Barcelona Supercomputing Center
Centro Nacional de Supercomputación

R user meeting

An-Chi Ho and Eva Rifà

contributor: Núria Pérez-Zanón

02/03/2023

Agenda

1. Ice-breaker: Fun package “cowsay”
2. News
   - General R
   - s2dv
   - ClimProjDiags
   - CSTools
   - CSIndicators
3. User presentation: Use case of PlotWeeklyClim() [Núria]
4. Q&A

Ice-breaker

What does the cowsay

A fun pacakge “cowsay” that allows printing of character strings as messages/warnings/etc. with **ASCII animals**, including cats, cows, frogs, chickens, ghosts, and more.

```
> sort(names(animals))
> say("time", by = "trilobite")
> say(what = "boo!", by = "ghost", what_color = "cyan", by_color =
"brown")
> say("catfact", "cat")
> say("fortune", by = "monkey")
> say("It’s Thanksgiving!!\nRun Run!!", by = "turkey", type = "warning")
```

General R

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

R-community activity

Upcoming R conferences:

- **useR!** 2023/2024 hybrid conference (no news, past years was during summer)
- **Barcelona R community**: Summer 2023 (tentative)
- **posit::conf(2023)** in Chicago from September 17 to September 20, 2023.
- **Shiny in Production** on 12th - 13th October 2023

and

- **R Barcelona community meetup** at BSC April 20th at 18.30 h !

@RLadiesBCN
@BarcelonaR_user
https://www.meetup.com/rladies-barcelona/

license change

We need to change all the licenses from Apache 2.0 to **GPL-3**.

We will do it in the next release of each package.

s2dv

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

cross validation development

New parameter in RPS(), RPSS(), ROCSS():

- “**cross.val**”: TRUE/FALSE to decide if use cross-validation or not

status: RPS and RPSS in master; ROCSS in branch *develop-ROCSS*

startR

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Bugfix of time attribute retrieval

If the units of time is “month since” & calendar type is “standard”, (e.g., `/esarchive/obs/ukmo/hadisst_v1.1/monthly_mean/tos/`), the calculation of time was wrong for certain case. Data itself is correct.

```
path <- '/esarchive/obs/ukmo/hadisst_v1.1/monthly_mean/$var$/$var$_$date$.nc'
date_vec <- c("196011", "196012", "196101", "196102")
data_obs <- Start(dat = path, var = 'tos',
               date = date_vec,
               latitude = indices(100), longitude = indices(100),
               synonims = list(longitude = c('lon', 'longitude'),
                               latitude = c('lat', 'latitude')),
               return_vars = list(time = c('date'),
                                  latitude = NULL, longitude = NULL),
               retrieve = TRUE)
attr(data_obs, 'Variables')$common$time
[1] "1960-11-01 UTC" "1961-01-31 UTC" "1961-02-01 UTC" "1961-02-01 UTC"
```

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

status: fixed in master

ClimProjDiags

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

New release 0.3.0

- On CRAN: https://cran.rstudio.com/web/packages/ClimProjDiags/index.html
- Check the NEWS: https://cran.rstudio.com/web/packages/ClimProjDiags/news/news.html
- Installed in workstation & Nord3v2, R/4.1.2-xxx.

WeightedMean

New parameters:

- `na.rm = TRUE` and `ncores = NULL`

Changes:

- `londim` and `latdim` indicate the dimension names (not the position)
- Added `multiApply::Apply`
- Improved `.WeightedMean`

status: In master and on CRAN

Check issue: https://earth.bsc.es/gitlab/es/ClimProjDiags/-/issues/13

SelBox

Changes:

- `londim` and `latdim` indicate the dimension names (not the position)
- Added initial checks
- Created unit test: test-Selbox

status: In master and CRAN

Check issue: https://earth.bsc.es/gitlab/es/ClimProjDiags/-/issues/14

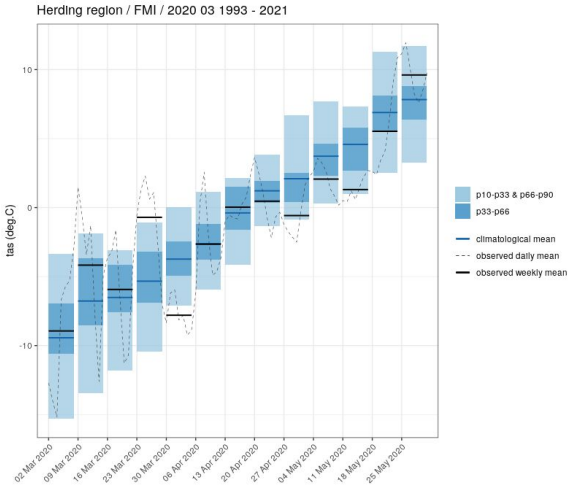
CSTools

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

PlotWeeklyClim

Created from Nuria's code. It compares the weekly climatology in a reference period to the observed conditions during the target period analyzed in the case study using ggplot

- **The necessary arguments:**
  - `data`: multidimensional array with at least 'sdate' and 'time' dimensions
  - `first_date`: first date of the target period of the observations
  - `ref_period_ini/end`: initial and end years for the reference period
- **The other arguments:**
  - `time_dim = 'time'`
  - `sdate_dim = 'sdate'`
  - Usual plot params: title, palette, fileout, width, …

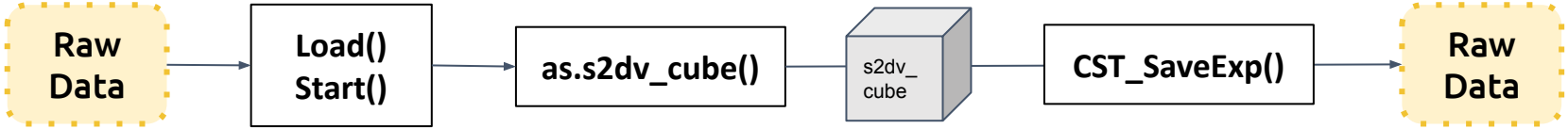


status: In master

Check issue:
https://earth.bsc.es/gitlab/external/cstools/-/issues/101

SaveExp

Writes a multidimensional data array to NetCDF.




```
CST_SaveExp(data, destination = "./", sdate_dim = 'sdate',
            ftime_dim = 'time', dat_dim = 'dataset',
            var_dim = 'var', memb_dim = 'member',
            single_file = TRUE, extra_string = NULL)

SaveExp(data, destination = "./", Dates = NULL, coords = NULL,
            varname = NULL, metadata = NULL, Datasets = NULL,
            startdates = NULL, dat_dim = 'dataset', sdate_dim = 'sdate',
            ftime_dim = 'time', var_dim = 'var', memb_dim = 'member',
            single_file = TRUE, extra_string = NULL)
```

status: In develop-new_s2dv_cube

Check issue: https://earth.bsc.es/gitlab/external/cstools/-/issues/110

SaveExp

Changes: **(1)** Works for multiple variables and datasets. **(2)** It can accept any dimension.

- `if (single_file = TRUE)` The multidimensional array and the metadata are saved in 1 file.
- `if (single_file = FALSE)` Saved in multiple files with Apply through dat_dim and var_dim.

Example:

1

```
repos <- '/esarchive/exp/ecmwf/system5_m1
/monthly_mean/$var$_f6h/$var$_$sdate$.nc'
data <- Start(dat = repos,
              var = c('sic', 'tas'),
              sdate = c('20170101', '20180101'),
              ensemble = indices(1:2), time = 'all',
              latitude = values(list(0, 10)),
              latitude_reorder = Sort(),
              [...])
```

2

```
data1 <- as.s2dv_cube(data)
```

3

```
CST_SaveExp(data1, memb_dim = 'ensemble',
            dat_dim = 'dat', single_file = FALSE)
```

```
.
├── sic
│   ├── sic_20170101.nc
│   └── sic_20180101.nc
└── tas
    ├── tas_20170101.nc
    └── tas_20180101.nc
```

```
netcdf tas_20170101 {
dimensions:
        longitude = 37 ;
        latitude = 36 ;
        ensemble = 2 ;
        time = UNLIMITED ; // (7 currently)
variables:
        double longitude(longitude) ;
                longitude:units = "degrees_east" ;
                longitude:long_name = "longitude" ;
                longitude:ndims = "1" ;
                longitude:standard_name = "longitude" ;
                longitude:axis = "X" ;
        double latitude(latitude) ;
                latitude:units = "degrees_north" ;
                latitude:long_name = "latitude" ;
                latitude:ndims = "1" ;
                latitude:standard_name = "latitude" ;
                latitude:axis = "Y" ;
```

CSIndicators

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Changes in CST functions

Current code

```
# when subsetting is needed, dimensions are also needed:
if (!is.null(start) && !is.null(end)) {
  if (is.null(dim(data$Dates$start))) {
    if (length(data$Dates$start) != dim(data$data)[time_dim]) {
      if (length(data$Dates$start) ==
          prod(dim(data$data)[time_dim] * dim(data$data)[sdate_dim])) {
        dim(data$Dates$start) <- c(dim(data$data)[time_dim], dim(data$data)[sdate_dim])
      } else {
        warning("Dimensions in 'data' element 'Dates$start' are missed and all data would be used.")
      }
    }
  }
}
```

New code

```
# Dates subset
if (!is.null(start) && !is.null(end)) {
  if (is.null(dim(data$attrs$Dates))) {
    warning("Dimensions in 'data' element 'attrs$Dates' are missed and all data would be used.")
    start <- NULL
    end <- NULL
  }
}
```

status: In develop-new_s2dv_cube

Check issue: https://earth.bsc.es/gitlab/es/csindicators/-/issues/25

shapefile

Why this development

- Many people and projects have their own scripts to use shapefile (e.g., S2S4E, MEDGOLD), but there is no a common function in our department.
- NUTS shapefile: Created by eurostats, GISCO. The NUTS are a hierarchical system divided into 3 levels. NUTS 1: major socio-economic regions, NUTS 2: basic regions for the application of regional policies, NUTS 3: small regions for specific diagnoses.
- Methods to read shapefile in R:
  - raster
  - rgdal/rgeos/sp
  - maptools/sp
  - terra/stars/sf (newer packages)
  - etc.

Function details

- Steps:
  (1) Take one .shp file and one .nc file as grid reference
  (2) Specify the desired shape regions by IDs or names
  (3) Compare the shape and the reference grid to find the corresponding grid points for each region
  (4) Create a mask array with the value 0 (no shp), 1 (1st region), 2 (2nd region), etc.
- Current used packages: rgdal, rgeos, sp
- Use easyNCDF package to load .nc file
- Only consider NUTS and ADM for now
- Problem: The required modules (GEOS, GDAL, PROJ) are complicated and have conflicts often. The modules don't work well on workstation.

issue: https://earth.bsc.es/gitlab/es/s2dv/-/issues/75#note_200522

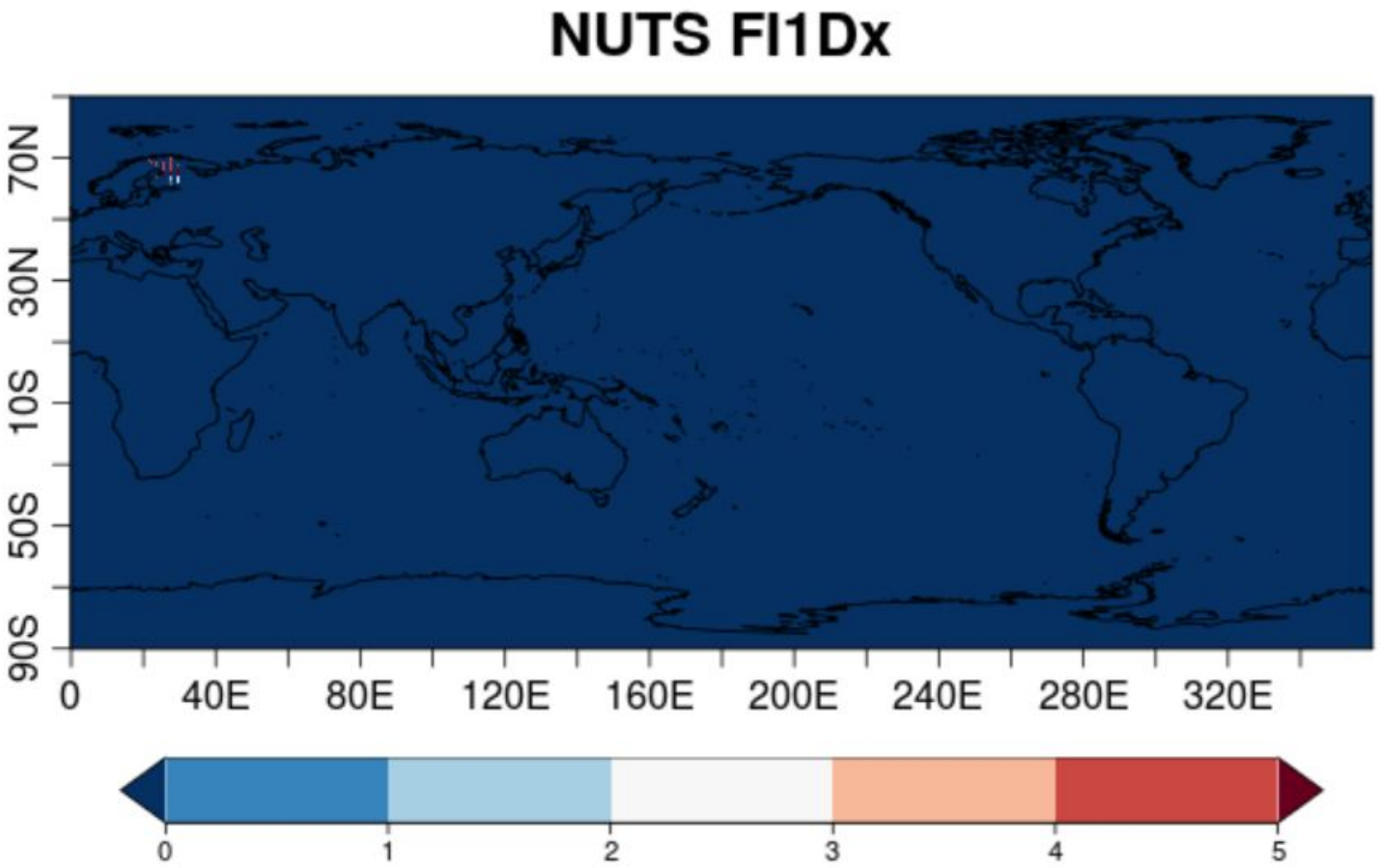
Minimum example

```
library(easyNCDF)
library(sp)
library(rgeos)
library(rgdal)
### source the function
source("/esarchive/scratch/aho/tmp/shp_mask.R")
###

shp.file <- '/esarchive/shapefiles/NUTS3/NUTS_RG_60M_2021_4326.shp/NUTS_RG_60M_2021_4326.shp'
# 0.1x0.1 grid
ref.file <- '/esarchive/recon/ecmwf/era5land/monthly_mean/tas_f1h/tas_201006.nc'

NUTS.id <- paste0("FI1D", c(1:3, 5, 7:9))
NUTS.name <- list(FI = c('Lappi', 'Kainuu'), SI = c('Pomurska', 'Podravska'))

mask <- shp_mask(shp.file, ref.file, reg.ids = NUTS.id)
```

Minimum example

```
> str(mask)
 num [1:3600, 1:1801] 0 0 0 0 0 0 0 0 0 0 ...
 - attr(*, "lon")= num [1:3600] 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 ...
 - attr(*, "lat")= num [1:1801] 90 89.9 89.8 89.7 89.6 89.5 89.4 89.3 89.2 89.1 ...
 - attr(*, "index")=List of 7
  ..$ 1: chr "FI1D1"
  ..$ 2: chr "FI1D2"
  ..$ 3: chr "FI1D3"
  ..$ 4: chr "FI1D5"
  ..$ 5: chr "FI1D7"
  ..$ 6: chr "FI1D8"
  ..$ 7: chr "FI1D9"
 - attr(*, "shapefile")=List of 7
  ..$ data       :'data.frame': 7 obs. of  9 variables:
  .. ..$ NUTS_ID   : chr [1:7] "FI1D1" "FI1D2" "FI1D3" "FI1D5" ...
  .. ..$ LEVL_CODE : int [1:7] 3 3 3 3 3 3 3
  .. ..$ CNTR_CODE : chr [1:7] "FI" "FI" "FI" "FI" ...
  .. ..$ NAME_LATN : chr [1:7] "Etel\303\244-Savo" "Pohjois-Savo" "Pohjois-Karjala" "Keski-Pohjanmaa" ...
  .. ..$ NUTS_NAME : chr [1:7] "Etel\303\244-Savo" "Pohjois-Savo" "Pohjois-Karjala" "Keski-Pohjanmaa" ...
  .. ..$ MOUNT_TYPE: int [1:7] 4 4 4 4 4 4 4
  .. ..$ URBN_TYPE : int [1:7] 3 3 3 3 3 3 3
  .. ..$ COAST_TYPE: int [1:7] 3 3 3 1 1 3 1
  .. ..$ FID       : chr [1:7] "FI1D1" "FI1D2" "FI1D3" "FI1D5" ...
  ..$ polygons   :List of 7
  ..
```

Minimum example



Future work

- Replace the packages that read shapefile because the critical ones (rgeos, rgdal) are going to retire by the end of this year (reference)
- Find a common place to save the shapefile, for common use, project, or operational.

```
> ls /esarchive/shapefiles/
```

  - `ATLAS  IPCC-WGI-reference-regions-v4_shapefile.zip  LAU  monitor  NUTS  NUTS3  shapefiles_barcelona_distrito  timezones`
- Save the mask as a netCDF file.
- Make PlotEquiMap() compatible with the mask array.

PlotWeeklyClim() use case

by Núria

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Successful research

- PlotWeeklyClim() code was originally developed for the H2020 **S2S4E** project
  - ➢ It is used to explore specific events: **case studies**
  - ➢ For each case study they did a Factsheet
    - Factsheet guide: https://s2s4e.eu/climate-services/case-studies/factsheets_guide
    - Factsheet examples: https://s2s4e.eu/index.php/climate-services/case-studies
- PlotWeeklyClim() function has been developed in the EURO-PolarNet-2 **ArticXchange** project
- New developments are being done in the context of **Decathlon** contract

ArticXchange

STORY #1

Anticipating the harvest time (Summer)

Period of **warmer than normal** conditions from **May to June** that triggers the start of summer earlier than normal

2018 and 2020

Hay making in mid-summer has become a normal activity of reindeer herders. In the southern herding associations, supplementary/corral feeding in winter is necessary to compensate for the loss of arboreal lichens associated with forest regeneration.

STORY #2

Anticipating the release of reindeer for outdoor feeding (Spring)

Not provided

Early snowmelt in Spring allows herders to release reindeer from fences earlier to feed in the forest.

STORY #3

Anticipating inopportune backwinter (Spring)

Period of **colder than normal** weather from **mid-April to mid-May** (chech sub-seasonal prediction - week by week).
Know it 2-3 weeks in advance

2004 and 2014

Backwinter (what Finnish call takatalvi) is a situation when it is very cold in spring or summer, when the calves are just born, and consequently it can affect the animal's survival.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

ArcticXchange

STORY #1

Anticipating the harvest time (Summer)

Period of **warmer than normal** conditions from **May to June** that triggers the start of summer earlier than normal

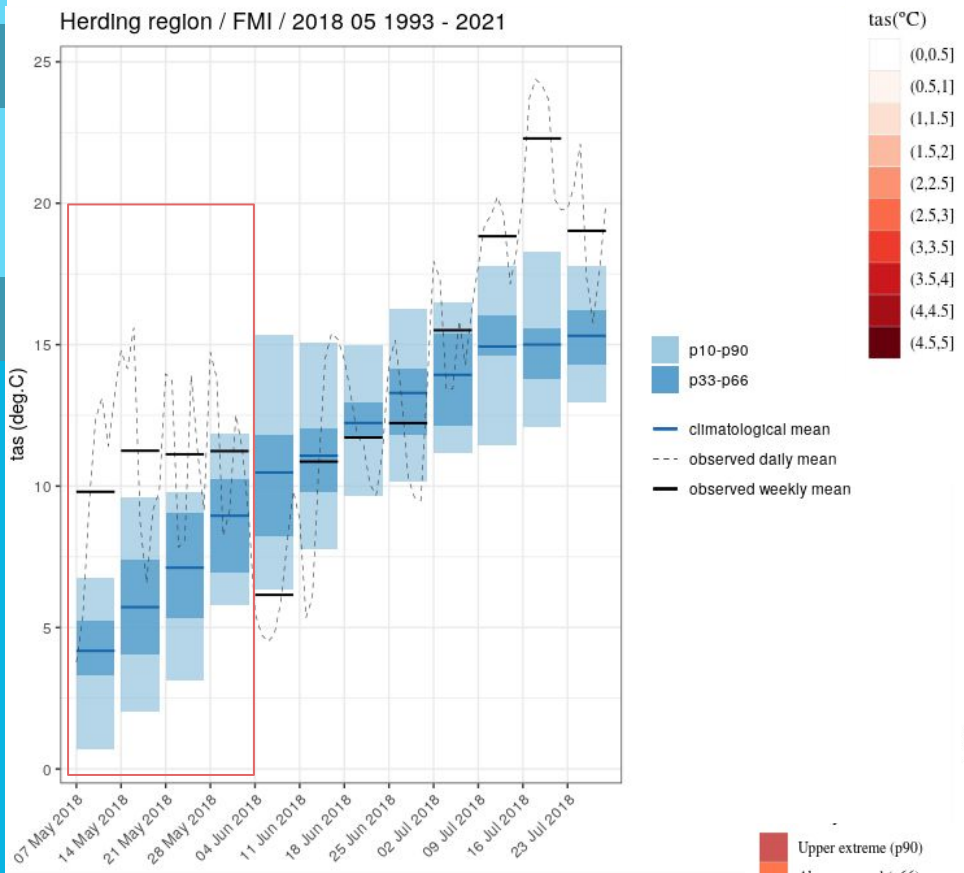
2018

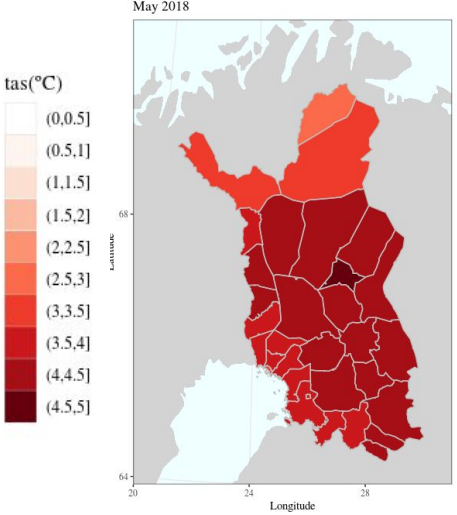
Past Observations

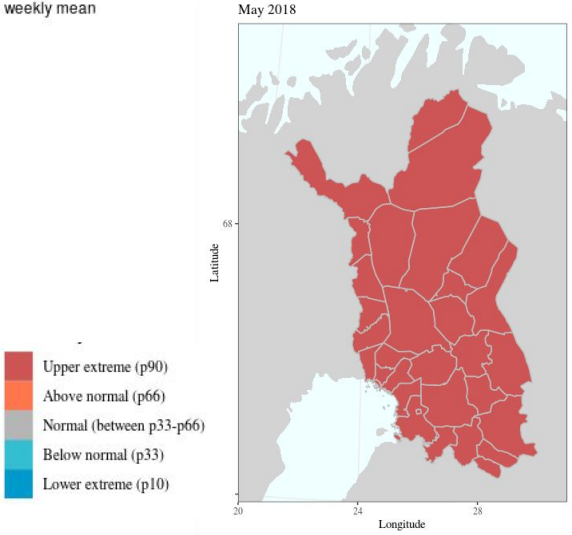
Four weeks during **May** show anomalous warm temperatures compared to the reference period 1993-2021 using FMI Climgrid dataset.

The monthly anomaly of May 2018 shows the spatial distribution of the temperatures that reach 5°C in the Pelkosenniemi municipality. These temperature anomalies beat the 90th percentile of the reference period making May 2018 an upper extreme event of heat.

ArcticXchange

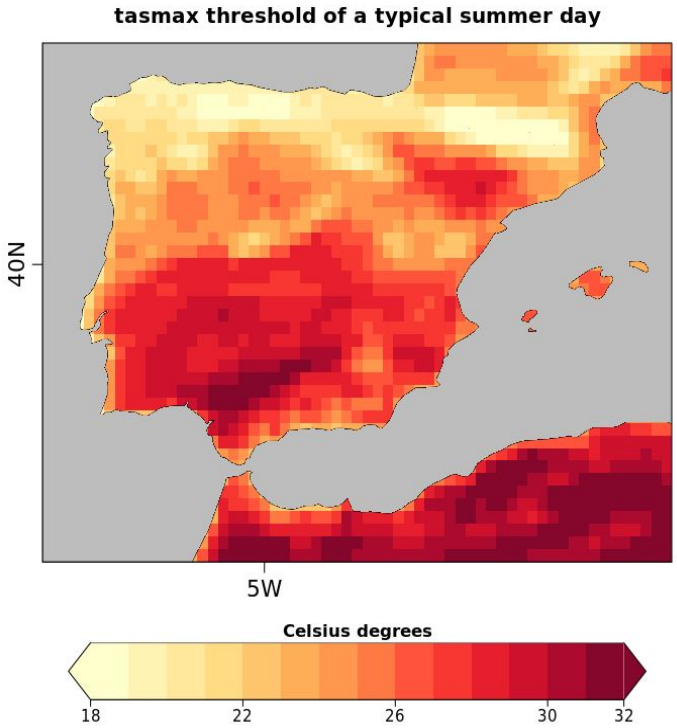






Decathlon application and further developments

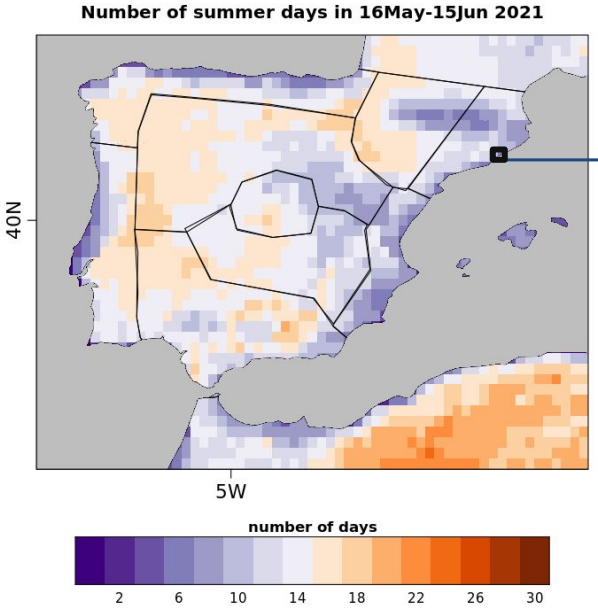
UMBRAL DE TEMPERATURA MÁXIMA: DÍAS TÍPICOS DE VERANO
Período de referencia: 1999-2020

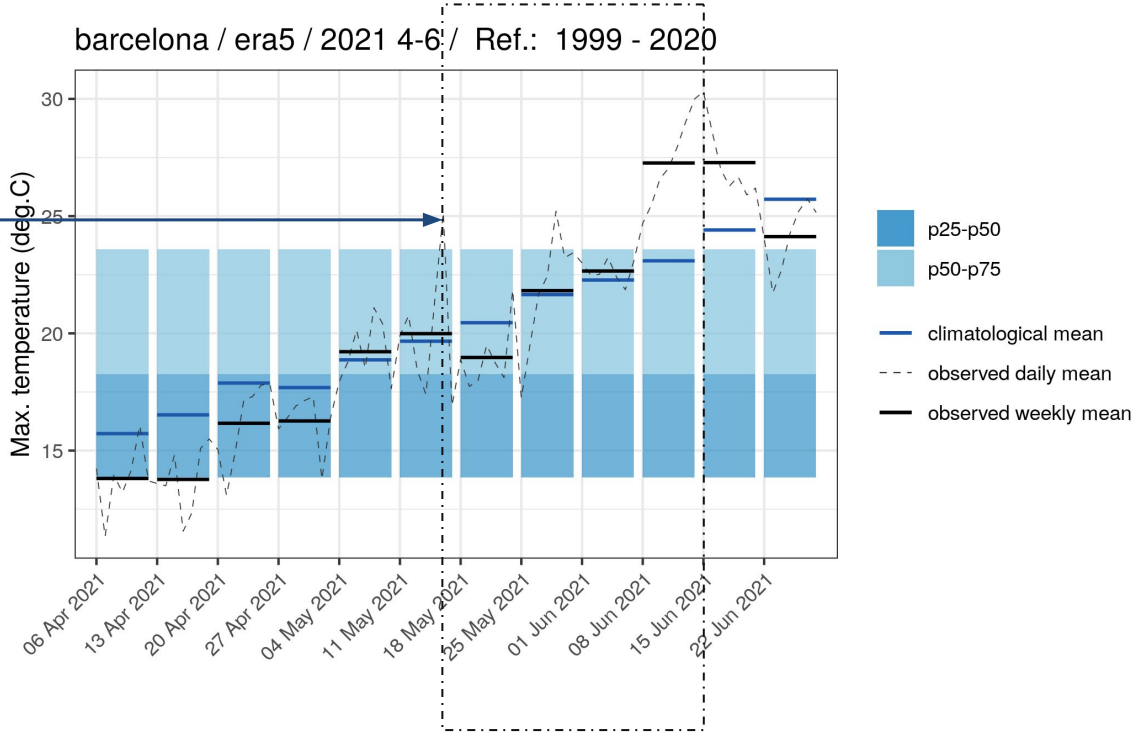


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Decathlon application and further developments

Días típicos de verano entre el 16 de mayo y el 15 de junio de 2021, Barcelona

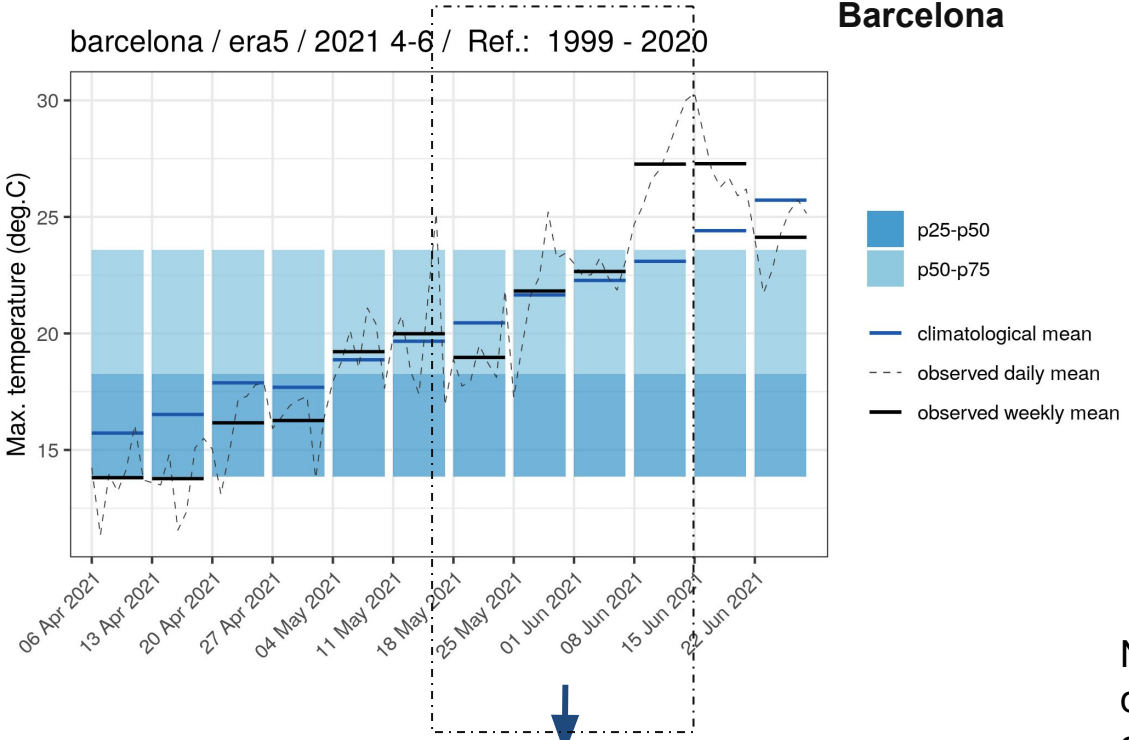


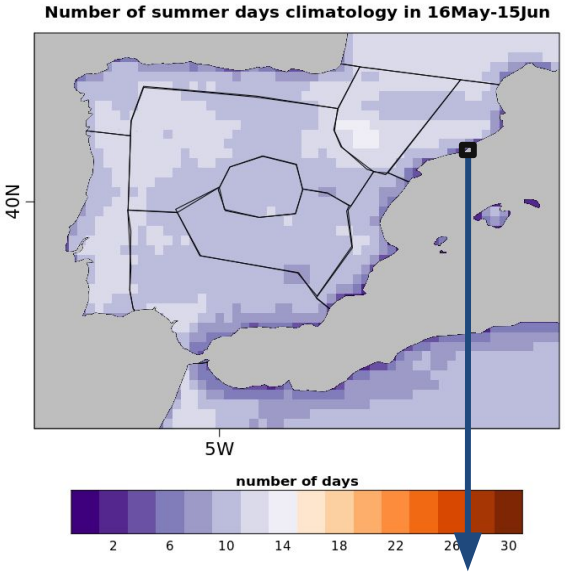


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Decathlon application and further developments

Días típicos de verano en 16mayo-15junio 2021 vs climatología del número de días Barcelona





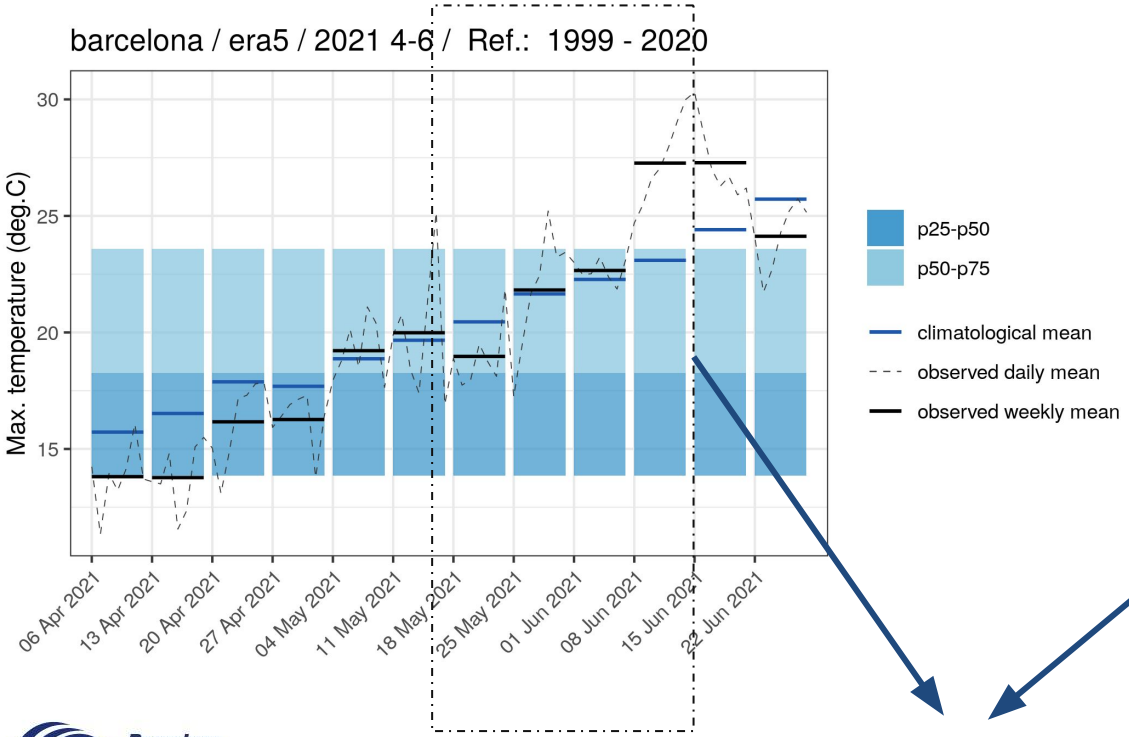
10 días con temperatura máxima típica de verano durante 16May-15Jun de 2021

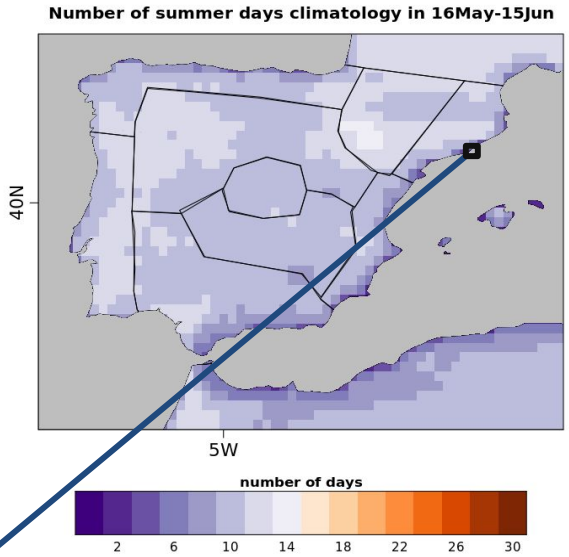
Número de días habituales en los que la temperatura máxima es típica de verano para el periodo que va entre el 16May-15Jun en Barcelona→ **7.7**

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Decathlon application and further developments

Días típicos de verano en 16mayo-15junio 2021 vs climatología del número de días Barcelona

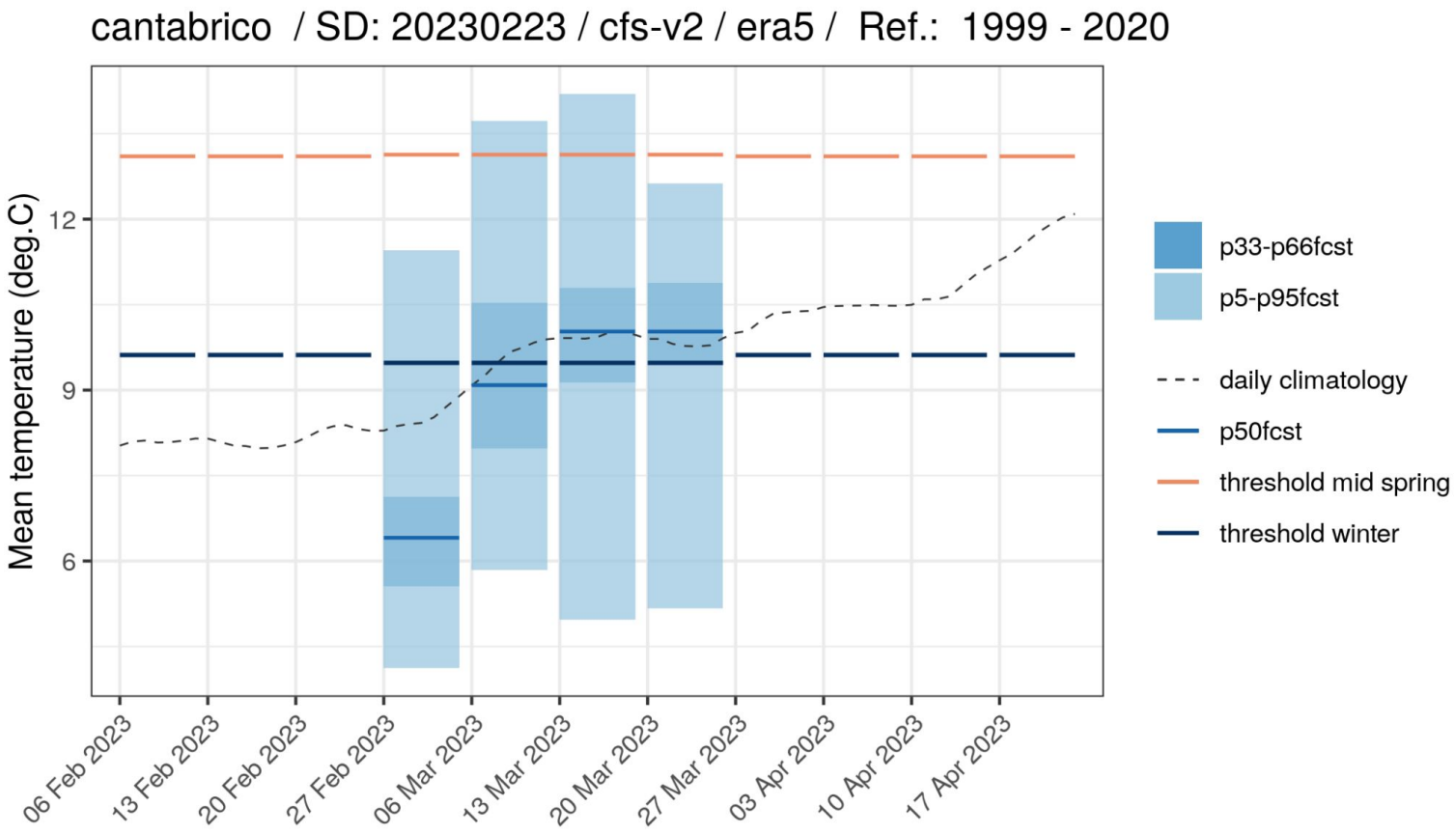




Barcelona Supercomputing Center
Centro Nacional de Supercomputación

En el periodo del 16 de mayo al 15 de junio de 2021 hubieron **2.3 días** con temperaturas máximas típicas de verano **más de lo habitual en Barcelona**

Decathlon application - including not only observations



Barcelona Supercomputing Center
Centro Nacional de S

Q & A

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Thanks for joining

Next meeting: 6th April, 12h