# New tools for daily data

C. Prodhomme

Informal presentation, IC3, Barcelona, 2nd October

SPECS

Seasonal-to-decadal climate Prediction for the improvement of European Climate Services

# Why creating a new tool especially for daily and 6hourly data ?

- Huge files: Between 30 and 120 times bigger than monthly files
  → s2dverification is not yet ready for treating this kind of data

- Need of statistics over longer period
  (season, multi-years season, years,
  climatology over the full period) => even bigger files

- Work on extremes, intra-seasonal variability, daily index:
  - need of daily climatology, anomalies, filtered data,
  index calculated from more than one variable....

# Requirements

## Efficiency

## Flexibility:

→ be able to work on different kind of data

- observations, different model sources, type of files, frequency, grb/ncdf, ncdf4

→ be able to work on different time period:

season, multi-years, full period, few weeks, days....

→ be able to perform different statistics:

Sd, number of days under/over threshold, number of days under/over climatology, percentile, calculation of index, anomalies, filtering, climatology

## User Friendly:

→ Many potential users

# Why python instead of bash ?

**Against**:

→ need to learn a new language for some users

**For (compared to bash):**

→ Easier modularisation
→ Many useful libraries (date, ncdf...) and type of object *(dict, list...)*
→ Object oriented programming
→ Better security checks
→ Easy file handling for cdo and nco *(issue with /tmp to be solved)*
→ More efficient to call cdo from python
→ module for ncdf4 *(not yet installed)*
→ Direct manipulation of *numpy array*
→ Easy way to create, read and write ncdf

# Class Data

**Data object:**

**dataname**: *b0cm*
**varname**: *tas*
**inputdir**: /cfunas/exp/ecearth/b0cm/daily_grb/tas
**outputdir**: /cfunas/exp/ecearth/b0cm/monthly_statistics/
**filelist**: [tas_19810501.grb, tas_19820501.grb.....]
**sdates**: [date(1981,5,1), date(1982,5,1)....]
**filetype**: “grb”
**freq**: 6hourly (timedelta)
**nmenber**: 10

***select_leadtime:***
*select the time period between leadtime1 and leadtime2 in all files of the Data object*

***ready:***
*Make the data ready to calculate statistics*

***set_outputdirAuto:***
*set the output directory*

***format_file_nc:***
*Set files for cdo*

***monthly_percentile***

***cdo_statistics***

***monthly_nbdays***

***seasonal_nbdays***

***seasonal_percentile***

**Constructor**:
*exp=Data(“b0cm”, “tas”)*

**Set and get to access all Variables**:
*set_filelist(filelist)*
*get_filelist()*
*.....*

***see***
*To print the object*

***duplicate***
*To create a new object with the same variables, except sdates and filelist*

***copy***
*Copy the variable of an object in an other object*

# Class DataMod(Data)

**DataMod object:**

**dataname**: *b0cm*
**varname**: *tas*
**inputdir**: /cfunas/exp/ecearth/b0cm/daily_grb/tas
**outputdir**: /cfunas/exp/ecearth/b0cm/monthly_statistics/
**filelist**: [tas_19810501.grb, tas_19820501.grb.....]
**sdates**: [date(1981,5,1), date(1982,5,1)....]
**filetype**: “grb”
**freq**: 6hourly (timedelta)
**nmenber**: 10

***duplicate***
*To create a new object with the same variables, except sdates and filelist*

**And all functions of Data!**

***set_files:***
*Look for the file between year1 and year2 for a given startdate*

***create_monthly_list:***
*create a list of DataMod object containing only one month*

***write_output:***
*Write the output files with the right format*

# Class DataObs(Data)

**DataMod object:**

**dataname**: *b0cm*
**varname**: *tas*
**inputdir**: /cfunas/exp/ecearth/b0cm/daily_grb/tas
**outputdir**: /cfunas/exp/ecearth/b0cm/monthly_statistics/
**filelist**: [tas_19810501.grb, tas_19820501.grb.....]
**sdates**: [date(1981,5,1), date(1982,5,1)....]
**filetype**: “grb”
**freq**: 6hourly (timedelta)
**nmenber**: 10

***duplicate***
*To create a new object with the same variables, except sdates and filelist*

**And all functions of Data!**

***set_files:***
*Look for the file between year1 and year2 for a given startdate*

***write_output:***
*Write the output files with the right format*

***set_InpudirAuto:***
*Set the input directory according to the observation name*

***monthly_nbdays***

***list_season_obs(dataname, varname, sdates1, sdates2,seaslen)***

# Class DataENS(DataMod) and DataEC23(DataMod)

**DataMod object:**

**dataname**: *b0cm*
**varname**: *tas*
**inputdir**: /cfunas/exp/ecearth/b0cm/daily_grb/tas
**outputdir**: /cfunas/exp/ecearth/b0cm/monthly_statistics/
**filelist**: [tas_19810501.grb, tas_19820501.grb.....]
**sdates**: [date(1981,5,1), date(1982,5,1)....]
**filetype**: “grb”
**freq**: 6hourly (timedelta)
**nmenber**: 10

***copy***
*To create a new object with the same variables, except sdates and filelist*

**And all functions of Data and DataMod!**

**Different !**

***set_inputdirAuto:***
*Look for the inputdir*

***format_file_nc:***
*transform the grb or nc file into ncdf file readable for cdo*

***ready:***
*Make the data ready to calculate statistics*

# Auxiliary Function

get_last_date

get_freq

quit_leadtime

diff_month

set_output_name

settimeaxis

write_time

write_realization

write_lonlat

write_insti

write_id

write_source

concat_monthly_data

ponderate_mean

# Future works

**Revise the code**:

- Complete documentation
- see if some optimization could be done
- implement the standard output convention name
- bugfixes ?

**Include new statistics:**

- standard deviation
- onset, mjo, weather regime...
- .....

**Write documentation on the wiki**

- tutorial needed ?

**Put the code on gitlab**

**Other suggestions and comments are welcome !!!!**

# Suggestion of name convention

Name of variables:
**var**_leadtime1-leadtime2_**stat**_sd1-sd2

**-monthly_statistics**

*tasmax_q90_sdate.nc*
*tasmax_nbdays_q90_19810601-20100831_sdate.nc*

**-seasonal_statistics**

*tasmax_01-04_q90_sdate.nc*
*tasmax_01-04_nbdays_q90_19810601-20100831_sdate.nc*

**-subseas_statistics**

*tasmax_0000-0050_sd_sdate.nc*
*tasmax_0000-0050_nbdays_q90_19810601-20100831_sdate.nc*

**-climatologies**

clim_*tasmax_0000-0050_nbdays_q90_19810601-20100831.nc*