



---

# CMIP6 Data Usage Report

A Fresh Eyes on CMIP project

J. LAVOIE, E. ZIEGLER, A. CARRERIC, G.  
CHELLINI, A. DUFFEY, R. JUNOD

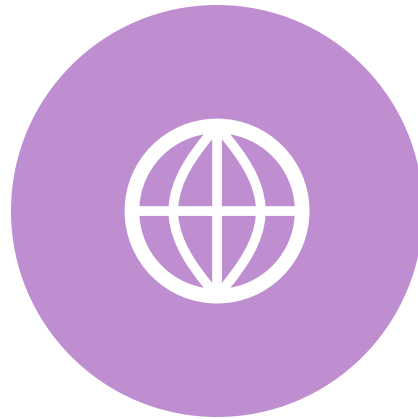
FEOC WORKSHOP, MAY 21 2025

# CMIP6 Data

---



16 PB OF DATA  
PRODUCED IN CMIP6



DOWNLOADED  
AROUND THE WORLD



MULTIPLE SOURCES

## Goal of the project

---

Analyse existing statistic from the CMCC Dashboard

---

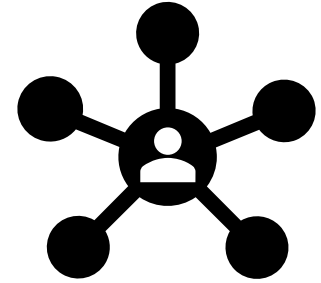
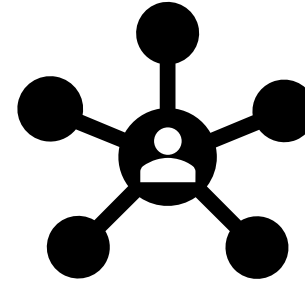
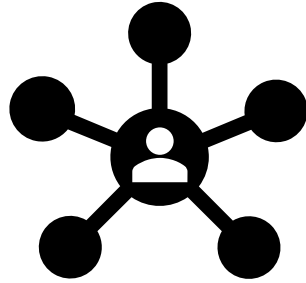
Gather statistic from other sources

---

Develop a report on global pattern usage of CMIP6 data

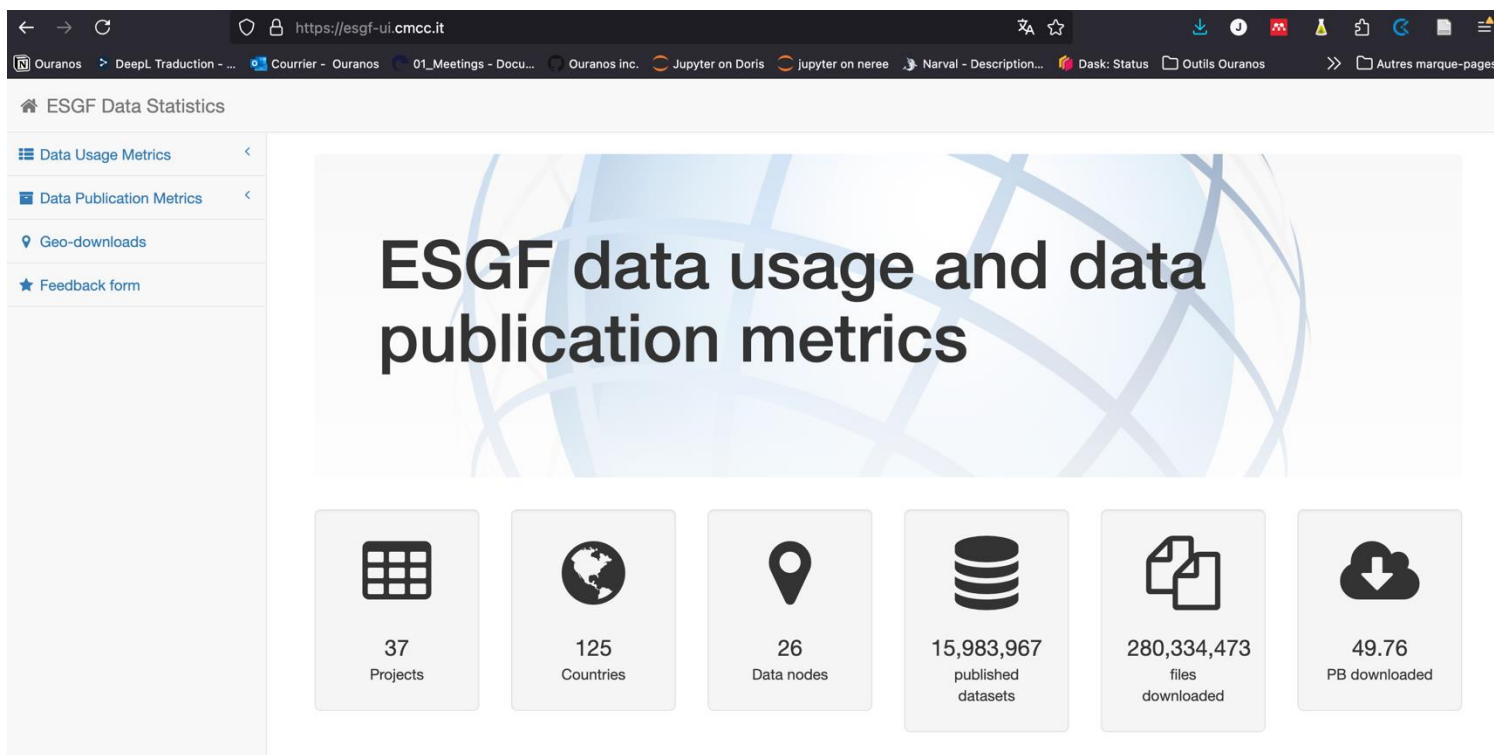
---

# Sources



- ESGF
- Pangeo curated subset on the cloud
  - Amazon bucket
  - Google bucket
- Copernicus Climate Change Service
- Copies on institutional servers
  - JASMIN/UK, DKRZ/Germany, NCI/Australia, DIAS/Japan, CAFE/China, NERSC/USA, NCAR/Glade/US
- Downstream datasets
  - dynamically downscalled (CORDEX)
  - Statistically downscalled

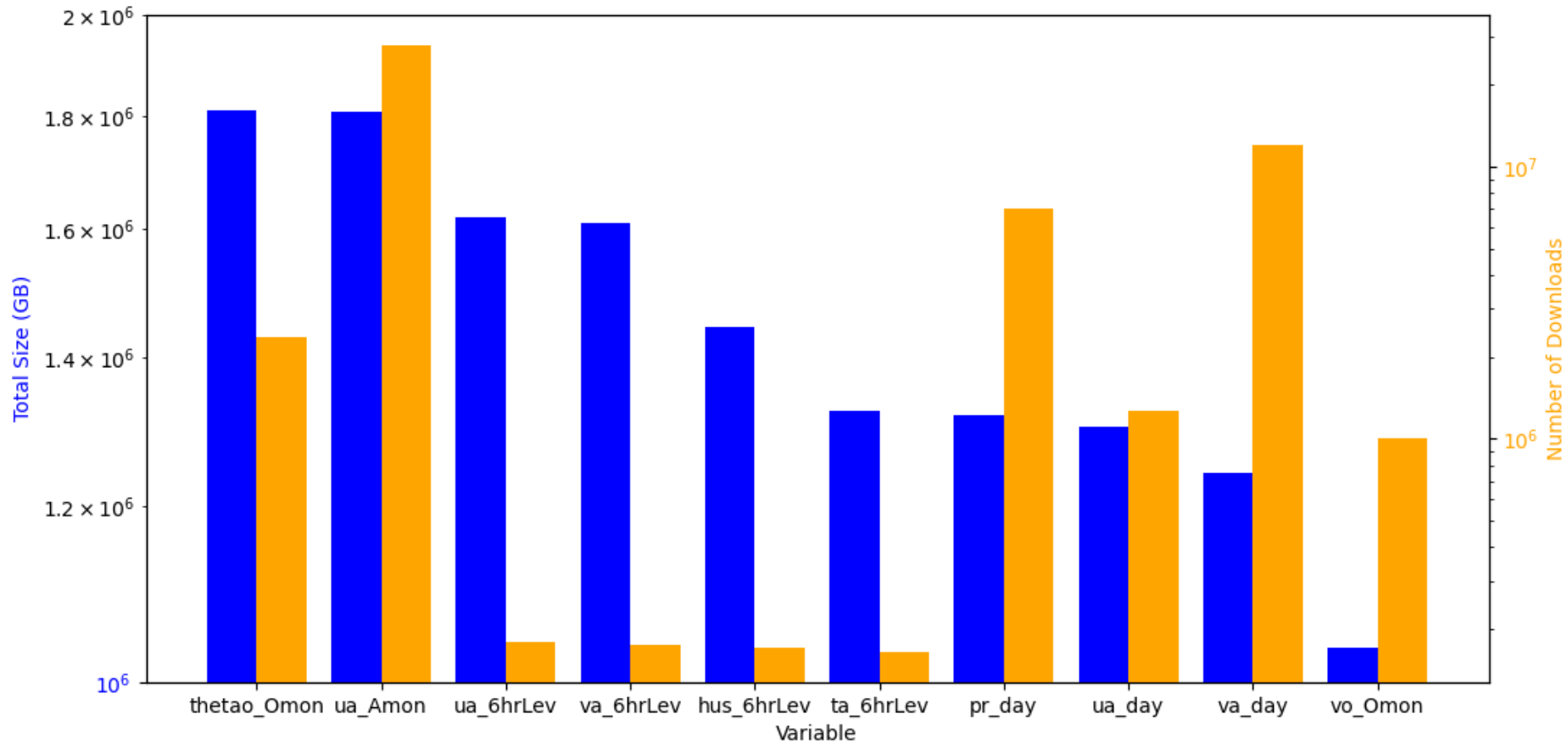
# CMCC Dashboard



# Work in progress

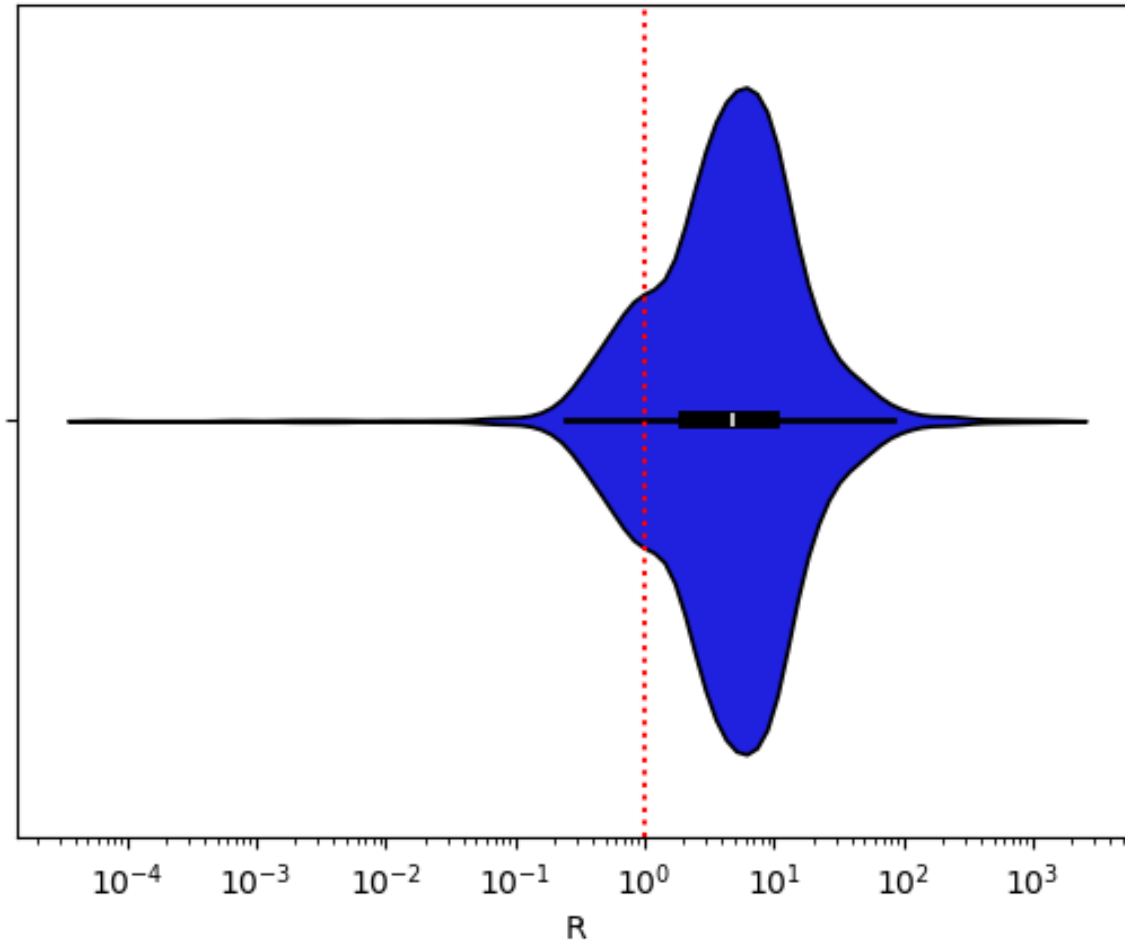


# Top variables



- Most used variables will depend on the metric used.
- Dimensionality, grids, availability and frequency will interfere with pure usage.

# Least used variables



$$R = \frac{\text{Total size of downloads}}{\text{Total size of uploads}}$$

- 249 unique variables under 1
- Average is 21.

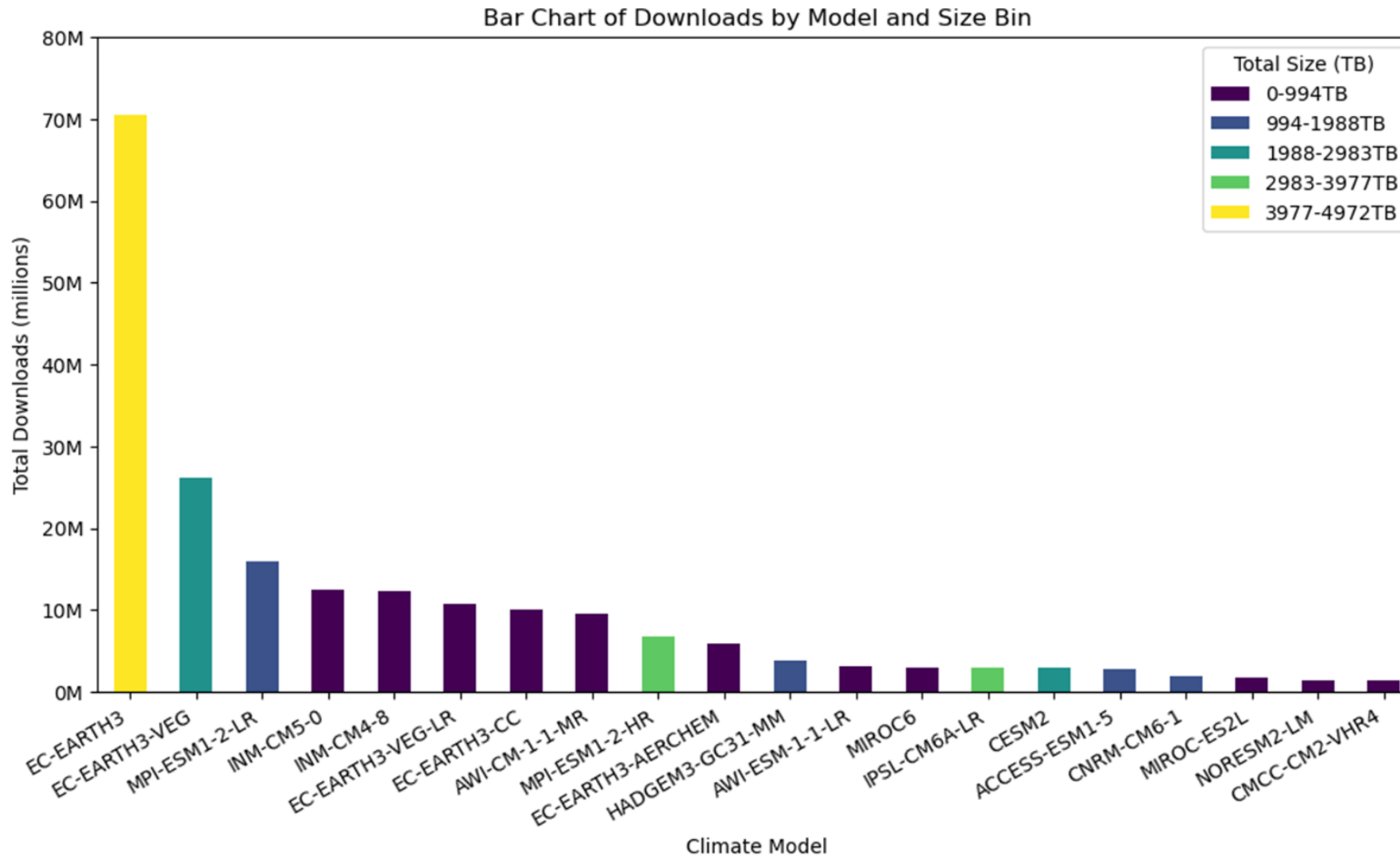
---

# Clean data

- In the CMIP6 controlled vocabulary, there are 43 accepted tables and 2062 accepted individual variables. The CMCC dashboard has gathered valid data on 1811 variables.
- Some data was not valid. E.g. `tos_Amon` or `tas_Amon` with frequency `day`.
- 782 830 erroneous files (~ 1000 TB) downloaded by users
  - QA/QC project for CMIP7 could help avoid this issue.



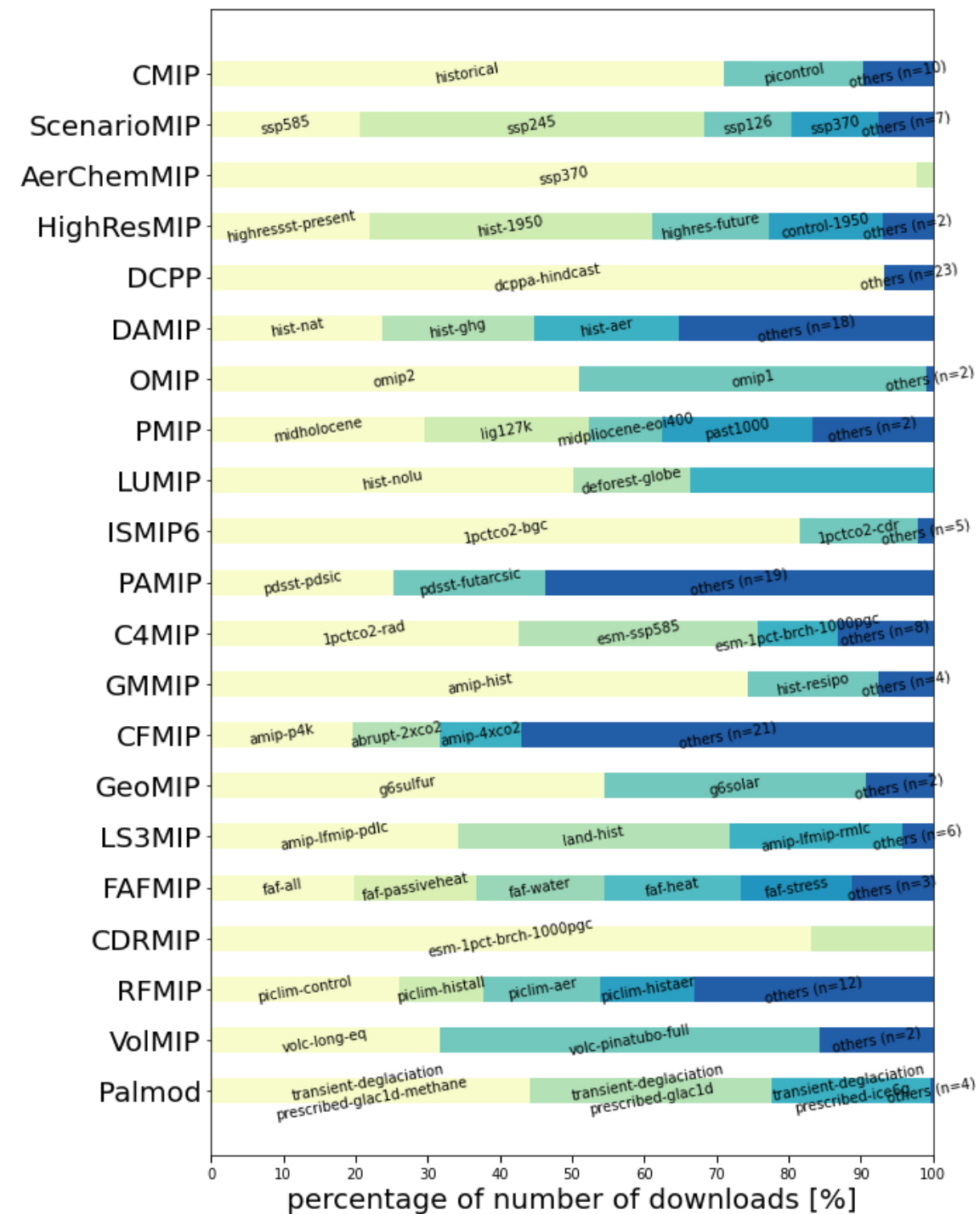
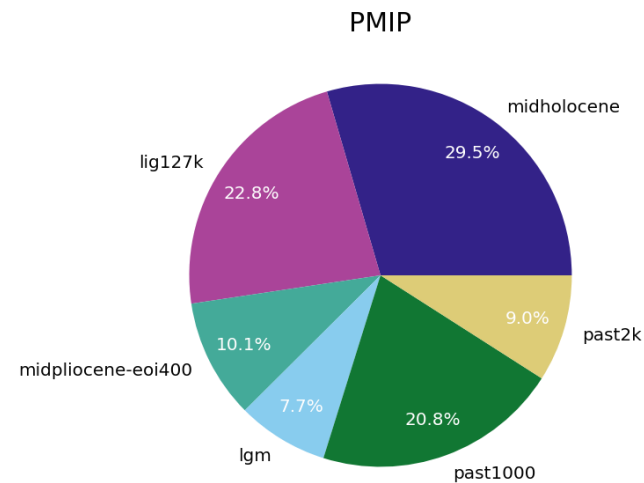
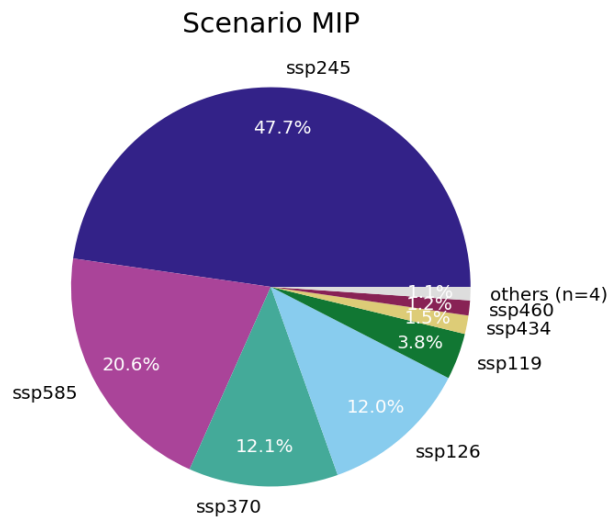
# Most downloaded models



- EC-Earth uploads a large number of files.

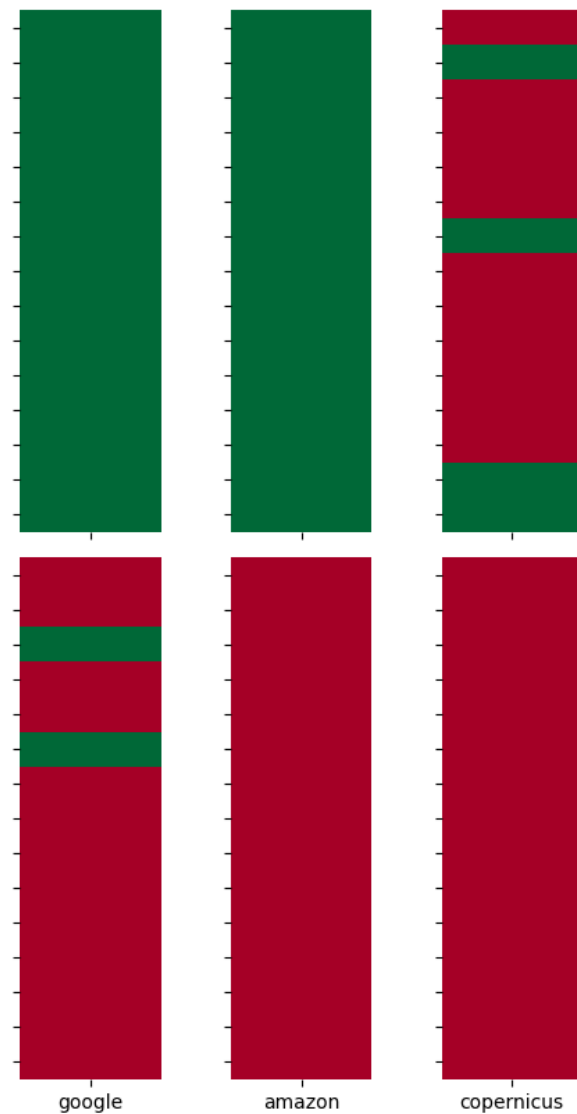
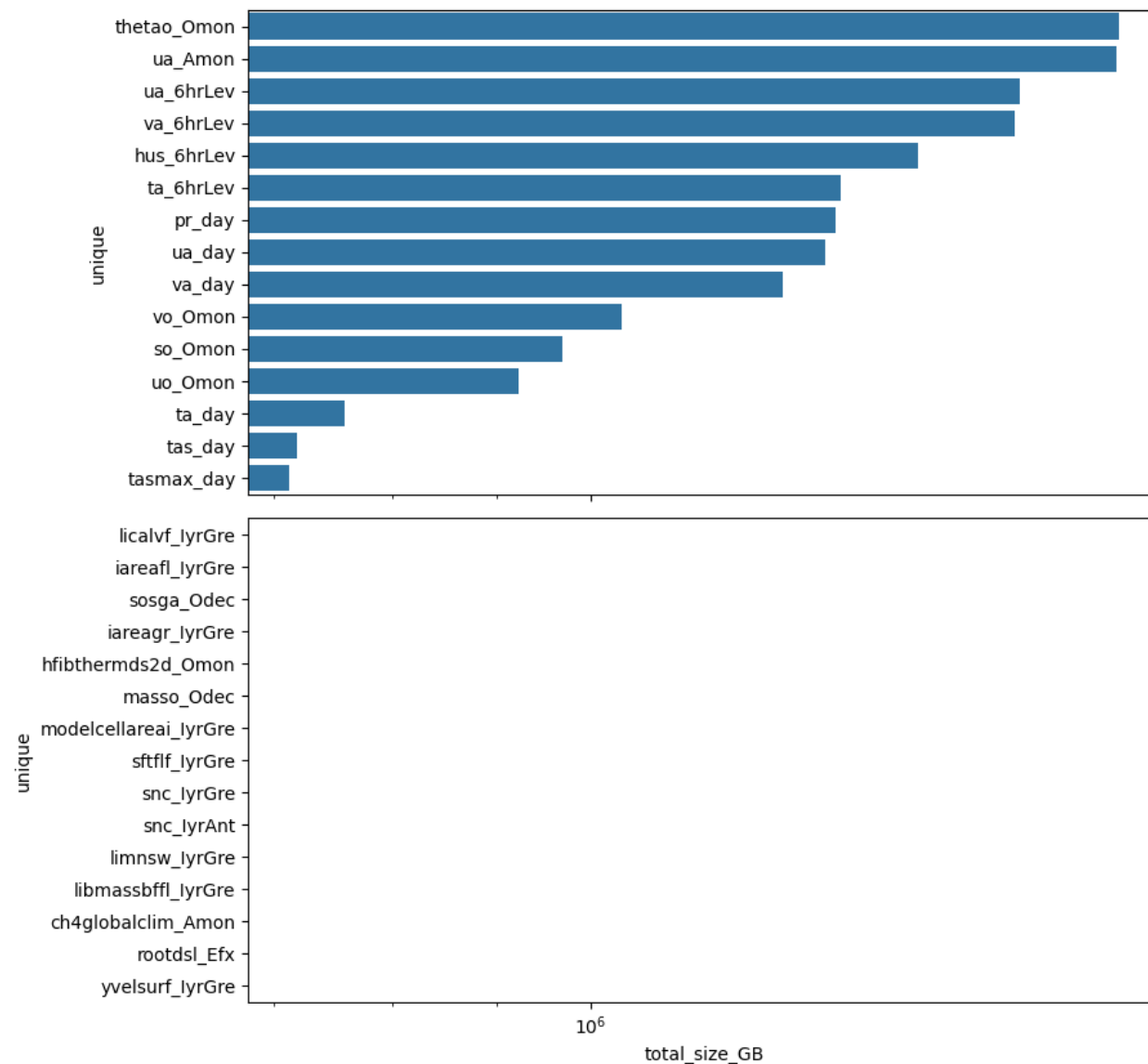
# Experiment per MIP

- Popularity of experiments on each MIPs.
- For some MIPs individual experiments dominate, while it is more balanced for other MIPs



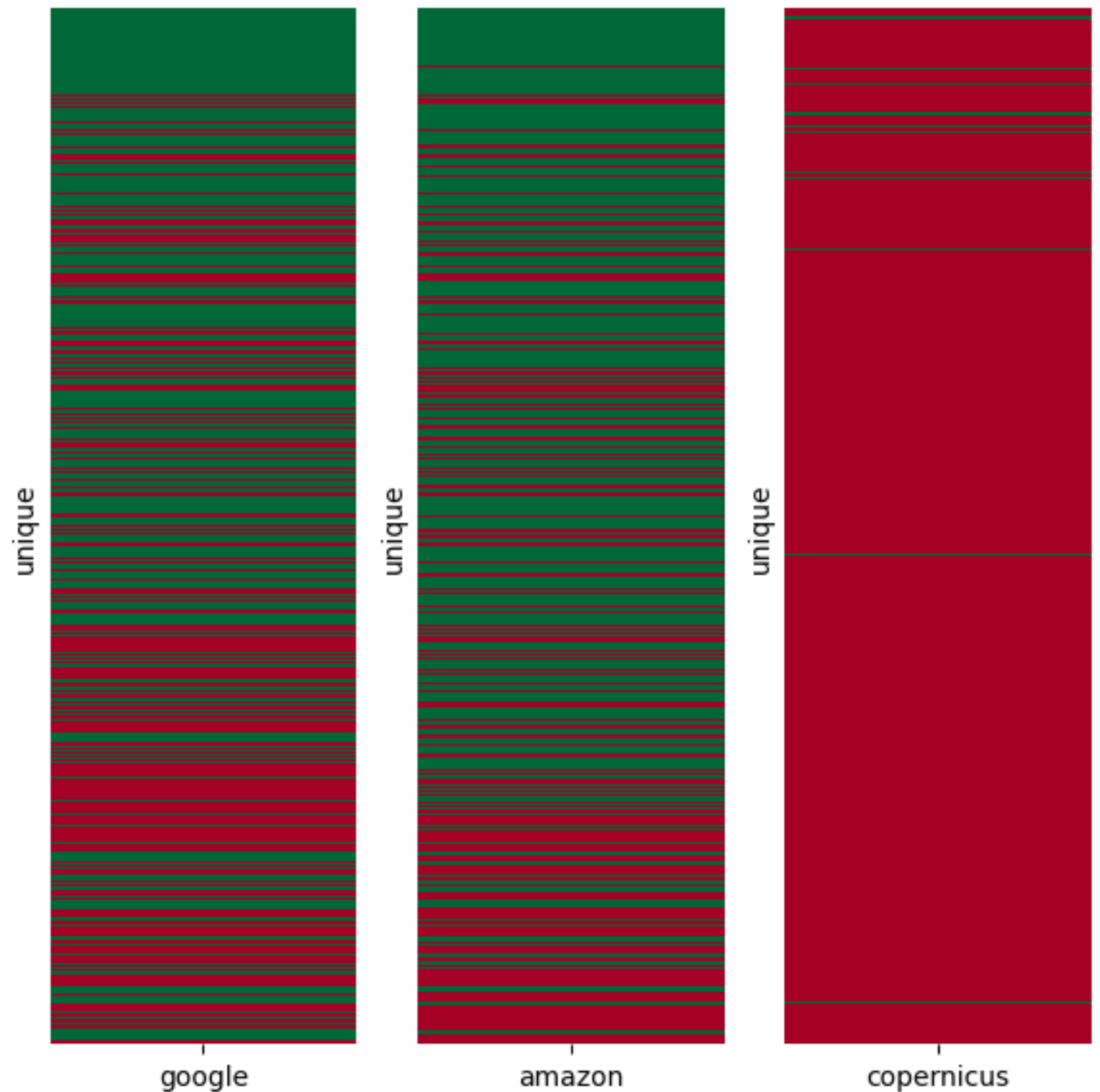
# Other sources

- Open Call:  
[lavoie.juliette@ouranos.ca](mailto:lavoie.juliette@ouranos.ca)
- Availability compared usage on esgf:



## Other sources

- Open Call:  
[lavoie.juliette@ouranos.ca](mailto:lavoie.juliette@ouranos.ca)
- Availability  
compared  
usage on esgf





# Next steps

- More complete data in collaboration with CMCC (if possible)
- Data from citation
- Geography data
- Disentangle usage from other patterns
- Expand to other sources
- Document challenges and caveats on the results
- Complete the report



# Thank you!

- Paul Durack and Matthew Mizielinski from the WIP
- Alessandra Nuzzo, Sandro Luigi Fiore, Fabrizio Antonio, Paola Nassisi from the CMCC
- Past and present member of the infrastructure task team and of the project