Harnessing machine learning to quantify the CNRM-CM6-1 climate model parametric and structural errors

Romain Roehrig

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
contact: romain.roehrig@meteo.fr

1. Climate model calibration

$$\frac{\partial \boldsymbol{x}}{\partial t} = \mathcal{D}(\boldsymbol{x}) + \sum_p \mathcal{P}_p(\boldsymbol{x}, \boldsymbol{\lambda_p})$$

Climate model = a software
+ external forcings
+ horizontal/vertical grids
+ a scientific content (i.e., fluid dynamics equations, parameterizations)
+ values for model internal/uncertain parameters >> ***calibration***

Calibration (or tuning)

- Common to most modelling frameworks
- Can be seen as an optimisation procedure under constraints (or ***metrics***), possibly with priorities.
- Need for high-quality ***references/observations***, with well-quantified uncertainties
- Typically, +1 W m^{-2} at TOA ~ +0.5–1.5 K of global mean near-surface temperature
  - *Given current uncertainties, present-day global-mean temperature in a climate model is mostly a result of tuning.*

A bottleneck for climate model development

- ***High dimensionality*** of the parameter space ~O(10)
- Climate model numerical simulations are ***computationally expensive***
  - An exhaustive exploration of the parameter space is not directly possible.
- Large number and variety of metrics O(10-100++) , sometimes subjective
- ***Overfitting*** issue, treatment of ***uncertainties***

Calibration of CNRM-CM6-1 (Voldoire et al. 2019, Roehrig et al. 2020)

- Manual calibration, 1 or 2 parameters at the same time, mixing well-defined metrics and more subjective considerations
- Calibration of stand-alone components before coupling, priorities among metrics
- Often questioning the model physical content. But difficult to disentangle true model structural limits from "just" a poor calibration?

2. Framework

i. Define targeted (scalar) ***metrics*** f, their ***reference values*** r_f and associated ***uncertainties*** $\sigma_{r,f}$
ii. Identify the relevant model ***parameters*** $\boldsymbol{\lambda}$, and their "acceptable" range >> ***input parameter space*** Λ
iii. Define a simulation strategy, build an experimental design, run simulations >> ***learning dataset***
iv. ***Emulate*** $f(\boldsymbol{\lambda})$ for each metric (Gaussian Processes)
v. Identify the sub-space of Λ which is compatible with references for all metrics
>> ***Not-Ruled-Out-Yet – NROY – space***
considering
- The reference uncertainty
- The emulator uncertainty
- The model structural error (tolerance to error) $\sigma_{d,f}$

>> ***Implausibility*** measure I_f, ***cutoff*** T

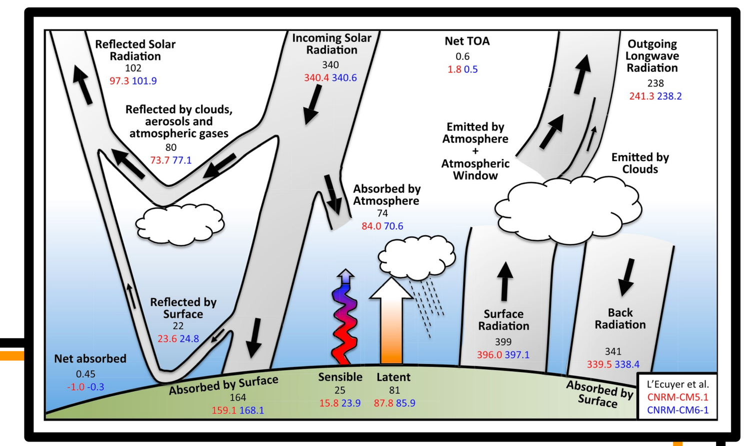
$$I_f(\boldsymbol{\lambda}) = \frac{|r_f - \mathrm{E}\left[f(\boldsymbol{\lambda})\right]|}{\sqrt{\sigma_{r,f}^2 + \sigma_{d,f}^2 + \mathrm{Var}\left[f(\boldsymbol{\lambda})\right]}}.$$

$$\mathrm{NROY}_f^1 = \{\lambda \mid I_f(\lambda) < T\}$$

$$\mathrm{NROY}^1 = \bigcap_f \mathrm{NROY}_f^1 = \{\lambda \mid I_f(\lambda) < T, \text{for all } f\}$$

vi. Iterate over several waves to reduce the emulators' uncertainty in NROY^{N-1}, until convergence

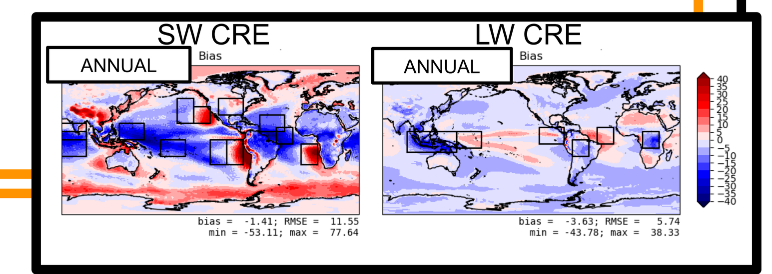
Couvreux et al. (2020), Hourdin et al. (2020) and reference therein



i. 3 classes of metrics

- ***Global averages of the energy budget components***
  - at TOA: OLR, OSR, Net, SW/LW CRE / at the surface ocean: Net, SWdn, LWdn
  - Values from CERES-EBAF, uncertainties based on the literature
  - Except Net at surface/TOA = 0 +/- 0.1 W m^{-2} : the model has to be equilibrated.
  - Tolerance to error: 0.5 W m^{-2}
- ***Zonally average profiles*** of SW/LW CRE + Temperature at 200 hPa
  - SW/LW CRE: CERES-EBAF with uncertainty of 2 W m^{-2} , tol. of 1 W m^{-2}
  - T200: based on ERA5/JRA55/MERRA/CFSR ensemble mean and std, tol. 1.5 K
- ***Regional and seasonal averages*** of SW/LW CRE and precipitation
  - SW/LW CRE: CERES-EBAF, uncertainty of 2 W m^{-2}, tolerance of 5 W m^{-2}
  - Precipitation: MSWEP/GPCP/TRMM 3B42 ensemble mean and std, tol. between 0.5 and 1 mm day^{-1}

>> ***63 metrics***

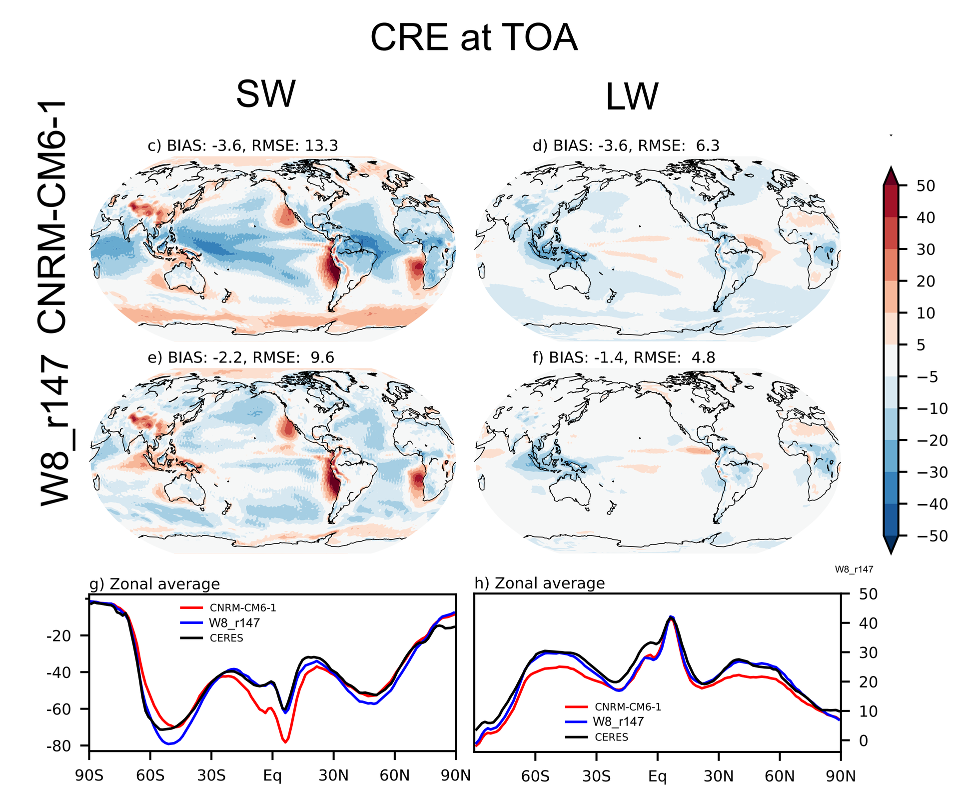


ii. 46 parameters

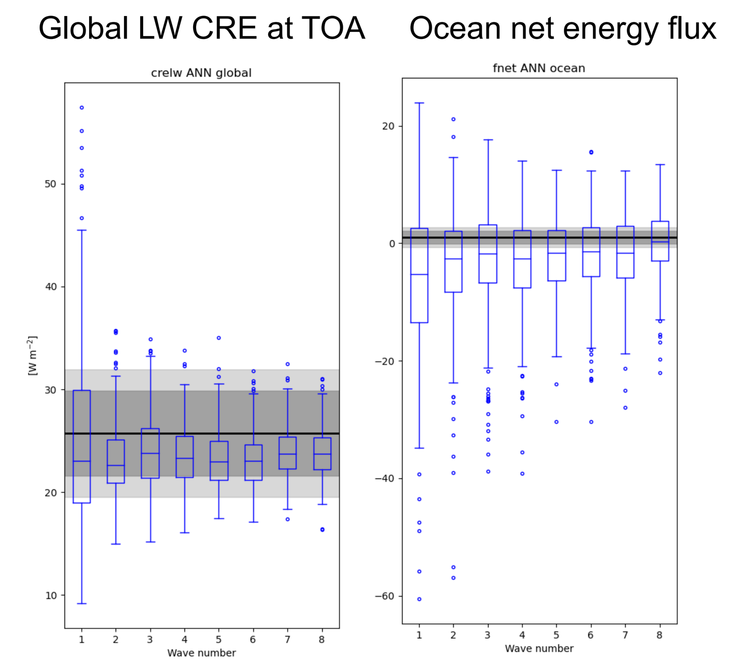
- 7 from turbulence (TKE scheme + PBL-top entrainment)
- 16 from microphysics (1-moment, 5 hydrometeors)
- 19 from the unified dry, shallow and deep convection scheme
- 4 from cloud radiative properties (heterogeneity)

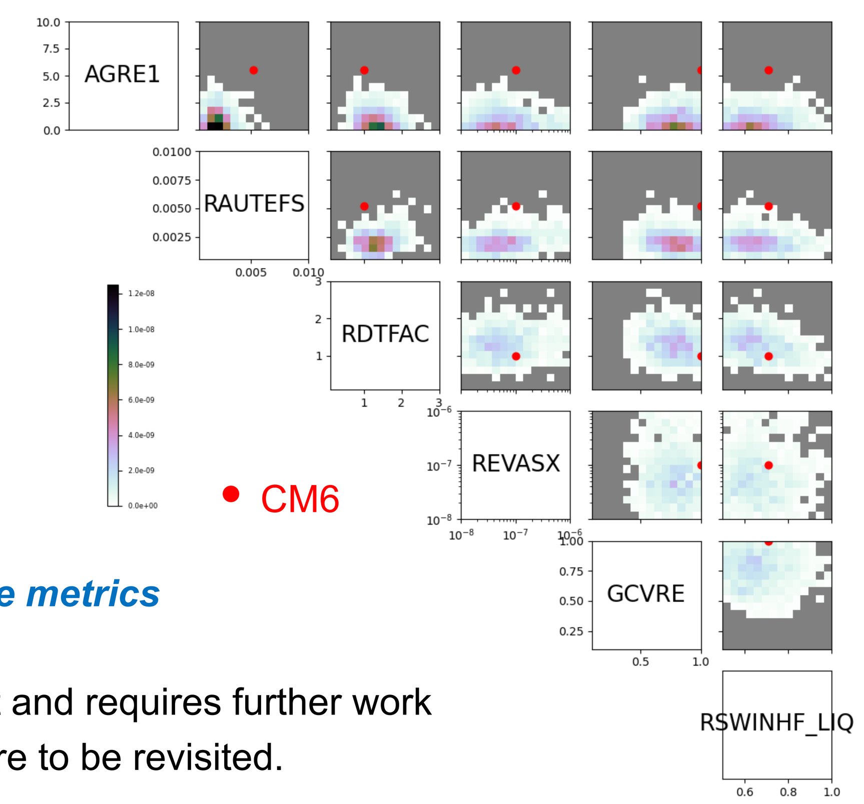
iii. Waves of 400 simulations

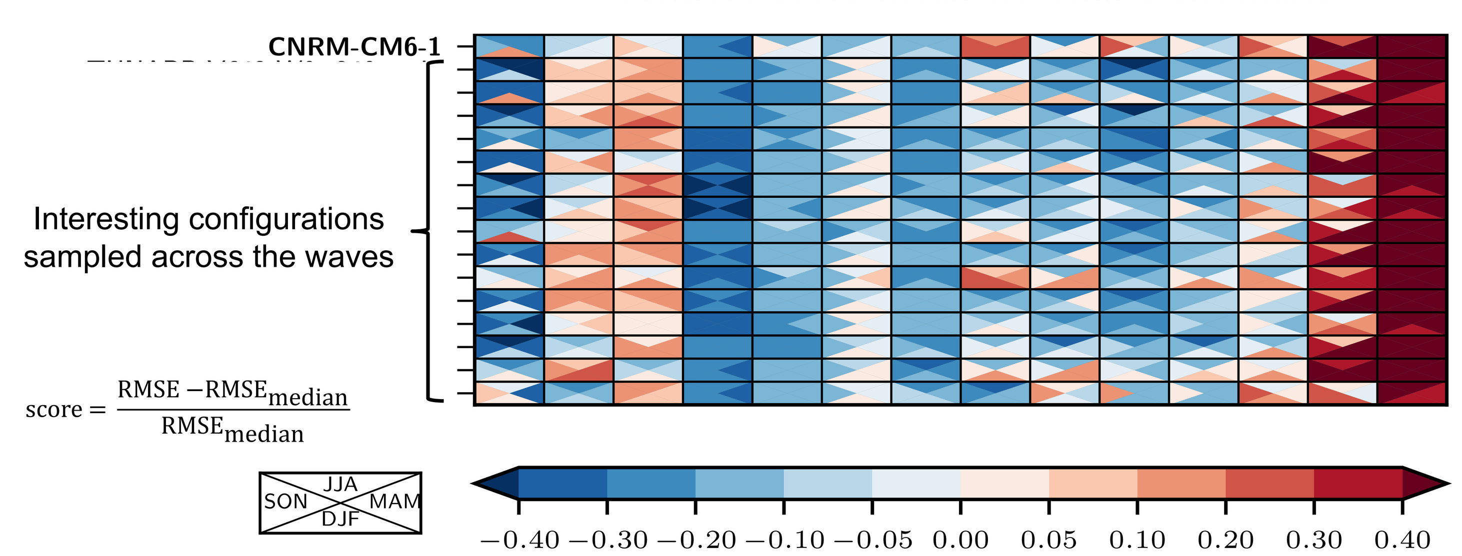
- 1-year ***sstclim*** simulations + 3-month spin-up
- sstclim vs amip correction of the reference target
- Consideration of ***internal variability uncertainty*** based on a 100-year sstclim simulation with CNRM-CM6-1.
- Latin Hypercube sampling for 1st wave



3. Toward a new calibration of CNRM-CM6-1







Unfortunately, none of Wave 1-8 simulations fulfils all the metrics

- Convergence is not yet achieved
- Appropriate sampling of small NROY spaces is difficult and requires further work
- Some tolerances to error are likely too weak and require to be revisited.

Nevertheless

- A few simulations fulfil all the metrics but one (for a cutoff of 3)
- A few have interestingly low RMSEs for targeted variables

4. Conclusions

- ***A better calibration of CNRM-CM6 can be achieved***
  - Several mean state features are ***improved or of similar quality***
  - Some errors truly ***structural***: clouds/radiation over eastern part of ocean basins, upper-tropospheric temperatures
  - Some ***trade-offs*** are required
- ***History matching with iterative refocussing***
  - Provides a ***relevant and efficient framework for model calibration in the presence of uncertainties***
  - Can help accelerate model development by comparing calibrated model version
    - *Assessing the true added value of a new development*
  - Helps better ***identify and quantify model structural errors***, and thereby helps focus on bias understanding/model development
  - ***Overall questions the scientific content of a climate model***
- ***Next steps***
  - Play with tolerances to error to better identify/quantify model ***structural errors*** and trade-offs
  - Add new metrics, e.g., variability: can we get both mean state and variability right?
  - ***Pre-conditioning*** with cheaper model configurations, e.g., 1D/LES for preserving process-level performance (Couvreux et al. 2020, Hourdin et al. 2020).
  - Towards calibration of ***ocean-atmosphere coupled configurations***
  - Develop ***physical interpretations*** of what is happening in the calibration process.

References

Couvreux *et al.* 2021: Process-based climate model development harnessing machine learning: 1. A calibration tool for parameterization improvement. *JAMES,* doi: 10.1029/2020MS002217
Hourdin *et al.*, 2020: Process-based climate model development harnessing machine learning: 2. Model calibration from single column to global. *JAMES,* doi: 10.1029/2020MS002225
Roehrig *et al.*, 2020: The CNRM Global Atmosphere Model ARPEGE-Climat 6.3: Description and Evaluation. *JAMES*, doi: 10.1029/2020MS002075
Voldoire *et al.*, 2019: Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1. *JAMES*, doi: 10.1029/2019MS001683.

Acknowledgements

This study has received funding from Agence Nationale de la Recherche - France 2030 as part of the PEPR TRACCS programme under grants number ANR-22-EXTR-0007 and ANR-22-EXTR-0008