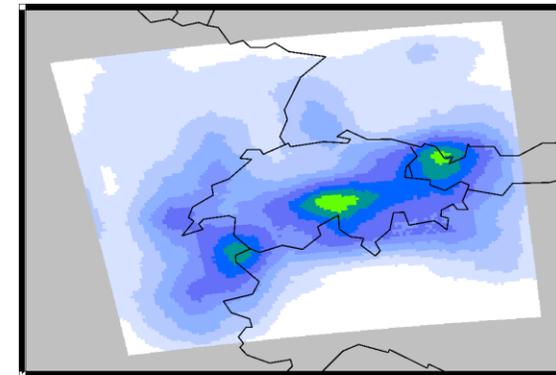
WRF upscaled to 0.64°



Downscaling
“standard”



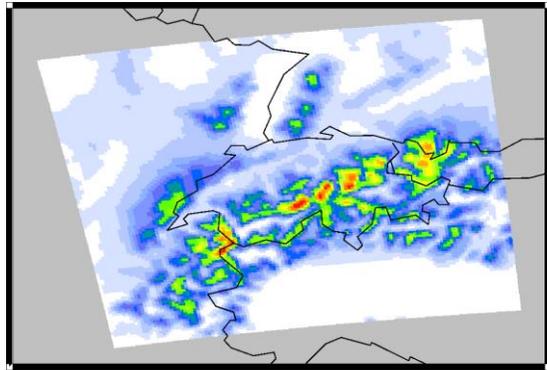
RainFARM
climatology



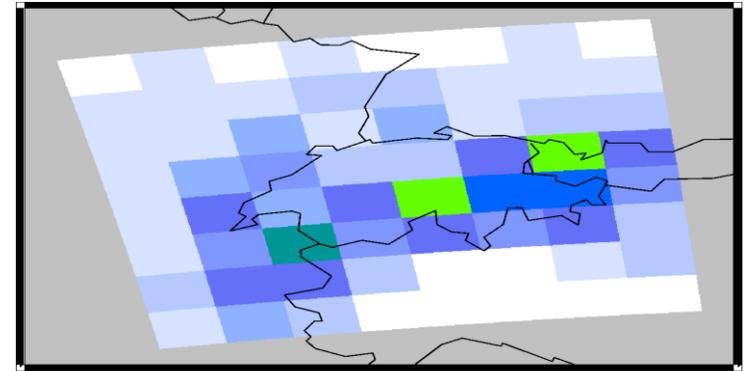
A “perfect model” experiment

WRF precip. (1979-2008, forced by ERA Interim), 0.04° res.

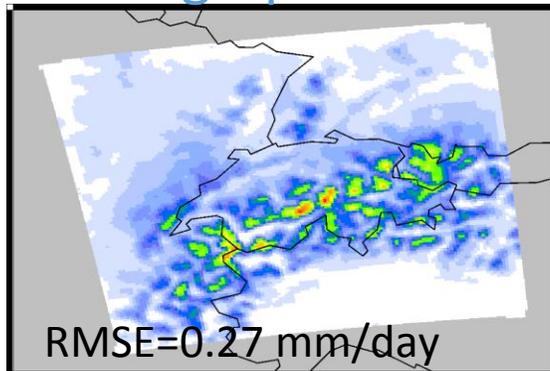
WRF at 0.04° (1979-2008)



WRF upscaled to 0.64°

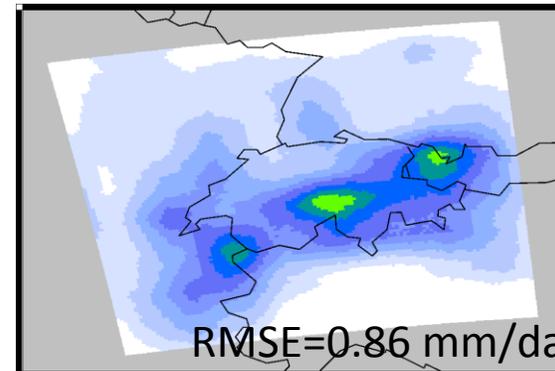


Downscaling
with orographic correction



RMSE=0.27 mm/day

Downscaling
“standard”



RMSE=0.86 mm/day

RainFARM
climatology



CST-Analogs

The analogs are days within the database which have a similar circulation to the day of interest. The temperature (or precipitation) of the analogs are then compared to the temperature (precipitation) of the day of interest

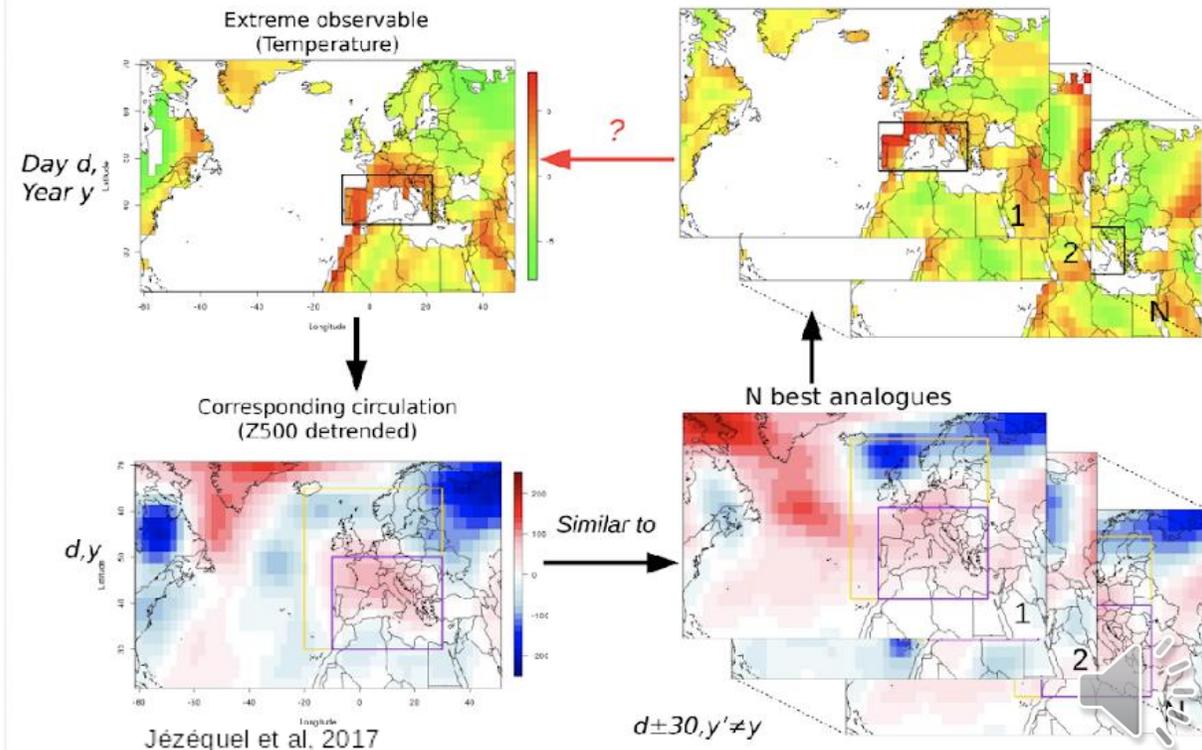
At least 2 variables: large scale (i.e. SLP) and local scale (i.e. temperature)

- No regional constraints
- Resolution depends on obs/reanalyses
- Method of Yiou et al, 2013:

Min euclidean distance
and max correlation

date	dist	corr
19860507	146	0.90
19920520	356	0.78
19980511	425	0.74
20010515	478	0.65
20130510	553	0.61
20090523	740	0.59

Yiou, P., T. Salameh, P. Drobinski, L. Menut, R. Vautard, and M. Vrac, 2013 : Ensemble reconstruction of the atmospheric column from surface pressure using analogues. *Clim. Dyn.*, 41, 1419-1437.Å



What else?

- Function to improve NAO forecasts
- Quantile mapping function
- Calibration function
- Ensemble combination for probabilistic forecasts



In future release:

- Weather Regimes
- Statistical Model for Orographic Precipitation
 - spatial distribution and intensity of precipitation over complex terrain
- ADAMONT
 - hourly time series of temperature, precipitation, ...
- And more...



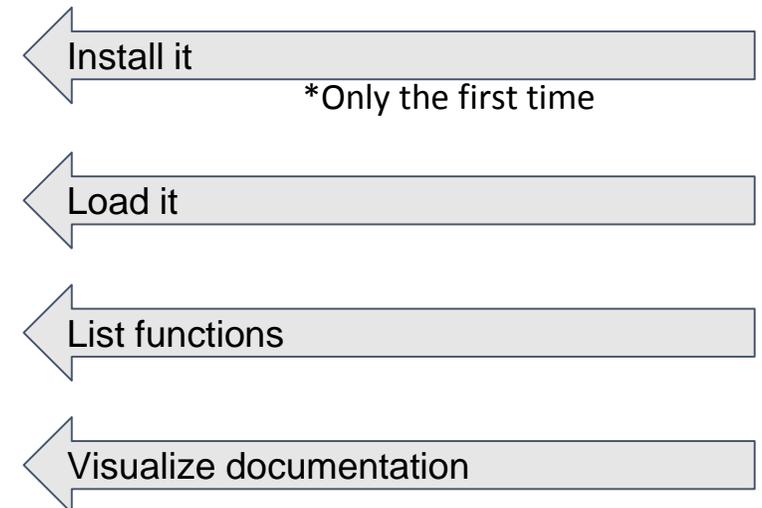
CSTools in R session



The CSTools R package v1.0.0 is published on **CRAN** (*The Comprehensive R Archive Network*).

You can use it in R

```
> install.packages("CSTools")  
  
> library(CSTools)  
Loading required package: maps  
  
> ls("package:CSTools")  
  
> ?PlotForecastPDF
```



Getting Started

CST_Load

read files and return an object 's2dv_cube'



Getting Started

CST_Load

read files and return an object 's2dv_cube'

s2dv_cube

it is an object containing the data and metadata:

- \$data an array with named dimensions
- \$lon a vector
- \$lat a vector
- \$Variable ... for instance 'tas', 'psl', ...
- \$Datasets
- \$Dates
- \$source_files
- ...



Getting Started

CST_Load

read files and return an object 's2dv_cube'

CST_Load is able to:

- load simultaneously experimental and observational data
- load monthly and daily data
- select a region
- select a period
- regrid data (it uses cdo internally)
- adjust the number of members loaded
- adjust the start dates, etc.

s2dv_cube

it is an object containing the data and metadata:

- \$data an array with named dimensions
- \$lon a vector
- \$lat a vector
- \$Variable ... for instance 'tas', 'psl', ...
- \$Datasets
- \$Dates
- \$source_files
- ...



Getting Started

CST_Load

read files and return an object 's2dv_cube'

... extra steps ...
CST_Anomaly
CST_Season

these intermediate steps will work on and
return a 's2dv_cube' object

CST_function1
CST_function2
CST_function3

Your CST function will work on a 's2dv_cube'
object and it can return a 's2dv_cube' too

s2dv_cube

it is an object containing the data and metadata:

- \$data an array with named dimensions
- \$lon a vector
- \$lat a vector
- \$Variable ... for instance 'tas', 'psl', ...
- \$Datasets
- \$Dates
- \$source_files
- ...



Vignette is an **instructive tutorial** demonstrating **practical uses** of the software **with discussion of the interpretation of the results**.

Using the `CST_Load` function from **CSTool package**, the data available in our data store can be loaded. The following lines show how this function can be used. Here, the data is loaded from a previous saved `.RData` file: Ask [nuria.perez at bsc.es](mailto:nuria.perez@bsc.es) for the data to run the recipe.

CST_Load
loads temperature data

CST_Load
loads precipitation data

```
require(zeallot)

glosea5 <- list(path = '/esnas/exp/glosea5/specs-seasonal_ilp1/$STORE_FREQ$/mean/$VAR_NAME$-allmemb/$VAR_NAME$/$S

c(exp_T, obs_T) %<-%
  CST_Load(var = temp, exp = list(glosea5),
    obs = obs, sdates = dateseq, leadtimemin = 2, leadtimemax = 4,
    latmin = 25, latmax = 75, lonmin = -20, lonmax = 70, output = 'lonlat',
    nprocs = 1, storefreq = "monthly", sampleperiod = 1, nmember = 9,
    method = "bilinear", grid = paste("r", grid, sep = ""))

c(exp_P, obs_P) %<-%
  CST_Load(var = precip, exp = list(glosea5),
    obs = obs, sdates = dateseq, leadtimemin = 2, leadtimemax = 4,
    latmin = 25, latmax = 75, lonmin = -20, lonmax = 70, output = 'lonlat',
    nprocs = 1, storefreq = "monthly", sampleperiod = 1, nmember = 9,
    method = "bilinear", grid = paste("r", grid, sep = ""))

#save(exp_T, obs_T, exp_P, obs_P, file = "./tas_prlr_toydata.RData")

# Or use the following line to load the file provided in .RData format:
load(file = "./tas_prlr_toydata.RData")
```

There should be four new elements loaded in the R working environment: `exp_T`, `obs_T`, `exp_P` and `obs_P`. The first two elements correspond to the experimental and observed data for temperature and the other are the equivalent for the precipitation data. It's possible to check that they are of class `sd2v_cube` by running:

```
class(exp_T)
class(obs_T)
class(exp_P)
class(obs_P)
```



...checks for the users to follow the steps

Vignette is an **instructive tutorial** demonstrating **practical uses** of the software **with discussion of the interpretation of the results**.

Latitudes and longitudes of the common grid can be saved:

```
Lat <- exp_T$lat
Lon <- exp_T$lon
```

CST_Anomaly
cross validation option

The next step is to compute the anomalies of the experimental and observational data using `CST_Anomaly` function, which could be applied over data from each variable, and in this case it's compute applying cross validation technique over individual members:

```
c(ano_exp_T, ano_obs_T) %<-% CST_Anomaly(exp = exp_T, obs = obs_T, cross = TRUE, memb = TRUE)
c(ano_exp_P, ano_obs_P) %<-% CST_Anomaly(exp = exp_P, obs = obs_P, cross = TRUE, memb = TRUE)
```

The original dimensions are preserved and the anomalies are stored in the `data` element of the correspondent object:

```
> str(ano_exp_T$data)
num [1, 1:9, 1:21, 1:3, 1:35, 1:64] -1.647 1.575 2.77 0.048 -1.886 ...
- attr(*, "dimensions")= chr [1:6] "dataset" "member" "sdate" "ftime" ...
> str(ano_obs_T$data)
num [1, 1, 1:21, 1:3, 1:35, 1:64] 0.0235 1.546 1.3885 -0.344 -5.972 ...
- attr(*, "dimensions")= chr [1:6] "dataset" "member" "sdate" "ftime" ...
```

checks to make sure
everything is ok

Two lists containing the experiment `ano_exp`, and the observation, `ano_obs`, lists should be put together to serve as input of the function to compute multivariate RMSEs.

Furthermore, some weights can be applied to the difference variables based on their relative importance (if no weights are given, a value of 1 is automatically assigned to each variable). For this example, we'll give a weight of 2 to the temperature dataset and a weight of 1 to the precipitation dataset:

```
ano_exp <- list(ano_exp_T, ano_exp_P)
ano_obs <- list(ano_obs_T, ano_obs_P)
weight <- c(2, 1)
```

Setting inputs
for next step



Vignette is an **instructive tutorial** demonstrating **practical uses** of the software **with discussion** of the **interpretation** of the **results**.

CST_MultivarRMSE
cross validation option

2.- Computing and plotting multivariate RMSEs

The multivariate RMSE gives an indication of the forecast performance (RMSE) for multiple variables simultaneously. Variables can be weighted based on their relative importance. It is obtained by running the `CST_MultivarRMSE` function:

```
mvrmse <- CST_MultivarRMSE(exp = ano_exp, obs = ano_obs, weight)
```

The function `CST_MultivarRMSE` returns the multivariate RMSE value for 2 or more variables. The output is a `CSTool` object containing the RMSE values in the `data` element and other relevant information:

```
> class(mvrmse)
> str(mvrmse$data)
num [1, 1, 1, 1:35, 1:64] 0.764 0.8 0.67 0.662 0.615 ...
> str(mvrmse$Variable)
Named chr [1:2] "tas" "prlr"
- attr(*, "names")= chr [1:2]
```

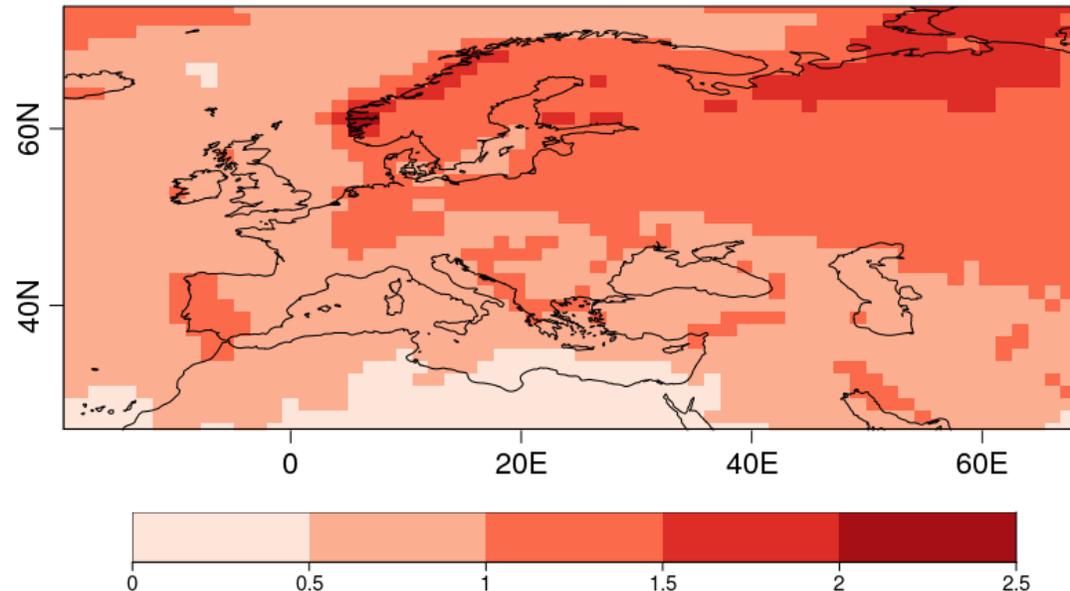
checks for the users to
follow the steps

The following lines plot the multivariate

```
PlotEquiMap(mvrmse$data, lon =
  toptitle = "Multiv
  bar_limits = c(0,2
  fileout = "./Multi
```

Visualization

Multivariate RMSE tas, prlr 1992 - 2012



Downscaling seasonal precipitation forecasts with RainFARM

A R vignette is available:

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

Downscaling seasonal precipitation forecasts with RainFARM

Preliminary setup

In order to run the examples in this vignette, the *CSTools* package and some other support R packages need to be loaded by running:

```
install.packages('CSTools')
library(CSTools)
```

We use test data provided by *CSTools* to load a seasonal precipitation forecast:

```
exp <- lonlat_prec
```

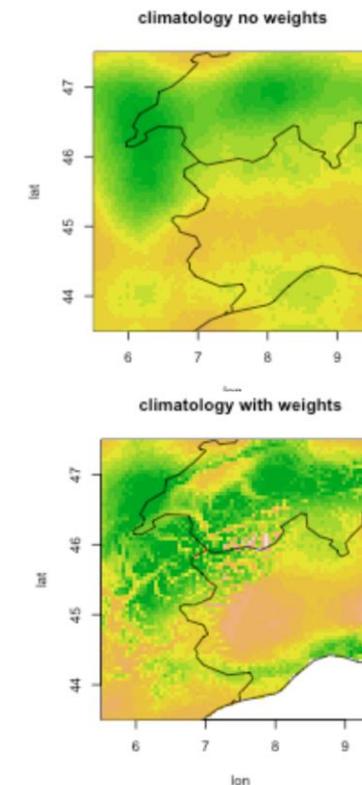
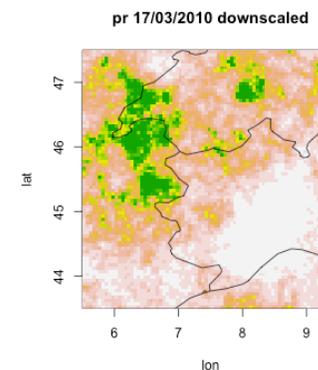
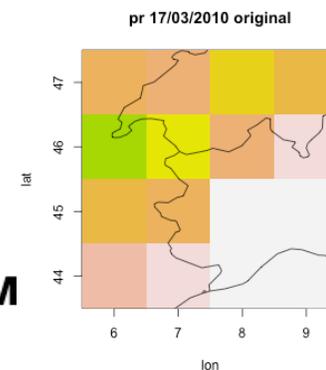
This gives us a *CSTools* object `exp`, containing an element `exp$data` with dimensions:

```
dim(exp$data)
#dataset member  sdate  ftime  lat  lon
#      1      6      3    31    4   4
```

```
ww <- CST_RFWeights("./worldclim.nc", nf = 20, lon = exp$lon, lat = exp$lat)
```

The result is a two-dimensional weights matrix with the same `lon` and `lat` dimensions as requested. The weights (varying around an average value of 1) encode how to distribute differently precipitation in each stochastic realization of RainFARM. We call again `CST_RainFARM()`, this time passing climatological weights:

```
exp_down_weights <- CST_RainFARM(exp, nf = 20, kmin = 1, nens = 3,
                                weights = ww, time_dim = c("member", "ftime"))
```





CSTools on CRAN

Developing methodologies to extract usable information from predictions. We will produce tools for prediction verification, calibration, downscaling, ensemble member combination and selection that will be publicly released via a toolbox and shared among partners and users.

Package ‘CSTools’

June 19, 2019

Title Assessing Skill of Climate Forecasts on Seasonal-to-Decadal Timescales

Version 1.0.1

Description Exploits dynamical seasonal forecasts in order to provide information relevant to stakeholders at the seasonal timescale. The package contains process-based methods for forecast calibration, bias correction, statistical and stochastic downscaling, optimal forecast combination and multivariate verification, as well as basic and advanced tools to obtain tailored products. This package was developed in the context of the ERA4CS project MEDSCOPE.

Doblas-Reyes et al. (2005) <doi:10.1111/j.1600-0870.2005.00104.x>.

Mishra et al. (2018) <doi:10.1007/s00382-018-4404-z>.

Terzago et al. (2018) <doi:10.5194/nhess-18-2825-2018>.

Torralba et al. (2017) <doi:10.1175/JAMC-D-16-0204.1>.

D’Onofrio et al. (2014) <doi:10.1175/JHM-D-13-096.1>.

Depends R (>= 3.2.0), maps

Imports s2dverification, rainfarmr, multiApply, ncd4, plyr, abind, data.table, reshape2, ggplot2, graphics, grDevices, stats, utils

Suggests zeallot, testthat, knitr, rmarkdown

VignetteBuilder knitr

License Apache License 2.0

Encoding UTF-8

LazyData true

RoxygenNote 5.0.0

NeedsCompilation no

R topics documented:

areave_data	2
CST_Anomaly	3
CST_BiasCorrection	5
CST_Calibration	6
CST_Load	7
CST_MultiMetric	8
CST_MultivarRMSE	10
CST_RainFARM	11
CST_RFSlope	13
CST_RFWeights	14
lonlat_data	16
lonlat_prec	17
PlotCombinedMap	18
PlotForecastPDF	20
PlotMostLikelyQuantileMap	21
RainFARM	24
RFSlope	26

<https://cran.r-project.org/web/packages/CSTools/index.html>





CSTools on CRAN

Developing methodologies to extract usable information from predictions. We will produce tools for prediction verification, calibration, downscaling, ensemble member combination and selection that will be publicly released via a toolbox and shared among partners and users.

Package ‘CSTools’

June 19, 2019

Title Assessing Skill of Climate Forecasts on Seasonal-to-Decadal Timescales

Version 1.0.1

Description Exploits dynamical seasonal forecasts in order to provide information relevant to stakeholders at the seasonal timescale. The package contains process-based methods for forecast calibration, bias correction, statistical and stochastic downscaling, optimal forecast combination and multivariate verification, as well as basic and advanced tools to obtain tailored products. This package was developed in the context of the ERA4CS project MEDSCOPE.

Doblas-Reyes et al. (2005) <doi:10.1111/j.1600-0870.2005.00104.x>.
Mishra et al. (2018) <doi:10.1007/s00382-018-4404-z>.
Terzago et al. (2018) <doi:10.5194/nhess-18-2825-2018>.
Torralba et al. (2017) <doi:10.1175/JAMC-D-16-0204.1>.
D’Onofrio et al. (2014) <doi:10.1175/JHM-D-13-096.1>.

Depends R (>= 3.2.0), maps

Imports s2dverification, rainfarmr, multiApply, ncd4, plyr, abind, data.table, reshape2, ggplot2, graphics, grDevices, stats, utils

Suggests zeallot, testthat, knitr, rmarkdown

VignetteBuilder knitr

License Apache License 2.0

Encoding UTF-8

LazyData true

RoxygenNote 5.0.0

NeedsCompilation no

R topics documented:

areave_data	2
CST_Anomaly	3
CST_BiasCorrection	5
CST_Calibration	6
CST_Load	7
CST_MultiMetric	8
CST_MultivarRMSE	10
CST_RainFARM	11
CST_RFSlope	13
CST_RFWeights	14
lonlat_data	16
lonlat_prec	17
PlotCombinedMap	18
PlotForecastPDF	20
PlotMostLikelyQuantileMap	21
RainFARM	24
RFSlope	26

<https://cran.r-project.org/web/packages/CSTools/index.html>

Thank you for your attention!

