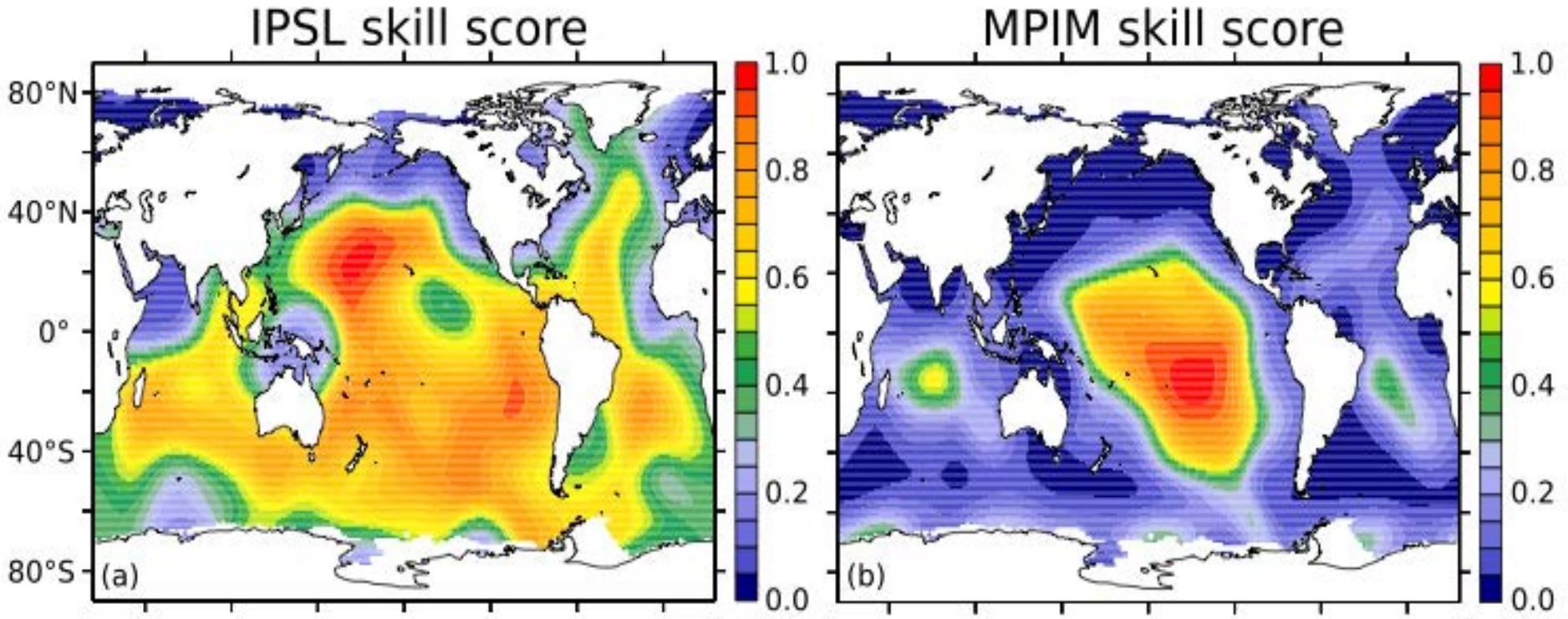
Inconsistent strategies to spin up models in CMIP5 and effects on model performance assessment

Roland Séférian, Laurent Bopp, Marion Gehlen, Laure Resplandy, James Orr, Olivier Marti, Scott C. Doney, John P. Dunne, Paul Halloran, Christoph Heinze, Tatiana Ilyina, Jerry Tjiputra, Jörg Schwinger

CMIP5: the age of the skill-score metrics…

2000-2007 (Stow et al., 2009)
~63 % of the reviewed paper provided a very simple evaluation

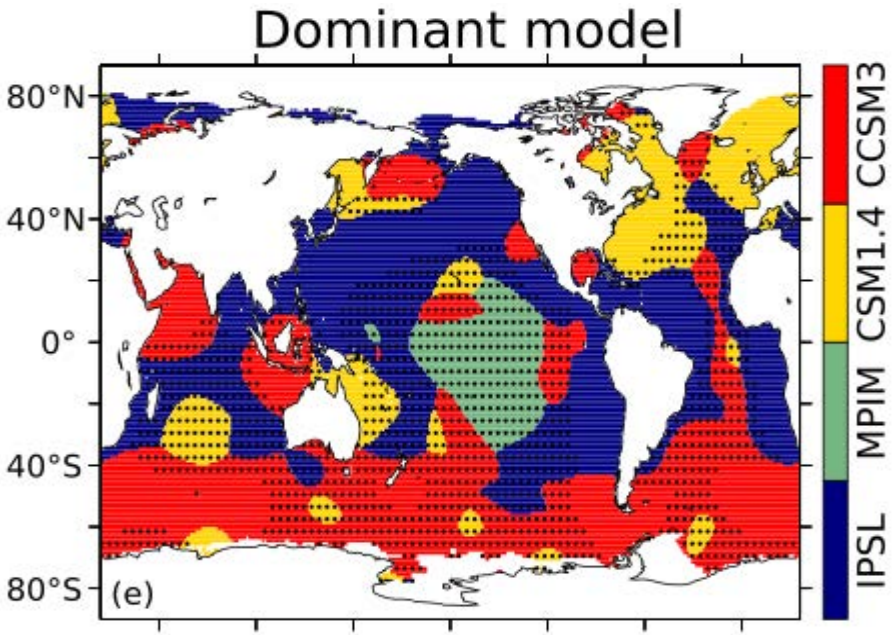
2009-onward (e.g., Frölicher et al., 2009, Steinarcher et al., 2011)
Ensemble model evaluation (cross evaluation) + model weighted solution

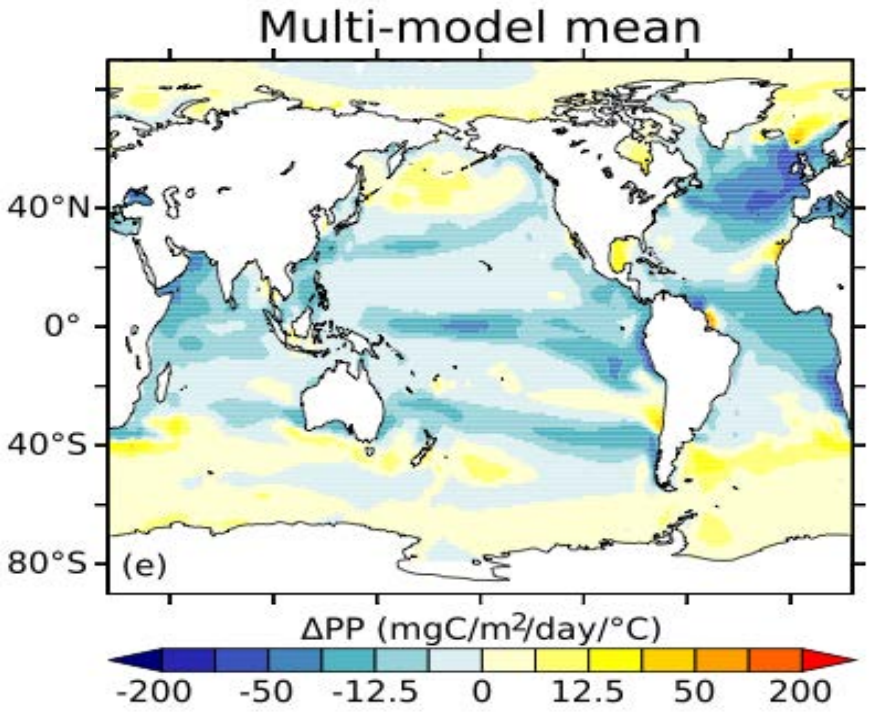


CMIP5: the age of the skill-score metrics…

2000-2007 (Stow et al., 2009)
~63 % of the reviewed paper provided a very simple evaluation

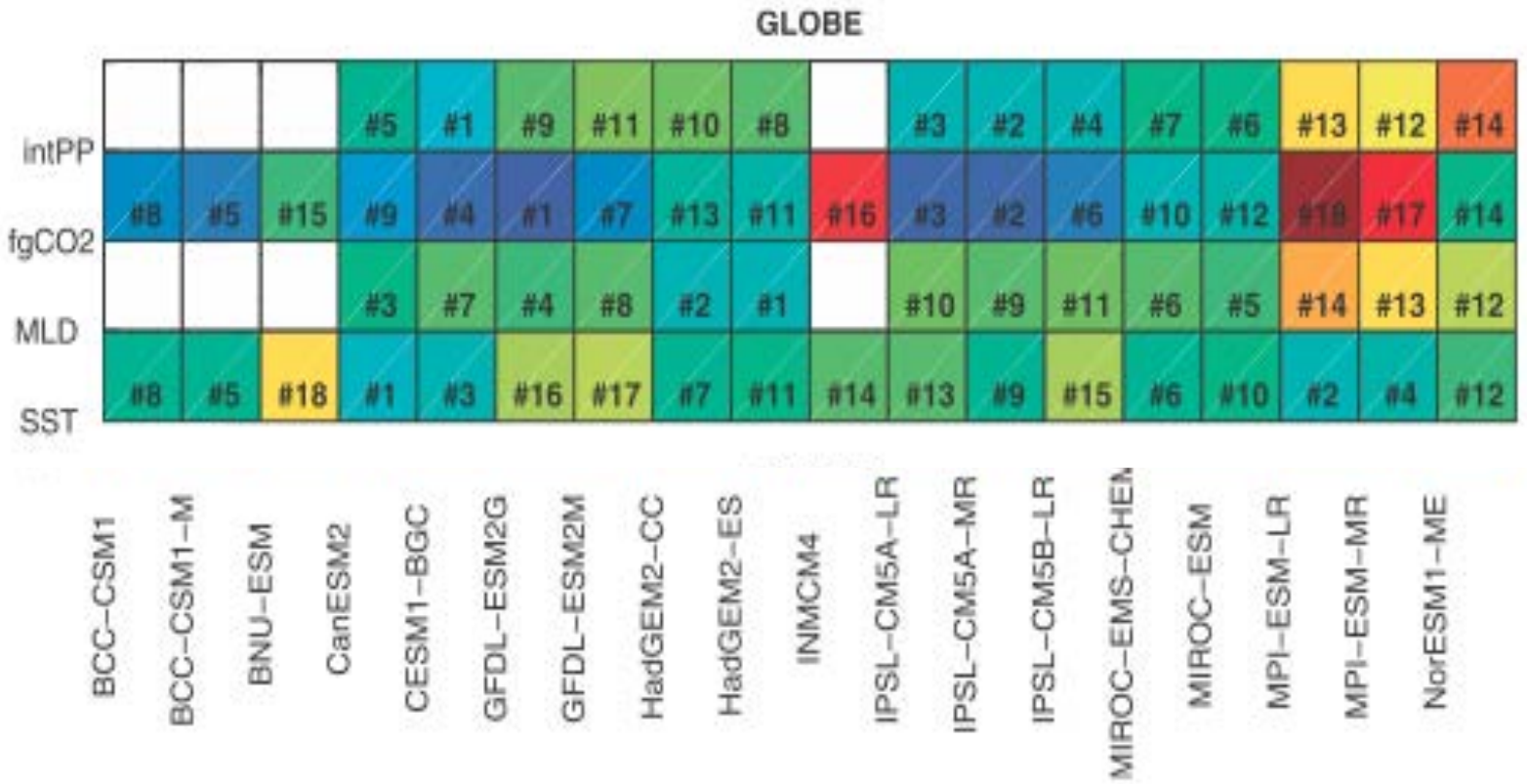
2009-onward (e.g., Frölicher et al., 2009, Steinarcher et al., 2011)
Ensemble model evaluation (cross evaluation) + model weighted solution





CMIP5: the age of the skill-score metrics…

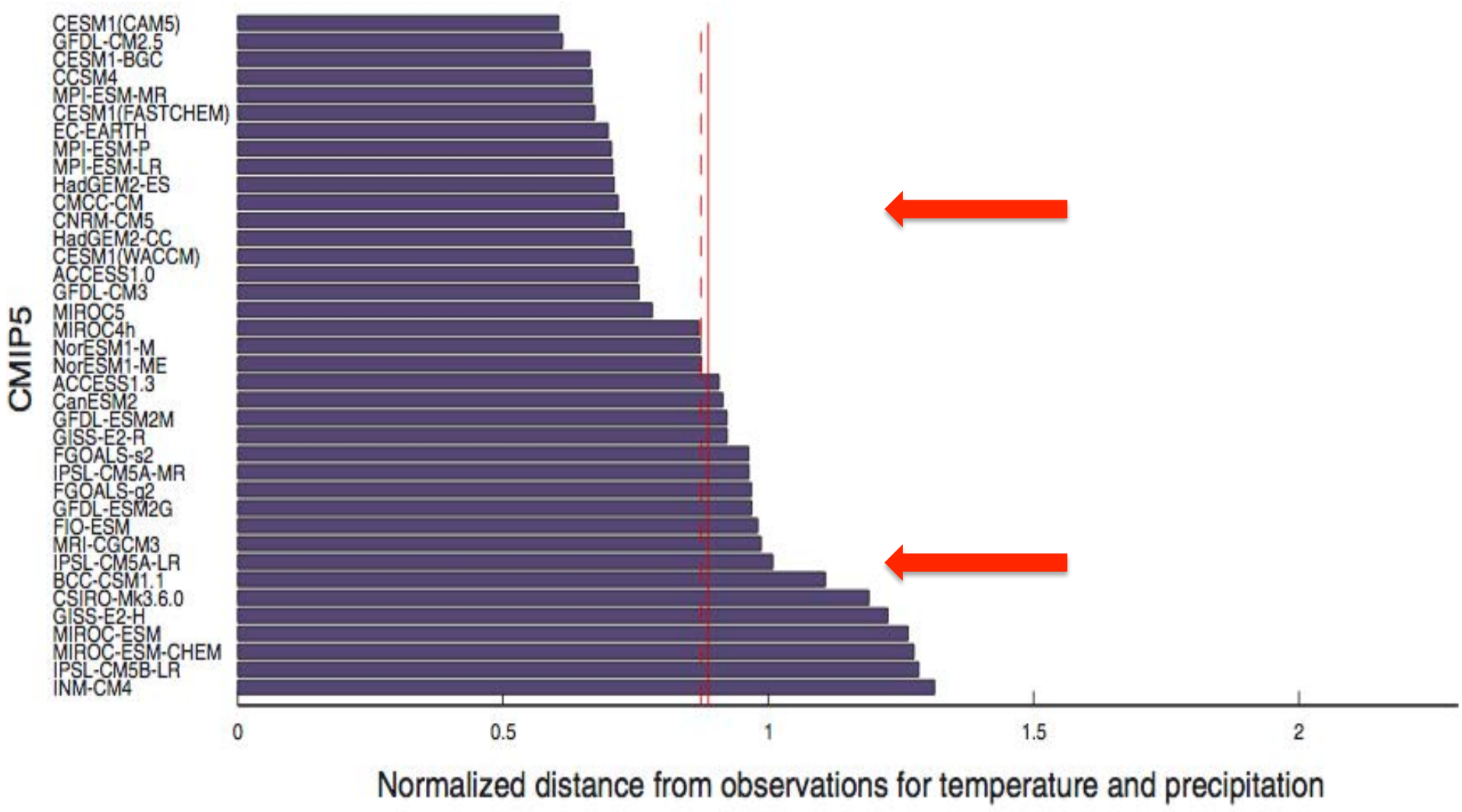
2012-onward (e.g., Anav et al., 2013)
Statistical metrics on seasonal cycle are used to rank models between each other



2013-onward (e.g., Cox et al., 2013, Wenzel et al., 2014, Massonet et al.)
Observational constrains as resonnable guess to weight model prediction

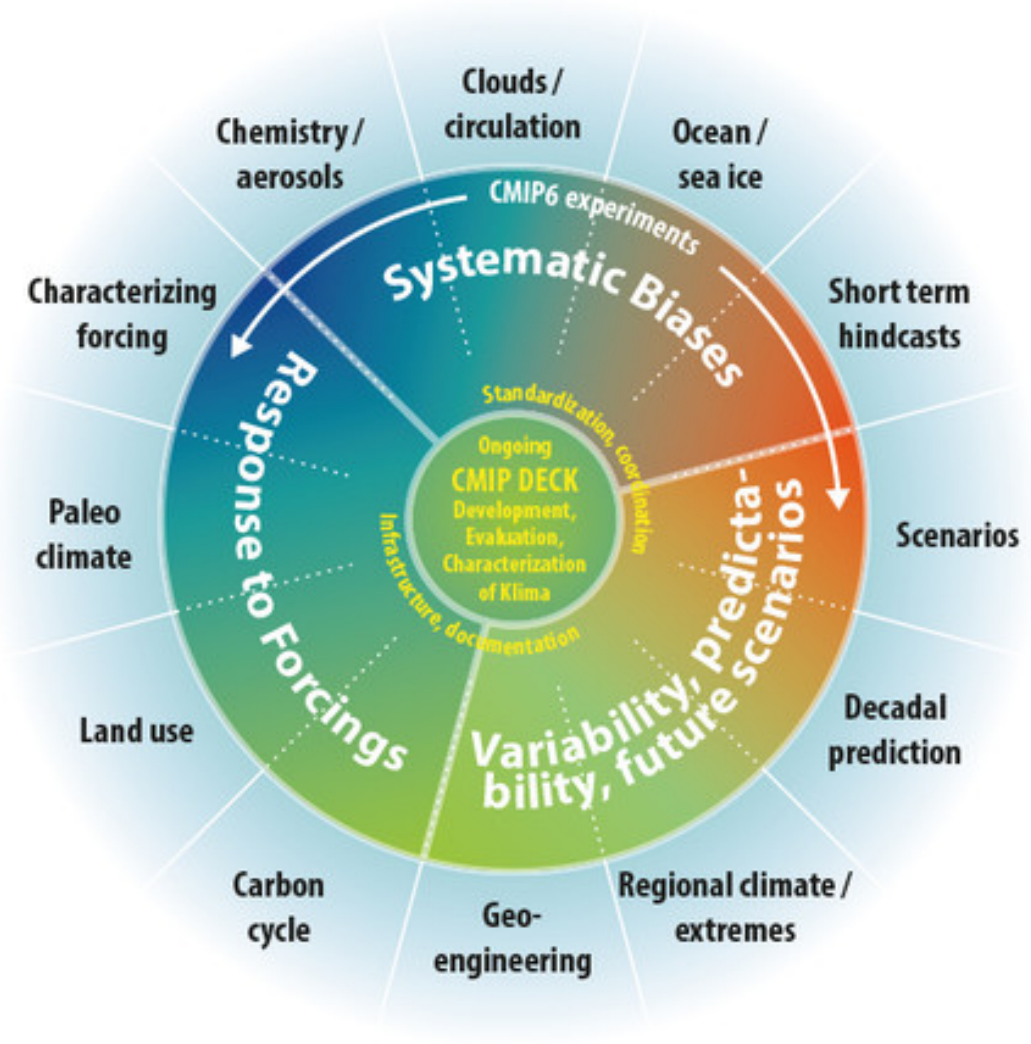
CMIP5: the age of the skill-score metrics…

2013-onward (e.g., Knutti et al., 2013)
Combination of of variables to rank models between each other



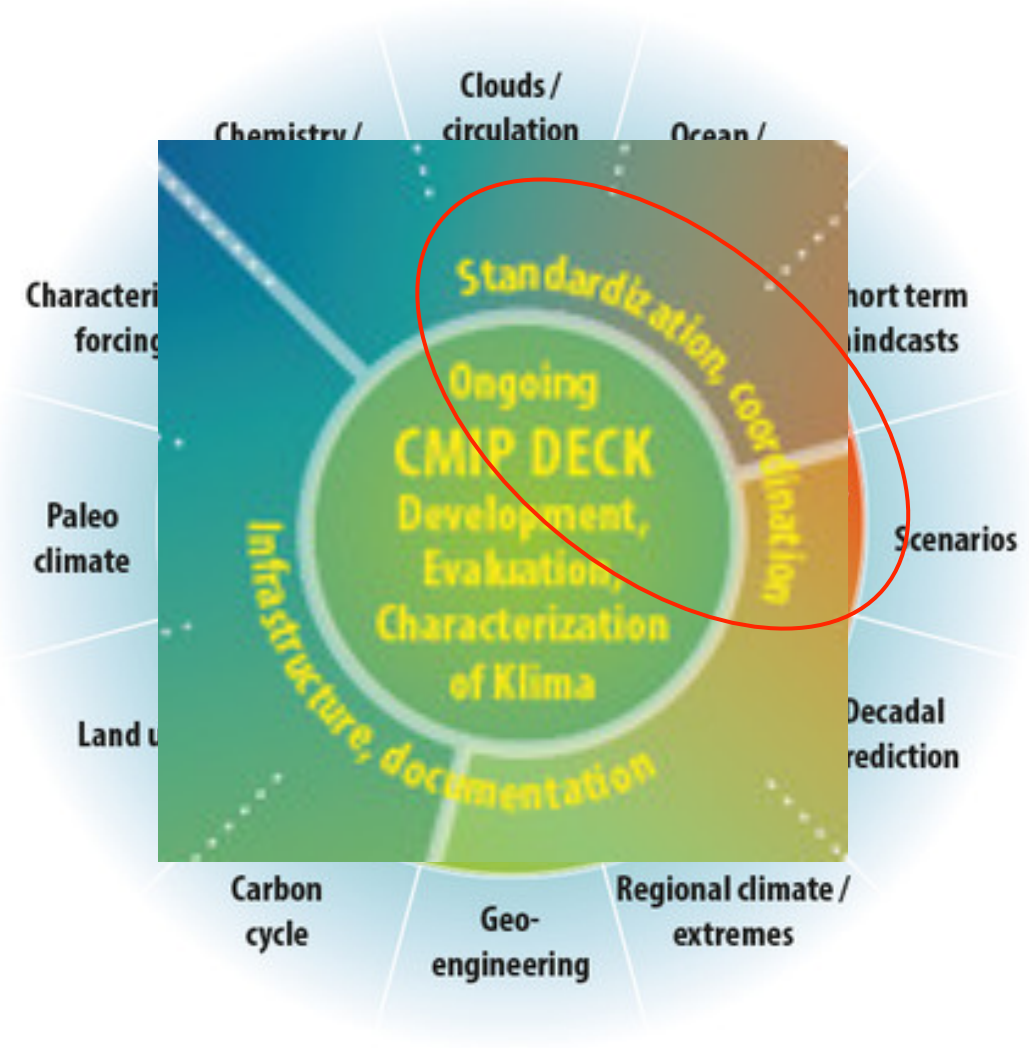
CMIP6: skill-score metrics climax…

2014-onward (e.g., Eyring et al., 2014)
Development of metrics package as unified framework to benchmark models (for **CMIP6**)

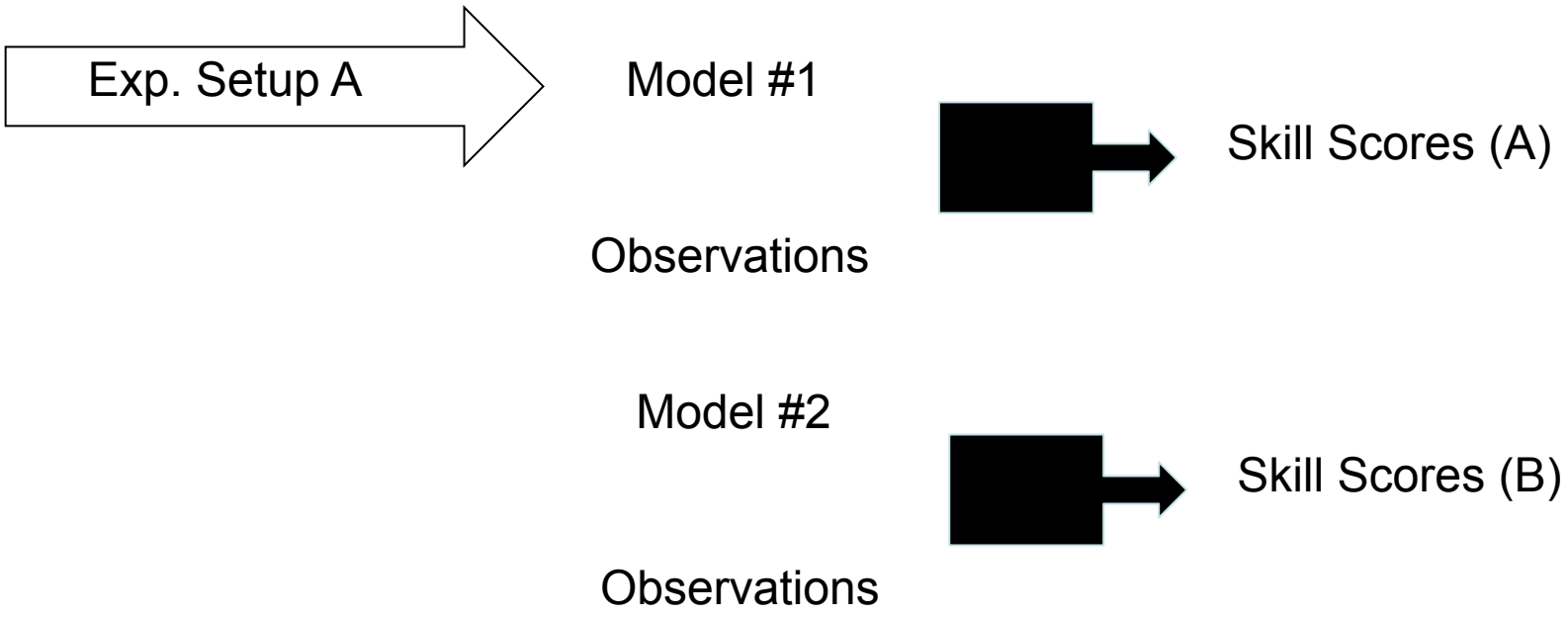


CMIP6: skill-score metrics climax…

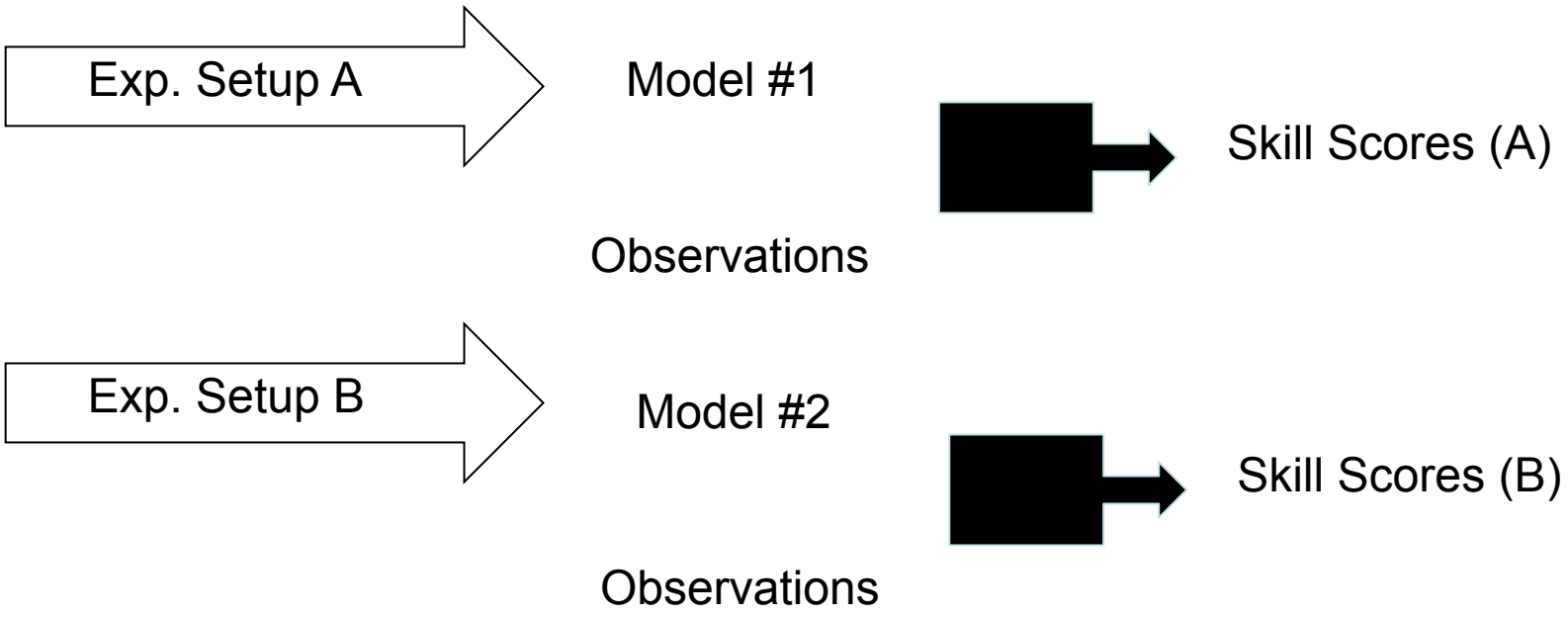
2014-onward (e.g., Eyring et al., 2014)
Development of metrics package as unified framework to benchmark models (for **CMIP6**)



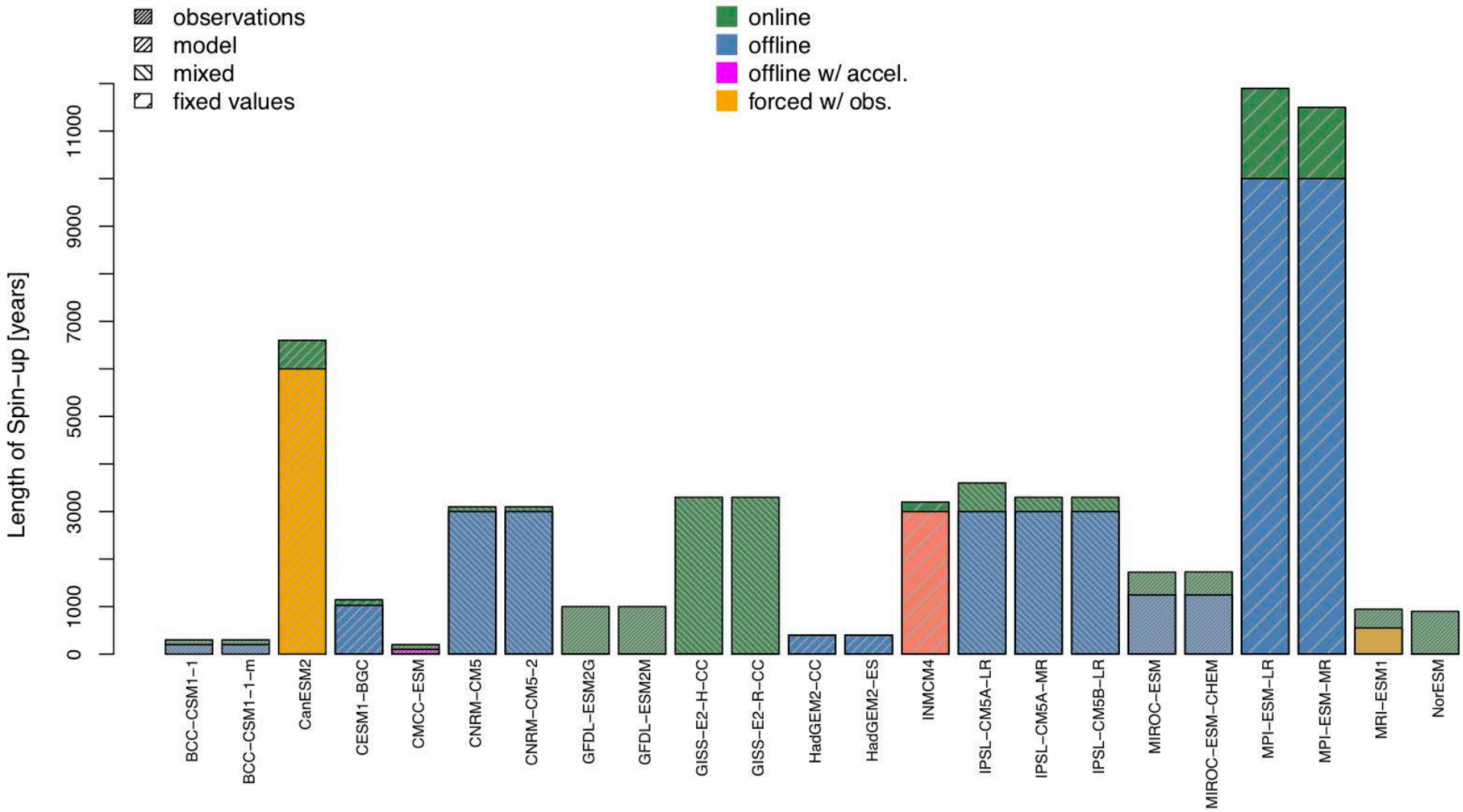
What do we call intercomparison ?



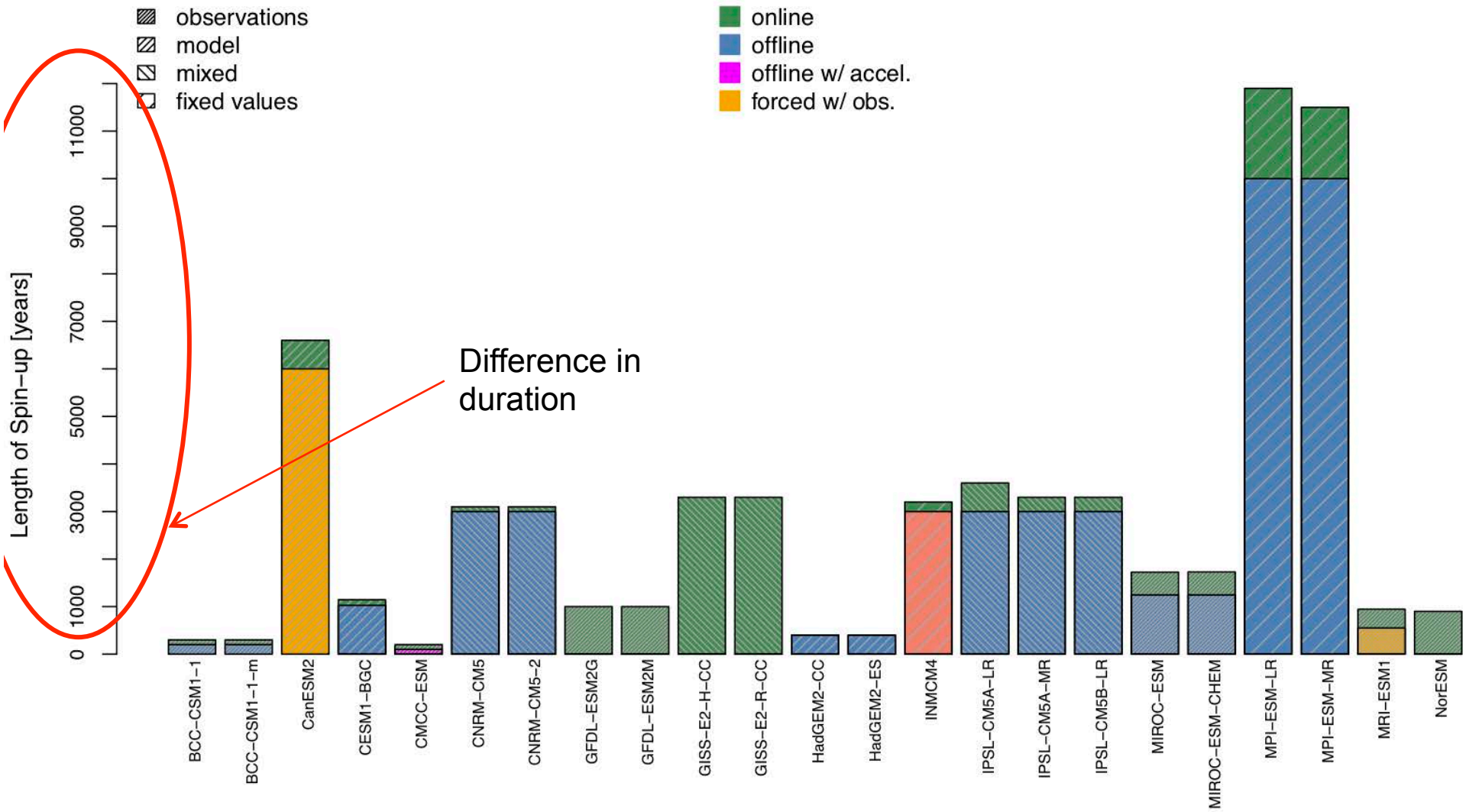
What do we call intercomparison ?



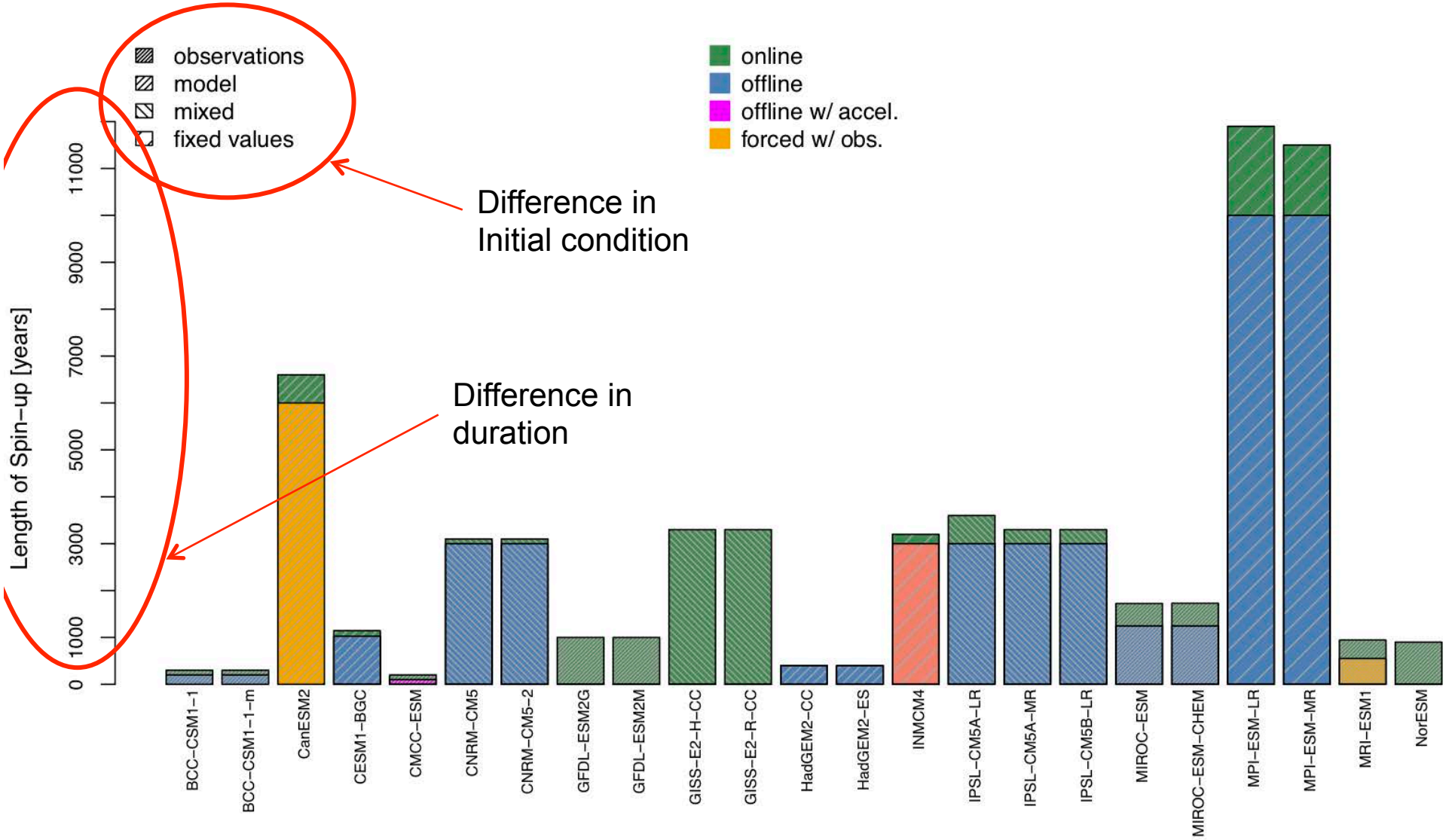
Can we really speak of intercomparison if experimental setup differs between models ?



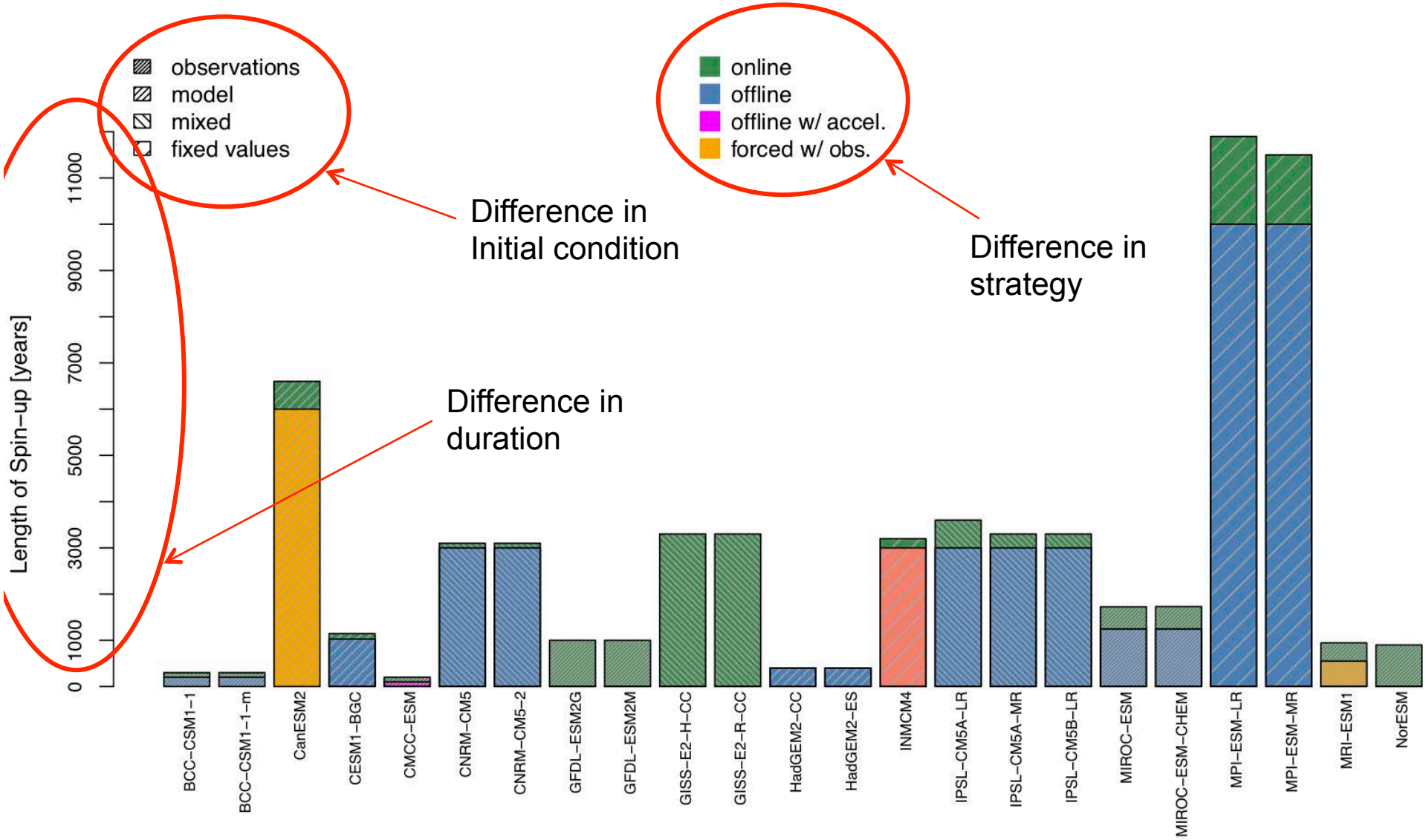
Can we really speak of intercomparison if experimental setup differs between models ?



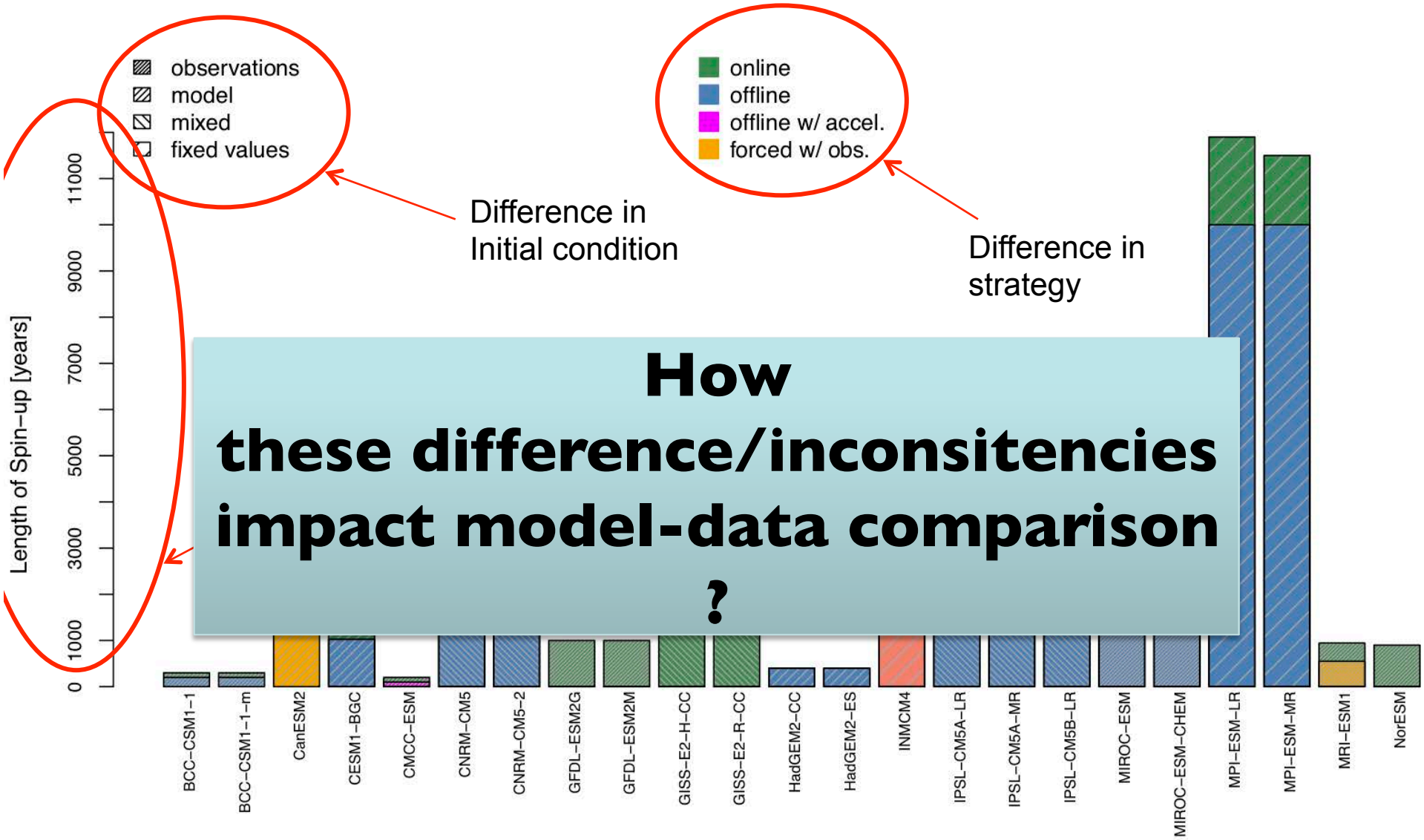
Can we really speak of intercomparison if experimental setup differs between models ?



Can we really speak of intercomparison if experimental setup differs between models ?



Can we really speak of intercomparison if experimental setup differs between models ?



How these difference/inconsitencies impact model-data comparison ?

Evaluating the impact of spin-up duration with IPSL-CM5A-LR

⇒ **No information available from metafor**

Parent_id: N/A

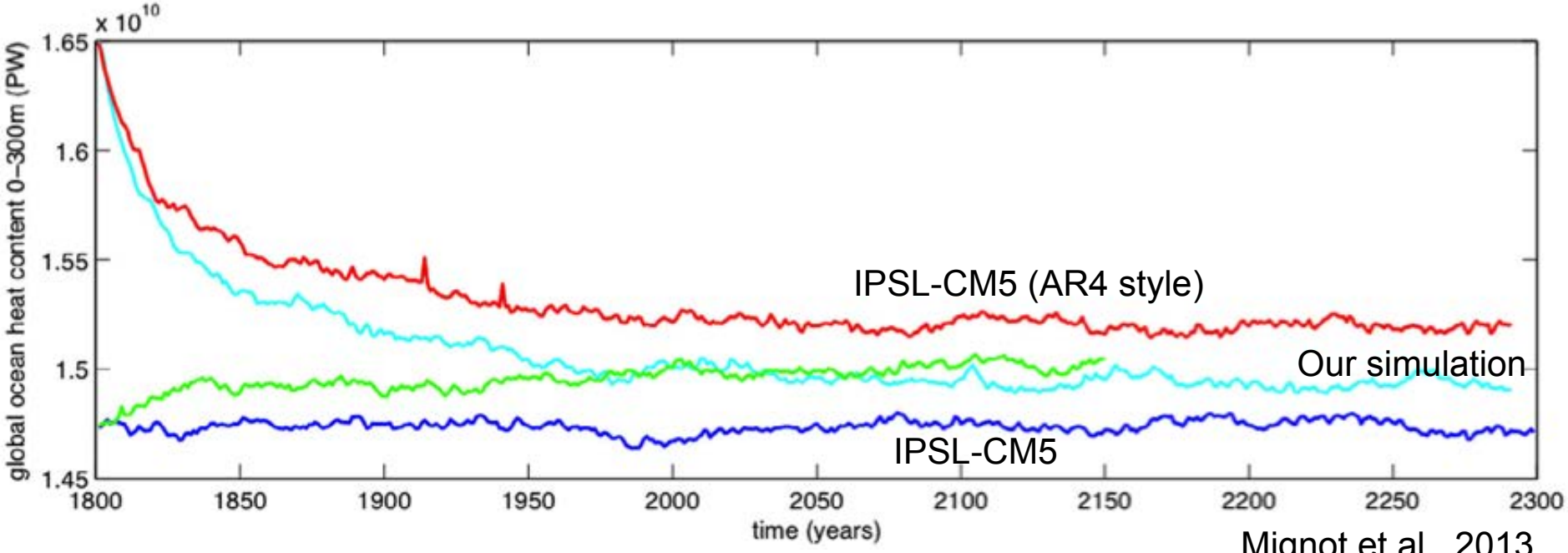
⇒ **No spin-up simulation distributed to the CMIP5 archive**

⇒ **Need to re-do simulation in a very naïve experimental setup:**

- Initialize model (IPSL-CM5A-LR) at rest with observations (WOA, GLODAP)
- Determine model skill-scores (correlation, biais, RMSE) along the spin-up time [500 yrs] with the same datasets

Focus on O_2 proxy of physical air-sea fluxes, circulation

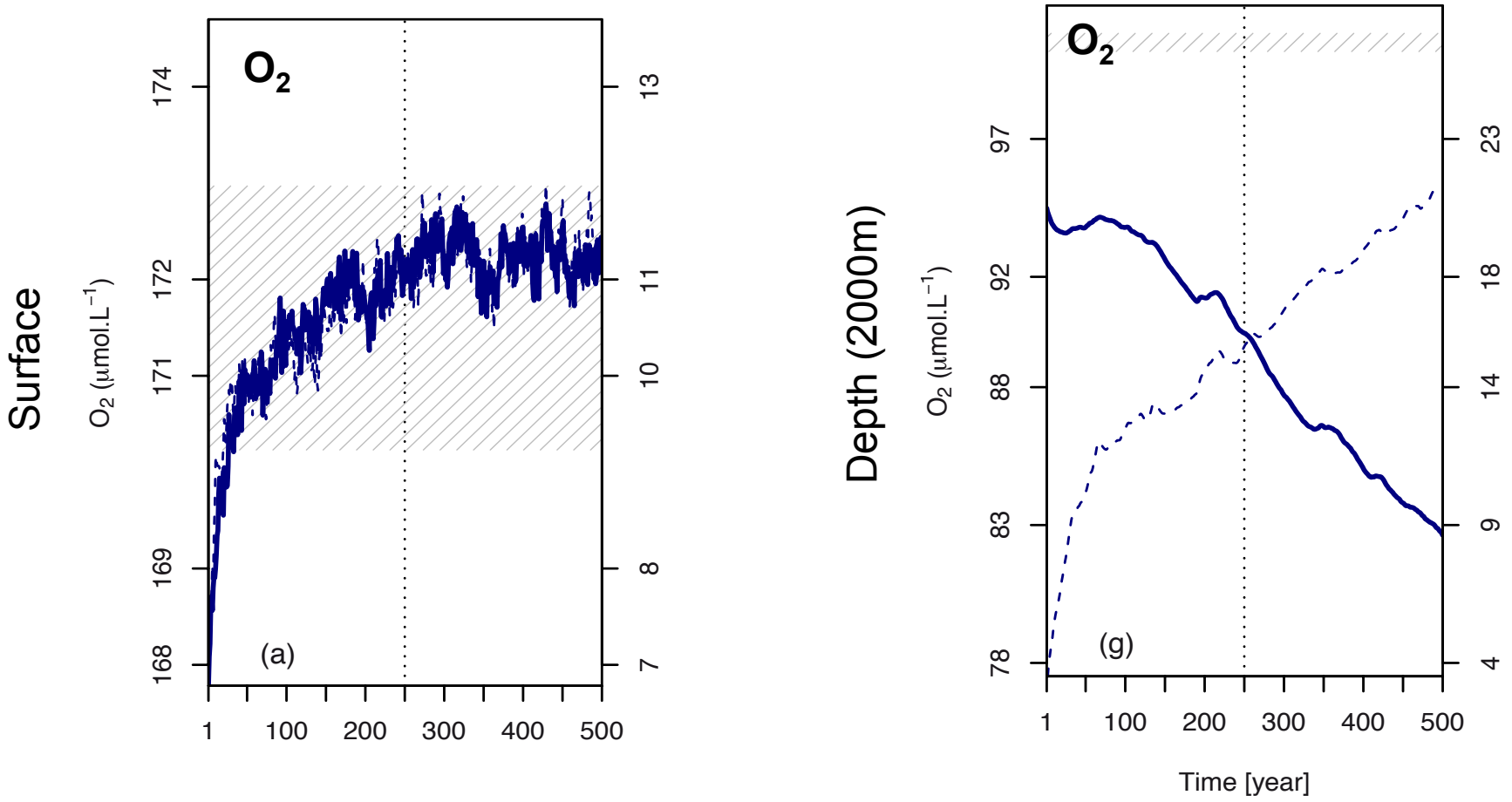
Evaluating the impact of spin-up duration with IPSL-CM5A-LR



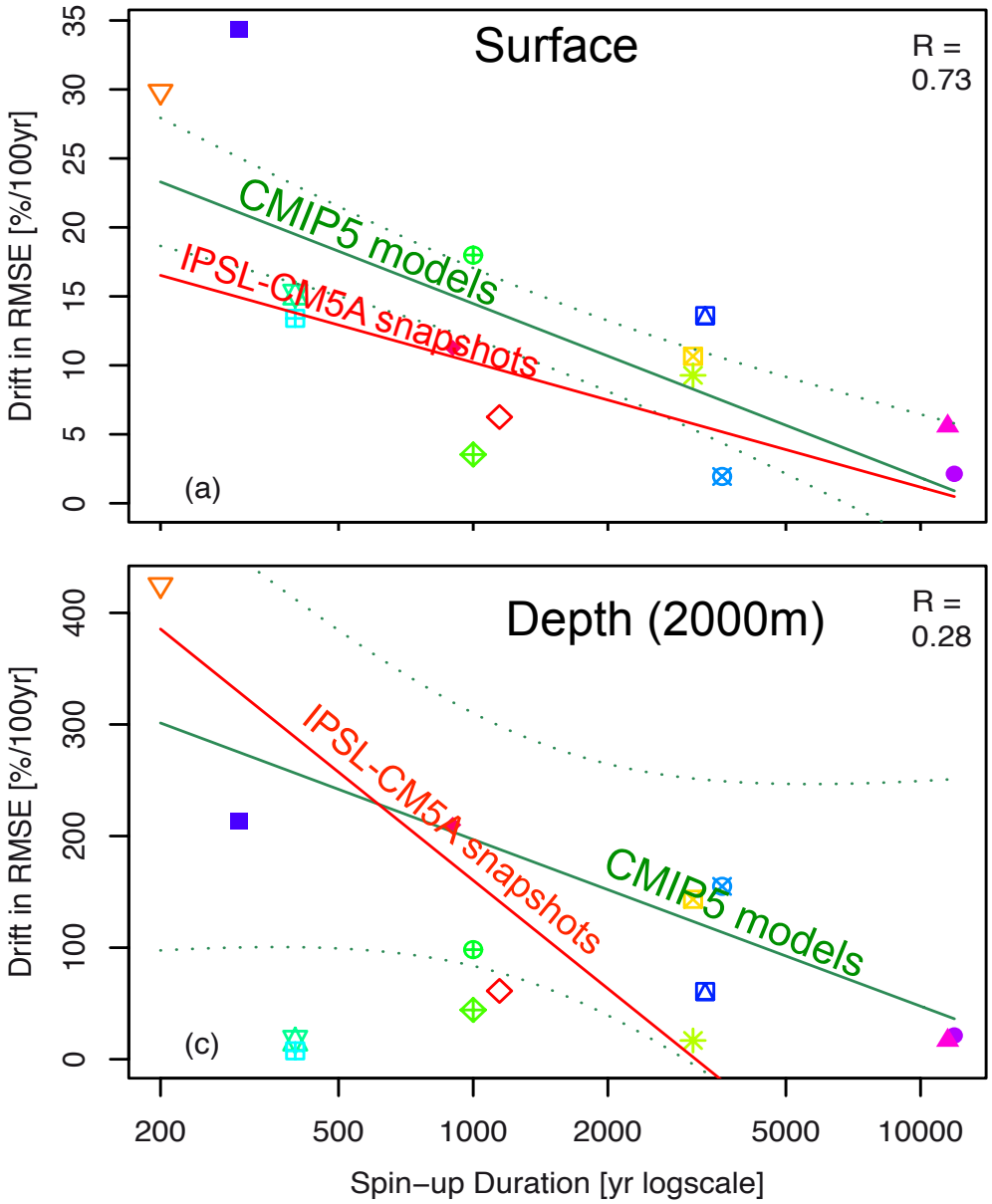
Mignot et al., 2013

⇒Drift in OHC weak and comparable to other CMIP5 models after 250 years of spin-up

Evaluating the impact of spin-up duration with IPSL-CM5A-LR



Tracking the drift in the CMIP5 archive

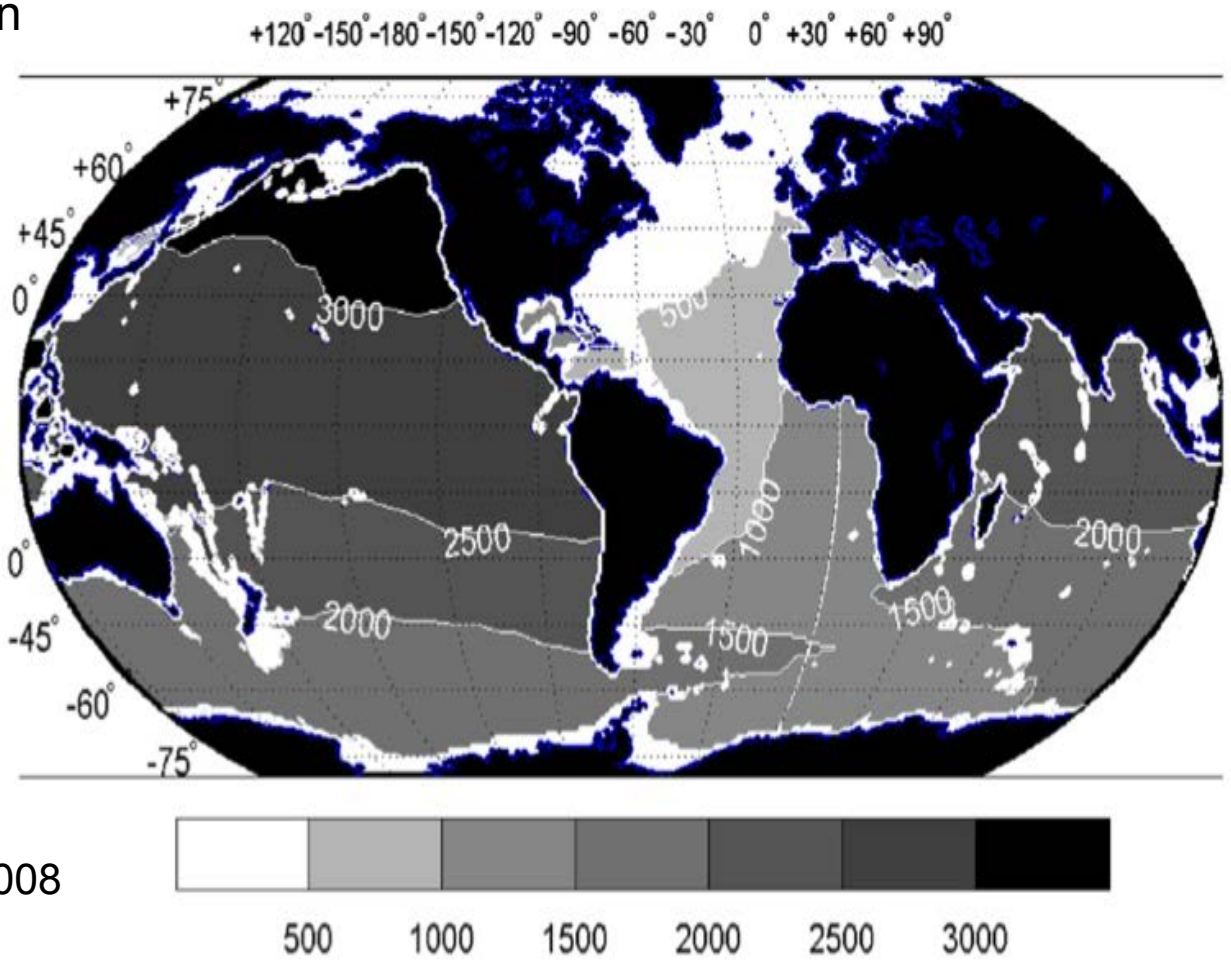


- CESM1−BGC
- CMCC−CESM
- CNRM−CM5
- CNRM−CM5−2
- GFDL−ESM2G
- GFDL−ESM2M
- HadGEM2−CC
- HadGEM2−ES
- IPSL−CM5A−LR
- IPSL−CM5A−MR
- IPSL−CM5B−LR
- MPI−ESM−LR
- MPI−ESM−MR

Tracking the drift in the CMIP5 archive

Not so surprising…
(1) Simple computation:
Ocean volume 3 x 10^{18} m^3
Deep water mass formation rate ~20 Sv
====
Mixing time of the ocean ~2000 years

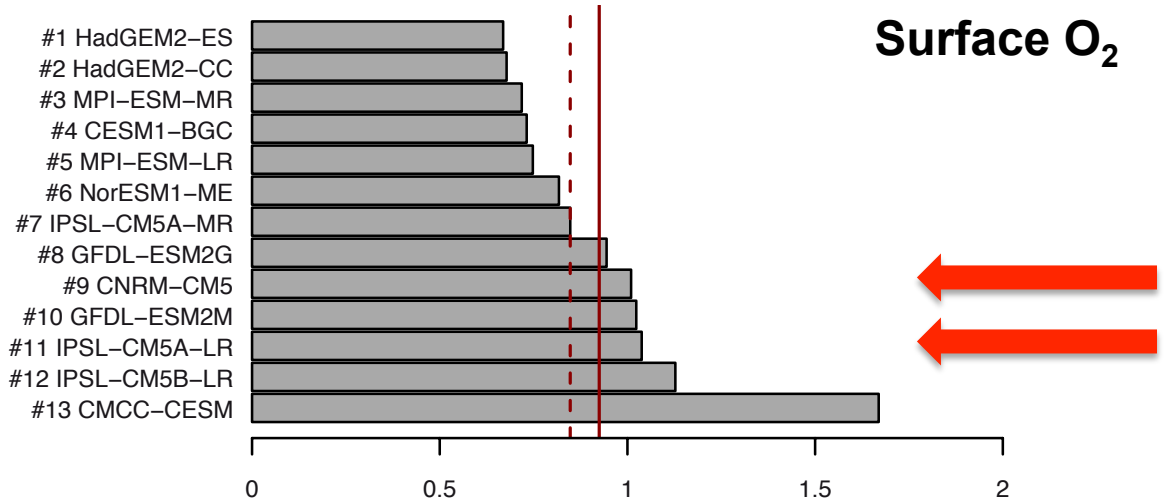
(2) Model simulation with data assimilation

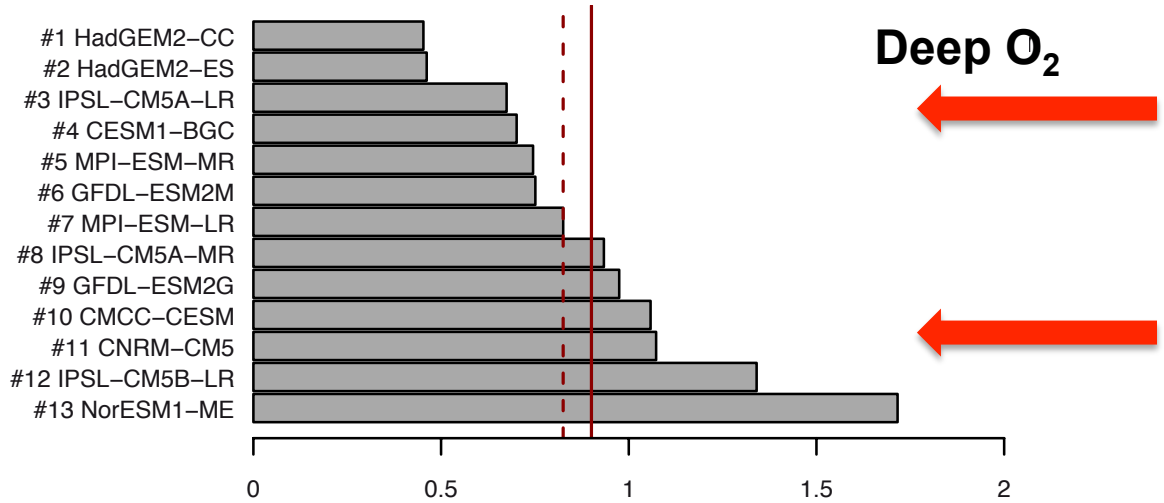


Wunch et al., 2008

Revisit model ranking accounting for model drift

Standard framework : $RMSE = E((m-o)^2)$

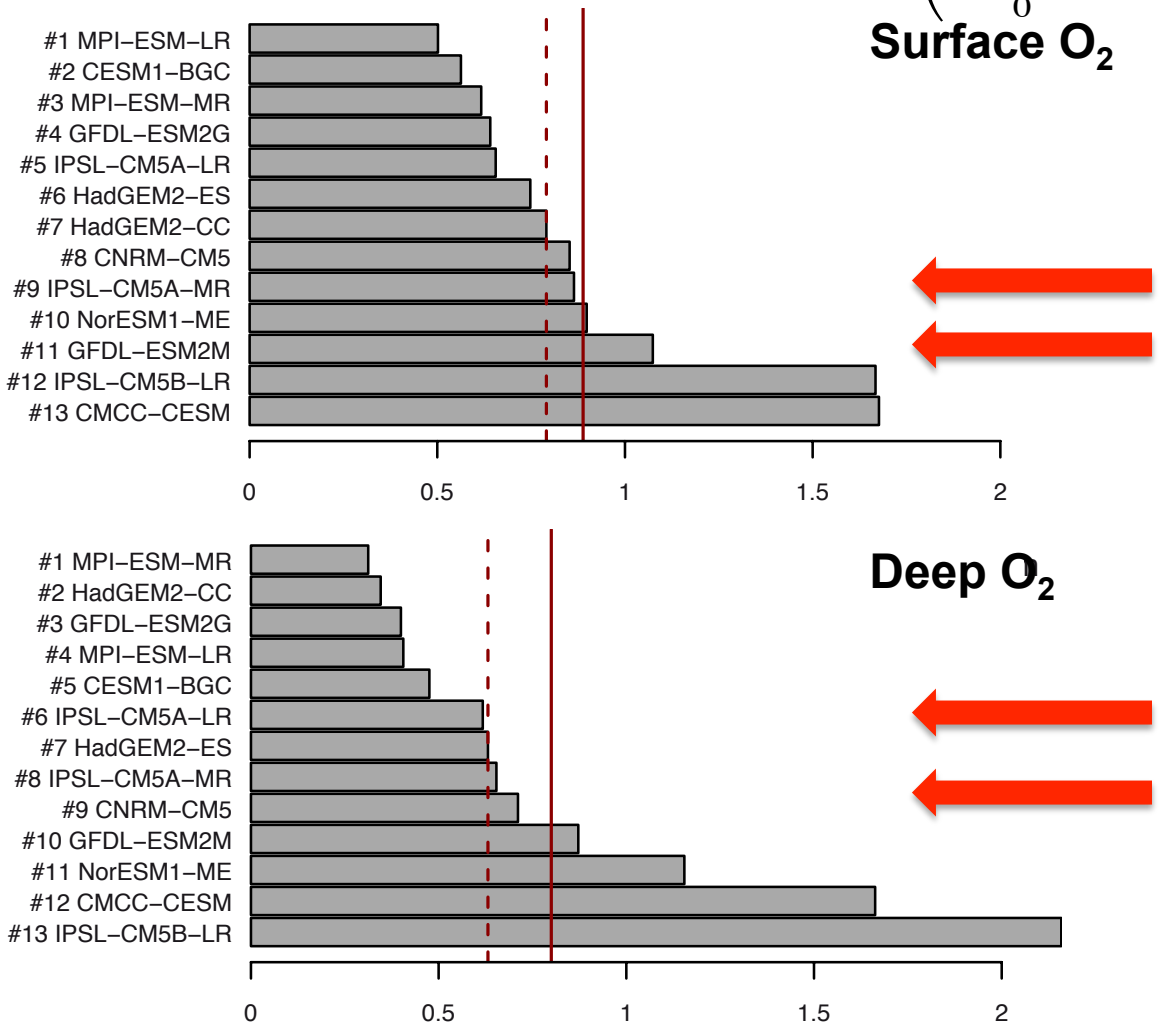




Normalized distance from observations for O_2

Revisit model ranking accounting for model drift

Penalized framework : $RMSE + \Delta RMSE\left(= \int_0^T drift(t=0) \times \exp(-\frac{1}{\tau} t)dt \right)$



Normalized distance from observations for O_2 – penalized with drift

Perspectives

⇒ **Need to define a common framework to run ocean/bgc simulations**

… as OCMIP2 (requiring 2000 years of spin-up simulation)

⇒ **Need to expand model metadata (no information on the spin-up is available on metafor)**

… Now: branchtime of piControl = N/A (not transparent at all !)

⇒ **Provide some recommendations for model weighing and model ranking**

… Skill score metrics are a 'snapshot' of the model and do not show is the model's fields are drifting or not…

⇒ **Further work will be done in CRESCENDO**

…use drifts to define confidence level on model results