Making Climate Model Evaluation available for End Users: an example with ENSO metrics

Eric Guilyardi, Yann Planton, Peter Gleckler, Jiwoo Lee, Andrew Wittenberg

and the CLIVAR ENSO Research Focus

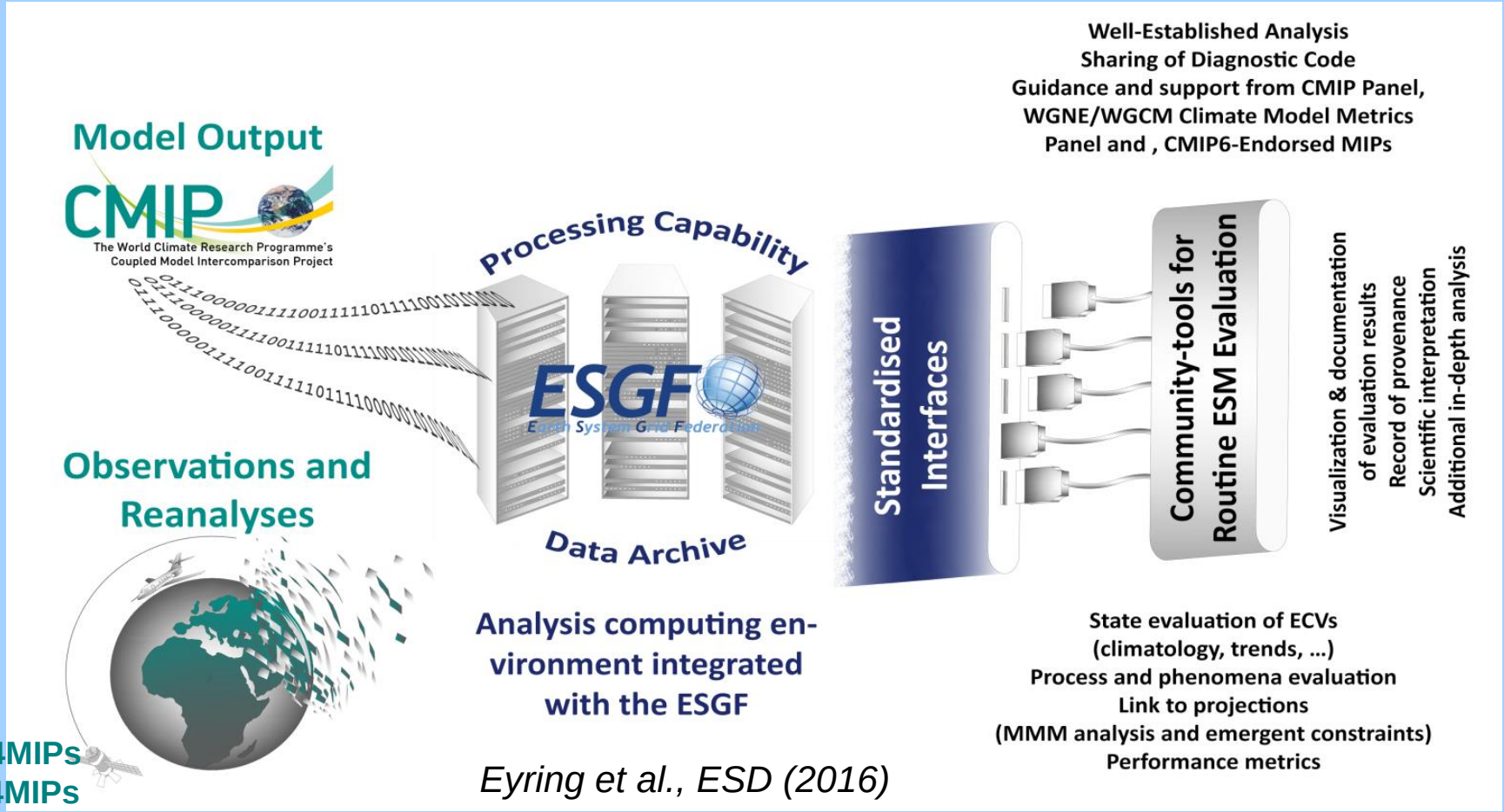
IPSL/LOCEAN, NCAS-Climate, PCMDI, GFDL

IV International Conference on El Niño Southern Oscillation: ENSO in a warmer Climate
16-18 October 2018. Guayaquil - Ecuador

Evaluating climate models

- Model evaluation:
  - **Exploratory** = understand mechanisms in models
  - **Consensus** = evaluate « the fitness for purpose »
- Metrics vs. diagnostics
- In scope here: consensus model evaluation metrics and diagnostics

Strategy for routine Earth system model evaluation in CMIP



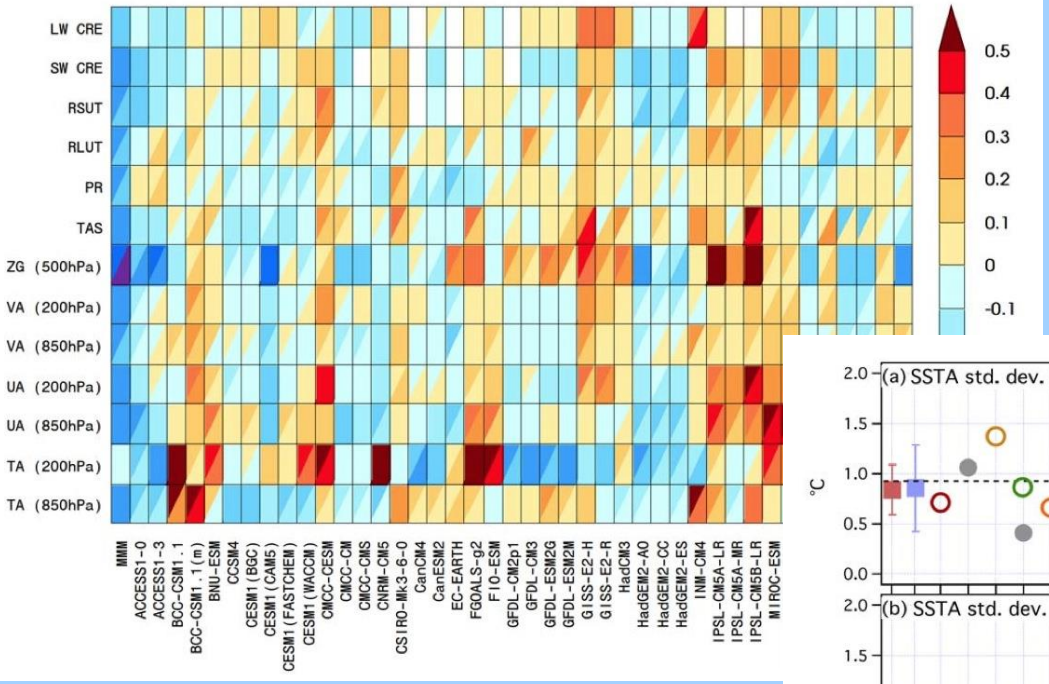
Eyring et al., ESD (2016)

Community experience gathered via pilot projects running alongside the ESGF for CMIP

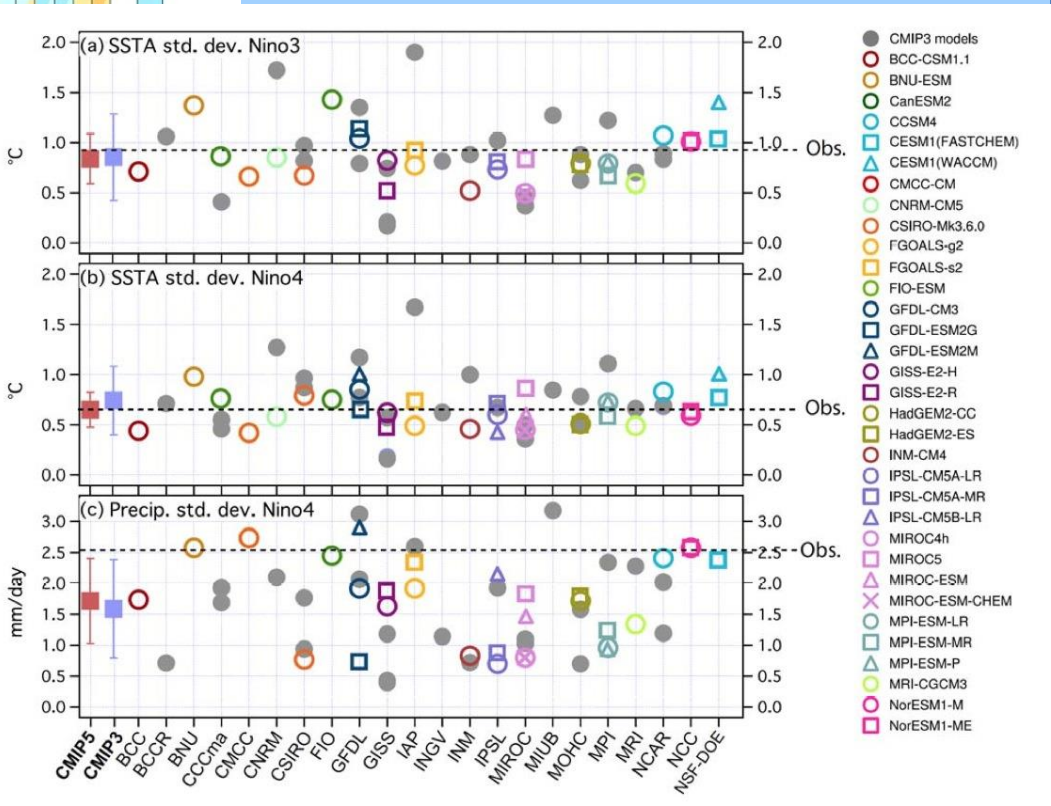
ESMValTool
Earth System Model eValuation Tool

PCMDI Metrics Package (PMP)

Examples of consensus climate model evaluation



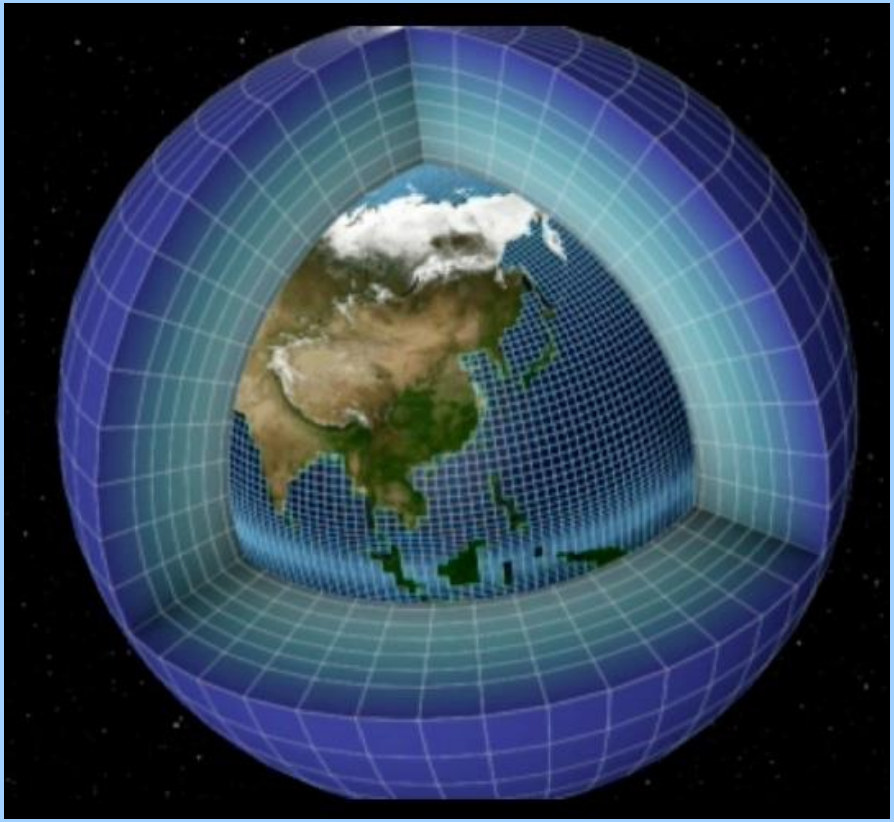
Produce such figures
Cf. Chap 9 IPCC AR5



Fitness for purpose: can we use these models to assess climate change ?

Making sense of climate simulations

- Huge amount of data
- Complex models and processes
- Hard to get to
- Hard to analyse



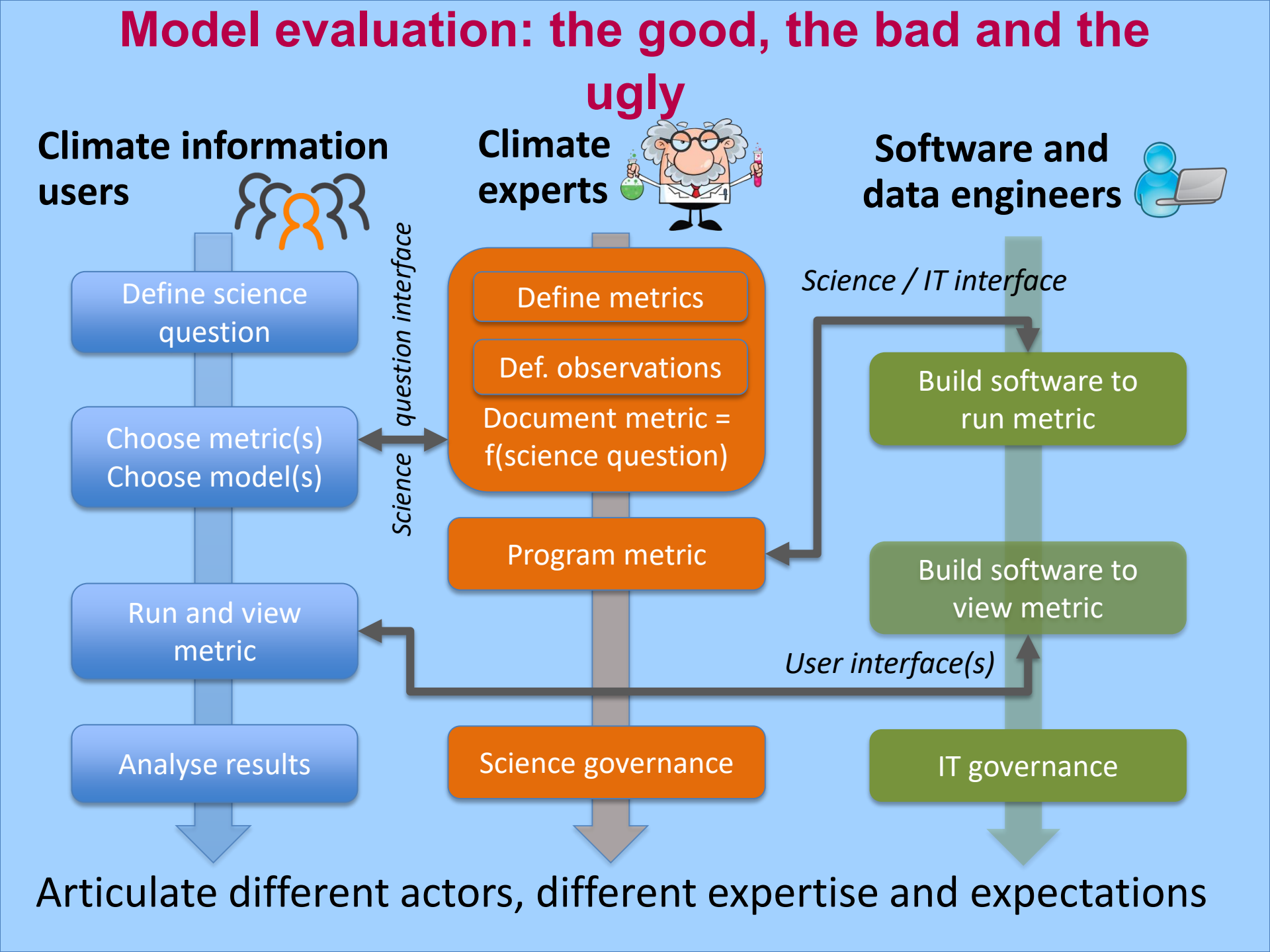
Can be intimidating…

.. risk of wrong usage

Climate models end users need best advice/support

- WCRP new Overarching Scientific Objectives
  1. Fundamental understanding of the climate system
  2. Advancing predictive skill on timescales up to a decade
  3. Constraining projections on decadal to centennial timescales
  4. **Connecting climate science with policy and services**
- Climate services (GFCS/WMO)
- Co-construction with non-experts in climate science and models

Model evaluation: the good, the bad and the ugly



Articulate different actors, different expertise and expectations

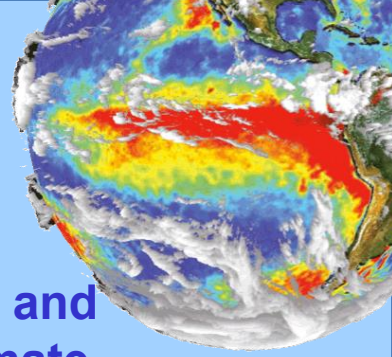
Separation of concerns for routine and consensus model evaluation

- User needs state-of-the-art :
  - **Science** of model evaluation
  - **Software** tools for model evaluation

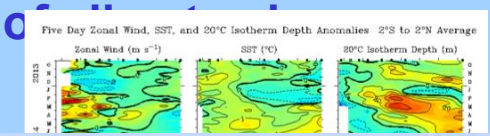
- Different experts -> different governance
- Otherwise one of them becomes obsolete
  - High risk of mis-use
  - Loss of trust, wasted ressources
- Articulation/modularity via clear interfaces
- e.g. lessons learned for CMIP, ESGF, ES-DOC,…

ENSO and tropical Pacific metrics for CMIP6

Eric Guilyardi (IPSL & NCAS/Climate) and Andrew Wittenberg (GFDL)
On behalf of the CLIVAR Reseach Focus "ENSO in a changing climate"

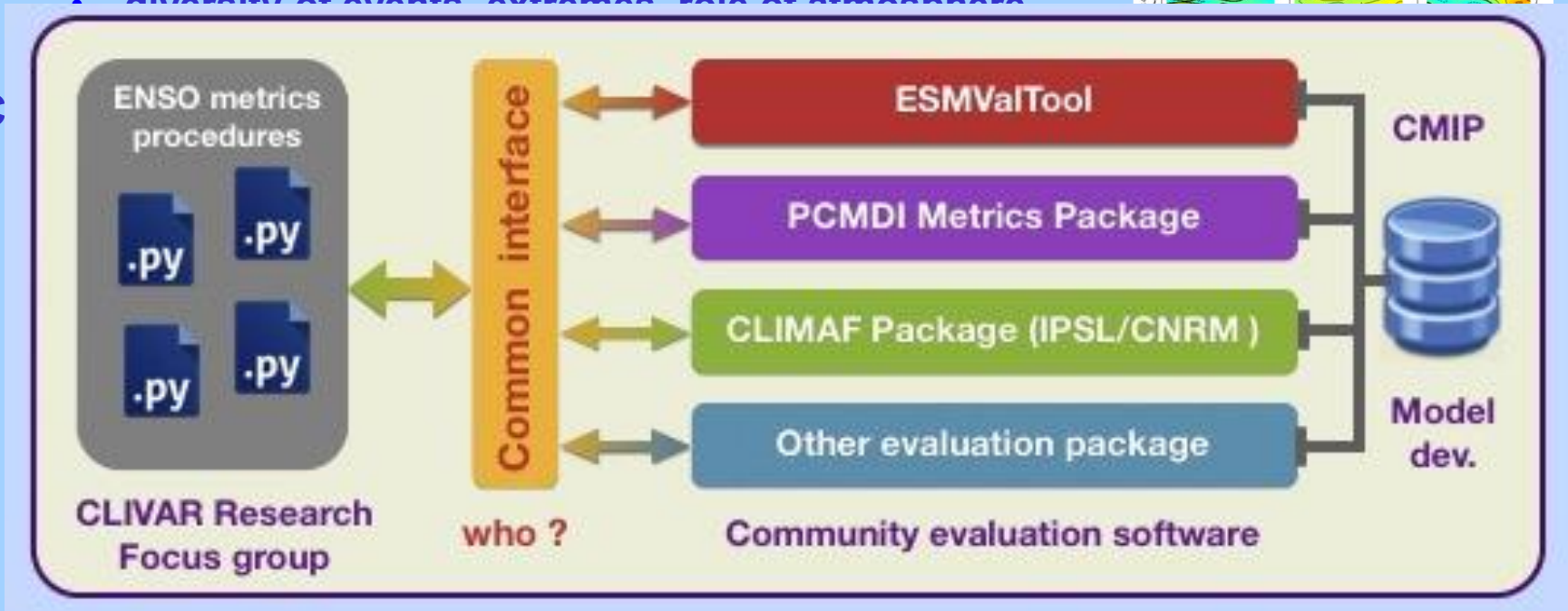


- Despite 30 years of progress, ENSO continues to surprise us and challenge our assumptions - It remains a major unsolved climate puzzle
- It is the "elephant in the room" for regional impacts o
- ENSO research very active field
  - diversity of events, extremes, role of atmosphere



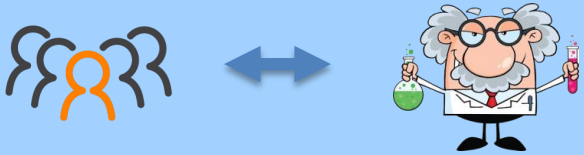
C

-
-
-
-



The climate information users

- What is the scientific question ?
- What are the related metrics?
- What are the reference “observations” ?



- Introduce concept of a Metric collection (MC) to address one specific science question

	Metric 1	Metric 2	Metric 3	Metric 4	Metric 5…
Collection A					
Collection B					
Collection C					
…					

Software for model evaluation

- How to best compute the metrics (get data, run on computer, visualize,…)
- Challenging project: define workflow, process, development, modularity, agility, beta testing, operational, funding, etc.

- Who has the knowledge for governance/trust?
  - IT and data experts (i.e. NOT climate experts nor users)
  - e.g. ENES, WIP, ESGF, ES-DOC…

Having a go at the separation of concerns for ENSO

Climate information users

Climate experts

Software and data engineers

Define science question

Choose metric(s)
Choose model(s)

Run and view metric

Analyse results

question interface

Science

Define metrics

Def. observations

Document metric = f(science question)

Program metric

Science governance

Science / IT interface

Build software to run metric

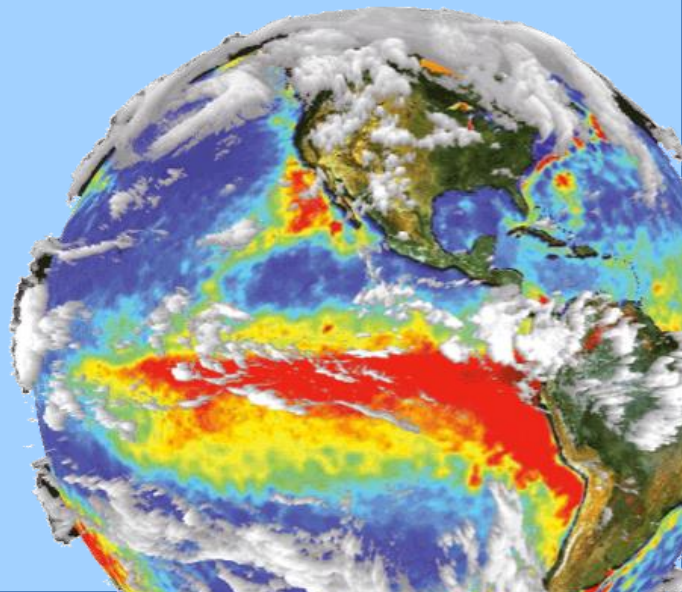
Build software to view metric

User interface(s)

IT governance

Work for CLIVAR ENSO experts

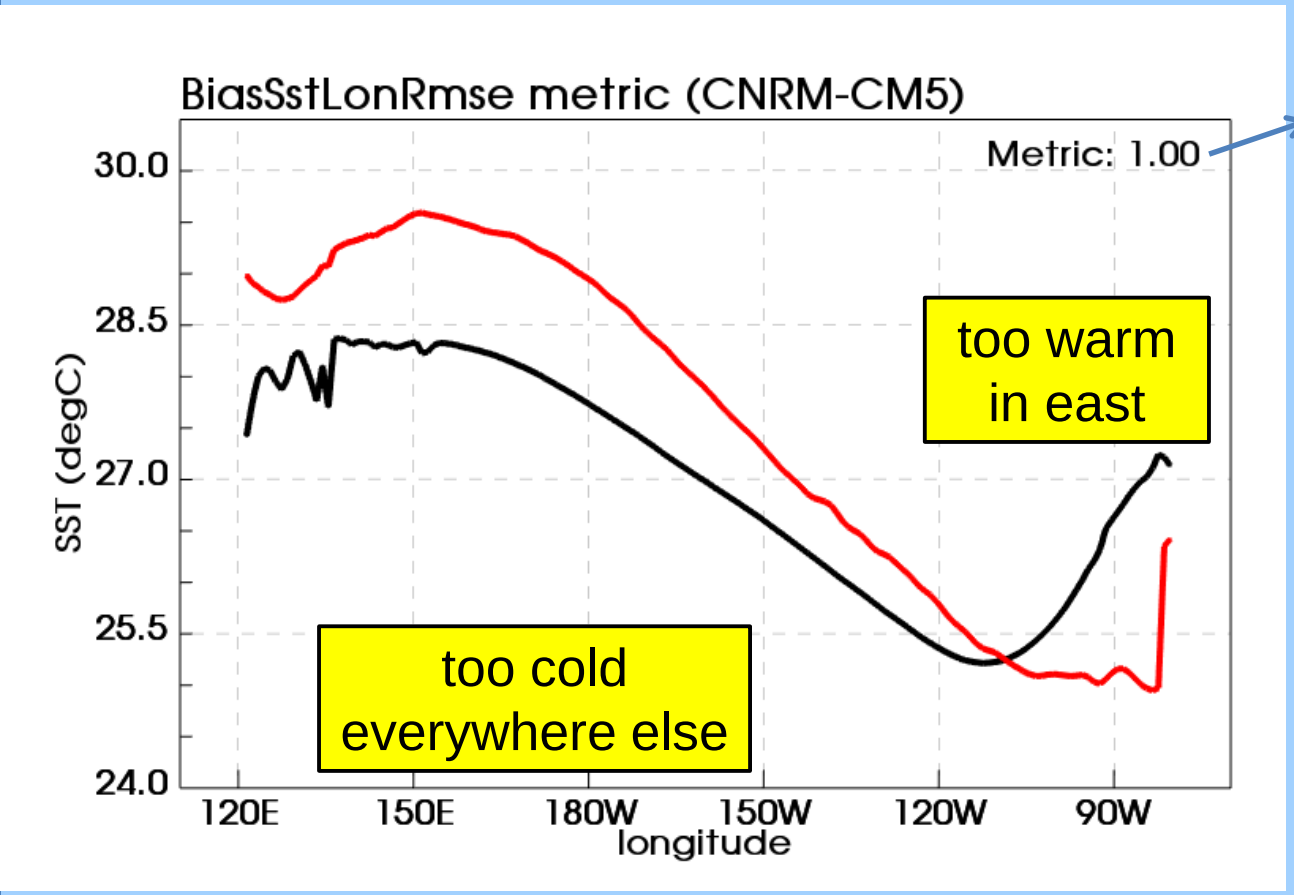
- Identify science questions (performance, 20th century teleconnections, processes, …)
- Associate a metrics collection to each
- Define metrics (e.g. python files) and associated science choices (reference observation, duration,…)
- Release versions and maintain science contents over time (do not worry about end-to-end technical implementation)
- Document MC and metrics and engage with users



List of science questions

- **ENSO performance in historical simulations**
  - Mean state, seasonal cycle, ENSO characteristics space/time, diversity
- **ENSO teleconnections in historical simulations**
  - Temperature and precipitation, land/ocean regional/global
- **ENSO processes**

Example of metric: annual-mean climate



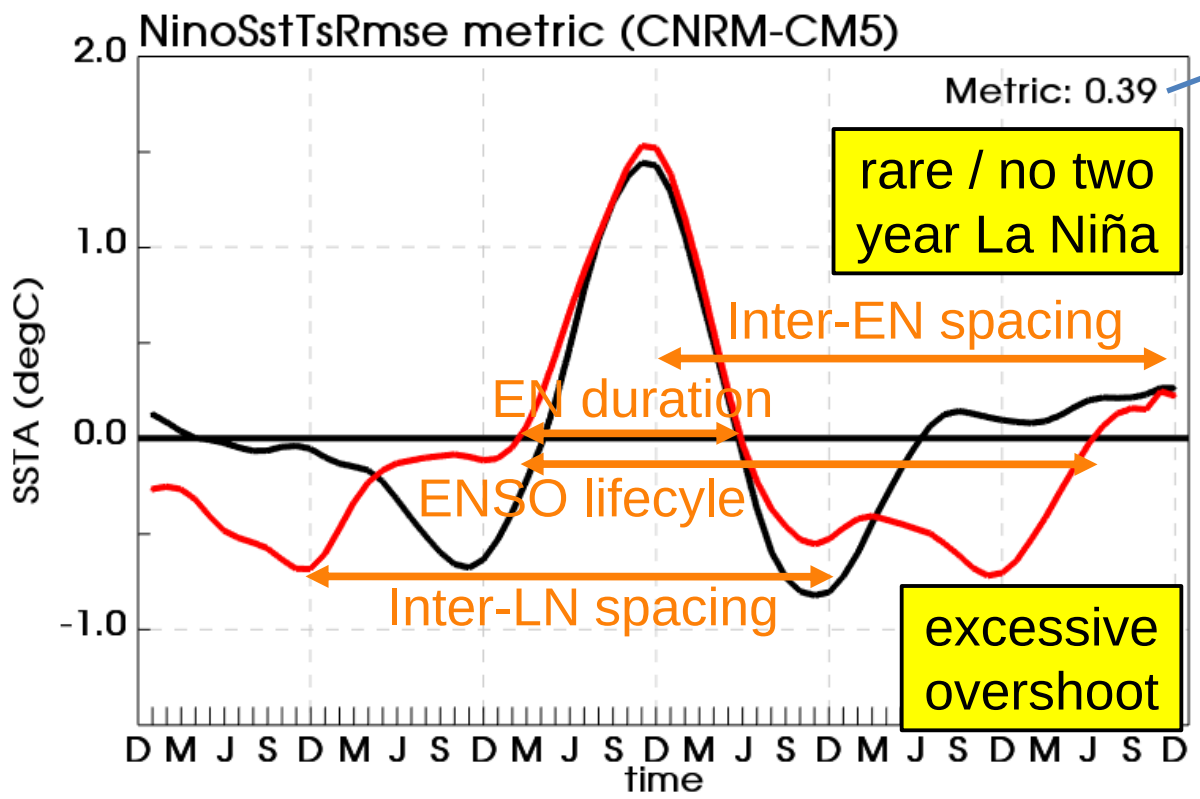
RMSE(**obs**, **model**) in °C

Same diagnostics in latitude in the eastern Pacific

Same diagnostics using:

- Taux
- Pr

Example of metric: El Niño life cycle



RMSE(**obs**, **model**) in °C

Diagnostic in longitude along the equator for SSTA during ENSO peak

Same diagnostics for La Niña events

Structure of pilot ENSO package

Engage with an IT infrastructure (driver)

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

https://github.com/eguil/ENSO_metrics

User chooses metric collection and models

PMP Driver

Understand work

(vars, obs…)

Execute work

(loop on models)

Collect results

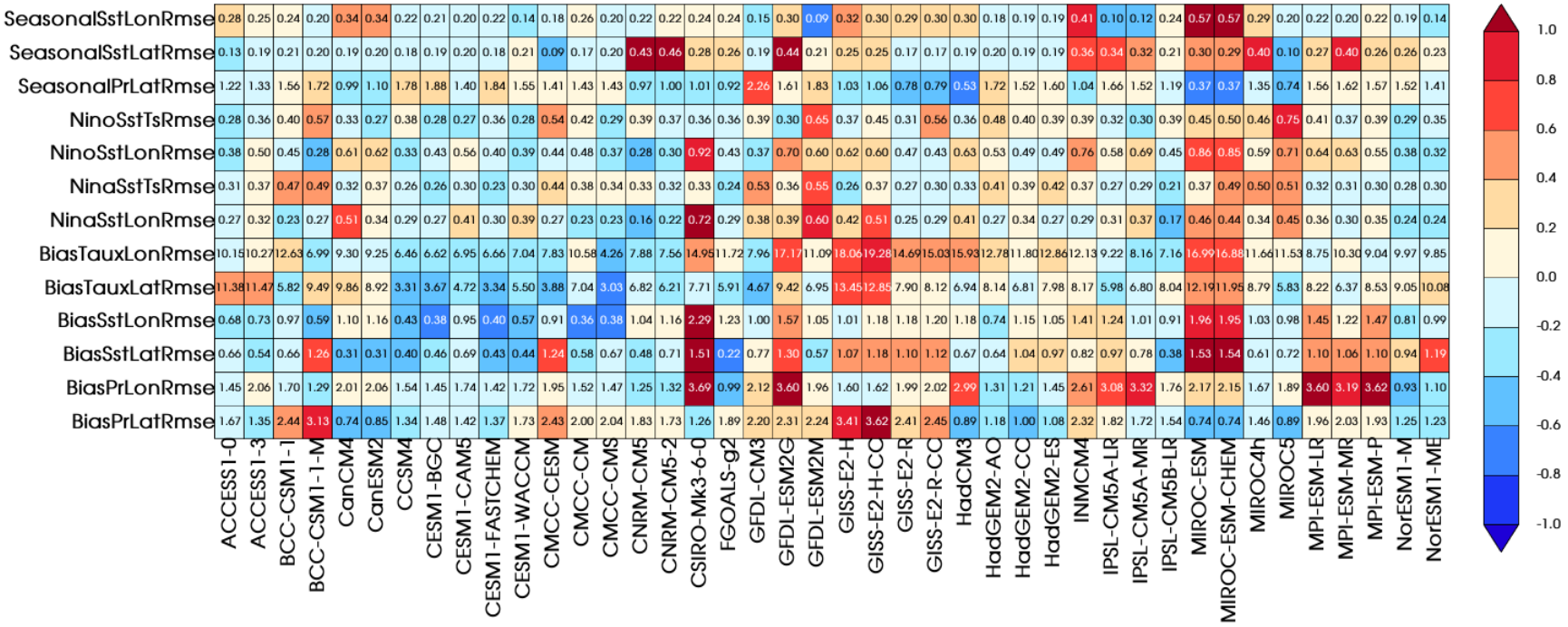
View results

Science / IT interface

User analyses results

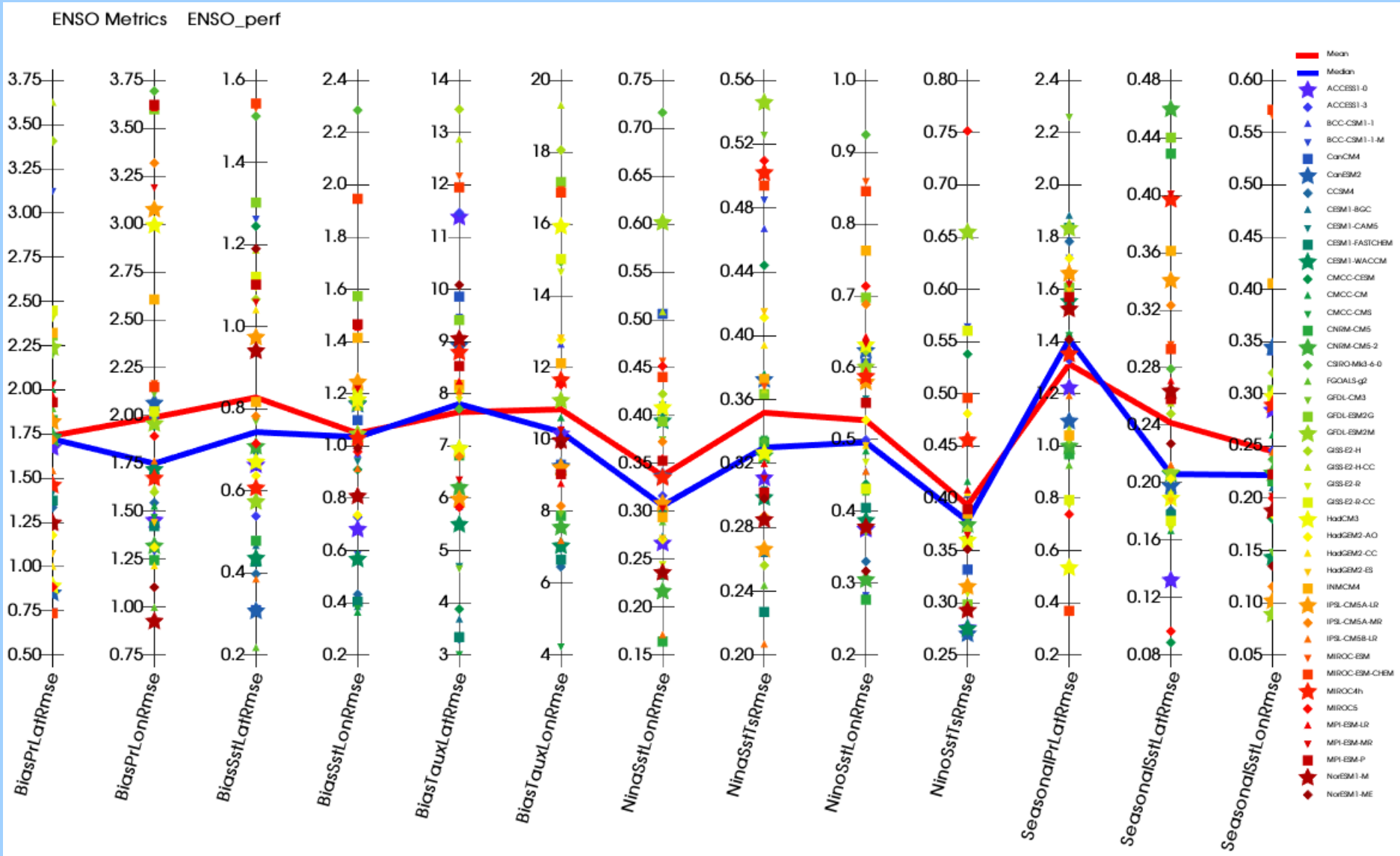
First results – ENSO performance in CMIP5 historical

Shading : relative performance wrt MME
Values : actual metric value (here RMSE)



CDAT

Another view: the parallel plot



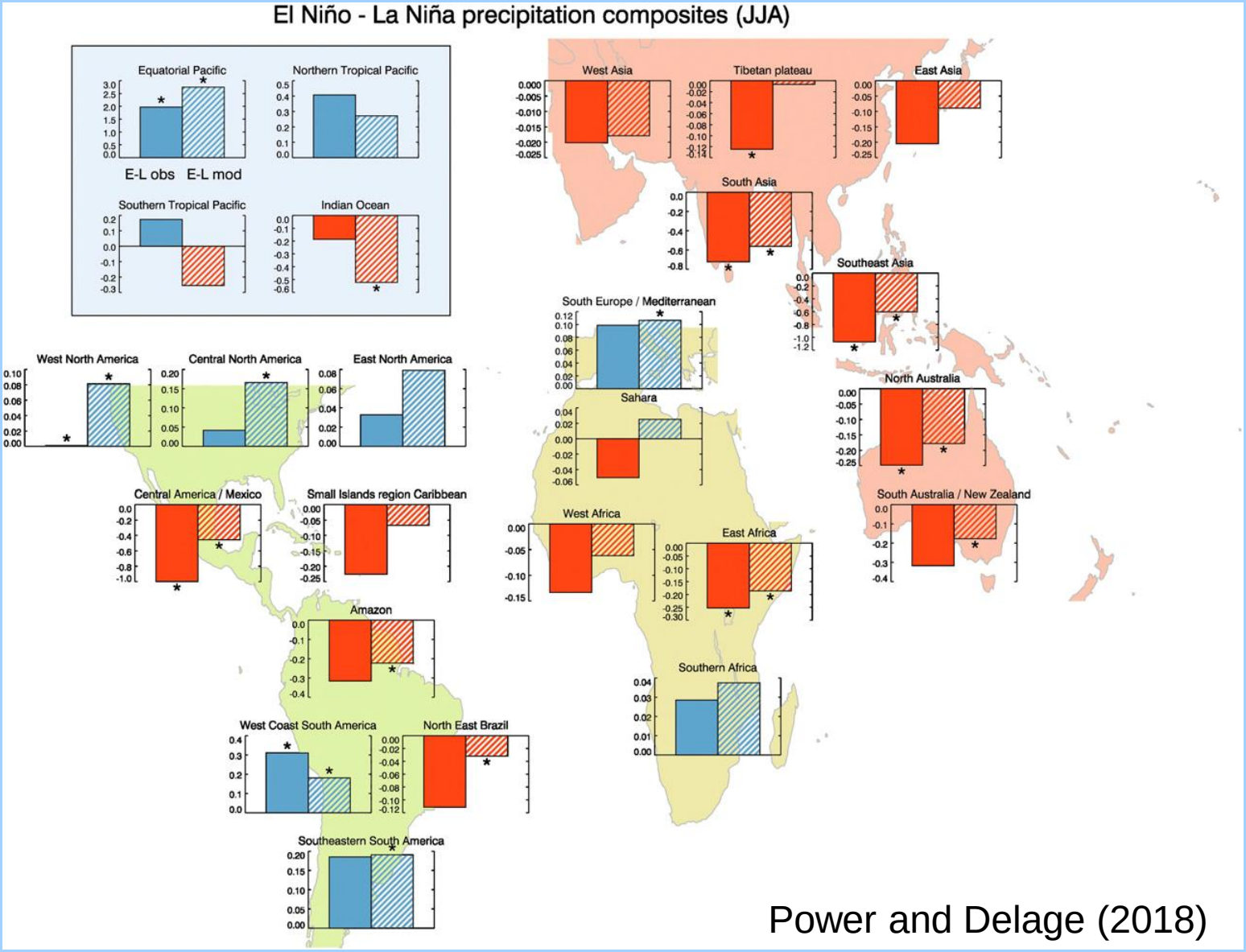
Documenting the science provenance

- For each metric, provide:
  - Documentation: whys and hows of the metric Math
  - Definition of the metric
  - Frequency (DA, MO,…)
  - Observations (as many as possible), including ref. period
  - Reference to show the robustness of metric
  - Minimum number of realisations (for metric to make sense)
  - Minimum lenght of simulation needed (for metric to make sense)
  - « Dive down » diag (map of RMS,…)
  - Normalisation wrt sharing single color bar in a MME env.

ENSO teleconnections

- Evaluation of ENSO-related teleconnection in historical simulations
- 10 metrics proposed :
  - 6 on mean ENSO temperature / precipitation patterns (Perry et al. 2017)
  - 4 on regional ENSO precipitation teleconnection (Power & Delage 2018)

EN minus LN precipitation (JJA)



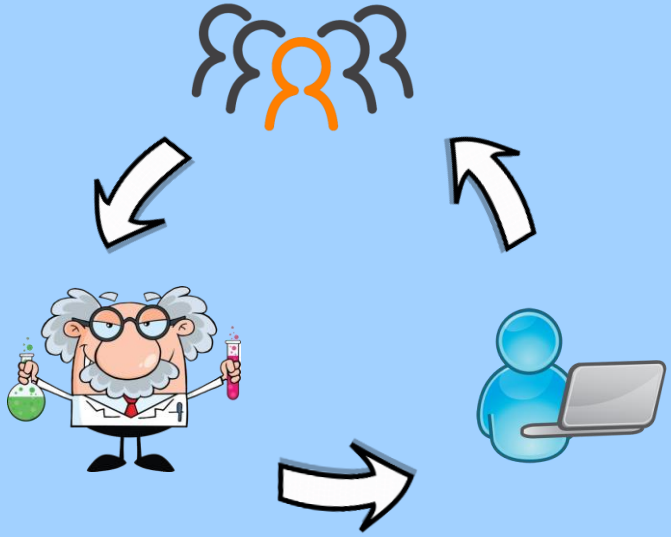
Power and Delage (2018)

Science governance: call for a « WCRP model evaluation panel »

- Need to label consensus model evaluation for community trust
- Like for MIPs proposed by the community, a panel is needed to define quality criteria and label packages
- Criteria should include process, maturity, level of documentation, review etc. maybe following IPCC-like process
- An ad-hoc group would oversee the technical standard for the science provenance (e.g. WIP for WGCM)

Summary - model evaluation

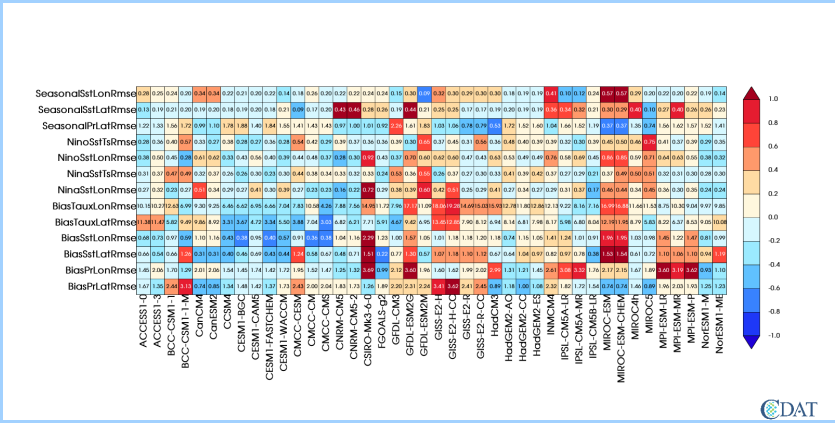
- Evaluating models for uses beyond the climate modelling experts is a new challenge
- Community ready to organise model evaluation to another level, separation of concerns, standards and interfaces
- Governance (end users, WCRP experts, data and software)
- Opportunity with new WCRP strategy and infrastructure projects (e.g. IS-ENES3)



ENSO metrics – stay tuned

Pilot implementation for ENSO metrics

- Contribution to CMIP6/IPCC AR6 evaluation
- Suite of model exploration tools for a wide range of users
- Long lasting contribution of CLIVAR ENSO experts
- Get involved !



Reducing CO_2 emissions: what can we do as a community ?

Having discovered the issue of climate change and writing IPCC reports is key… but does not preclude actions to reduce emissions

What can we do,
avoiding both too strong
advocacy and denial ?

Act as a community !
For us it means: **less air travel for scientific meetings**

Could we devise best practice for the climate community ?

E.g:

- Manage a shared carbon budget/emissions targets per team/group
- Favour early career scientists travel
- Influence conference organisation to develop two-way remote participation (give the choice)
- Fund videoconferencing equipment (bandwith, dedicated rooms, local/regional hubs)
- Offset (only) if travel cannot be avoided

Collective approach allows more long lasting effects, shared values and many co-benefits

Summary - model evaluation

- Evaluating models for uses beyond the climate modelling experts is a new challenge
- Community ready to organise model evaluation to another level, separation of concerns, standards and interfaces
- Governance (end users, WCRP experts, data and software)
- Opportunity with new WCRP strategy and infrastructure projects (e.g. IS-ENES3)