

EGU2020-9686, updated on 02 Jul 2022

<https://doi.org/10.5194/egusphere-egu2020-9686>

EGU General Assembly 2020

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



## Python-based Multidimensional and Parallel Climate Model Data Analysis in ECAS

**Regina Kwee**<sup>1</sup>, Tobias Weigel<sup>1</sup>, Hannes Thiemann<sup>1</sup>, Karsten Peters<sup>1</sup>, Sandro Fiore<sup>2</sup>, and Donatello Elia<sup>2</sup>

<sup>1</sup>DKRZ, Data Management, Germany (kwee@dkrz.de)

<sup>2</sup>CMCC

This contribution highlights the Python xarray technique in context of a climate specific application (typical formats are NetCDF, GRIB and HDF).

We will see how to use in-file metadata and why they are so powerful for data analysis, in particular by looking at community specific problems, e.g. one can select purely on coordinate variable names. ECAS, the ENES Climate Analytics Service available at Deutsches Klimarechenzentrum (DKRZ), will help by enabling faster access to the high-volume simulation data output from climate modeling experiments. In this respect, we can also make use of “dask” which was developed for parallel computing and can smoothly work with xarray. This is extremely useful when we want to exploit fully the advantages of our supercomputer.

Our fully integrated service offers an interface via Jupyter notebooks (ecaslab.dkrz.de). We provide an analysis environment without the need of costly transfers, accessing CF standardized data files and all accessible via the ESGF portal on our nodes (esgf-data.dkrz.de). We can analyse the data of e.g. CMIP5, CMIP6, Grand Ensemble and observation data. ECAS was developed in the frame of European Open Source Cloud (EOSC) hub.