

Broadening access to supercomputers for CMIP6 and CORDEX multimodel comparisons

EGU2020 Sharing Geoscience Online, moreno@dkrz.de

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK
FOR EARTH SYSTEM MODELLING



IS-ENES3 is funded by the EU
Horizon 2020 Research and
Innovation program under the
grant agreement No 824084





New free of charge server-side data-near computing service

The ENES e-infrastructure project IS-ENES3 [H2020-INFRAIA](#) addresses the Coupled Model Intercomparison Project ([CMIP](#)) challenges by developing, documenting, and deploying new and advanced models and tools, standards, and services to deal with unprecedented data volumes and model complexity.

Data-near server-side processing capabilities at the IS-ENES3 High Performance Computing centers are now made accessible to a broader user community.

New free of charge server-side data-near computing service

client-side workflow



~~search & discovery~~



~~heavy data
download~~



~~desktop-based data
analysis~~



server-side workflow



search & discovery

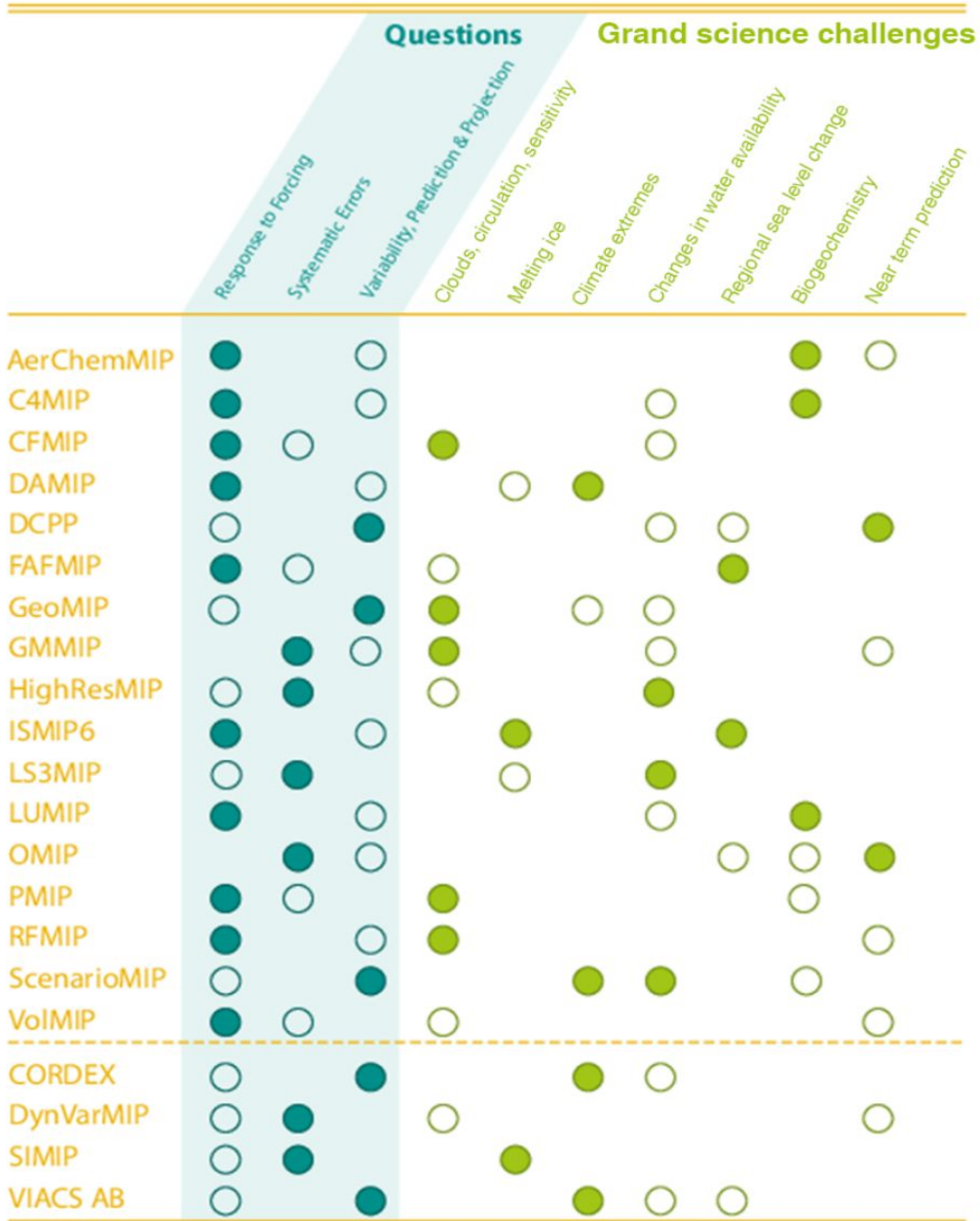


data near parallel
data analysis



lightweight data
download





→ to directly access 7+ petabytes of CMIP5, CMIP6, and CORDEX data, the complete available pool, continuously growing and maintained

Eyring et al. (2016)

<https://portal.enes.org/data/data-metadata-service/analysis-platforms>

→ to minimize data transfer because the supercomputer is also holding the data and/or can replicate them



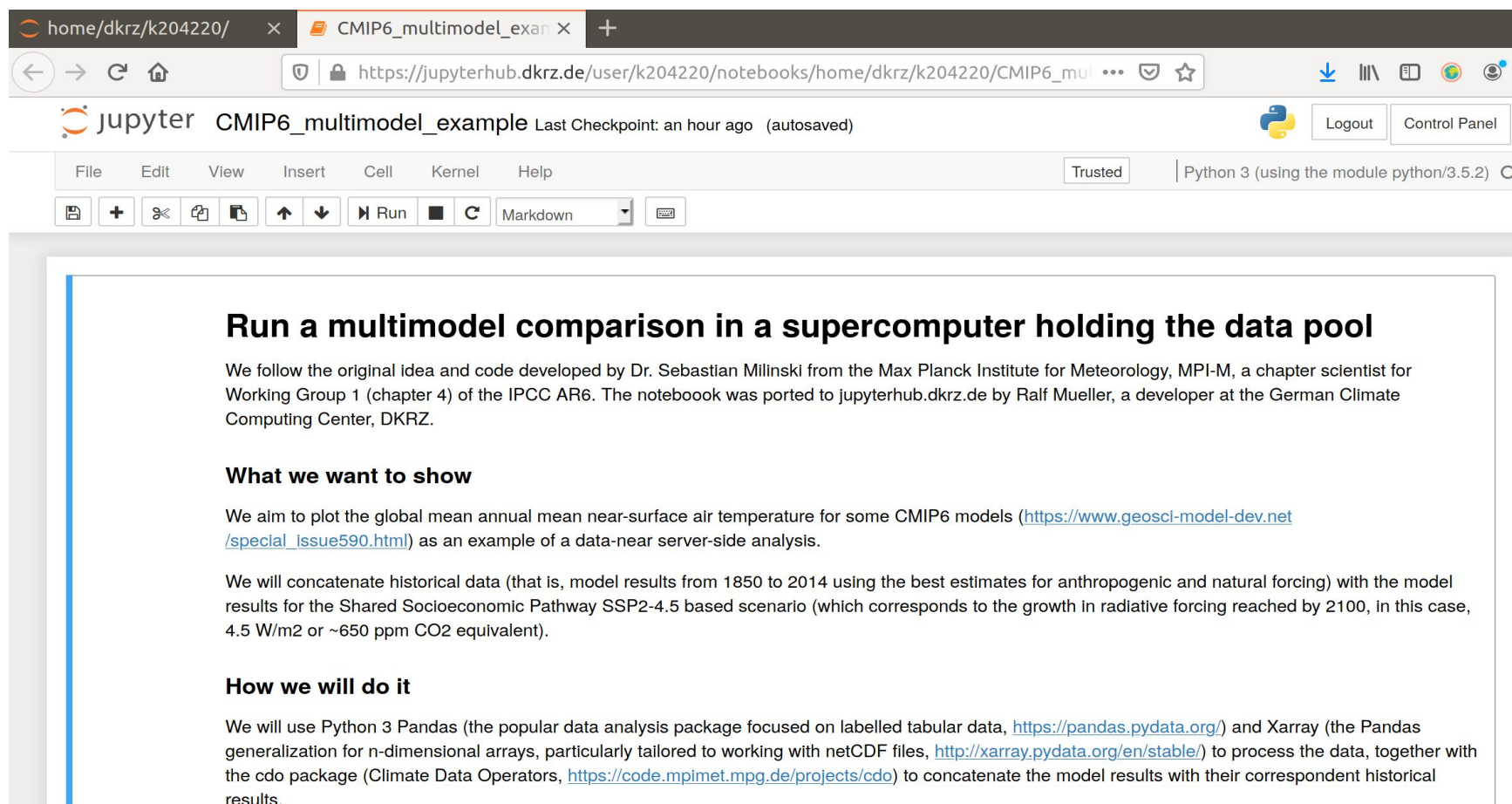
→ to speed up the computation running your multimodel analysis in one of the IS-ENES3 world-class high performance computers



→ to run your own scripts and Jupyter notebooks

Example running at <https://jupyterhub.dkrz.de>

Repo of test cases: https://gitlab.dkrz.de/k204220/multimodel_comparison_CMIP6/



Run a multimodel comparison in a supercomputer holding the data pool

We follow the original idea and code developed by Dr. Sebastian Milinski from the Max Planck Institute for Meteorology, MPI-M, a chapter scientist for Working Group 1 (chapter 4) of the IPCC AR6. The notebook was ported to jupyterhub.dkrz.de by Ralf Mueller, a developer at the German Climate Computing Center, DKRZ.

What we want to show

We aim to plot the global mean annual mean near-surface air temperature for some CMIP6 models (https://www.geosci-model-dev.net/special_issue590.html) as an example of a data-near server-side analysis.

We will concatenate historical data (that is, model results from 1850 to 2014 using the best estimates for anthropogenic and natural forcing) with the model results for the Shared Socioeconomic Pathway SSP2-4.5 based scenario (which corresponds to the growth in radiative forcing reached by 2100, in this case, 4.5 W/m² or ~650 ppm CO₂ equivalent).

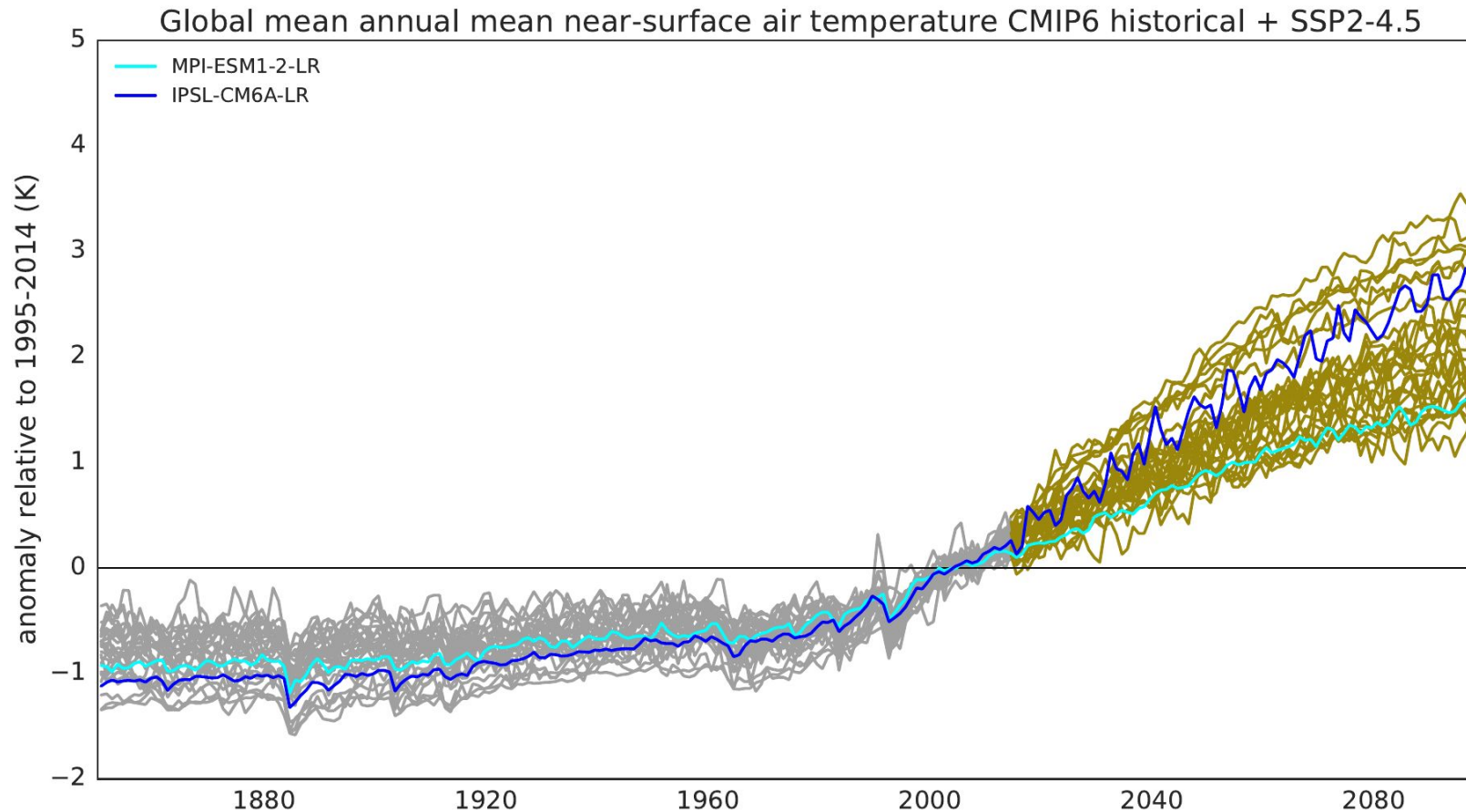
How we will do it

We will use Python 3 Pandas (the popular data analysis package focused on labelled tabular data, <https://pandas.pydata.org/>) and Xarray (the Pandas generalization for n-dimensional arrays, particularly tailored to working with netCDF files, <http://xarray.pydata.org/en/stable/>) to process the data, together with the cdo package (Climate Data Operators, <https://code.mpimet.mpg.de/projects/cdo>) to concatenate the model results with their correspondent historical results.

→ to run your own scripts and Jupyter notebooks

Example running at <https://jupyterhub.dkrz.de>

Repo of test cases: https://gitlab.dkrz.de/k204220/multimodel_comparison_CMIP6/



**OPEN CALL
APPLY HERE**

moreno@dkrz.de <https://portal.enes.org/data/data-metadata-service/analysis-platforms>