Possible next steps in community-based model evaluation

Peter Gleckler (PCMDI) and Eric Guilyardi (IPSL)

with inputs from many colleagues

WGCM22

BSC, Barcelona, 25-29 March 2019

Context

Peer-reviewed publication: primary mechanism for documenting CMIP research.
But, pressing reasons to more efficiently define, produce, summarize, and make relevant model evaluation results available, e.g.:

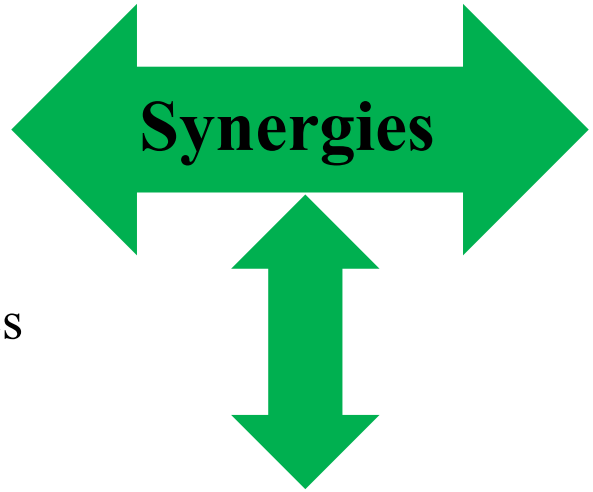
- Inform national assessments, the IPCC process, stakeholders, and public
- Advance science more efficiently (provide high level summaries; less re-inventing)
- More directly contribute to model development (via useful quick feedback)

Community-based model evaluation capabilities are becoming a reality, thanks to the design target provided by the CMIP conventions and standards

A few examples

Integrating capabilities

- ESMValTool (CVDP, *others*)
- CMEC (PMP, ILAMB, TECA)
- NOAA MAPP process diagnostics



Expert teams

- CLIVAR ENSO group
- WGNE MJO task force
- CFMIP diagnostics

There are also many scientifically focused independent tools under development

The WGNE/WGCM Metrics and Diagnostics Panel

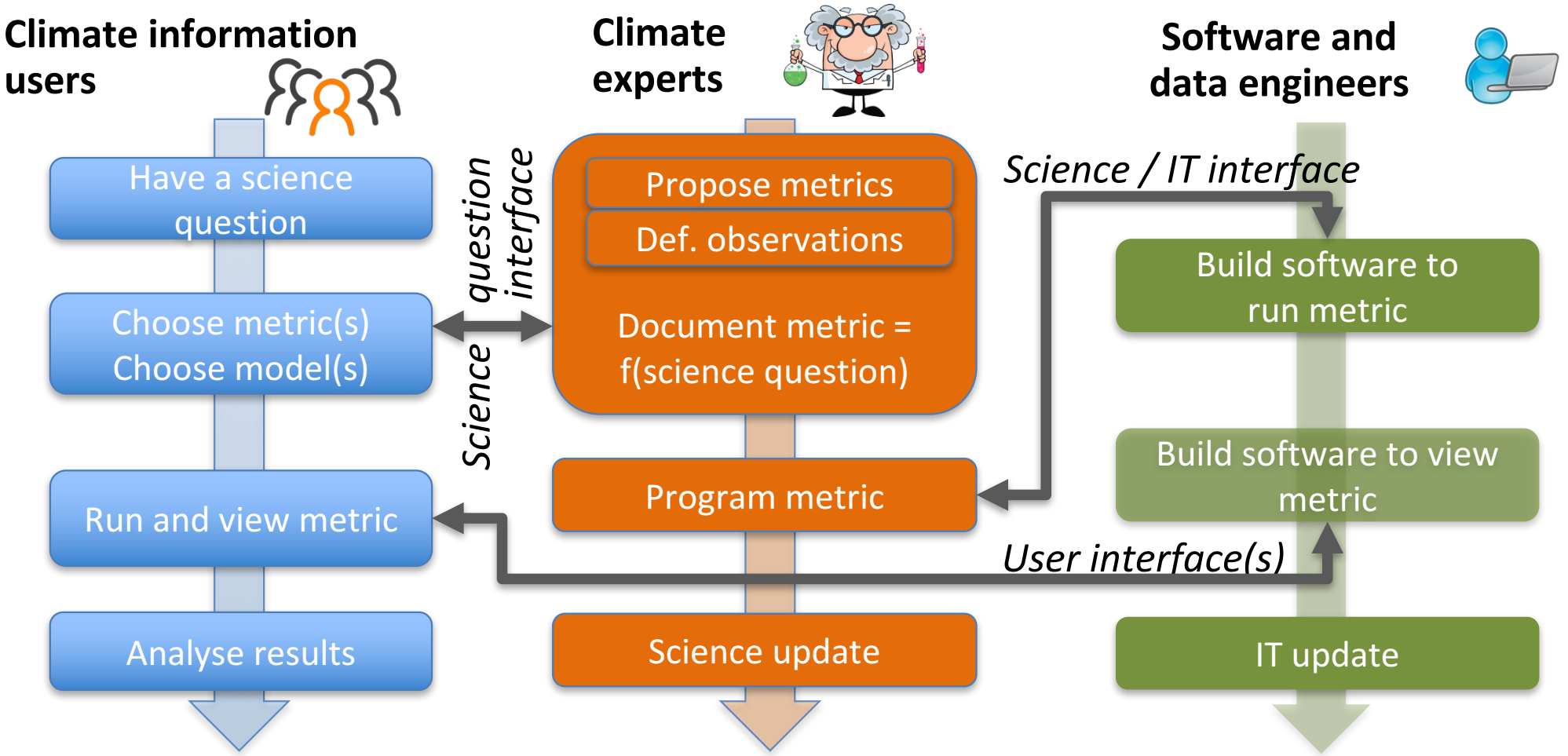
Beth Ebert, Veronika Eyring, Pierre Friedlingstein, Peter Gleckler (chair), Simon Marsland, Robert Pincus, Karl Taylor, Keith Williams

- Has helped draw attention to metrics and stimulate research
- The panel has been relatively inactive the last few years and in light of the the WCRP Strategic and Implementation plans it is a good time to rethink

Anticipating future needs and expectations in CMIPx

- Individual research still fundamental to CMIP related science
- Building on that, routine and systematic evaluation will only become increasingly important
- Increasing involvement by expert teams would help ensure advancing science is progressively incorporated into community evaluation capabilities
- Nurturing a set of standards for how these tools can be linked (yet remain independent) will be more complicated than establishing data standards but it is essential

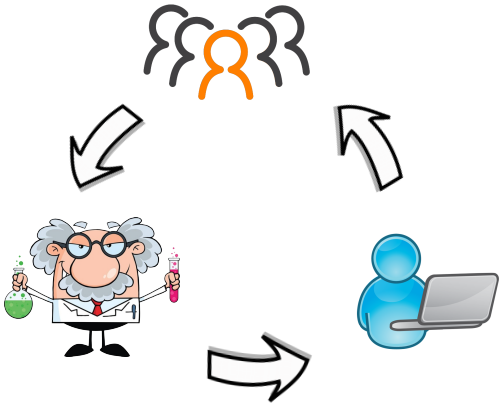
Model evaluation workflow



Articulate different actors, different expertise and expectations

Separation of concerns + co-construction

- Climate information users need state-of-the-art :
  - **Science** of model evaluation
  - **Software** tools for model evaluation
- Different experts -> different workflow
- Otherwise one of them becomes obsolete
  - High risk of mis-use
  - Loss of trust, wasted ressources
- Articulation/modularity via clear interfaces
- e.g. lessons learned for CMIP, ESGF, ES-DOC,…



Pilot ENSO package to test ideas

Engage with an IT infrastructure (driver), here PMP

ENSO Metrics Package

ENSOCollectionsLib.py

ENSOComputeMetricsLib.py

ENSOMetricsLib.py

Define pre-processing

ENSO Support libs

dict{}

Documentation of every step in calculation carried through with the results

Science / IT interface

PMP Driver

User chooses metric collection and models

Understand work

(vars, obs…)

Execute work

(loop on models)

Collect results

View results

User analyses results

https://github.com/eguil/ENSO_metrics

1) Engage with PMP/PCMDI

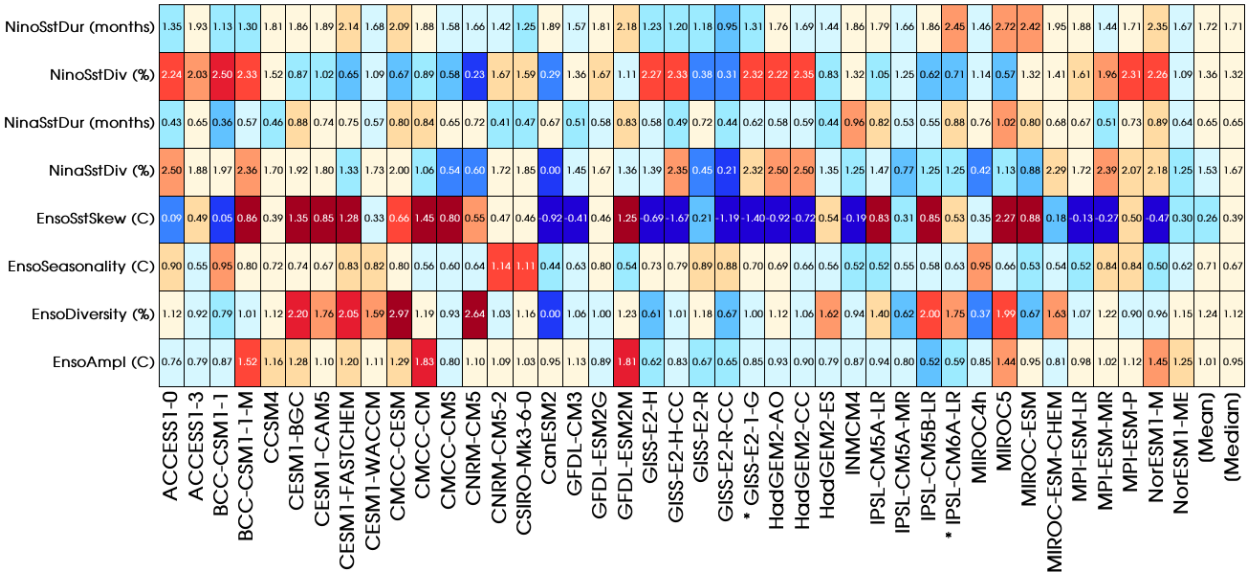
2) Engage with ESMValTool (IS-ENES3)

First results – ENSO performance in CMIP5 historical

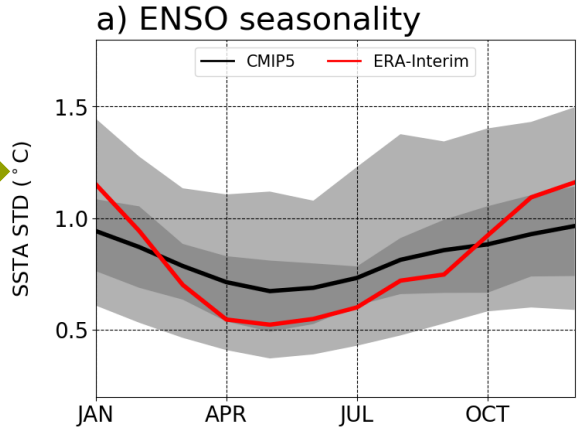
Address these specific questions:

- ENSO performance in historical simulations
- ENSO teleconnections in historical simulations
- ENSO processes

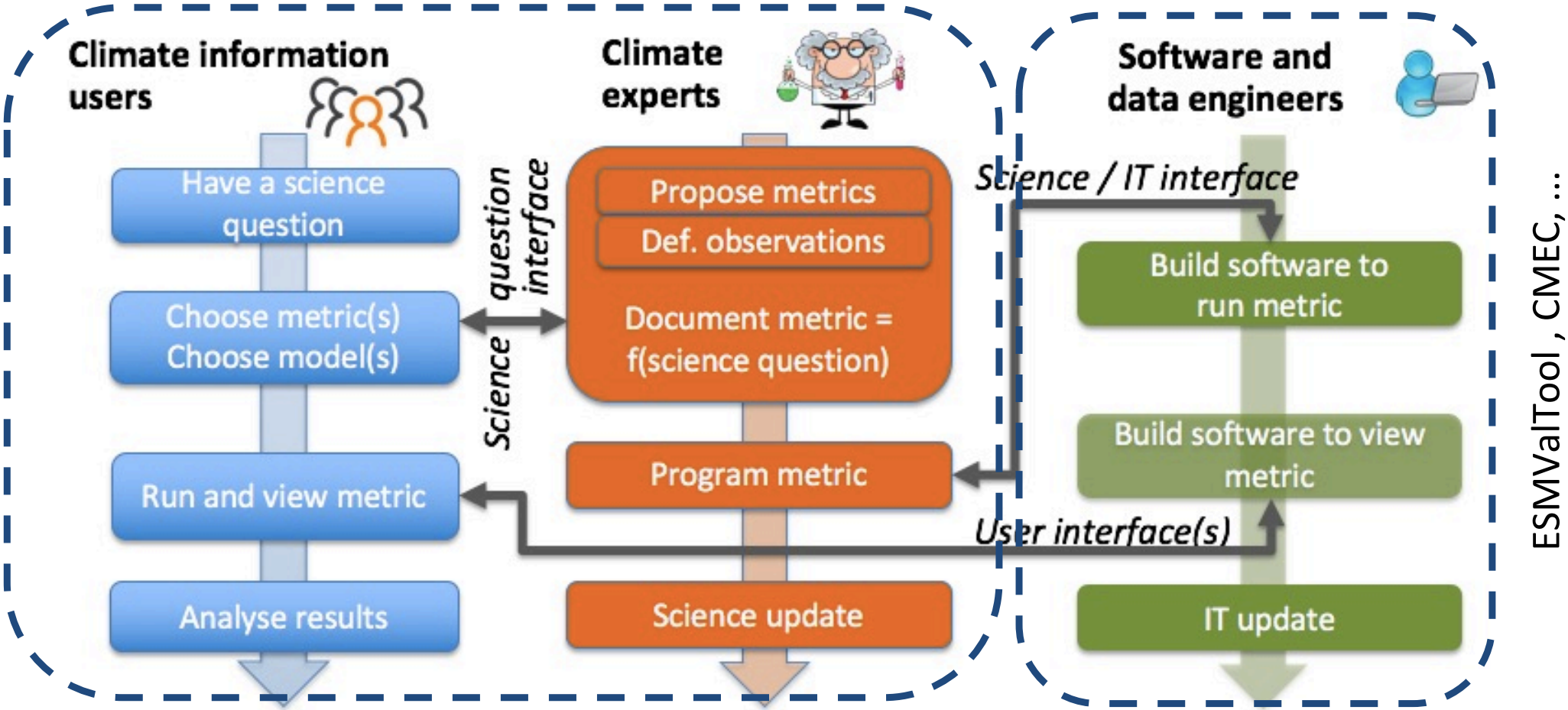
Shading : relative performance wrt MME
Values : actual metric value



"Dive down" diagnostics



How can we move forward ?



Nudge via community organisation

Development via specific funding

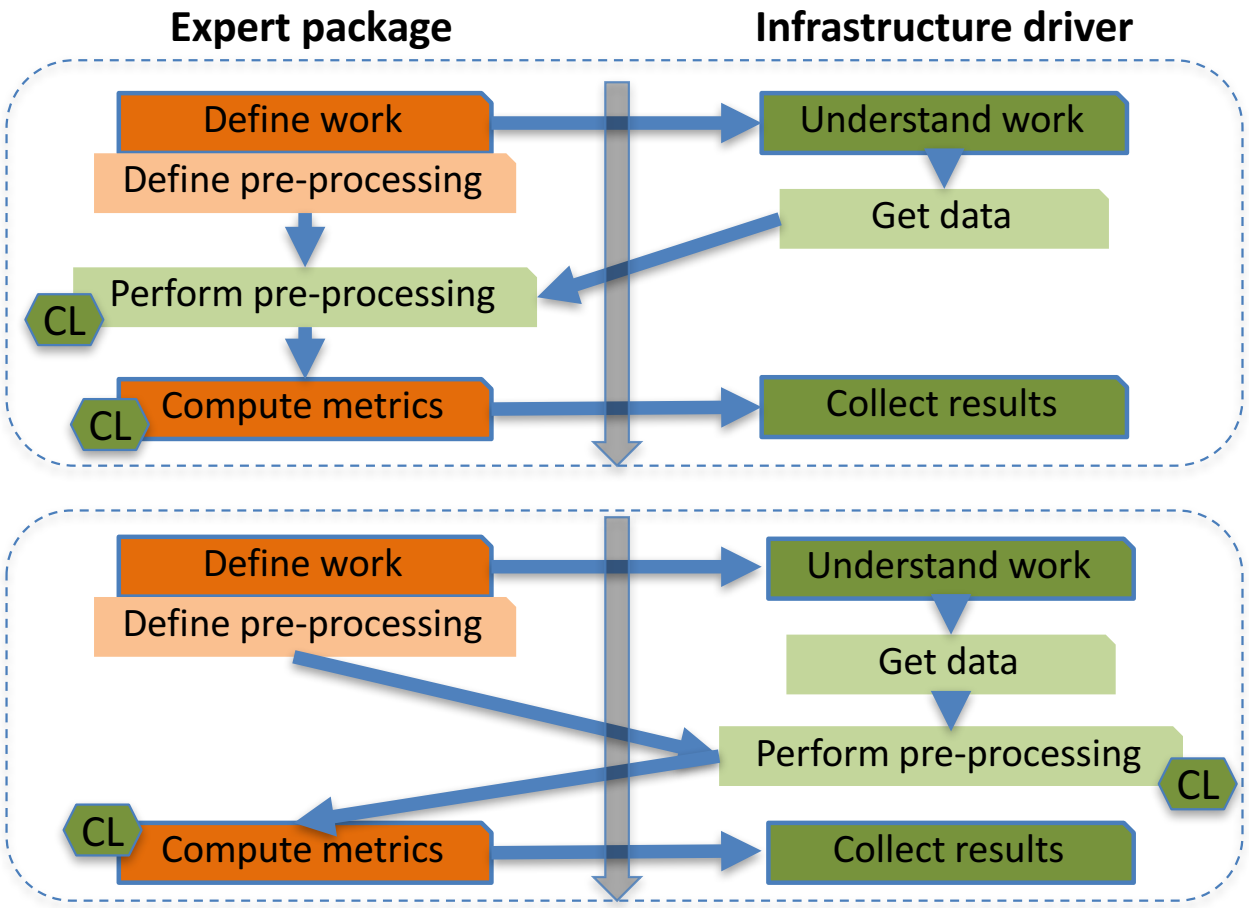
Possible work plan for a potential "WCRP model evaluation panel or work group"

- 2019: WCRP and existing panels help refine scope and vision and identify members
- First year (2019-2020): analysis of existing model evaluation efforts and identify opportunities/gaps
- Year 2: select and advance a few pilot areas (beyond ENSO)
- Year 3: unveil consensus model evaluation framework and process for these first areas
- Year 4: review activity via WCRP process (tdb)

Summary and discussion

- Community–based model evaluation involves 3 pillars that need to be articulated
- Viable process proposed (pilot study) that requires further community discussion (e.g. include NWP/SF ?)
- Because of our community organisation and funding, resilience requires modularity and diversity of software tools
- Model evaluation standards and framework may help many groups develop diagnostics towards interoperability
- Next steps

Example of interfaces choices



Enable different compute libraries <CL> kernels (e.g. CDAT, IRIS, …) ?