TRANSLATE

climaTe Risk informAtion from eNSambLe weAther and climaTe prEdictions

Deliverable Information	
Deliverable Number	**2.2 : A tool to extract the UNSEEN events that will be made available in open-source repositories and a report with technical documentation.**
Work Package	WP2: Tools and methodologies for the evaluation of UNSEEN events and their socio-economic impact
Task	2.2: Development of a tool to extract UNSEEN events from predictions
Responsibility	University of Bologna, CNR
Authors	Elena Bianco, Paolo Davini, Agostino Manzato, Giuseppe Zappa, Paolo Ruggieri
Project Information	
Project full name	climaTe Risk informAtion from eNSambLe weAther and climaTe prEdictions
Main ERC field	Physical Sciences and Engineering
Principal Investigator	Paolo Ruggieri
Substitute Principal Investigator	Giuseppe Zappa

Finanziato dall'Unione europea
NextGenerationEU

Part C: Comparison between the historical EFAS and 100 UNSEEN EFAS timeseries

This part will describe the program called **"EFAS_surrogate_explorer.py"**, which main purpose is to read the historical EFAS river discharge (RD) data in one point (default is "Bomporto" hydrological station) of a river watershed (default is "Panaro" river). The choice of river and station point are defined in the external library called **"region_details.py"**, which must be in the program root directory, called BINDIR, same as the main program.

The main Input data are the saved in the DATADIR directory. In particular, the historical EFAS netCDF RD timeseries is saved under DATADIR+"EFAS_historical/" (e.g.

EFAS_historical/postpro/day/watersheds/efas_panaro_masked.nc), while the 100 surrogate timeseries are saved under DATADIR+"EFAS_surrogate/". In particular, in our implementation the 100 surrogates are saved in 4 monthly subdirs (named after the initial month of the timeseries, i.e. Apr, May, Jun or Jul), each one containing 25 members (named from "00" to "24"). For example, our surrogate RD timeseries are saved in:

EFAS_surrogate/surrogate-v3/trimestral/Panaro/Apr/EFAS5_surrogate_Panaro_dis24_trimestral_Apr_00.nc

...

EFAS_surrogate/surrogate-v3/trimestral/Panaro/Jul/EFAS5_surrogate_Panaro_dis24_trimestral_Jul_24.nc

Please note that the control variable **ReadExisting** can be used to force the creation of postprocessed files even if they are already existing (in case it is set to False).

It is important that the program run in a virtual environment with all the needed libraries. Most of them are automatically loaded when the virtual environment efas is created with the command: **conda env create -f efas.yml,** where the file **efas.yml** configuration file is made available.

The program workflow is the following:

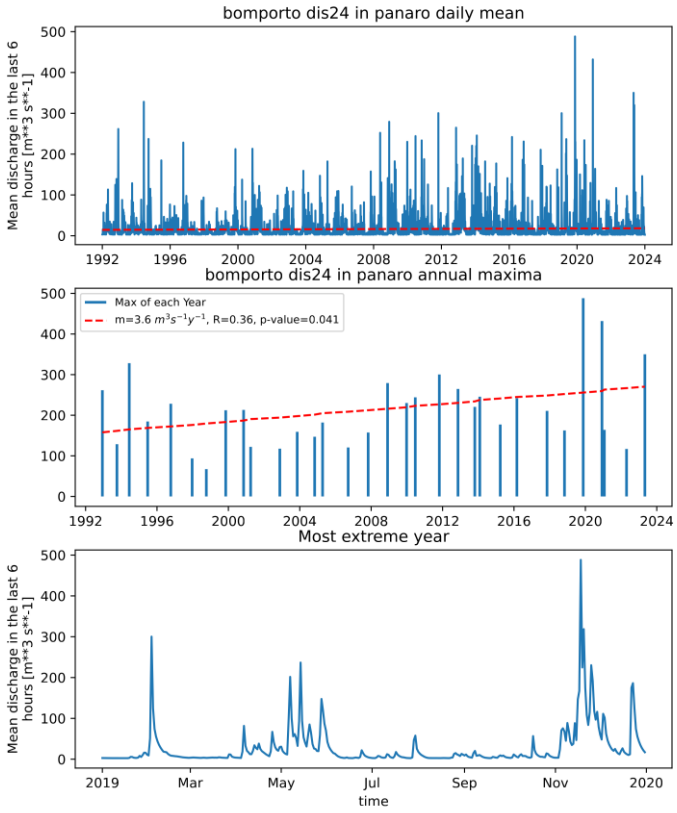
- The first thing that the program does is to read the watershed RD dataset and save the timeseries of RD only on the chosen location. For example for Panaro river and Bomporto station it reads:
  DATADIR/EFAS_historical/postpro/day/watersheds/efas_panaro_masked.nc

  and saves the smaller NetCDF file containing ED data only for one location:

DATADIR/EFAS_historical/postpro/day/single_gp/efas_panaro_bomporto_day.nc

- Then it computes the annual maxima for this timeseries and saves these data in *DATADIR/EFAS_historical/postpro/day/single_gp/efas_panaro_bomporto_day_annual_maxima.nc*

- The timeseries of daily mean RD and that of annual maxima are then fitted linearly and correlation with time is computed. A figure is generated to plot the full timeseries of daily mean RD (upper panel), the timeseries of annual maxima (mid panel) and the timeseries of daily mean RD for the year having the absolute maxima. The figure for Bomporto is the following:

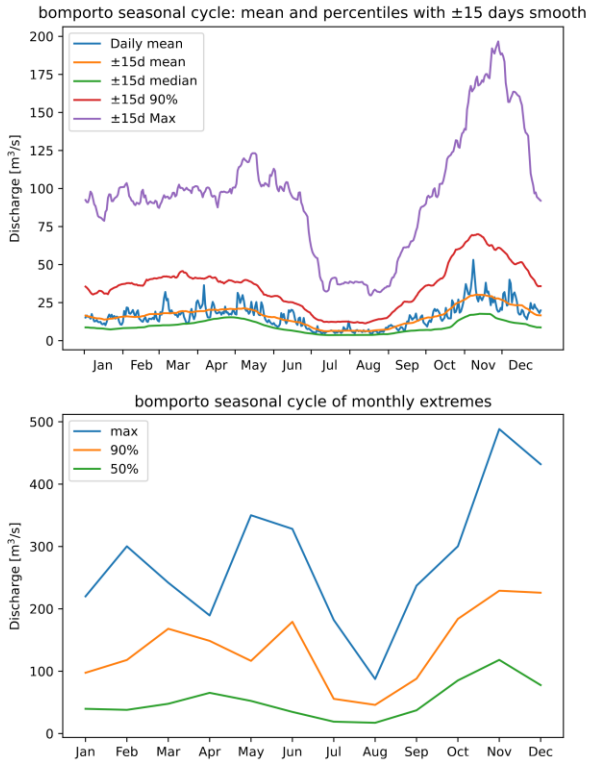


BINDIR/figures/panaro/day/bomporto_timeseries_fitted.pdf

- After that, it computed some statistics of the timeseries and of its annual cycle and saves these data in

DATADIR/EFAS_historical/postpro/day/watersheds/stats/efas_panaro_seasonalcycle_win31_pctx50.nc

At the same time, it plots the figure with the statistics of the daily mean RD annual cycle (of the RD smoothed by a 31-days moving average) and the statistics of the monthly annual-maxima cycle (RD not smoothed):



BINDIR/figures/panaro/day/bomporto_seasonal_cycle.pdf

- The next part of the program repeats the same flow, but on the 100 surrogate EFAS reanalysis. In particular, since these data are organized in four different subdirs (Apr, May, Jun and Jul), the program uses a *multiprocessing* environment to run simultaneously four parallel processes on four (or more) different CPU cores.
  The input files are like this one:

EFAS_surrogate/surrogate-v3/trimestral//Panaro/Jul//EFAS5_surrogate_Panaro_dis24_trimestral_Jul_00.nc

While the output files are like these netCDF files and PDF figures:

DATADIR/EFAS_surrogate/surrogate-v3/trimestral//Panaro/Jul/efas_surrogate_panaro_bomporto_Jul_24_masked.nc

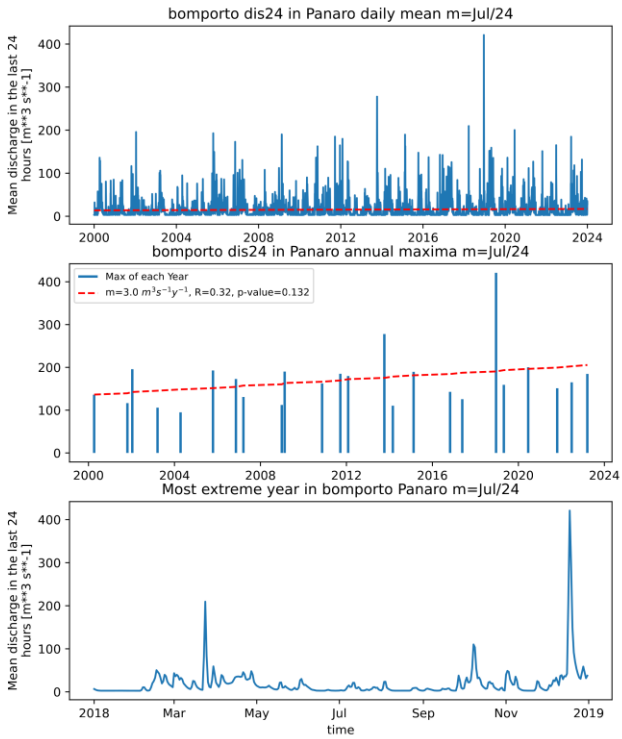
DATADIR/EFAS_surrogate/surrogate-v3/trimestral//Panaro/Jul/ efas_surrogate_Panaro_bomporto_Jul_24_masked_annual_maxima.nc

BINDIR/EFAS_historical/postpro/day/watersheds/stats/efas_panaro_seasonalcycle_win31_pctx50_Jul_24.nc

BINDIR/figures/panaro/day/surrogate /EFAS_surrogate_bomporto_timeseries_fitted_Jul_24.pdf

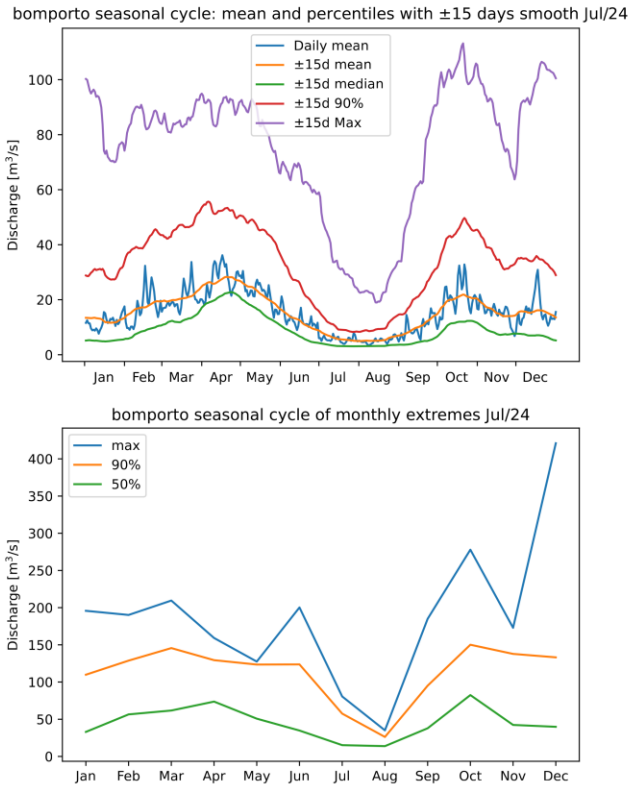
BINDIR/figures/panaro/day/bomporto_seasonal_cycle_surrogate _Jul_01.pdf

An example of this kind of figures are the following two (month=Jul, member=24):



BINDIR/figures/panaro/day/surrogate /EFAS_surrogate_bomporto_timeseries_fitted_Jul_24.pdf

and



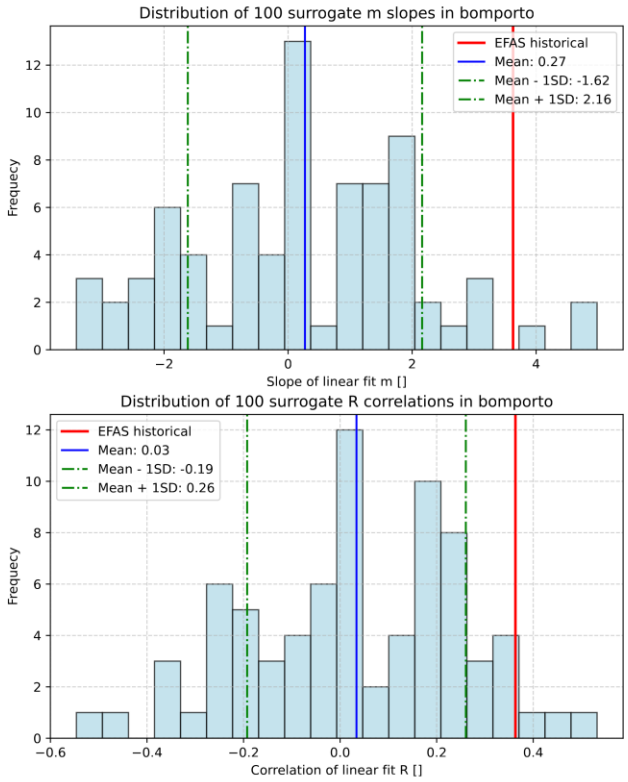
BINDIR/figures/panaro/day/bomporto_seasonal_cycle_surrogate _Jul_01.pdf

- All the parameters relative to the absolute maximum of each timeseries and relative to the linear fit with time for both the historical EFAS as well for all the 25 surrogate EFAS timeseries of a given initial month are saved in four text files, called:
  EFAS_surrogate_bomporto_timeseries_fit_results_Apr.txt
  ...
  EFAS_surrogate_bomporto_timeseries_fit_results_Jul.txt

- At the end of the (four) multiprocessing blocks, the program reads these four text files and creates these comparison figures between historical EFAS and statistics of the 100 surrogates:

  BINDIR/figures/panaro/day/bomporto_annual_maximum_surrogate_100_members.pdf

  which shows the distribution of the 100 slopes m and correlation coefficients R:

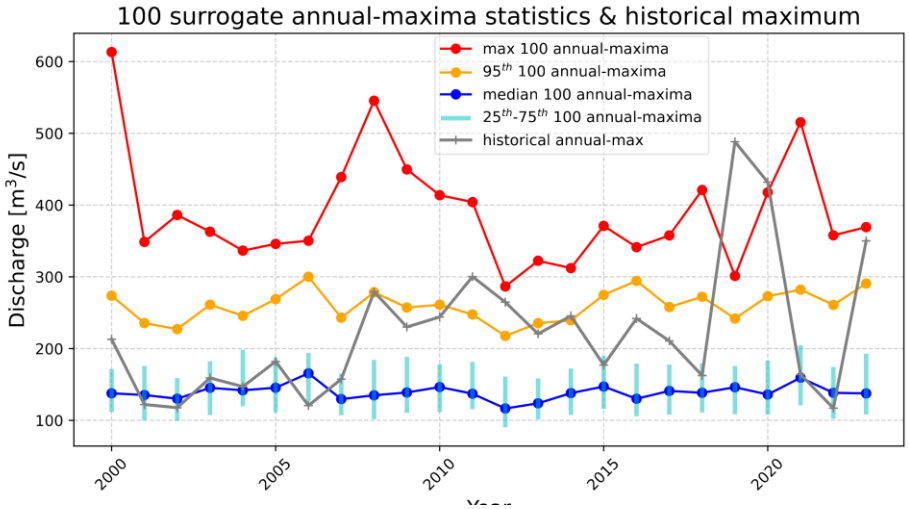


BINDIR/figures/panaro/day/bomporto_m_and_R_distro_for_annual_maximum_surrogate_100_members.pdf

And

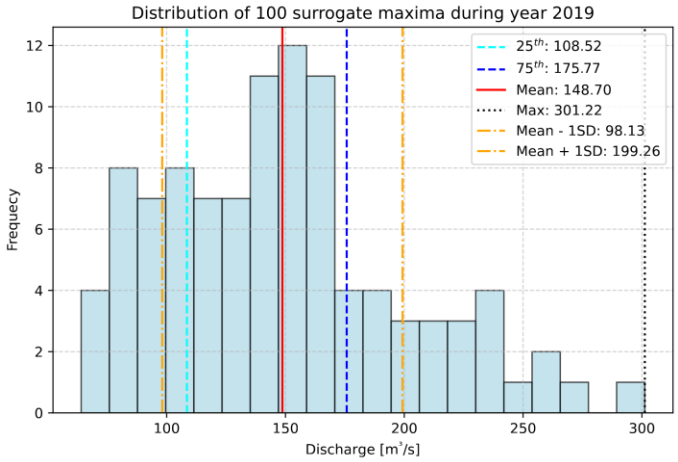
BINDIR/figures/panaro/day//bomporto_annual_maximum_surrogate_statistics_quantile.pdf

Which shows the timeseries of historical EFAS annual maxima and of statistics of 100 surrogate EFAS annual maxima per year:



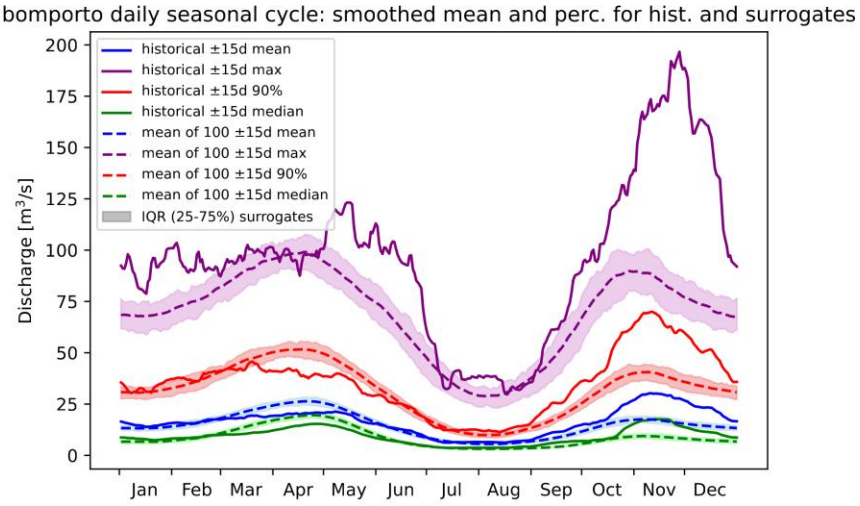
BINDIR/figures/panaro/day//bomporto_annual_maximum_surrogate_statistics_quantile.pdf

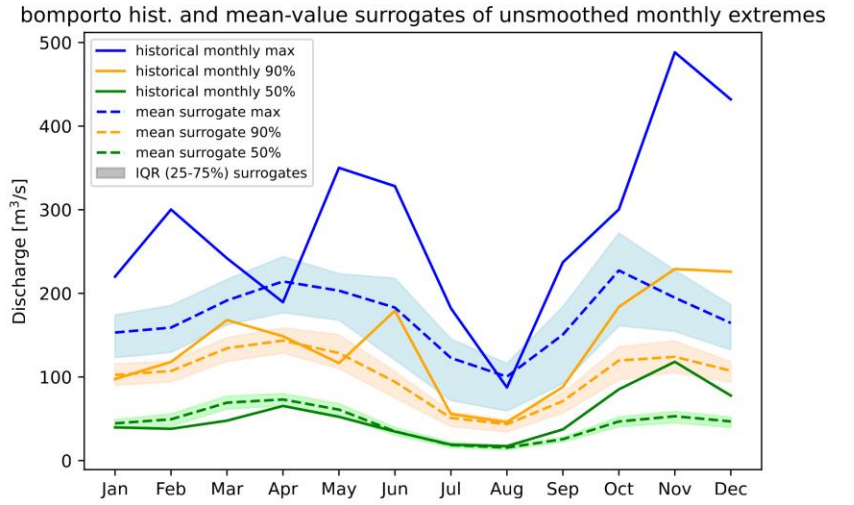
- For a given year (default is 2019) it plots also the distribution of the 100 maxima, to show how much skewed is this kind of distribution:



BINDIR/figures/panaro/day/bomporto_annual_maximum_surrogate_statistics_2019.pdf

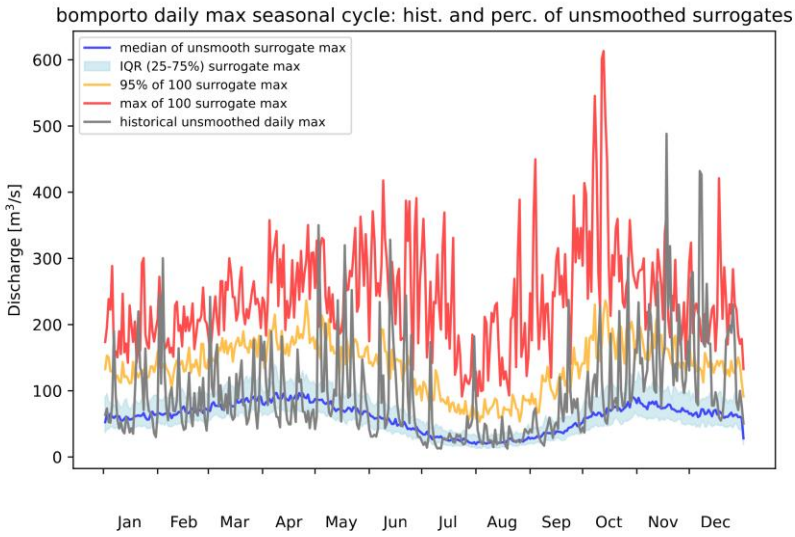
- Lastly, it adds two composite figures for the seasonal cycle: the first one shows the (31-days moving-averaged) historical EFAS cycle (from all daily mean) and statistics for the 100 surrogate cycles:





BINDIR figures/panaro/day/bomporto_seasonal_cycle_surrogate_statistics.pdf

while the second figure shows only the cycle of annual maxima of historical EFAS and of the statistics of absolute maxima of all 100 surrogates timeseries (no smoothing):



BINDIR/figures/panaro/day/bomporto_seasonal_cycle_daily-max_surrogate_statistics.pdf