Barcelona Supercomputing Center
Centro Nacional de Supercomputación

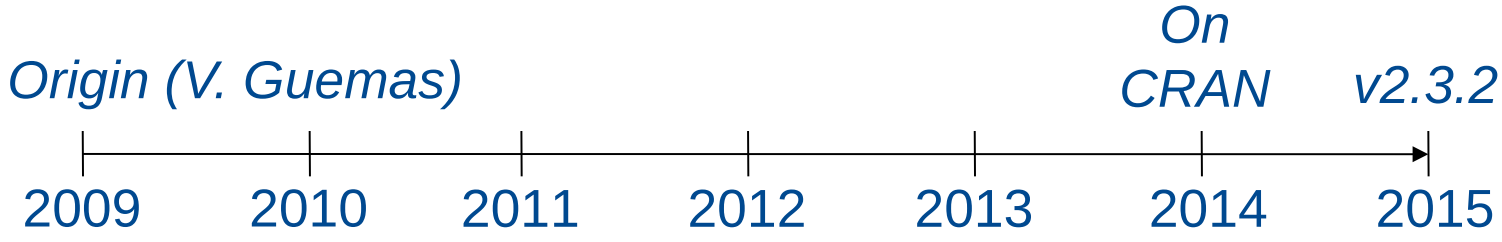
s2dverification

Seasonal to decadal forecast verification in R
Overview

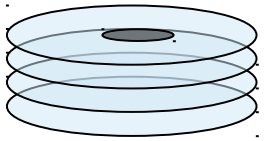
Nicolau Manubens

- Outline
  - Introduction
  - Module diagram
  - Example of use
  - BigData issues
  - BigData approaches

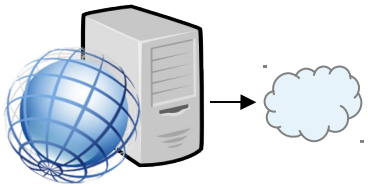
- Introduction
  - **Forecast verification**: evaluating model performance by comparing its output with observational data.
  - **s2dverification** (seasonal to decadal verification) is an R package that gathers various forecast verification tools coded by scientists, aiming to score models which run forecasts from a few seasons to few decades long.



- Module diagram



LOCAL STORAGE



ESGF NODE or OPeNDAP SERVER

s2dverification package

DATA RETRIEVAL

Fetches experimental and observational data sets stored in multiple conventions:

- NetCDF, NetCDF4
- File per member, file per starting date, ...
- Supports specific folder and file naming conventions

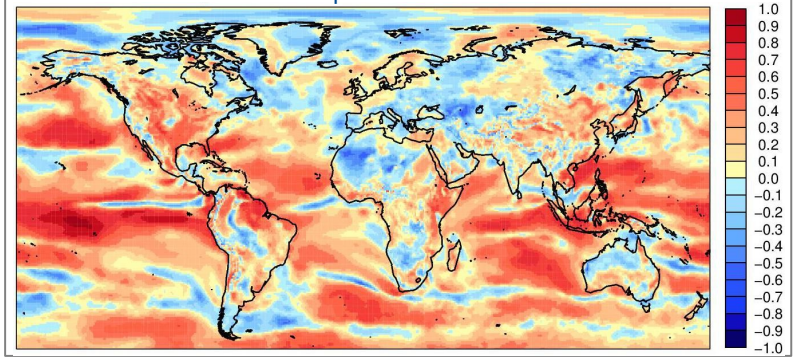
BASIC STATISTICS

SCORES

Correlation, ACC, RMSSS, CRPS, ...

PLOT

Anomaly Correlation Coefficient. 10M Wind Speed ECMWF S4 1 month lead with start dates once a year on first of November and Era-Interim in DJF from 1981 to 2011. Simple bias correction with cross-validation.



Most time consuming: SCORES and DATA RETRIEVAL

- Example of use

```
data <- Load('tas',
             exp = c('EnsEcmwfSeas', 'EnsMetfrSeas'),
             obs = c('ERAint'), nmember = 9,
             sdates = paste(1979:2005, '0501', sep = “”),
             leadtimemin = 2, leadtimemax = 4,
             latmin = 35, latmax = 75,
           lonmin = -25, lonmax = 70,
           storefreq = 'monthly', output = 'lonlat')
```

Pick monthly 2-meter air temperature in J-J-A over Europe from ECMWF and Meteofrance ensemble experiments and from ERA-interim observation, from may 1st starting dates from 1979 to 2005.

- Example of use

```
meanModEnsSeas <- Mean1Dim(Mean1Dim(data$mod, 4), 2)
meanModObsSeas <- Mean1Dim(Mean1Dim(data$obs, 4), 2)
```

Average model and observational data along J-J-A (4th dimension) for all starting dates, and then average all models and observations across ensemble members.

```
corr <- Corr(meanModEnsSeas, meanModObsSeas)
```

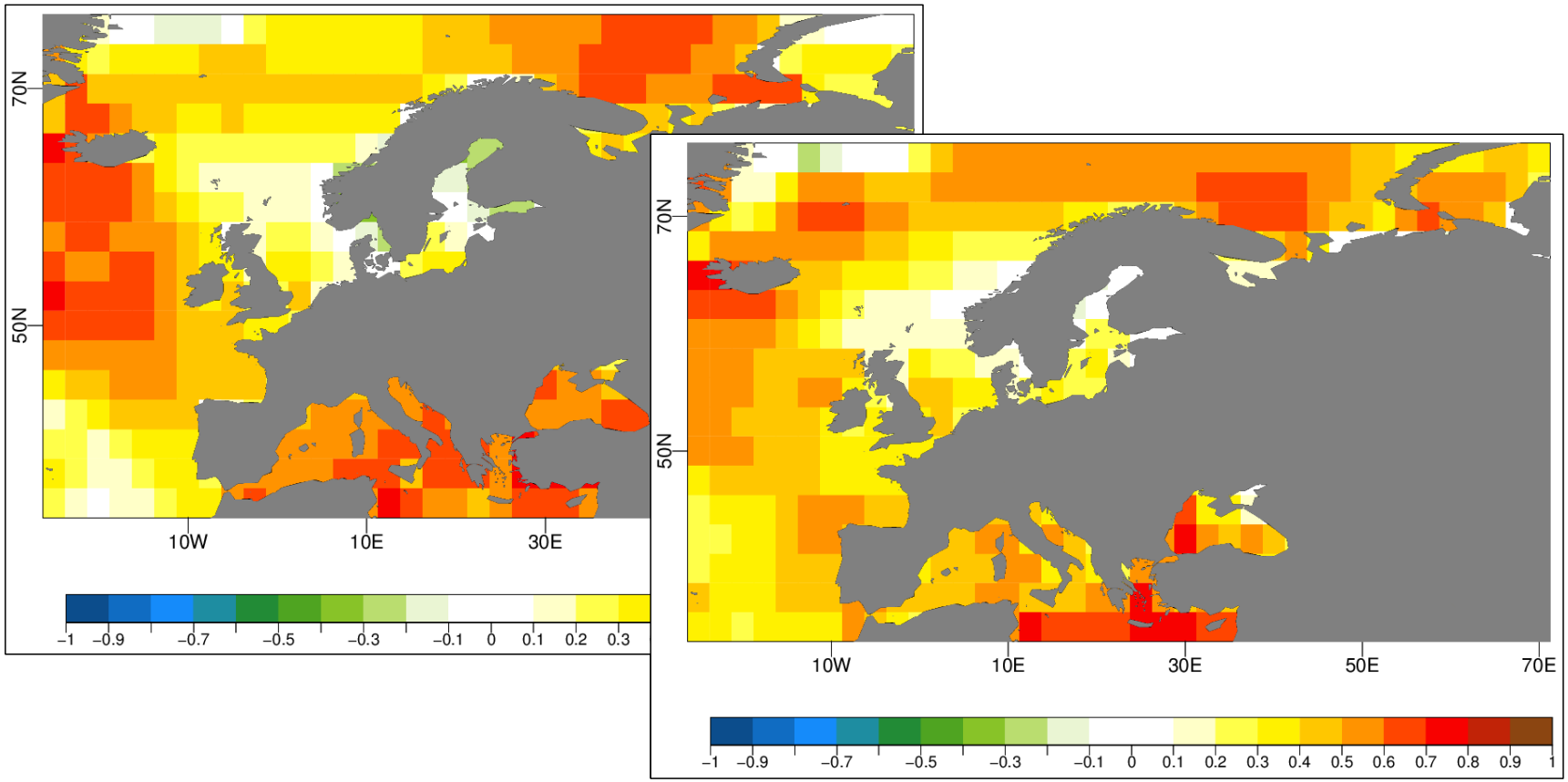
Calculate the time correlation between each grid point in the model data and in the observational data.

```
intervals <- seq(-1, 1, length.out = 21)
PlotEquiMap(corr[1, 1, , ], data$lon, data$lat, brks = intervals)
PlotEquiMap(corr[2, 1, , ], data$lon, data$lat, brks = intervals)
```

Plot a map of correlations in 1979-2005 between each experiment's and ERA-interim's data averaged in JJA and across ensemble members.

- # Example of use

ECMWF ensemble experiment and ERA-interim's 2-m temperature averaged along JJA and across members correlation in 1979-2005.



Meteofrance ensemble experiment and ERA-interim's 2-m temperature averaged along JJA and across members correlation in 1979-2005.

- BigData issues
  - **Computing time can raise** to several hours in score computation or data retrieval.
  - Involved data occupies in some cases far more than the **available main memory** and hangs the machine.
    - Example case:

      2 x 9 x 27 x 3 x 17 x 39 x 8 bytes → 7.7 Mbyte
    - Usual case:

      5 x 9 x 27 x 60 x 73 x 144 x 8 bytes → 6.1 Gbyte
    - Big case:

      1 x 50 x 36 x 120 x 144 x 288 x 8 bytes → 71.6 Gbyte

(*n. of datasets* x *n. of members* x *starting dates* x lead-times x *latitudes* x *longitudes*)

- BigData approaches
  - In progress
    - Avoid R memory duplications as much as possible.
    - Wrappers of Fortran functions are always faster.
    - Exploit multi-core.
  - Future work
    - Store active data in disk instead of main memory.
    - Parallelize verification tools to use on cluster.
    - Reduce R's default representation precision.

www.bsc.es

Thank you!

For further information please contact
nicolau.manubens@bsc.es