HPC Challenges for CMIP: lessons from CMIP6 and potential next steps

Jean-François Lamarque
Climate & Global Dynamics Laboratory (CGD)
National Center for Atmospheric Research
Boulder, CO

May 9, 2022

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

Background on CMIP

- The objective of the Coupled Model Intercomparison Project (CMIP) is to better **understand past, present and future climate changes** arising from natural, unforced variability or in response to changes in radiative forcing in a **multi-model context**.
- Started in 1995 -> first set of **common** experiments: comparing the model response to an **idealized** forcing - a constant rate of increase which was accomplished using a CO_2 increase of 1% per year compounded.
- Dual role of CMIP: **scientific enterprise** and **support for policymakers**.
- CMIP has played a **critical role** in supporting the IPCC Assessment Reports, as data & papers from CMIP simulations have been used extensively in the ARs.
- All CMIP activities are overseen by a coordinated pair of subcommittees: the CMIP Panel and the WGCM Infrastructure Panel (WIP).

Background on CMIP

- The objective of the Coupled Model Intercomparison Project (CMIP) is to better **understand past, present and future climate changes** arising from natural,

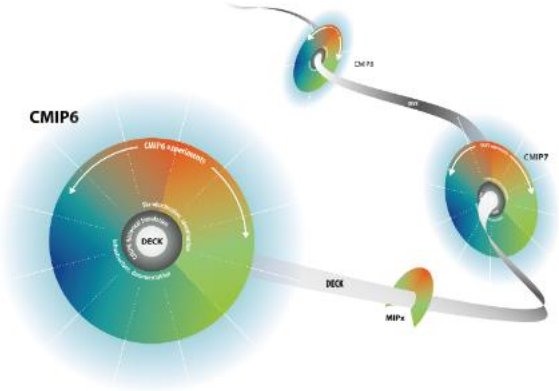
Based on the 2018 WCRP review, “A conservative estimate of the national investments in CMIP6 places their value in excess of US$3 bn, based on scientists’ time to develop and run the models and to design the experiments, and the supercomputing costs to deliver the simulations”

- Dual role of CMIP: **scientific enterprise** and **support for policymakers**.
- CMIP has played a **critical role** in supporting the IPCC Assessment Reports, as data & papers from CMIP simulations have been used extensively in the ARs.
- All CMIP activities are overseen by a coordinated pair of subcommittees: the CMIP Panel and the WGCM Infrastructure Panel (WIP).

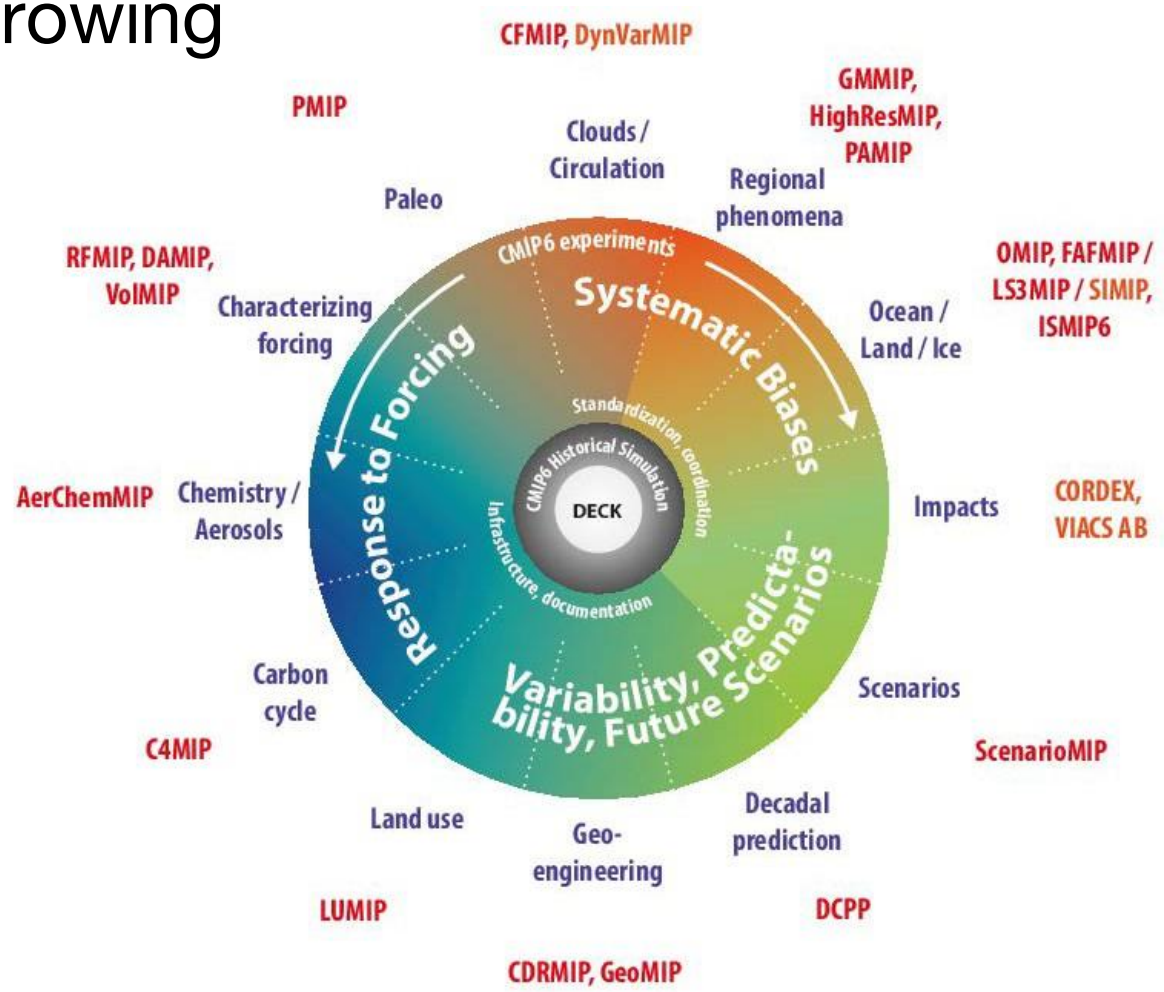
CMIP6

- Number of models/model versions on ESGF: 114 (44 centers)
- Generated data: > 11 PB and growing

Diagnostic, Evaluation, and Characterization of Klima (DECK)



- Pre-industrial control
- 1%CO2
- 4xCO2
- AMIP

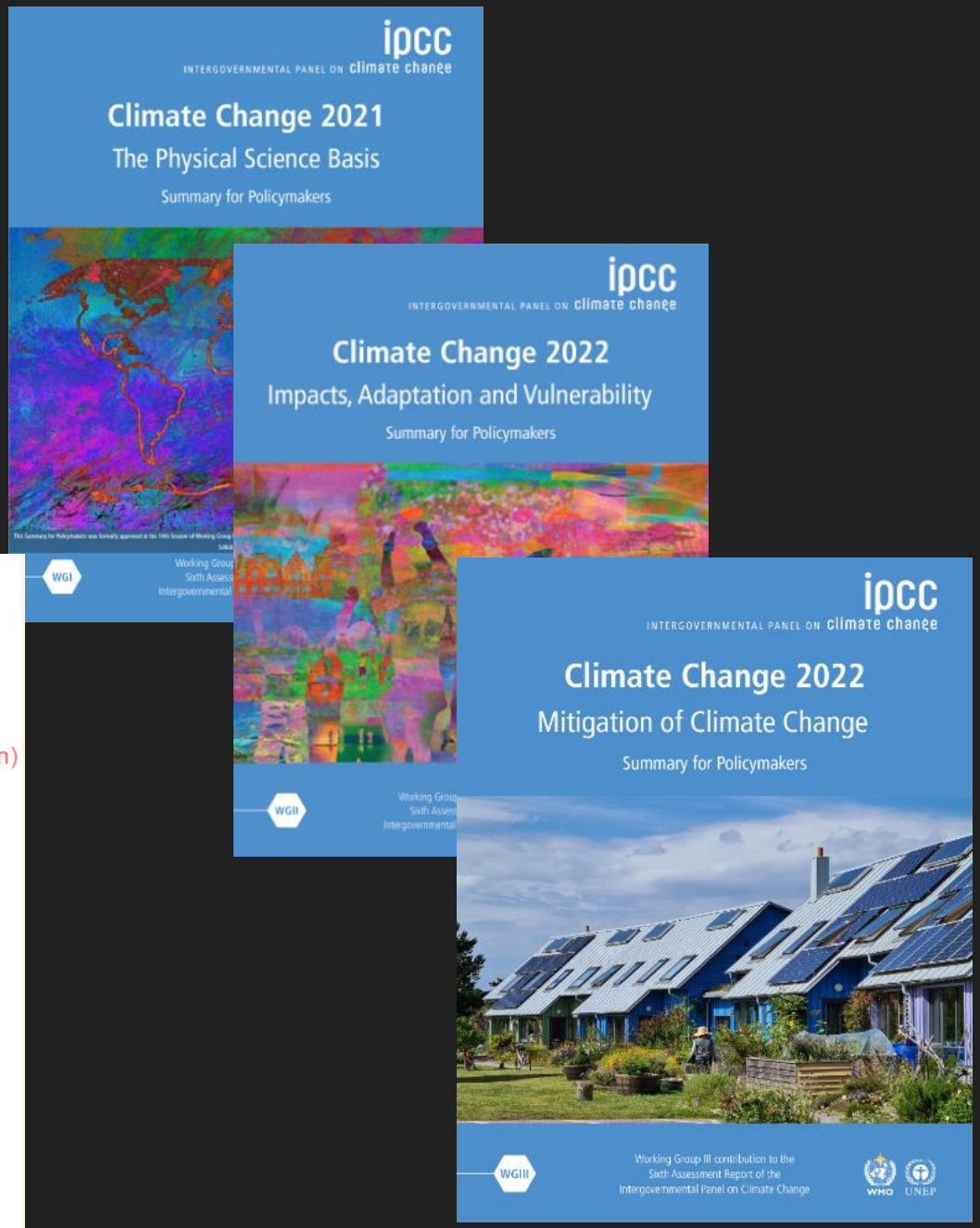


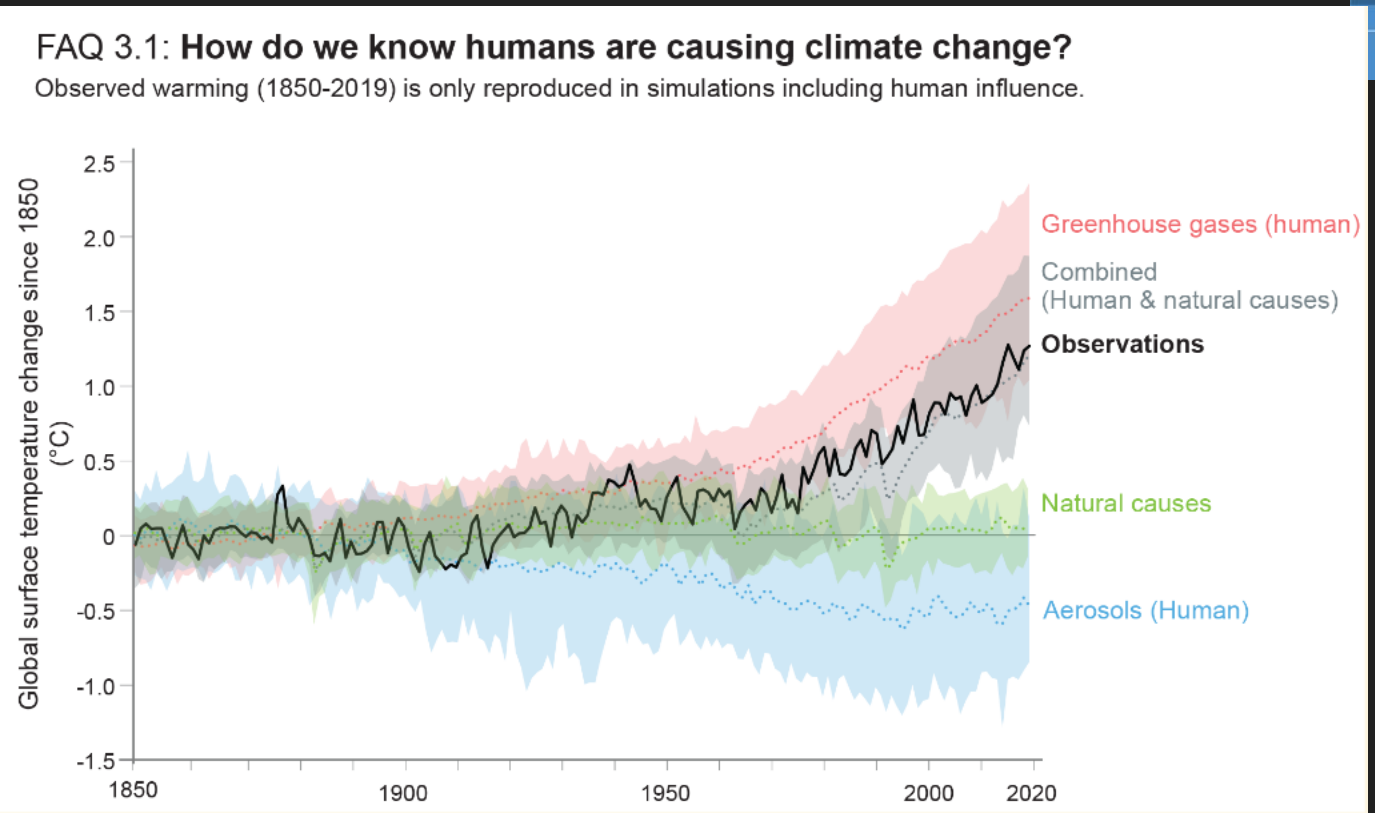
Three IPCC climate reports just released

August 2021

February 2022

April 2022





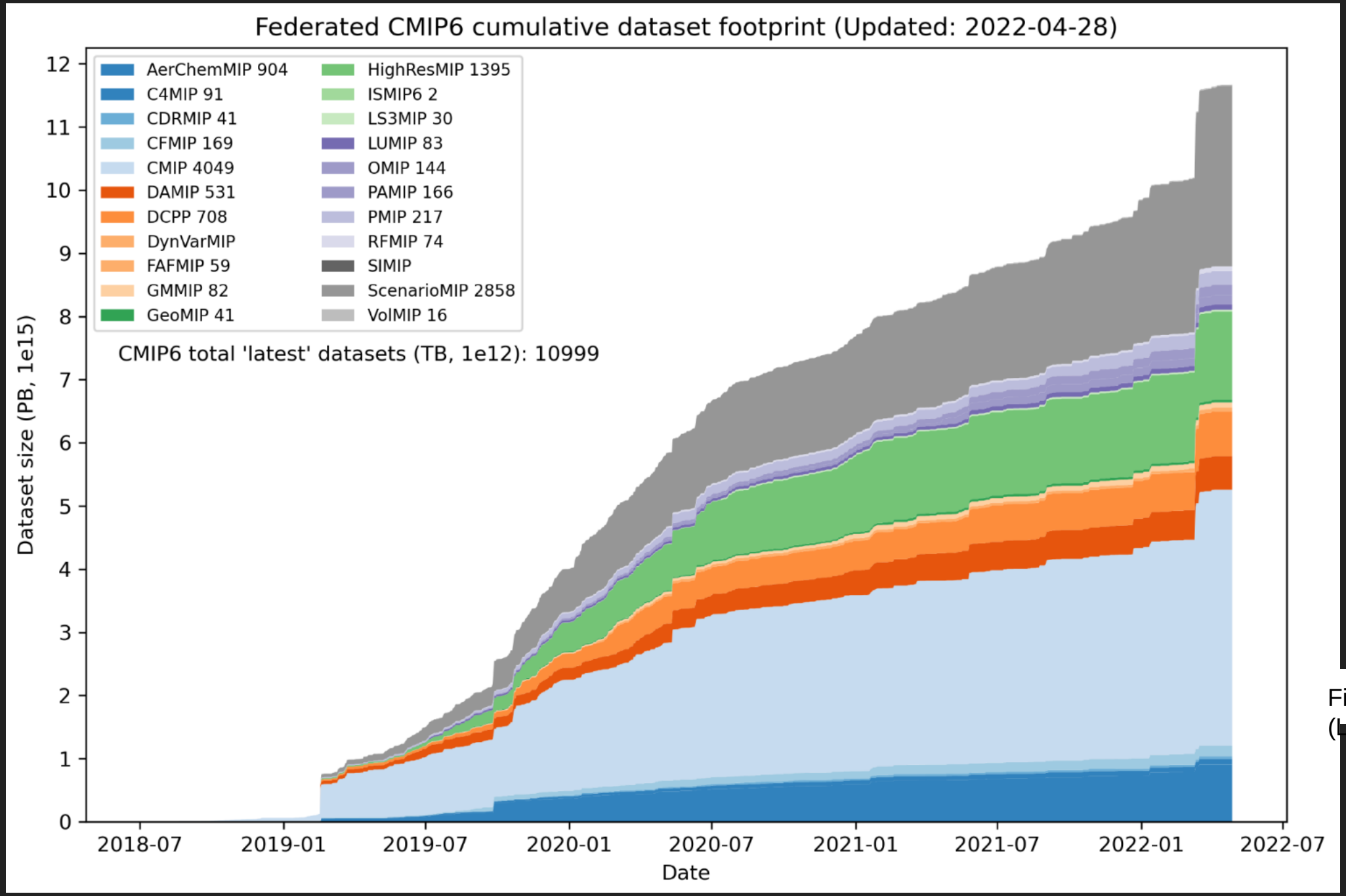
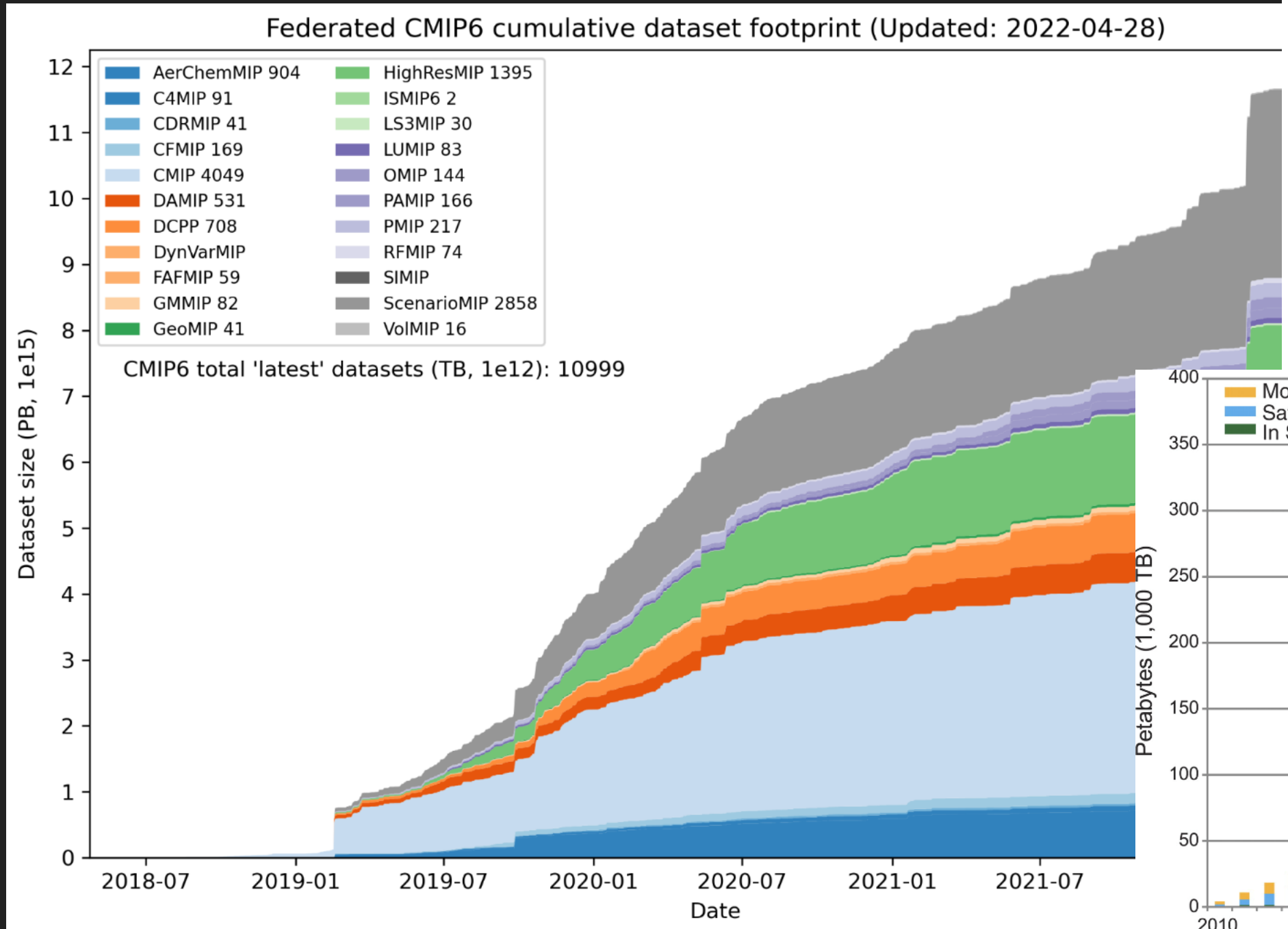
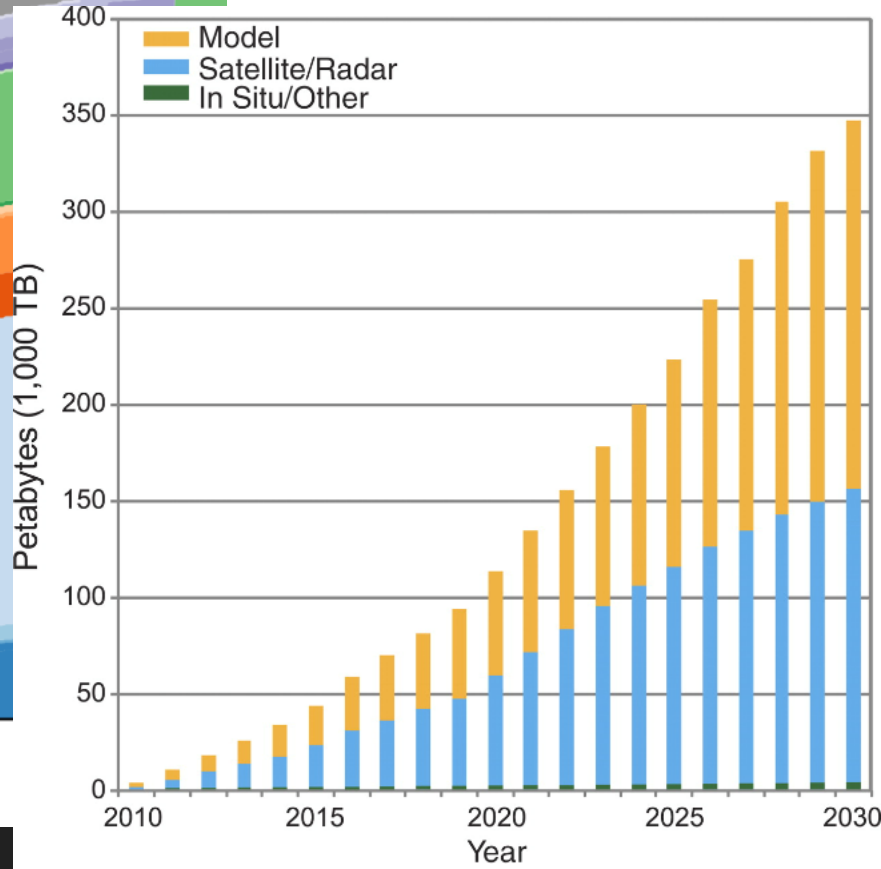


Figure from P. Durack (LLNL)

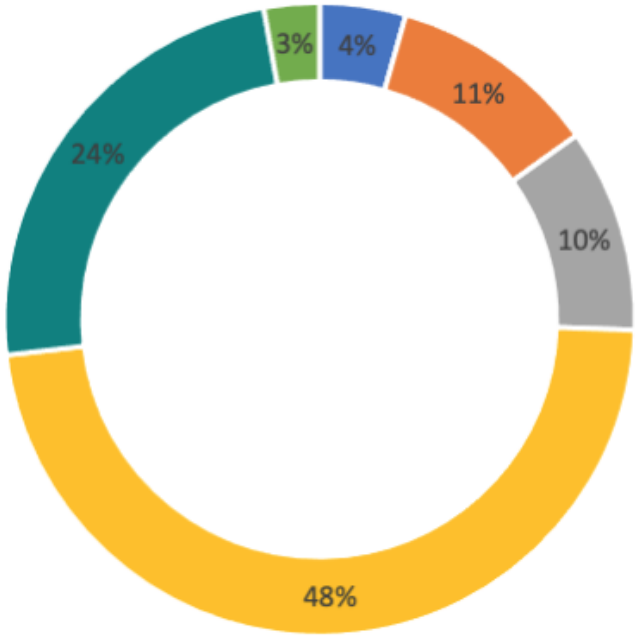


An increasingly daunting aspect of having tens and eventually hundreds of petabytes of climate data openly available for analysis is how to actually look at and use the data, all the while understanding uncertainties (Overpeck et al., 2011)

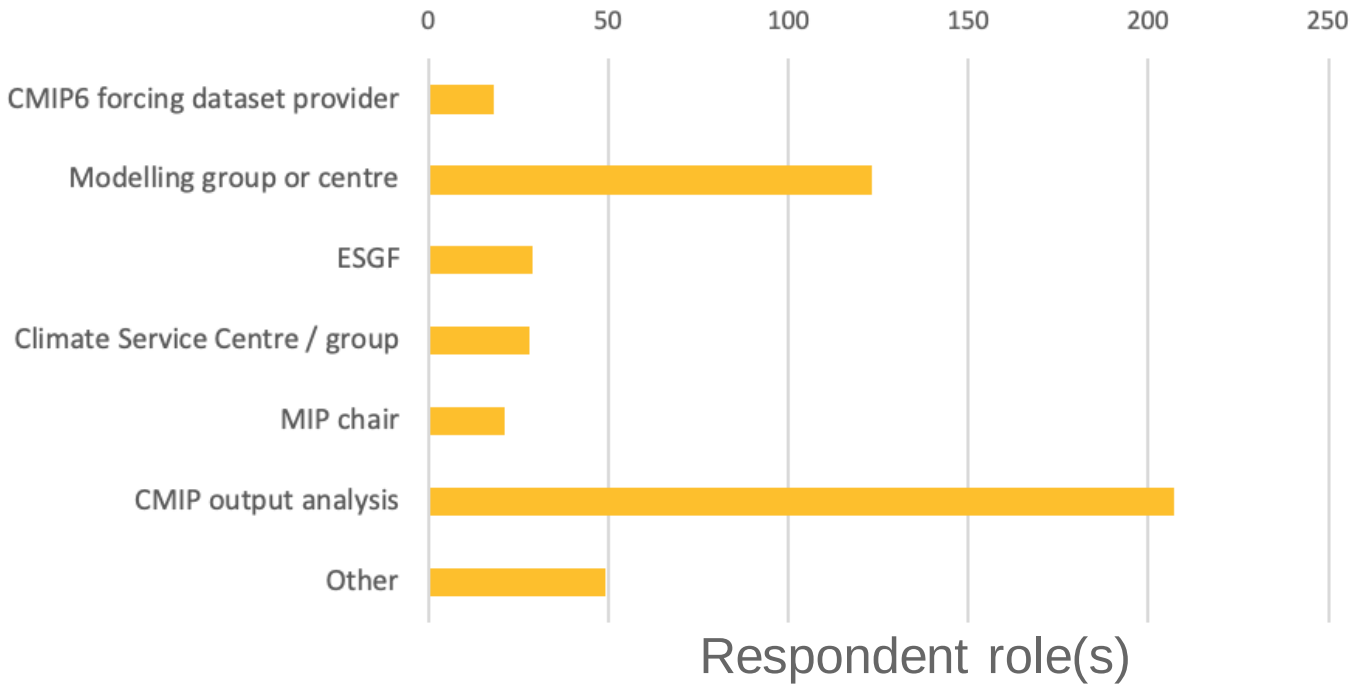


CMIP6 Survey

- Number of models/model versions on ESGF: 114 (44 centers)
- Generated data: > 11 PB and growing



Fraction of respondents

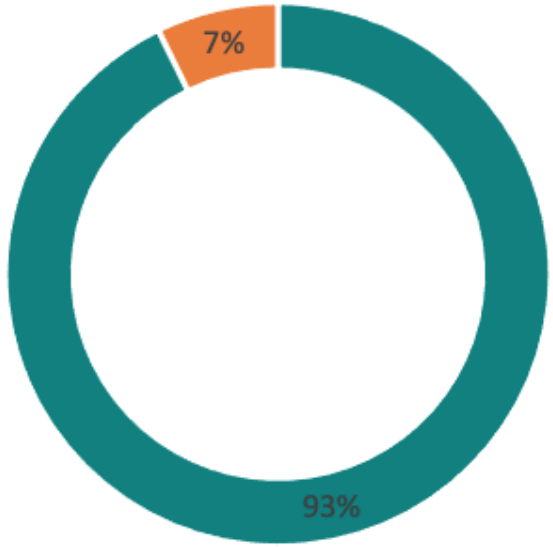


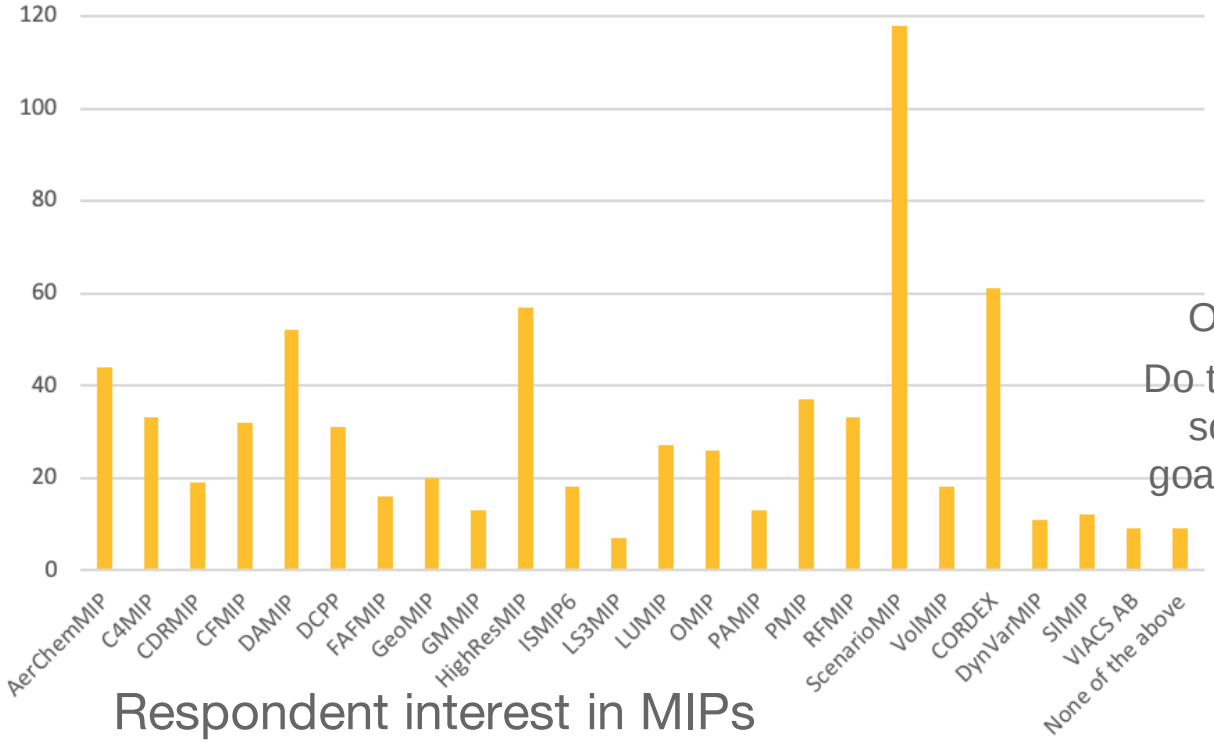
Respondent role(s)

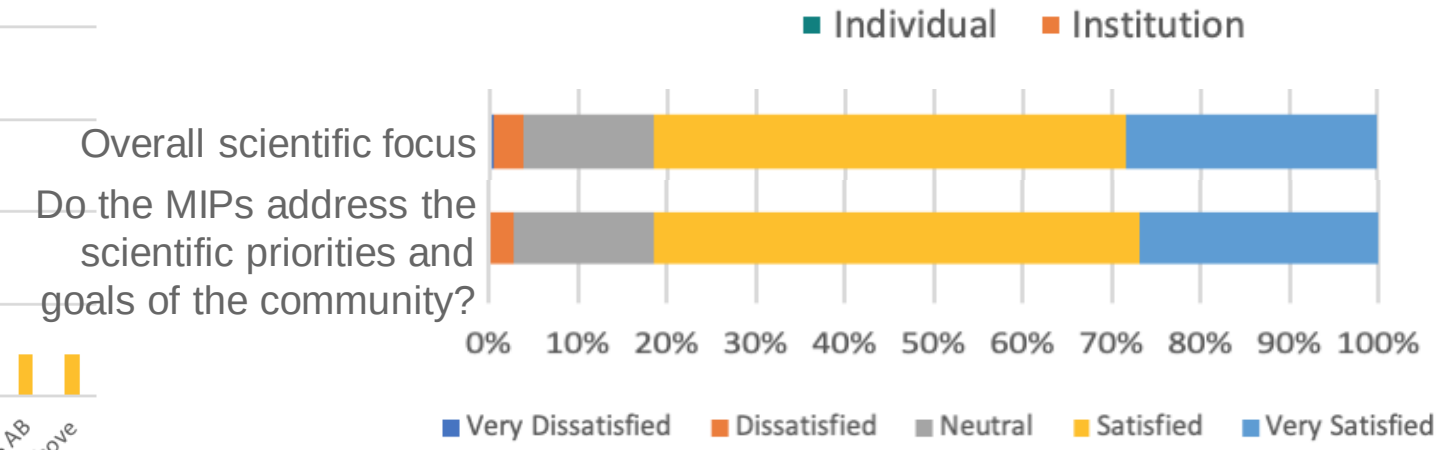
Figures from Eleanor O'Rourke (CMIP Project Office). Preliminary data, do not cite.

CMIP6 Survey

- Number of models/model versions on ESGF: 1
- Generated data: > 11 PB and growing







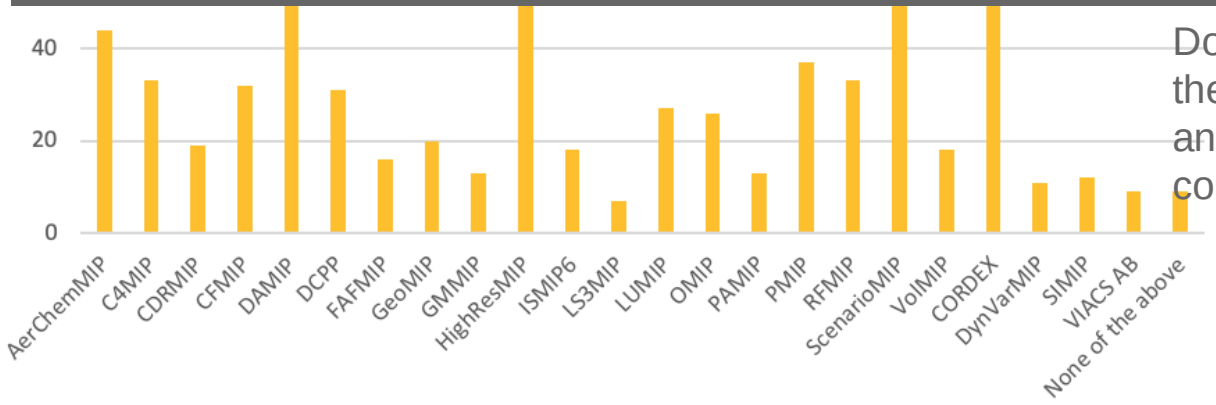
Figures from Eleanor O'Rourke (CMIP Project Office). Preliminary data, do not cite.

Key points

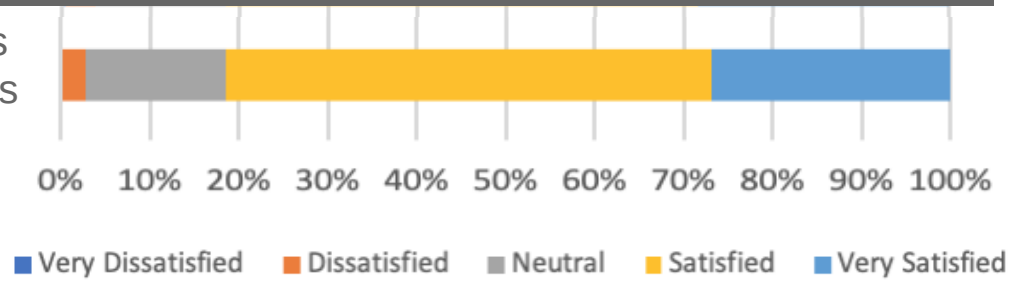
- CMIP is central to the use of multi-model approaches to answer scientific and societally-relevant questions, exemplified by its role in assessment reports
- CMIP6 survey indicates a strong support for the overall scientific focus
- CMIP6 represents a LARGE investment worldwide, with a growing participation of non-OECD countries

➢ How do we optimize the HPC resources?

➢ How do we maximize the relevance of the simulations and the generated data to support the broad international community needs, research and stakeholders?



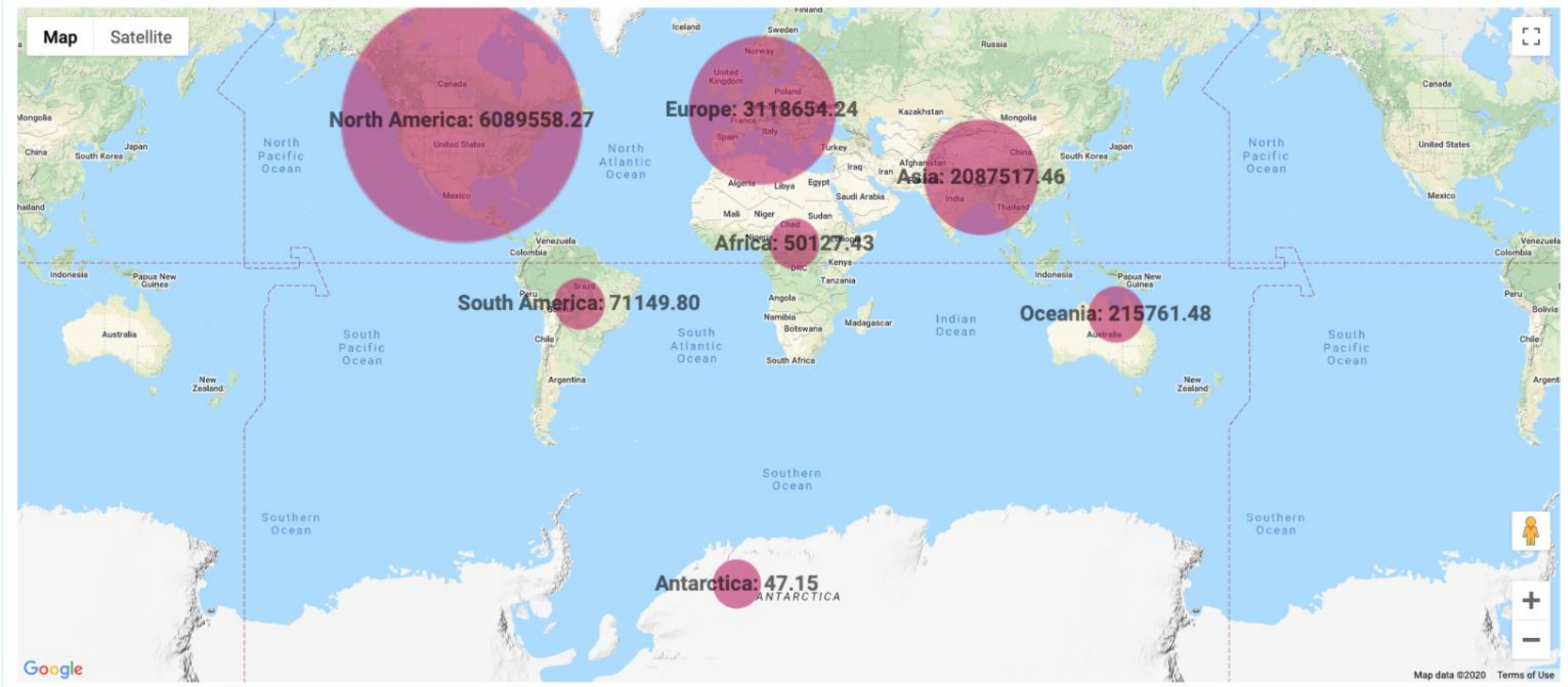
Do the MIPs address the scientific priorities and goals of the community?



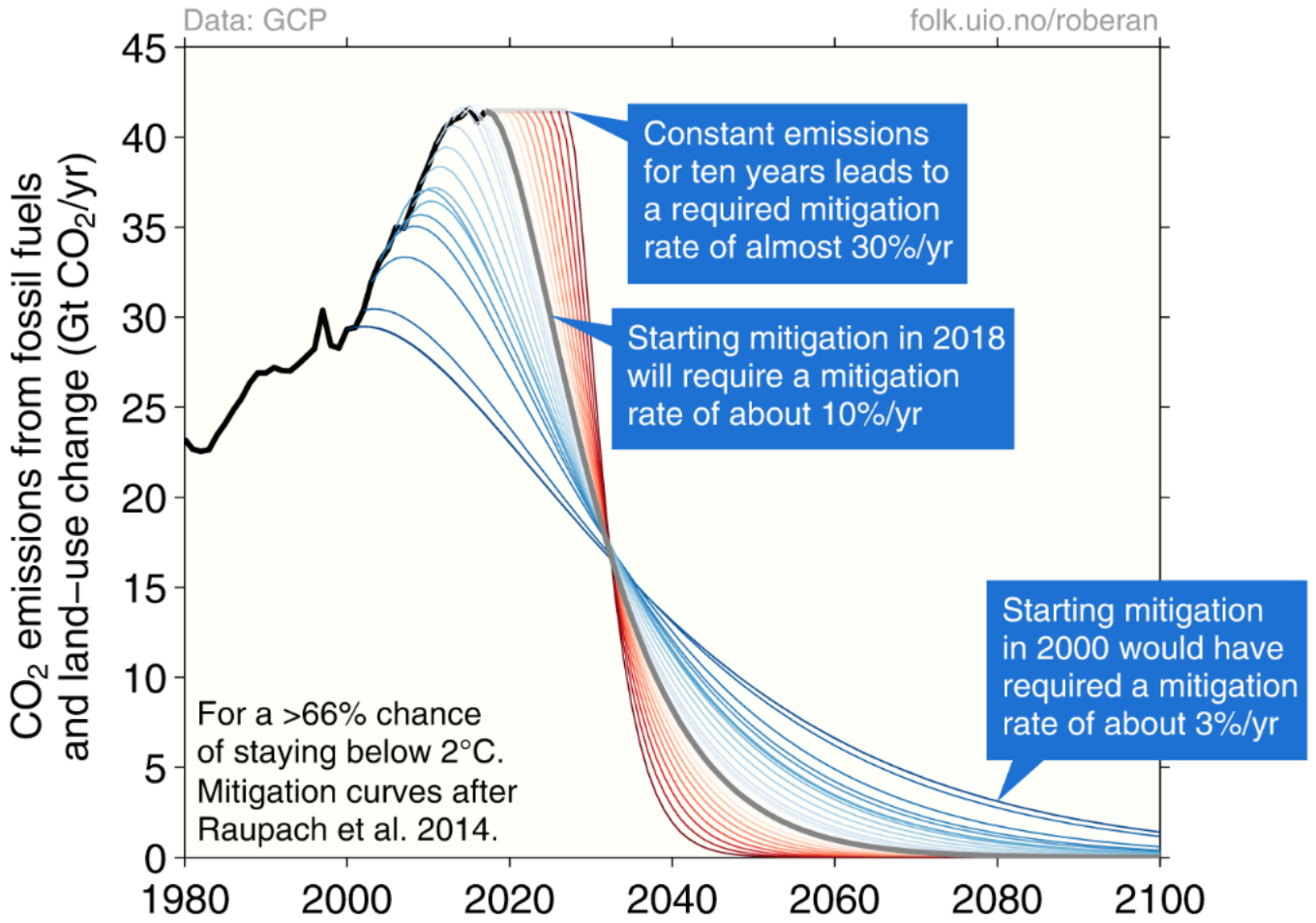
Figures from Eleanor O'Rourke (CMIP Project Office). Preliminary data, do not cite.

Are we reaching out the interested communities?

CMIP6 downloaded data volume by continent



What we need to do for below 2C target…



Strong mitigation needed globally:

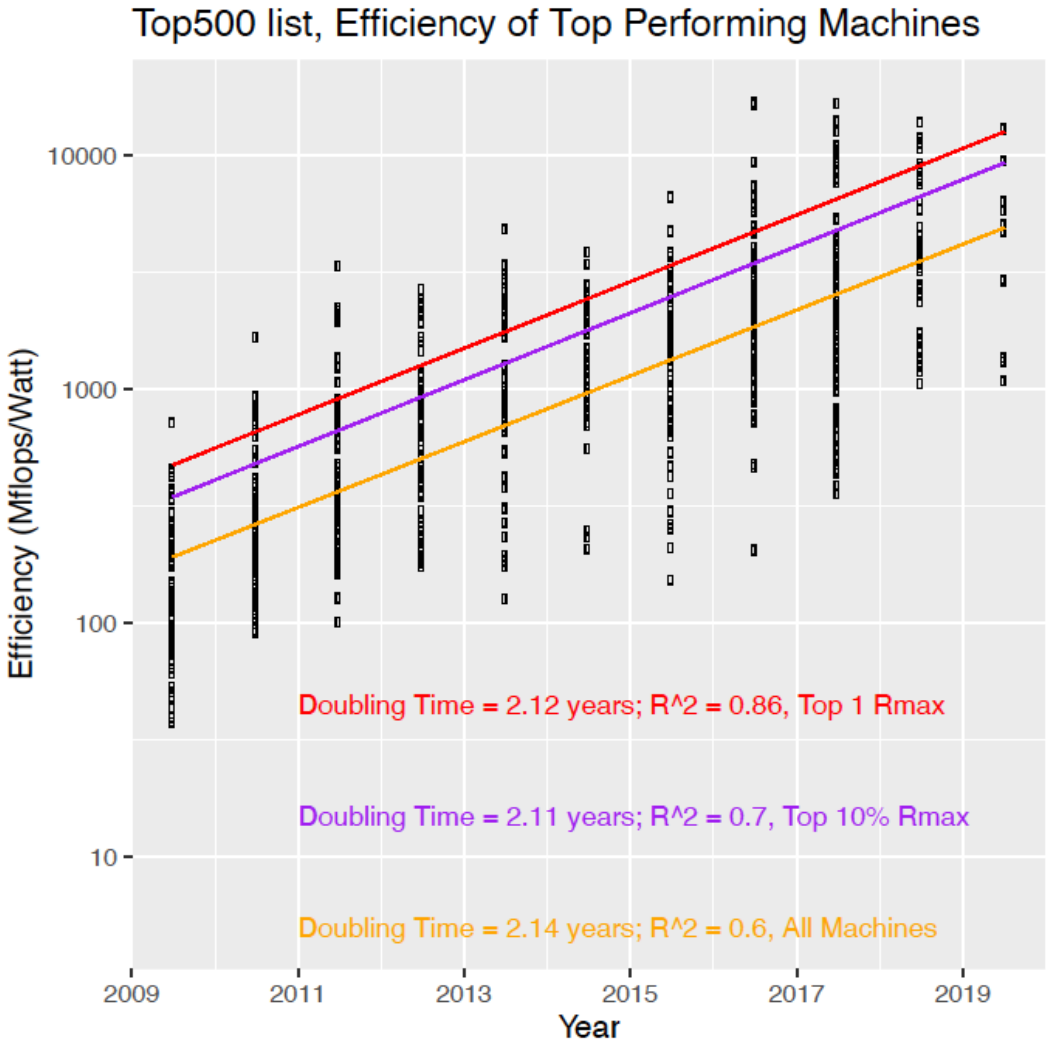
about 50% by 2030, 100% by 2050

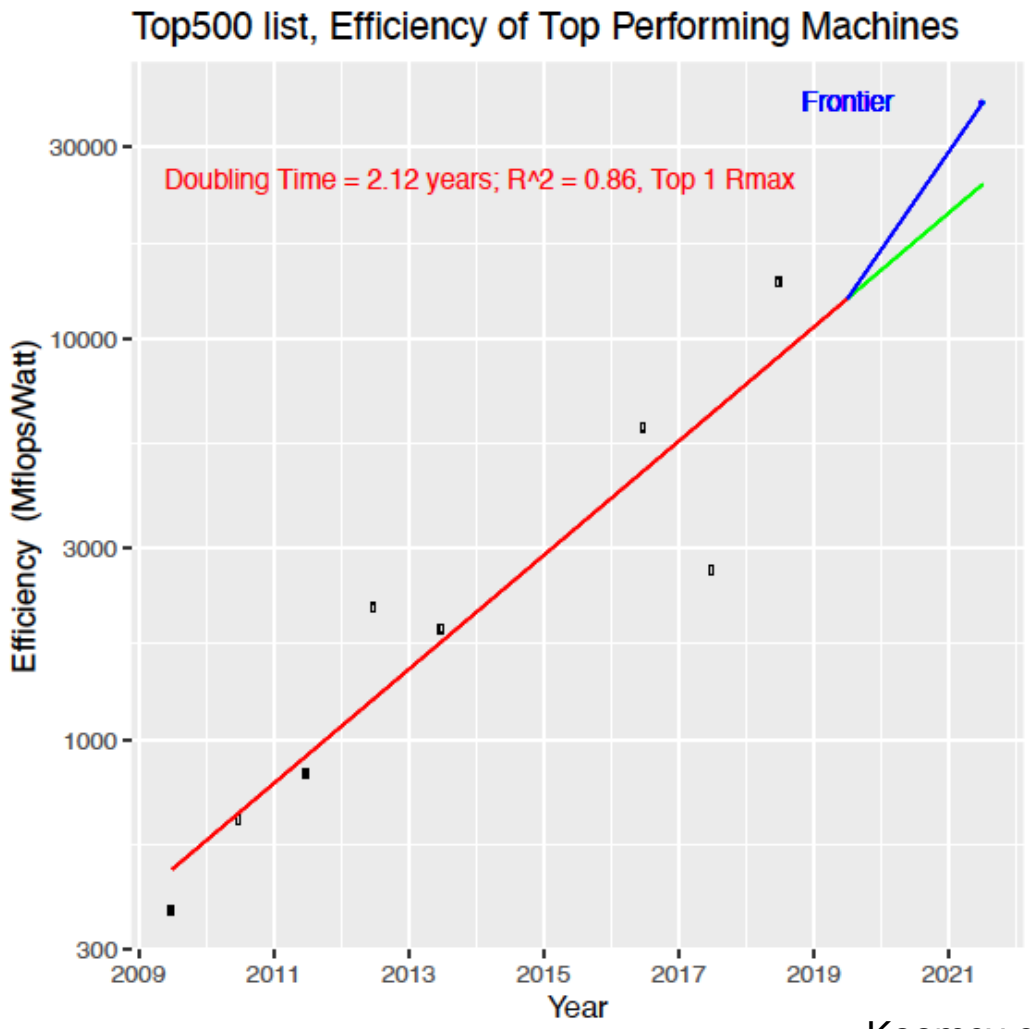
The Climate science sector **must** be leading the effort.

We should, **at least**, have the same ambitions.

Slide from P. Friedlingstein (U. Exeter)

How do we lower our carbon footprint?





Koomey et al., 2019

How do we lower our carbon footprint?

- Monitor the Carbon footprint of Computing and Data centers
- Use certified green energy supplier
- Use energy wisely

How do we lower our carbon footprint?

- Monitor the Carbon footprint of Computing and Data centers
- Use certified green energy supplier
- Use energy wisely

CMIP6 Experiments: Institutions/Models	Useful SY	Total SY	Useful Data Produced (PB)	Total Data Produced (PB)	Useful CH (Mh)	Total CH (Mh)	Total Energy Cost (Joules)	Carbon Footprint (CO2/KWh)
EC-Earth	17,598	27,568	0.73	1.34	27.2	41.8	$1.27x10^{12}$	162.6t
CNRM-CERFACS	23,620	72,000	1.2	1.98	106.4	325	3.13E+12	49.5t
IPSL	53,000	143,000	1.2	7	100	270	6.16E+12	122t
CMCC	965	NA	0.965	NA	1.99	NA	1.61E+12	
UKMO	23,431	NA	7.3	NA	473	NA	1.76E+13	572.5t
DKRZ	1,276	1,321	0.606	NA	5.52	5.90	4.09E+11	24.8t
NCC-NORESM2	6,484	NA	0.297	NA	11.7	NA	4.75E+11	
NERC	640	NA	0.460	NA	55.497	NA	2.17E+12	
MPI	24,175	35,000	1.9	NA	968.116	NA	6.20E+11	37.6t

M. Acosta, IS-ENES, 2020

How do we lower our carbon footprint?

- Monitor the Carbon footprint of Computing and Data centers
- Use certified green energy supplier
- Use energy wisely

Saving and storing 100 GB in the cloud per year would result in a carbon footprint of about 0.2 tons of CO_2, based on the usual U.S. electric mix.

https://medium.com/stanford-magazine/carbon-and-the-cloud-d6f481b79dfe

11 PB -> 20,000 tons per year

M. Acosta, IS-ENES, 2020

CMIP6 Experiments: Institutions/Models	Useful SY	Total SY	Useful Data Produced (PB)	Total Data Produced (PB)	Useful CH (Mh)	Total CH (Mh)	Total Energy Cost (Joules)	Carbon Footprint (CO2/KWh)
EC-Earth	17,598	27,568	0.73	1.34	27.2	41.8	1.27×10^{12}	162.6t
CNRM-CERFACS	23,620	72,000	1.2	1.98	106.4	325	3.13E+12	49.5t
IPSL	53,000	143,000	1.2	7	100	270	6.16E+12	122t
CMCC	965	NA	0.965	NA	1.99	NA	1.61E+12	
UKMO	23,431	NA	7.3	NA	473	NA	1.76E+13	572.5t
DKRZ	1,276	1,321	0.606	NA	5.52	5.90	4.09E+11	24.8t
NCC-NORESM2	6,484	NA	0.297	NA	11.7	NA	4.75E+11	
NERC	640	NA	0.460	NA	55.497	NA	2.17E+12	
MPI	24,175	35,000	1.9	NA	968.116	NA	6.20E+11	37.6t

How do we lower our carbon footprint?

- Monitor the Carbon footprint of Computing and Data centers
- Use certified green energy supplier
- Use energy wisely

Saving and storing 100 GB in the cloud per year would result in a carbon footprint of about 0.2 tons of CO_2, based on the usual U.S. electric mix.

https://medium.com/stanford-magazine/carbon-and-the-cloud-d6f481b79dfe

11 PB -> 20,000 tons per year

M. Acosta, IS-ENES, 2020

CMIP6 Experiments: Institutions/Models	Useful SY	Total SY	Useful Data Produced (PB)	Total Data Produced (PB)	Useful CH (Mh)	Total CH (Mh)	Total Energy Cost (Joules)	Carbon Footprint (CO2/KWh)
EC-Earth	17,598	27,568	0.73	1.34	27.2	41.8	1.27×10^{12}	162.6t
CNRM-CERFACS	23,620	72,000	1.2	1.98	106.4	325	3.13E+12	49.5t
IPSL	53,000	143,000	1.2	7	100	270	6.16E+12	122t
CMCC	965	NA	0.965	NA	1.99	NA	1.61E+12	
UKMO	23,431	NA	7.3	NA	473	NA	1.76E+13	572.5t
DKRZ	1,276	1,321	0.606	NA	5.52	5.90	4.09E+11	24.8t
NCC-NORESM2	6,484	NA	0.297	NA	11.7	NA	4.75E+11	
NERC	640	NA	0.460	NA	55.497	NA	2.17E+12	
MPI	24,175	35,000	1.9	NA	968.116	NA	6.20E+11	37.6t

How do we lower our carbon footprint?

- Monitor the Carbon footprint of Computing and Data centers
- Use certified green energy supplier
- Use energy wisely

➔ Can we reduce CMIP7 CO_2 emissions by 50% relative to CMIP6?
➔ How much of CMIP6 simulations/models can we reuse?
➔ How many/which MIPS/scenarios do we really need?
➔ How many ensembles do we really need?
➔ How many simulations at high resolution?
➔ Do we need all modelling groups to do everything with their State-of-the-Art model?
➔ Isn't there a more efficient way to get organised?
➔ How do we optimize data storage/analysis?

How do we lower our carbon footprint?

- Monitor the Carbon footprint of Computing and Data centers
- Use certified green energy supplier
- Use energy wisely

The next phase(s) of CMIP will only be able to achieve all its objectives (science & societal relevance, carbon footprint and accessibility) if all parties are involved in the planning and definition!

- ➔ Can we reduce CMIP7 CO_2 emissions by 50% relative to CMIP6?
- ➔ How much of CMIP6 simulations/models can we reuse?
- ➔ How many/which MIPS/scenarios do we really need?
- ➔ How many ensembles do we really need?
- ➔ How many simulations at high resolution?
- ➔ Do we need all modelling groups to do everything with their State-of-the-Art model?
- ➔ Isn't there a more efficient way to get organised?
- ➔ How do we optimize data storage/analysis?