Agenda

1. Ice-breaker

2. News

   - s2dv
   - startR
   -
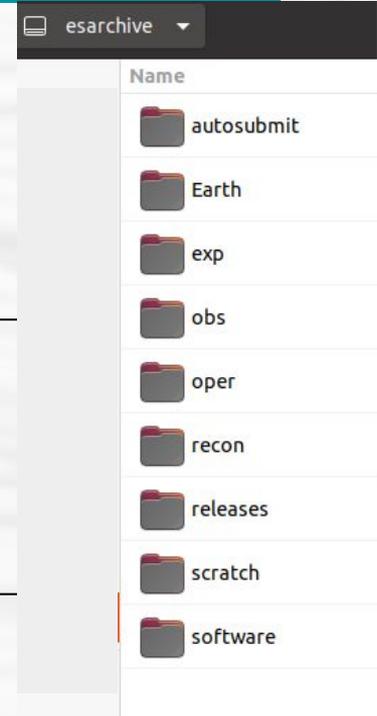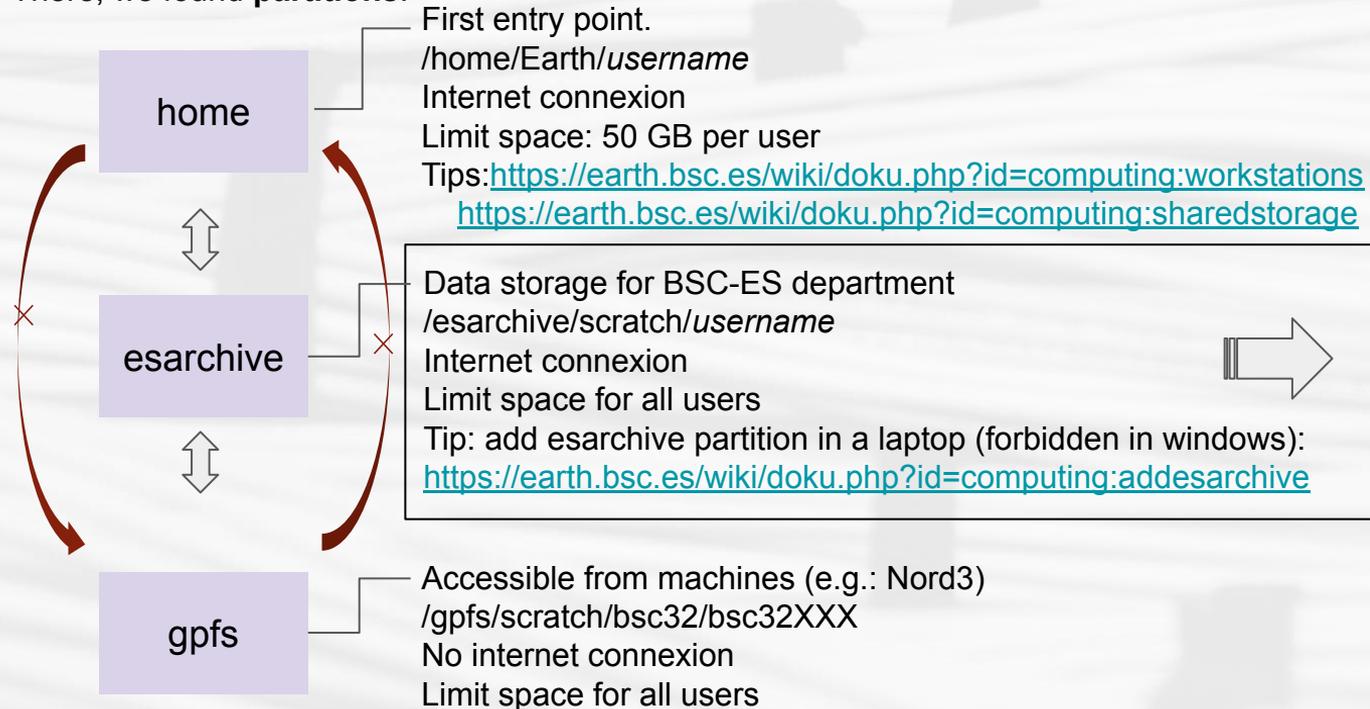3. ~~Visual Studio Code & R [Lluís]~~ (delay to the next meeting)
4. Q&A

Ice-breaker

What do we know about BSC-ES infrastructure?

We **access to BSC infrastructure**:
- When we use our workstations in the office
- When we connect to our bscearthXXX.int.bsc.es (with VPN on)
- https://docs.google.com/spreadsheets/d/1PXzbKlh0BVDfLg3HSoq64-3kb-DNgcw5_KeT5vdY3Z4/edit#gid=0
- For windows https://earth.bsc.es/wiki/doku.php?id=computing:sshwindows

There, we found **partitions**:

home

First entry point.
/home/Earth/*username*
Internet connexion
Limit space: 50 GB per user
Tips:https://earth.bsc.es/wiki/doku.php?id=computing:workstations
https://earth.bsc.es/wiki/doku.php?id=computing:sharedstorage

esarchive

Data storage for BSC-ES department
/esarchive/scratch/*username*
Internet connexion
Limit space for all users
Tip: add esarchive partition in a laptop (forbidden in windows):
https://earth.bsc.es/wiki/doku.php?id=computing:addesarchive

gpfs

Accessible from machines (e.g.: Nord3)
/gpfs/scratch/bsc32/bsc32XXX
No internet connexion
Limit space for all users

What do we know about BSC-ES infrastructure?

It is also possible to connect to BSC infrastructure through **servers** (physical machines):
- bscearth000.int.bsc.es/ bscearth001.int.bsc.es
  - download data
  - run the automatic package tests (GitLab CI, see e.g.:
    https://earth.bsc.es/gitlab/es/s2dv/-/pipelines)
- Shiny server
  - bscesshiny01.bsc.es
- bscesftp.bsc.es
  - share files https://earth.bsc.es/wiki/doku.php?id=computing:public_ftp

S
E
R
V
E
R
S

What do we know about BSC-ES infrastructure?

Software stack is a collection of programms and modules (including the operating system, architectural layers, protocols, runtime environments, …) that are installed in a machine.

★ The software stack at BSC could be different among machines and departments
★ We have access to:
  ○ BSC software stack (not managed by CES)
  ○ BSC-ES software stack (managed by CES)
    ■ workstations already using it
    ■ in some machines, we should edit the **bashrc** to use it (instructions are always in the wiki: https://earth.bsc.es/wiki/doku.php?id=library:computing)
    ■ It is built on **modules**, useful commands:
      ● *module list*
      ● *module load R*
      ● *module av*
    ■ other softwares like mendeley can be open in the workstation: */shared/earth/software/mendeley/latest/bin/mendeleydesktop*
★ Open an issue in **request** to ask for new software or R packages

S O F T W A R E   S T A C K

What do we know about BSC-ES infrastructure?

Which information we need to know for each machine?
- does it have bsc-es software?
- bidirectional connection
- internet
- slurm/lsf
- memory per node ….

Nord3
Marenostrum 4
workstations
Marenostrum 5
AMD cluster
Power 9
Nord3_v2

Find the information: https://earth.bsc.es/wiki/doku.php?id=library:computing

SOFTWARE STACK

What do we know about BSC-ES infrastructure?

workstations

- R 3.6.1
- To be use for debugging code (small data) or running startR workflow in remote machines
- BSC-ES software stack
- access esarchive
-

Nord3

- R 3.6.2
- To be use for running jobs
- **lsf** job scheduler
- No internet connexion
- no bidirectional connexion
- **will be deprecated**
- BSC-ES software stack
- X11 ssh -X

Nord3_v2

- R 3.6.2
- To be use for running jobs
- **slurm** job scheduler
- Internet connexion
- bidirectional?
- BSC-ES software stack

Marenostrum 4

- shared machine in BSC
- R 3.6.1
- To be use for running jobs
- BSC software stack

AMD cluster

- R 3.6.1
- To be use for running jobs
- BSC-ES software stack
- no acces to esarchive

Marenostrum 5

- coming soon?

Power 9

- BSC-ES software stack
- only for GPU users
- Can R use GPUs?
  e.g: gpuR

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

M
A
C
H
I
N
E
S

What do we know about BSC-ES infrastructure?

Recommendations

★   Work in the BSC environment (avoid local copies of files)
    Follow steps to **use workstation remotely**:
    https://earth.bsc.es/wiki/doku.php?id=computing:workstationsgraphicallyremotely

★   Save your scripts in GitLab (intermediate and final versions)
    ○   In an existing gitlab project
    ○   In a personal project
    ○   Documentation: https://earth.bsc.es/wiki/doku.php?id=library:computing#git
    ○   Clone the repo under /esarchive/scratch/*username*
        ■   You will have internet connexion to push your changes
        ■   The code will be accessible from workstation and Nord3
        ■   There is no backup copy of esarchive (that's a good reason to use gitlab)

★   Don't install local versions of R packages
    ○   if so, we cannot debug the code an reproduce the errors
    ○   better to open an issue in request to ask for the installation

★   Infrastructure in the wiki:
    https://earth.bsc.es/wiki/doku.php?id=library:best_practices#network_infrastructure

RECOMMENDATIONS

Problems [Núria has, who else? what else?]

➢ I would like to source() functions from gitlab directly
   > source("https://earth.bsc.es/gitlab/es/startR/-/blob/master/R/AddStep.R")
   Error in source("https://earth.bsc.es/gitlab/es/startR/-/blob/master/R/AddStep.R") :
    https://earth.bsc.es/gitlab/es/startR/-/blob/master/R/AddStep.R:1:1: unexpected '<'
   1: <
    ^

Solution:
   > source("https://earth.bsc.es/gitlab/es/s2dv/-/raw/master/R/PlotLayout.R")

What do we know about BSC-ES infrastructure?

Problems [Núria has, who else? what else?]

➢   When connecting to Nord3, I don't see the list of modules availables

        bsc32339@login3:~> module av
        �' for 'reading'ERROR:50: Cannot open file '�
        �' for 'reading'ERROR:50: Cannot open file '�
        �' for 'reading'ERROR:50: Cannot open file '�
        �' for 'reading'ERROR:50: Cannot open file '�
        �' for 'reading'ERROR:50: Cannot open file '�
        �' for 'reading'ERROR:50: Cannot open file '�
        �' for 'reading'ERROR:50: Cannot open file '�

➢   It works correctly for me in Nord3_v2.

➢   If you find a similar error, ask R team, Albert Vila or Kim.

P
R
O
B
L
E
M
S

What do we know about BSC-ES infrastructure?

Doubts [Núria has, who else? what else?]

- ➢ interactive session is possible from all machines?
- ➢ if so, which allow pop-up windows to display plots? Do we need always a X11 terminal?
  - ○ **type "xterm" to check**
- ➢ How we know if a machine allows bidirectional connexion? I don't find the answer in the documentation (e.g.: https://www.bsc.es/user-support/power.php)
  - ○ [Action] Núria to meet with Kim to clarify this point

P R O B L E M S

s2dv

PlotEquiMap

The shapefile is now plotted **on top of** the original figure. That is, you can use the parameter 'country.borders' and 'filled.continent' as before, and add additional shapefile through parameter 'shapefile' without sacrificing the default functions. You can find the enhancement in this branch.

```
PlotEquiMap(data, lon, lat, filled.continents = F,
            shapefile = '/esarchive/scratch/cdelgado/data/CountryBoundary.rds')
```
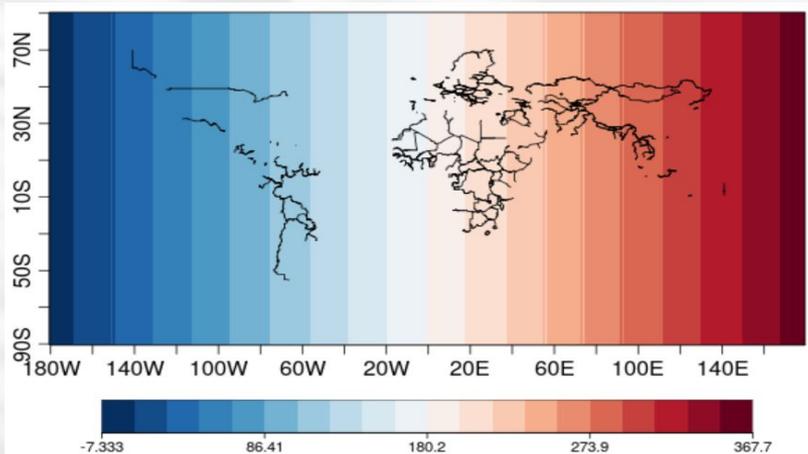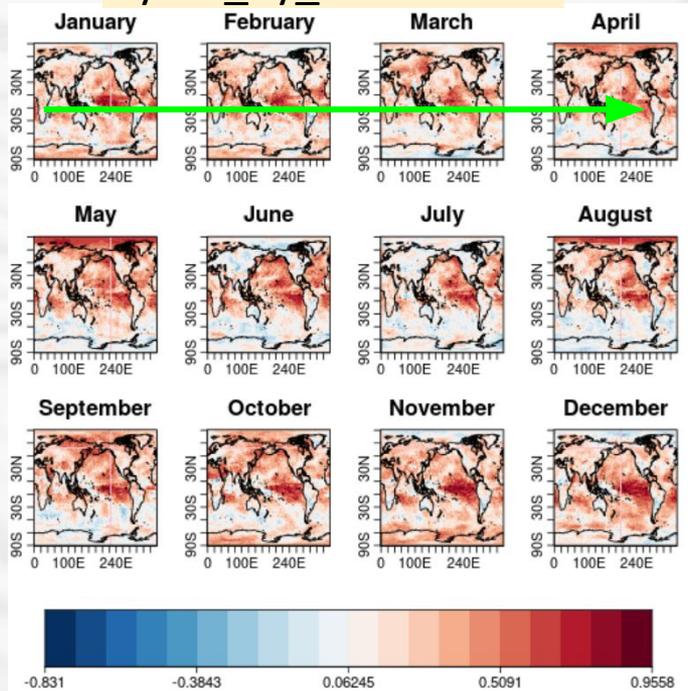
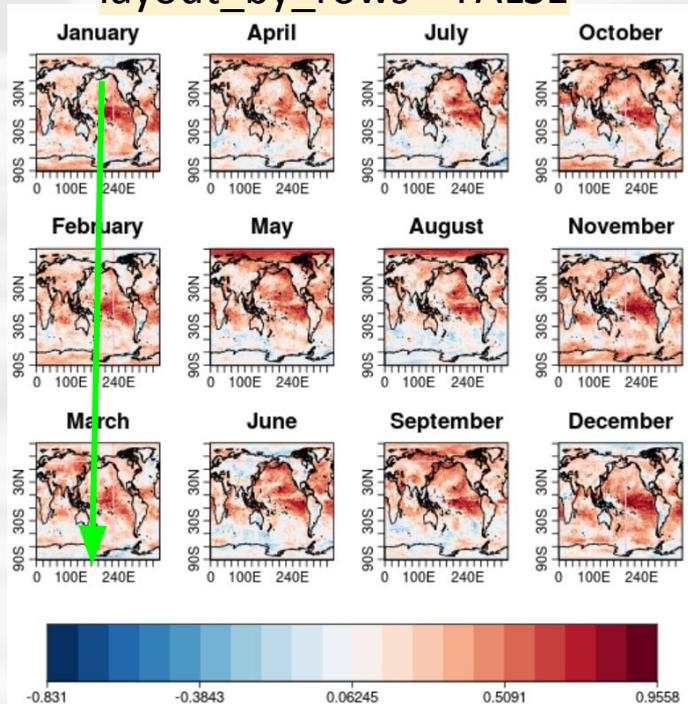**BEFORE**   **AFTER**
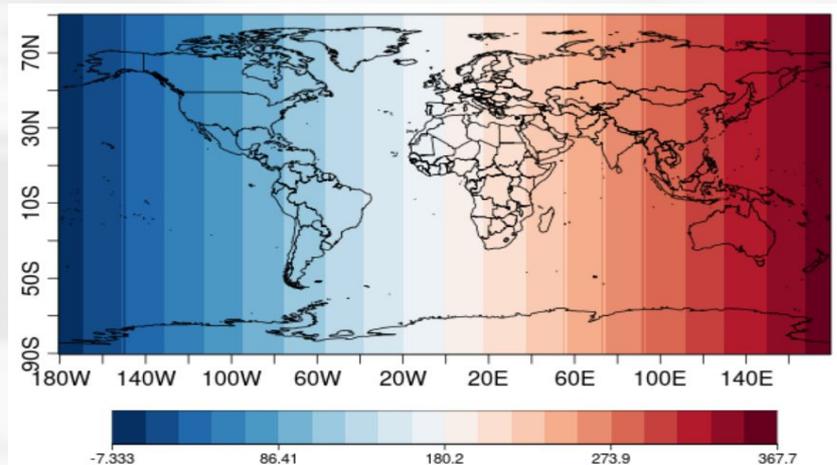
PlotLayout

New parameter 'layout_by_rows': Logical indicating whether the panels should be filled by columns (FALSE) or by raws (TRUE, default). It has been included in master branch.

MeanDims

New parameter 'drop': To choose whether to keep the averaged dim (FALSE) or drop it (TRUE). The default value is TRUE. You can find this development in [this branch](#).

data <- array(rnorm(120), dim = c(dat = 1, member = 5, sdate = 2, time = 12))

res <- MeanDims(data, 'member', drop = TRUE)

→ dim(res): [dat = 1, sdate = 2, time = 12]

res <- MeanDims(data, 'member', drop = FALSE)

→ dim(res): [dat = 1, **member = 1**, sdate = 2, time = 12]

CDORemap

New interpolation methods included ([issue 52](#)):

- con2
- laf
- nn

So, now the CDORemap() interpolation methods are 'con', 'bil', 'bic', 'dis', 'con2', 'laf', 'nn'.

→ Anything left? Please let us know as soon as possible.

The enhancement can be found here in gitlab master branch:
https://earth.bsc.es/gitlab/es/s2dv/-/blob/master/R/CDORemap.R

Discussion

1. PlotEquiMap(): The parameter 'dots' can plot dots or symbol on the figure. Is there a need to have the second 'dots' parameter?
    This is seen in a local function in S2S4E gitlab project. But I don't know if it is a common need. For now, we're not going to do this development. If you have this need, please contact us or open an issue on gitlab.

2. ACC(): area-weighted mean (issue 47)
    The development is required by some users but we need more information to do it. Please leave your comment in the issue, thanks!

startR

Discussion

1. **merge_across_dims** usage

   When an inner dimension (e.g., 'time') is continuous through multiple files (e.g., 'sdate'), we can merge files by this dimension. The lines to be used are : `time_across = 'sdate'` & `'merge_across_dims = TRUE'`. By this means, you will get a time dimension that the length is *time*sdate*, and 'sdate' dimension disappears.

   <mark>Question</mark>: Is there any case that `time_across = 'sdate'` is used but `'merge_across_dims = FALSE'`? Any specific reason to define Start() in this way?

   If you use '_across' + 'merge_across_dims = FALSE', please contact me to have a discussion.

Discussion

2. **'crop'** in parameter 'transform_param':

```
transform_params = list(grid = 'r360x181',

                        method = 'con',

                        crop = c(lons.min, lons.max, lats.min,
lats.max))
```

The value is the same as in CDO, i.e, can be <u>TRUE</u>, <u>FALSE</u>, or <u>four numbers of the border</u>.

HOWEVER, in startR, the three options lead to the same result. The output region is always the one defined by `lat` and `lon` dimensions.

Question: Can we deprecate 'crop' this argument? And startR will look for the correct region and crop automatically. Any concern?

→We are going to deprecate 'crop' in the next release. If you have any concern, please contact us soon.

RStudio

RStudio

1) From the workstation (or ssh -X user@bscearthXXX.int.bsc in a terminal):
- module load RStudio
- module load R/3.6.1-foss-2015a-bare
- rstudio &

2) From laptop:
- Follow steps to use workstation remotely:
  https://earth.bsc.es/wiki/doku.php?id=computing:workstationsgraphicallyremotely
- Follow steps 1 (above)

Q & A

Next meeting: 3rd Feb. 2021 (11 am)