xclim: xarray-based climate data analytics

Pascal Bourgault, David Huard, Trevor James Smith, Travis Logan, Abel Aoun, Juliette Lavoie, Éric Dupuis, Gabriel Rondeau-Genesse, Raquel Alegre, Clair Barnes, Alexis Beaupré Laperrrière, Sébastien Biner, David Caron, Carsten Ehbrecht, Jeremy Fyke, Tom Keel, Marie-Pier Labonté, Ludwig Lierhammer, Jwen-Fai Low, Jamie Quinn, Philippe Roy, Dougie Squire, Ag Stephens, Maliko Tanguy, and Christopher Whelan

1 Ouranos Consortium, Montréal, Québec, Canada 2 Centre européen de recherche et de formation avancée en calcul scientifique (CERFACS), Toulouse, France 3 University College London (UCL), London, United Kingdom 4 Deutsches Klimarechenzentrum (DKRZ), Hamburg, Germany 5 Environment and Climate Change Canada (ECCC), Canada 6 Helmholtz-Zentrum Hereon, Geesthacht, Germany 7 Commonwealth Scientific and Industrial Research Organisation (CSIRO), Adelaide, Australia 8 Centre for Environmental Data Analysis (CEDA), United Kingdom 9 UK Centre for Ecology & Hydrology (UKCEH), Wallingford, United Kingdom 10 Jakarta, Montréal, Québec, Canada 11 Institut de recherche d’Hydro-Québec (IREQ), Varennes, Québec, Canada 12 Hydro-Québec, Québec, Canada 13 Grantham Institute for Climate Change and the Environment, Imperial College London, London, United Kingdom 14 Independent Researcher, Canada 15 Independent Researcher, United States

Corresponding author

DOI: 10.21105/joss.05415
Software
- Review
- Repository
- Archive

Summary

xclim is a Python library that enables computation of climate indicators over large, heterogeneous data sets. It is built using xarray objects and operations, can seamlessly benefit from the parallelization handling provided by dask, and relies on community conventions for data formatting and metadata attributes. xclim is meant as a tool to facilitate both climate science research and the delivery of operational climate services and products. In addition to climate indicator calculations, xclim also includes utilities for bias correction and statistical adjustment, ensemble analytics, model diagnostics, data quality assurance, and metadata standards compliance.

Statement of need

Researchers and climate service providers analyse data from large ensembles of Earth System Model (ESM) simulations. These analyses typically include model comparisons with observations, bias-correction and statistical adjustment, computation of various climate indicators and diagnostics, and ensemble statistics. As the number of models contributing to these ensembles grows, so does the complexity of the code required to deal with model idiosyncrasies, outlier detection, unit conversion, etc. In addition, growing ensemble sizes and advancements in the spatiotemporal resolution of ESMs further raises the computational costs of running those analyses. xclim is designed to meet the operational needs of climate service providers by offering algorithms for over 150 climate indicators, multiple downscaling algorithms, ensemble statistics, and other associated utilities.

The development of xclim started in 2018 at Ouranos, a consortium on regional climatology and adaptation to climate change based in Montréal, Québec, from the need to deliver data for a pan-Canadian atlas of climate indicators. In-house specialists at Ouranos had different implementations for the same indicators, and there was a desire to adopt a common library
that would tie together investments in research and development with operational production capabilities. At the time, the package that was closest to meeting these requirements was icclim (Pagé, Aoun, et al., 2022), a library developed within the context of the European Climate Assessment & Dataset project, whose purpose was to monitor and analyze changes in climate extremes. It was not, however, designed to be easily extensible, and we believed the indicators they offered could be written more succinctly and computed more efficiently by relying on objects and primitives from xarray (Hoyer & Hamman, 2017), with distributed computation and scheduling via dask (Team, 2016). xclim started as a reimplementation of icclim with an xarray backend, drawing inspiration from projects like MetPy (May et al., 2022), and eventually grew to include other algorithms routinely used in climate data analysis, both simple and complex.

xclim is intended to be one component in a larger software ecosystem for climate data analysis. Other libraries often used in tandem with xclim are clisops (Stephens et al., 2022), a spatiotemporal subsetting and averaging library (originally a fork of xclim’s subsetting module), and xESMF (Zhuang et al., 2022), a PANGEO-developed library for spatial regridding.

**Key Features**

**Climate indicators calculations**

An Indicator class is built around a compute function defining a climate indicator. It performs health checks on input data (units, time frequency, outlier detection), handles missing values, and assigns attributes to the output, complying to the Climate and Forecast (CF) metadata Conventions (Hassell et al., 2017). Indicators can be customized using a context manager, by class inheritance, or through a YAML file—the latter allowing for the creation of custom collections of indicators for batch processing.

**Statistical adjustment and bias correction**

The xclim.sdba subpackage provides different algorithms to adjust the distribution of simulated variables to observed variables. It adopts a train / adjust paradigm, where corrections are first calculated, then applied to the target data or saved for later use. Most methods support additive or multiplicative corrections, different time groupings (seasonal, monthly, or daily with a rolling window). Correction factors can be interpolated between time groupings to avoid discontinuities in the corrected data.

**Ensemble analysis**

The xclim.ensembles subpackage bundles utilities to facilitate the analysis of results from multiple models. It includes functions to reduce the ensemble size using clustering algorithms, metrics of ensemble robustness, and significance of climate change signals.

**Spatial analogs**

The xclim.analogs subpackage offers tools to find spatial climate analogs using a selection of distribution comparison algorithms.

**Internationalization tools**

In order to better support the international community, xclim provides methods for building dynamic multilingual metadata translations via the xclim.corelocales module. While French is currently the only translation officially supported, other languages can be extended via JSON-based indicator field mappings.

Other utilities

Among the various modules within xclim, a few merit explicit mention:

- \texttt{xclim} is core component of Finch \cite{Finch}, a server hosting climate analytics services behind a Web Processing Services (WPS) interface. Finch itself is part of the computational backend of ClimateData.ca, an online data portal to access, visualize and analyze climate data over Canada. \texttt{xclim} is now also a core component of icclim from version 5.0, which itself is used in the climate4impact project \cite{climate4impact}. The statistical adjustment tools of \texttt{xclim} are also being used by the Climate Impact Lab to downscale and adjust CMIP6 simulations on HPCs for climate impact studies.

Projects using \texttt{xclim}

\texttt{xclim} is core component of Finch \cite{Finch}, a server hosting climate analytics services behind a Web Processing Services (WPS) interface. Finch itself is part of the computational backend of ClimateData.ca, an online data portal to access, visualize and analyze climate data over Canada. \texttt{xclim} is now also a core component of icclim from version 5.0, which itself is used in the climate4impact project \cite{climate4impact}. The statistical adjustment tools of \texttt{xclim} are also being used by the Climate Impact Lab to downscale and adjust CMIP6 simulations on HPCs for climate impact studies.

Acknowledgements

\texttt{xclim} was developed thanks to the financial and strategic contributions of the Canadian Center for Climate Services and the Ouranos Consortium. We also acknowledge the contributions from Marie-Pier Labonté, David Caron, Jwen-Fai Low, Raquel Alegre, Clair Barnes, Sébastien Biner, Philippe Roy, Carsten Ehbrecht, Tom Keel, Ludwig Lierhammer, Jamie Quinn, Dougie Squire, Ag Stephens, Maliko Tanguy, Jeremy Fyke, Yannick Rousseau, Christian Jaouvin, and Christopher Whelan, as well as our user base who regularly provide valuable bug reports and enhancement / support requests.

References


