

Processing large datasets using EGI-ACE EOSC resources for the climate community

Christian Pagé & Abel Aoun (CERFACS, France)

christian.page@cerfacs.fr <u>https://linkedin.com/in/pagechristian</u> <u>https://www.researchgate.net/profile/Christian_Page</u>

CECI, Université de Toulouse, CNRS, CERFACS, Toulouse, France abel.aoun@cerfacs.fr (CERFACS)



PDF POSTER

Impacts of Climate Change



2021 Germany Erftstadt, southwest of Cologne



- 2020 Hurricane Delta causes damage to Louisiana's Gulf Coast
- Urgent needs of impact assessments Identify mitigation solutions
- > Multiple domains: infrastructures, urban, agriculture, transportation, etc.
- > Easy to use tools are needed for very diverse users

- Python code developed@CERFACS since 2013
- Performance optimized
- > Fully compliant to CF and Metadata Standards
- >Validated against climpact & xclim
- **Easy install**: pip install icclim
- Implement the proper percentile indices calculations when calculation period overlaps reference period: bootstrapping method



III icclim: 49 Standard Indices



- Intra-period extreme temperature range [° C] **ETR**
- Warm days (days with mean temperature > 90th percentile of daily mean temperature) - **TG90p**
- Summer days (days with max temperature \geq 25 ° C) **SU**
- ...

>>> files = ['tasmax_day_CNRM-CM5_historical_r1i1p1_19950101-19991231.nc', 'tasmax_day_CNRM-CM5_historical_r1i1p1_20000101-20041231.nc', 'tasmax_day_CNRM-CM5_historical_r1i1p1_20050101-20051231.nc']

- >>> dt1 = datetime.datetime(1998,1,1)
- >>> dt2 = datetime.datetime(2005,12,31)
- >>> out_f = 'SU_JJA_CNRM-CM5_historical_r1i1p1_1998-2005.nc'
- # OUTPUT FILE: summer season values of SU
- >>> icclim.index(index_name='SU', in_files=files, var_name='tasmax', time_range=[dt1, dt2], slice_mode='JJA', out_file=out_f)

icclim: Code Architecture IV

II icclim: Climate Indices

- >Using many **xclim** climate indices functions as building blocks
- > xclim functions are using xarray, dask, pandas and numpy: optimized and parallel execution
- icclim v5 implements a specific API very similar to v4
- >Extended capabilities: userdefined indices, user-specific thresholds, etc.





climate4impact (C4I) V

> Flexible analysis features (Notebooks with icclim - Data Staging/Reduction Workflows) > Automated reproducibility mechanisms and documentation (Data/Analysis)

Take Home 🏠 Messages 🗹

VI EGI-ACE/EOSC Resources

> ENES Data Space resources

Workflows for data staging & remote subsetting-reduction (WPS) onto Customisable Notebooks



- 1. Wide Needs for tools to easily calculate climate indices
- **2.** icclim is a flexible, robust and fast python software for calculating climate indices
- 3. Creating pre-calculated standard indices datasets for CMIP6 is really a need
- **4. EGI-ACE through CMCC provides EOSC-based resources to generate** large datasets for the climate research community

provided by CMCC

Computing close to storage with data locally available

>Jupyter-Lab interface

>ssh/terminal access too

- Large Data Science environment (16 cores, 16GB RAM)
- Several TB of storage for output

Jupyter

Very fast and efficient tech

support



- Institutes/Climate Models
- Historical and SSPs
- > Members

- **Work Plan** VII
- Summer Days (SU) index already processed >Implementation of the generic processing script in progress







Time Periods (1850-2100)

Can be extended to CORDEX, CMIP5, ERA5

Especially complex with indices requiring a reference period (percentiles for example) Some inhomogeneity in model data needs to be addressed

>Need to choose the reference period: 1991-2020

Special thanks to the EGI-ACE and CMCC support! s-enes

IS-ENES: https://is.enes.org/

icclim 5.4 https://github.com/cerfacs-globc/icclim (pip install icclim)

climate4impact 2.0 beta: https://dev.climate4impact.eu/

The IS-ENES3 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°824084

https://rebrand.ly/egi icclim