

Climate4Impact (C4I) and icclim Tools to ease access to climate data



Christian Pagé
CERFACS, Toulouse, France



Royal Netherlands
Meteorological Institute
*Ministry of Infrastructure
and the Environment*

Abel Aoun

CERFACS Toulouse, France

Alessandro Spinuso

KNMI, De Bilt, Netherlands

IS-ENES3 Central&Eastern Europe Autumn School
From Climate Projections to Climate impacts via Regional
Downscaling

Prague, 28 November-2 December 2022



Max-Planck-Institut
für Meteorologie



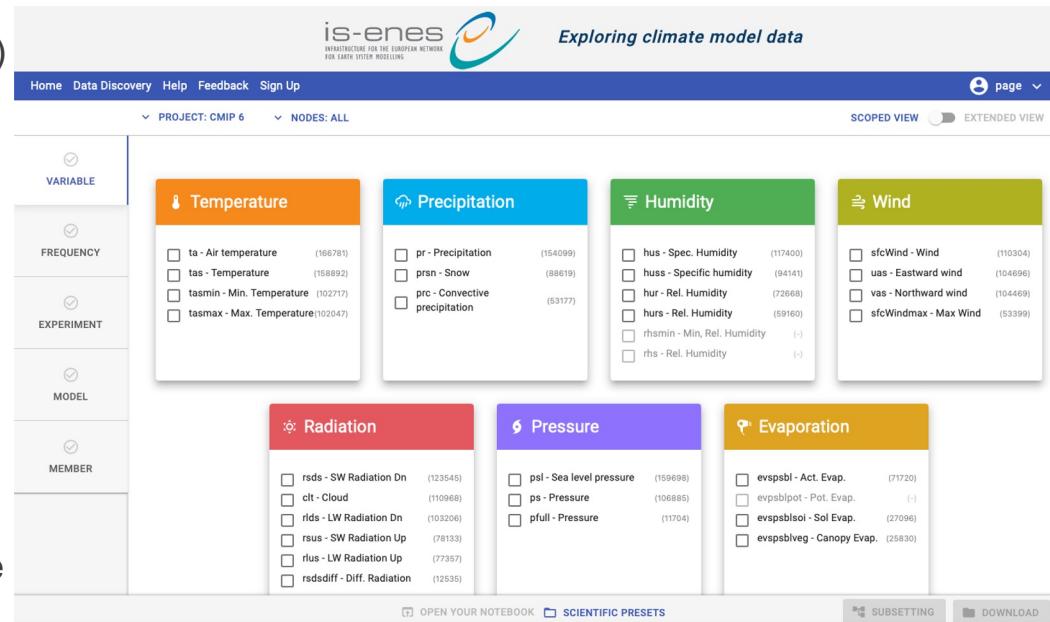
The University
of Manchester





Platform for researchers to explore climate data and perform analysis

- Front-end to climate data infrastructure (ESGF)
- Tailored Search Interface with view modes
- Jupyter-Lab enhanced environment
- Notebooks gallery
- Flexible analysis features
 - Climate indices calculation: **icclim**
 - Data Staging/Reduction Workflows
 - Personal store for processing outcome
- Automated reproducibility mechanisms and documentation (Data/Analysis)
- Modular Deployment & Decoupled Architecture
- Pages for Models Performance Comparison (**ESMValTool**)



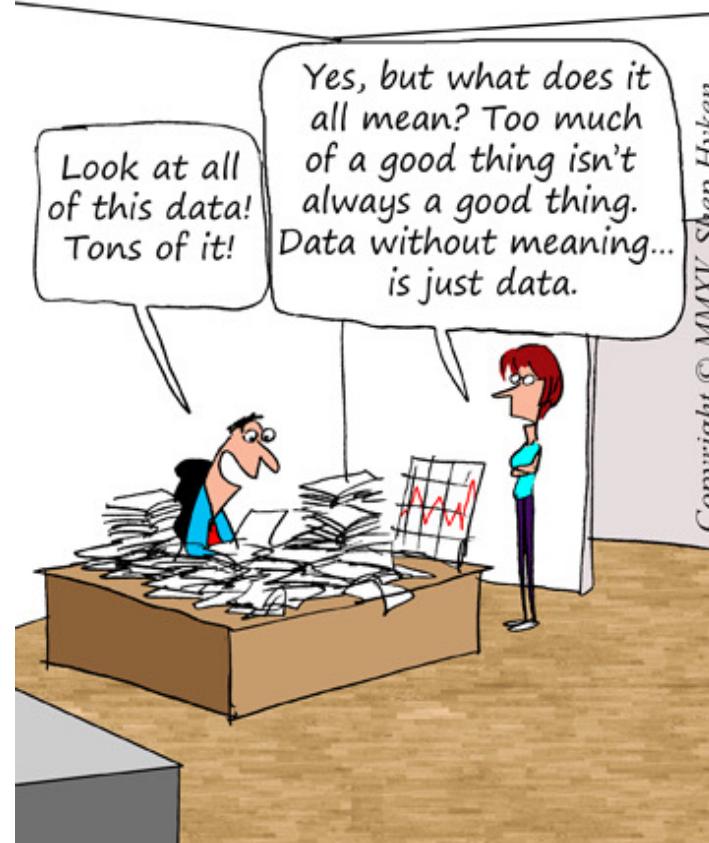
The screenshot shows the climate4impact (C4I) portal's search interface. On the left, a sidebar allows filtering by VARIABLE (e.g., ta, pr, rh), FREQUENCY (e.g., daily, monthly), EXPERIMENT (e.g., CMIP6, MIROC), MODEL (e.g., GFDL-CM3, MIROC4-UK3), and MEMBER (e.g., r1i1p1f2). The main area displays results in seven categories:

- Temperature:** ta - Air temperature (166781), tas - Temperature (158892), tasmin - Min. Temperature (102717), tasmax - Max. Temperature (102047)
- Precipitation:** pr - Precipitation (154099), prsn - Snow (88619), prc - Convective precipitation (53177)
- Humidity:** hus - Spec. Humidity (117400), huss - Specific humidity (94141), hur - Rel. Humidity (72688), hurs - Rel. Humidity (59160), rhsmn - Min. Rel. Humidity (-), rhs - Rel. Humidity (-)
- Wind:** sfcWind - Wind (110304), uas - Eastward wind (104696), vas - Northward wind (104469), sfcWindmax - Max Wind (53399)
- Radiation:** rsds - SW Radiation Dn (123545), clt - Cloud (110968), rlds - LW Radiation Dn (103206), rsus - SW Radiation Up (78133), rlus - LW Radiation Up (77357), rsdsdiff - Diff. Radiation (12535)
- Pressure:** psl - Sea level pressure (159698), ps - Pressure (106885), pfull - Pressure (11704)
- Evaporation:** evpsbli - Act. Evap. (71720), evpsblpot - Pot. Evap. (-), evpsbisoi - Sol. Evap. (27096), evpsbiveg - Canopy Evap. (25830)

At the bottom, there are links for OPEN YOUR NOTEBOOK, SCIENTIFIC PRESETS, SUBSETTING, and DOWNLOAD.

V2: Complete Redesign from V1
Beta version available <https://dev.climate4impact.eu>





Copyright © MMXV Shep Hyken.

Search Parametrisation made easier

Exploring climate model data

Home Data Discovery Help Feedback Sign Up

PROJECT: CMIP 6 NODES: ALL SCOPED VIEW

VARIABLE Frequency Experiment Model Member

variable:tasmax frequency:day experiment_id:ssp585 source_id:EC-Earth3 member_id:r1i1p1f1 member_id:r2i2p1f1

Temperature

- ta - Air temperature (9)
- tas - Temperature (6)
- tasmin - Min. Temperature (6)
- tasmax - Max. Temperature (6)

Precipitation

- pr - Precipitation (6)
- prsn - Snow (4)
- prc - Convective precipitation (1)

Radiation

- rsds - SW Radiation Dn (5)
- rlus - LW Radiation Up (4)
- rsus - SW Radiation Up (2)
- rlds - LW Radiation Dn (2)
- rsdssdiff - Diff. Radiation (-)
- clt - Cloud (2)

Pressure

- ps - Pressure (6)
- psi - Sea level (6)
- pfull - Pressure (6)

Nodes Selection by Service

Exploring climate model data

Home Data Discovery Help Feedback Sign Up

PROJECT: CMIP 6 NODES: ALL SCOPED VIEW

VARIABLE Frequency Experiment Model Member

variable:tasmax frequency:day experiment_id:ssp585 source_id:EC-Earth3 member_id:r1i1p1f1 member_id:r2i2p1f1

Available ESGF Nodes

Select & enable Rook WPS subsetting

Node	Subsetting Mode
esgf1.dkrz.de	Rook WPS
esgf3.dkrz.de	Rook WPS
aims3.llnl.gov	Opendap
cmip.dess.tsinghua.edu.cn	Opendap
cmip.fio.org.cn	Opendap
cordexesg.dmi.dk	Opendap
crd-esgf-drc.ec.gc.ca	Opendap
data.meteo.unican.es	Opendap
dataserver.nccs.nasa.gov	Opendap
dpesgf03.nccs.nasa.gov	Opendap
espg-cccr.tropmet.res.in	Opendap

Wind

- uas - Eastward wind (6)
- vas - Northward wind (6)
- sfcWind - Wind (5)
- sfcWindmax - Max Wind (4)

Evaporation

- evpsbl - Act. Evap. (-)
- evpsbjpot - Pot. Evap. (-)
- evpsbsol - Sol. Evap. (-)
- evpsbiveg - Canopy Evap. (-)

Climate4Impact Search for CMIP5/6 CORDEX Data

<https://dev.climate4impact.eu>

The screenshot shows a search interface for CMIP6 CORDEX data. The left sidebar lists categories: VARIABLE, FREQUENCY, EXPERIMENT, MODEL, and MEMBER. The main area displays variables grouped by category:

- Temperature:** tasmin - Min. Temperature (0), tasmax - Max. Temperature (6), ta - Air temperature (6), tas - Temperature (4).
- Precipitation:** pr - Precipitation (6), prsn - Snow (4), prc - Convective precipitation (1).
- Humidity:** hurs - Rel. Humidity (6), huss - Specific humidity (5), rhmin - Min. Rel. Humidity (1), rhmax - Max. Rel. Humidity (1), rha - Spec. Humidity (5), hur - Rel. Humidity (1).
- Wind:** sfcWind - Wind (6), sfcWindEastward - Eastward wind (6), sfcWindNorthward - Northward wind (6), sfcWindMax - Max Wind (4).
- Radiation:** rsds - SW Radiation Dn (5), rsus - SW Radiation Up (1), rlds - LW Radiation Dn (2), rlus - LW Radiation Up (2), rsddif - Diff. Radiation (2), cld - Cloud (2).
- Pressure:** ps - Pressure (1), cst - Sea level pressure (6).
- Evaporation:** evipobol - Act. Evap. (1), evipobolP - Pot. Evap. (1), evipobolS - Sol Evap. (1), evipobleg - Canopy Evap. (1).

At the bottom are buttons for "OPEN YOUR NOTEBOOK", "SCIENTIFIC PRESETS", "SUBSETTING", and "DOWNLOAD".

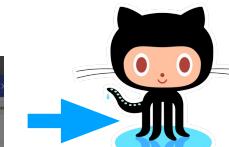
Workflows for data staging & reduction onto icclim powered Notebooks

The screenshot shows a Jupyter Notebook interface. The code cell contains Python code for calculating summer days and plotting them. The output cell shows two heatmaps of summer days across a geographic area.

```

# Set spatial extent and centre
central_lat = 47.8
central_lon = 10.0
extent = [-138, 38, 38, 56] # Western Europe
# Calculate time average
icclim = InvenioClimTime(), keep_attrs=True)
# Set plot projection

```

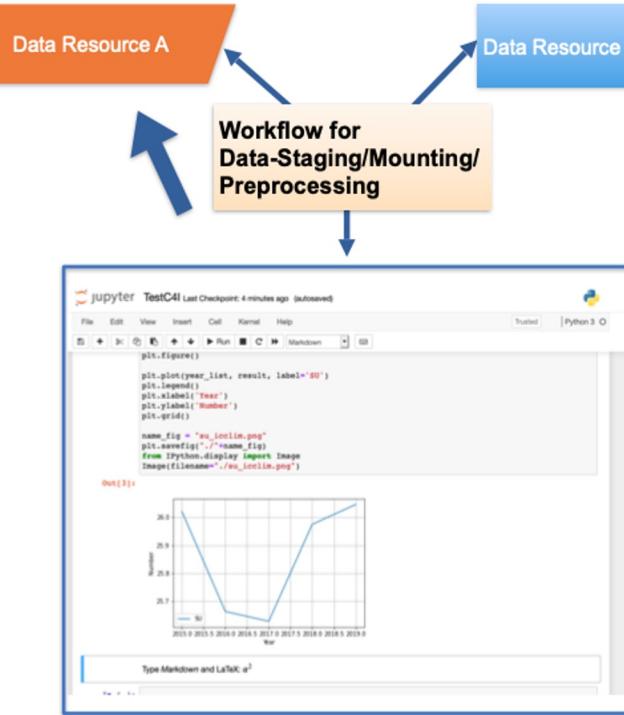


Save/Share Progress to Git

Reduced Data MyBinder Reproduce

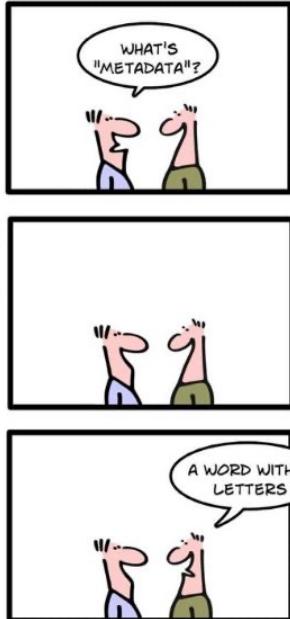
- Trace Changes to Restore, Recover Software and/or Data





- access distributed raw data
- develop, document and reuse methods for processing and visualisation.
- update/extend raw data and software
- Track changes and rollback
(Traceability/Recovery)
- keep old versions of the data after updates
(Reproducibility)
- snapshot and restore the state of a workspace software
(Reproducibility)

SIMPLY EXPLAINED:
METADATA



Workflow Monitoring

GitHub Authentication

Snapshot Controls

Data Staging Rollback

Activities History and Provenance

A screenshot of the IS-ENES Climate Data Infrastructure for Climate 4 Impact interface. It shows a sidebar with sections for Workflow Monitoring, GitHub Authentication, Snapshot Controls, Data Staging Rollback, and Activities History and Provenance. A red curly brace on the left groups the last three sections. To the right is a main panel showing a Jupyter Notebook interface with code and a plot, and a GitLab repository page for 'C4I Use Cases as Jupyter Notebooks'.

<https://gitlab.com/is-enes-cdi-c4i/notebooks>

Home Data Discovery Help Feedback Sign Up

PROJECT: CMIP 6 NODES: CUSTOM (2)

VARIABLE
 variable:ta
 variable:prsn
 variable:huss
 frequency:day
 experiment_id:ssp585

FREQUENCY
 FREQUENCY

EXPERIMENT
 EXPERIMENT

MODEL
 MODEL

MEMBER
 MEMBER

Model

- CanESM5 - CanESM5
- MPI-ESM1-2-LR - MPI-ESM1.2-LR
- UKESM1-0-LL - UKESM1.0-N96ORCA1
- CNRM-CM6-1 - CNRM-CM6-1
- MIROC6 - MIROC6
- CNRM-ESM2-1 - CNRM-ESM2-1
- MRI-ESM2-0 - MRI-ESM2.0
- GFDL-CM4 - GFDL-CM4
- HadGEM3-GC31-LL - HadGEM3-GC3.1-N96ORCA1
- MPI-ESM1-2-HR - MPI-ESM1.2-HR
- INM-CM4-8 - INM-CM4-8
- INM-CM5-0 - INM-CM5-0
- ACCESS-CM2 - Australian Community Climate and Earth System Model
- AWI-CM-1-1-MR - AWI-CM 1.1 MR
- NorESM2-LM - NorESM2-LM (low atmosphere-medium ocean resolution)
- BCC-CSM2-MR - BCC-CSM 2 MR
- CMCC-CM2-SR5 - CMCC-CM2-SR5
- FGOALS-g3 - FGOALS-g3

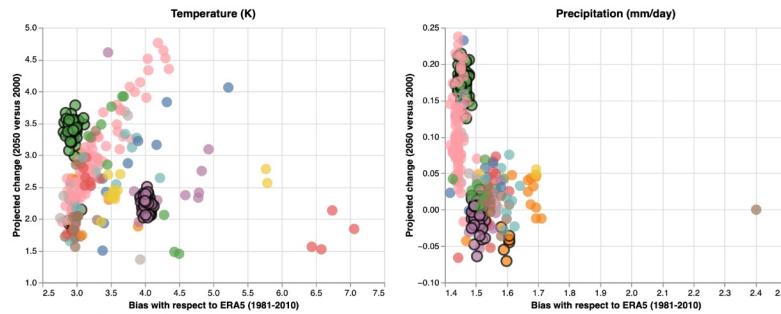
[COMPARE MODEL PERFORMANCE](#)

Climate impact result viewer

This application shows results from CMIP5 and CMIP6 models, calculated with ESMValTool. It is intended to provide some guidance for climate impact researchers, to select one or more datasets that adequately sample the spread of the CMIP ensemble.

- Bias is calculated with respect to the ERA5 reanalysis dataset over the period 1981–2015.
- Future change is calculated for 2036–2065 as compared to 1986–2015.
- Area is set to Europe (lon 0–39; lat 30–76.25)
- All data are taken from the RCP/SSP 8.5 scenario

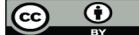
Hold ctrl to pan and zoom, hold alt to select a range (points will be highlighted in both graphs), then hold shift to select multiple points.



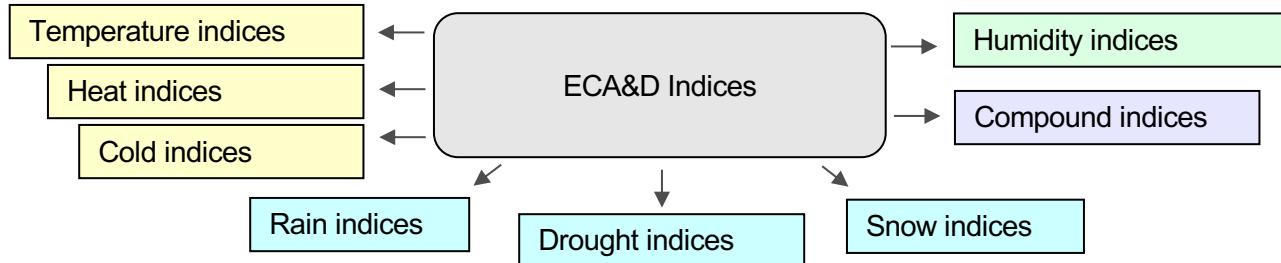
[View static recipe output](#)

Netherlands eScience Center

IS-ENES3 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084



icclim: Index Calculation for Climate



- 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 > 25 ° C) - **SU**
- ...

- Python code developed at Cefracs since September 2013
- Funded by EU FP7 IS-ENES2, FP7 CLIPC and H2020 IS-ENES3
- Generic and modular approach, can be reused in other environments
- New V5 completely rewritten and using underlying xclim functions, based on xarray and dask
- I/O interface is structured for optimal performance
- Implement the proper percentile indices calculations when calculation period overlaps reference period (called bootstrapping method)
- Available indices: https://icclim.readthedocs.io/en/latest/explanation/climate_indices.html#icclim-capabilities



Documentation: https://icclim.readthedocs.io/en/latest/python_api.html

Source code: <https://github.com/cerfacs-globc/icclim>

Current Version 6.1.2: <https://github.com/cerfacs-globc/icclim/releases/tag/v6.1.2>

`icclim.index(**kwargs)`

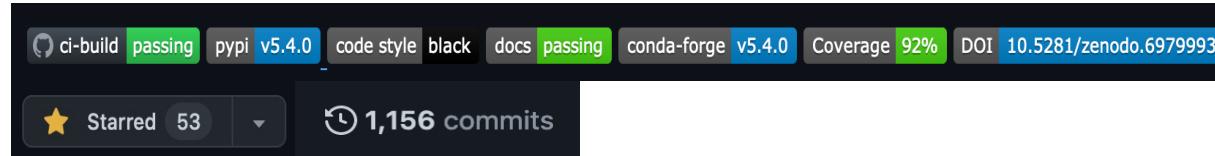
Parameters:

- `in_files` (`str / list[str] / Dataset / DataArray`) – Absolute path(s) to NetCDF dataset(s), including OPeNDAP URLs, or path to zarr store, or `xarray.Dataset` or `xarray.DataArray`.
- `index_name` (`str`) – Climate index name. For ECA&D index, case insensitive name used to lookup the index. For user index, it's the name of the output variable.
- `var_name` (`str / list[str] / None`) – `optional` Target variable name to process corresponding to `in_files`. If `None` (default) on ECA&D index, the variable is guessed based on the climate index wanted. Mandatory for a user index.
- `slice_mode` (`str`) – Type of temporal aggregation: {"year", "month", "DJF", "MAM", "JJA", "SON", "ONDJFM" or "AMJJAS"}. Default is "year". See `slice_mode` for details.
- `time_range` (`list[datetime.datetime, datetime]`) – `optional` Temporal range: upper and lower bounds for temporal subsetting. If `None`, whole period of input files will be processed. Default is `None`.
- `out_file` (`str / None`) – Output NetCDF file name (default: "icclim_out.nc" in the current directory). Default is "icclim_out.nc". If the input `in_files` is a `Dataset`, `out_file` field is ignored. Use the function returned value instead to retrieve the computed value. If `out_file` already exists, icclim will overwrite it!
- `threshold` (`float / list[float] / None`) – `optional` User defined threshold for certain indices. Default depend on the index, see their individual definition. When a list of threshold is provided, the index will be computed for each thresholds.



icclim v6.1.2

<https://github.com/cerfacs-globc/icclim>



ci-build passing pypi v5.4.0 code style black docs passing conda-forge v5.4.0 Coverage 92% DOI 10.5281/zenodo.6979993

Starred 53 1,156 commits

```
import icclim
summer_days = icclim.su("netcdf_files/tasmax_1990-2100.nc", out_file="summer_days.nc")
```

A F.A.I.R icclim



Findable:

- [Open source code](#), licensed under the permissive Apache 2 license.
- Try to follow [OpenSSF principles](#)
- Documentation freely available on [readthedocs](#)
- Dissemination by Christian in conferences (AGU, AMS, EGU...).
- Integrated in [C4I](#) and C3S.



Accessible

- Documentation on readthedocs.
- Github issues and pull requests.
- CERFACS support the development.



Interoperable

- Based on well knowns libraries: xarray, xclim, numpy.
- Support NetCDF and xarray.Dataset input formats.
- Indices are derivated from ECA&D and CF conventions standards.
- icclim custom parsing for non CF complying inputs.



Reusable

- NetCDF output, with enriched metadata.
- Easy to integrate within existing infrastructures.
- Easy to write small scripts with it.



icclim: examples

notebooks: <https://gitlab.com/is-enes-cdi-c4i/notebooks/>

```
## KNMI TX
tx_files = glob.glob(f"netcdf_files/knmi/clean/*tx*.nc")
bp = [datetime.datetime(1901, 1, 1), datetime.datetime(1921, 12, 31)]
icclim.index(index_name='tx90p',
              in_files=tx_files,
              base_period_time_range=bp,
              slice_mode='YS',
              out_file="netcdf_files/output/out.nc")
```

[new in 5.1.0] Rechunk how data is stored

Documentation: https://icclim.readthedocs.io/en/latest/how_to/dask.html#create-an-optimized-chunking-on-disk

```
import icclim

ref_period = [datetime.datetime(1980, 1, 1), datetime.datetime(2009, 12, 31)]
with icclim.create_optimized_zarr_store(
    in_files="netcdf_files/tas.nc",
    var_names="tas",
    target_zarr_store_name="opti.zarr",
    keep_target_store=False,
    chunking={"time": -1, "lat": "auto", "lon": "auto"},
) as opti_tas:
    icclim.index(
        index_name="TG90p",
        in_files=opti_tas,
        slice_mode="YS",
        base_period_time_range=ref_period,
        out_file="netcdf_files/output/tg90p.nc",
    )
```

[new in 5.1.0] Compute multiple indices at once

Documentation https://icclim.readthedocs.io/en/latest/how_to/recipes_ecad.html#multi-index-computation

```
bp = [datetime.datetime(1991, 1, 1), datetime.datetime(1999, 12, 31)]
tr = [datetime.datetime(1991, 1, 1), datetime.datetime(2010, 12, 31)]
res = icclim.indices(
    index_group=IndexGroup.HEAT,
    in_files=".//netcdf_files/climpact.sampledata.gridded.1991-2010.nc",
    base_period_time_range=bp,
    time_range=tr,
    out_file="pouetpouet.nc",
)
```

[new in 5.2.0] one function per index

Documentation: https://icclim.readthedocs.io/en/latest/references/ecad_functions_api.html

```
import glob
import icclim
summer_days = icclim.su(
    in_files=glob.glob("netcdf_files/tasmax*.nc"),
    out_file="summer_days.nc",
)
```

[new in 5.3.0] custom season between exact dates



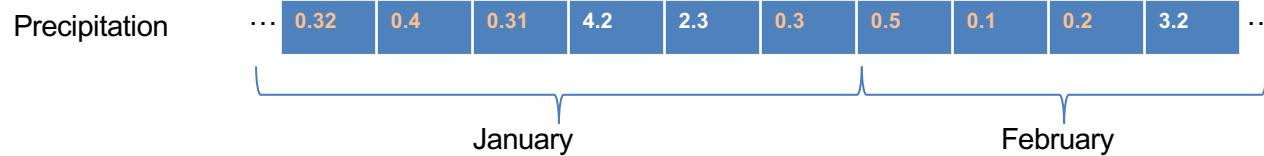
```
result = icclim.su(  
    in_files=xr.open_dataset(climp_file).tmax,  
    slice_mode=["season", ["19-07", "8 Aout"]],  
    ).compute()
```

icclim v6

Generic climate indices made easy

 CERFACS

[new in 6.0] Spells starting before the season bounds are properly counted



Say we want to compute cdd on a monthly basis...

CDD: number of consecutive where $\text{pr} < 1 \text{ mm/day}$

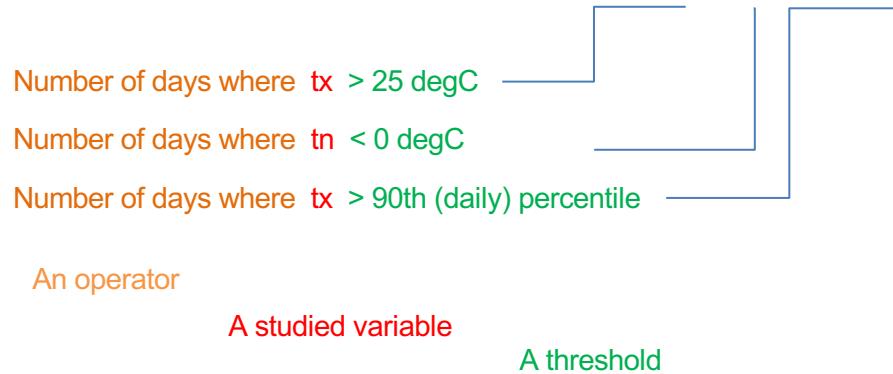
```
cdd = icclim.cdd(in_files=precipitation, slice_mode="month").CDD.compute()
```

Should February cdd be 3 days or 4 ?

With icclim 6 we count spells length before resampling into monthly values,
So CDD in February would be 4 here

[new in 6.0] Generic indices – What ?

What's the difference between SU, FD and Tx90p ?



For our output, we want the right computation to be executed and to have specific metadata to clearly identify :

- What was the computation configuration
- What was the initial data used

[new in 6.0] Generic indices – How ?

```
icclim.index(in_files=climp_file,  
             Operator index_name="count_occurrences",  
             Studied variable var_name="tmax",  
             Threshold threshold="≥ 20 deg_C",  
             ).compute()
```

Not CF proof!

standard_name: number_of_days_with_maximum_air_temperature_above_threshold

long_name: Number of days when maximum air temperature is greater or equal
to 293.15 K for each year.

units: d

cell_methods: time: sum over days

```

thresh = Threshold("≥ 75 period_per",
                    threshold_min_value="1 mm/day",
                    reference_period=['1991-01-01', "31 décembre 2000"])

r75p = icclim.index(climp_file,
                     pr_thresh = Threshold("≥ 75 doy_per", threshold_min_value="1 mm/day")
                     temp_thresh = Threshold("≥ 75 doy_per"))                                CW (ECAD)

c_w = icclim.index(climp_file,
                     pr_thresh = Threshold("≥ 75 doy_per", threshold_min_value="1 mm/day")
                     temp_thresh = Threshold("≥ 75 doy_per")                                     Cold, wet and windy
                     wind_thresh = Threshold("≥ 95 doy_per")                                    (Bretagne?)

c_w_w = icclim.index(ds,
                      index_name="count_occurrences",
                      var_name=["precip", "tmax", "sfcWind"],
                      threshold=[pr_thresh, temp_thresh, wind_thresh])
  
```

[new in 6.0] Generic indices – Operators

count_occurrences	SU, TR, TG90p, CD, R95p SD50cm ...
max_consecutive_occurrence	CSU, CFD, CDD, CWD
sum_of_spell_lengths	WSDI, CSDI
excess	GD4
deficit	HD17
fraction_of_total	R75pTOT, R95pTOT, r99pTOT
maximum	RX1day, TXx, TNx, custom_maximum
minimum	TXn, TNn, user_index - minimum
average	TG, TN, TX, SDII, SD, custom_mean
sum	PRCPTOT, custom_sum
standard_deviation	**new**
max_of_rolling_sum	RX5day, user_indexrolling_sum
min_of_rolling_sum	custom_rolling_sum
min_of_rolling_average	custom_rolling_mean
max_of_rolling_average	custom_rolling_mean
mean_of_difference	DTR
difference_of_extremes	ETR
mean_of_absolute_one_time_step_difference	vDTR
difference_of_means	custom_anomaly

[new in 6.0] Generic indices – Thresholds

Scalar	Threshold(query=" \geq ", value=25, unit="degC")
Day of year percentiles	Threshold(query=" \geq ", value=99, unit="doy_per")
Period percentiles	Threshold(query=" \geq ", value=99, unit="period_per")
Bounded period percentiles.	Threshold("math"> \geq 75 period_per", threshold_min_value="1 mm/day", reference_period=['1991-01-01', "31 décembre 2000"])
Per-grid cell threshold	Threshold(query=" \geq ", value="data.nc", threshold_var_name="tmin", unit="K")
Sequence of scalars	Threshold(query=" \geq ", value=[25, 29], unit="degC")



Thanks !

On behalf of the climate4impact and icclim teams

<https://dev.climate4impact.eu>

<https://icclim.readthedocs.io/en/stable/>

Christian Pagé

Alessandro Spinuso

Abel Aoun

christian.page@cerfacs.fr

alessandro.spinuso@knmi.nl

abel.aoun@cerfacs.fr



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



Our website

<https://is.enes.org/>



Follow us on Twitter !
[@ISENES_RI](https://twitter.com/ISENES_RI)



Contact us at

is-enes@ipsl.fr



Follow our channel
IS-ENES3 H2020

