# Addressing the calibration bottleneck using machine learning: Application to the CNRM-CM6-1 model

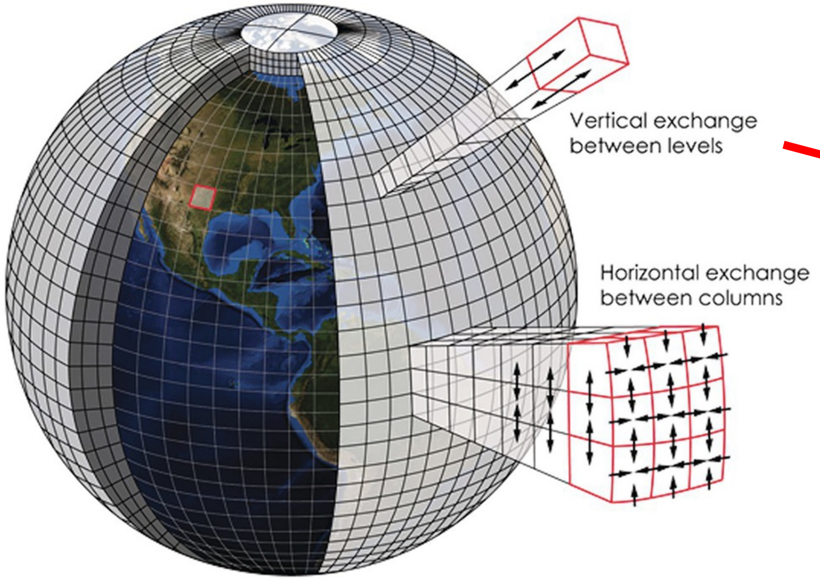
**Romain Roehrig**

CNRM, Météo-France and CNRS, Toulouse, France

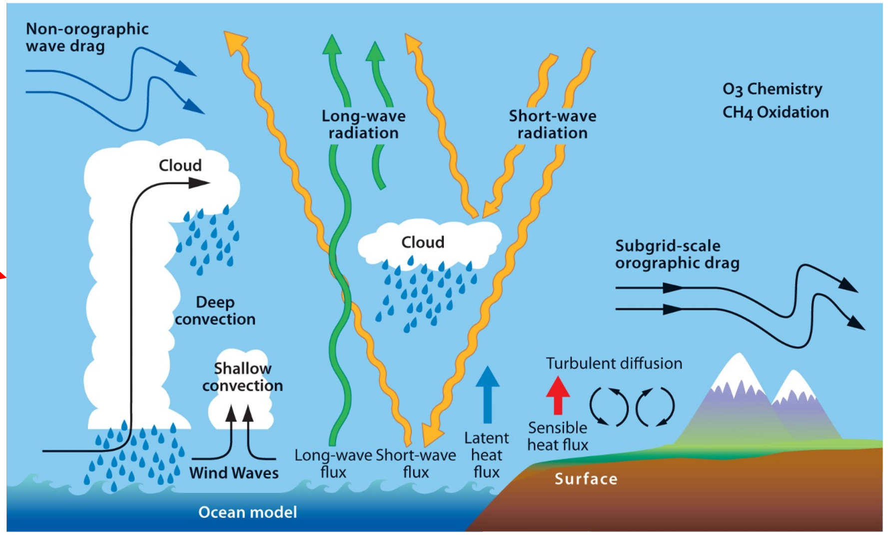
With contributions from D. Williamson, F. Couvreux, F. Hourdin, O. Audouin, J. Salter, N. Villefranque, V. Volodina, W. Xu…

Workshop on AI for Ocean, Atmosphere and Climate Dynamics, 11 April 2023, Brest, France

# Climate models - basics




$$\frac{\partial \boldsymbol{x}}{\partial t} = \mathcal{D}(\boldsymbol{x}) + \sum_{p} \mathcal{P}_p(\boldsymbol{x})$$




***Developing a climate model:***

- Choose a set of equations issued from fluid dynamics, thermodynamics, with some approximations.
- Discretise them on the sphere to solve them numerically
- Thereby introduce a model resolution:
  - ➢ some processes are explicitly resolved, some not >> parameterizations
  - ➢ note some processes need to be parameterized whatever the resolution (e.g., radiation, microphysics)
- Do it for each component of the climate system and couple individual models together

# Climate model calibration

***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***

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

# Climate model calibration

***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***

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

## *Calibration (or tuning)*

- Common to most modelling frameworks
- Can be seen as an optimisation procedure under constraints (or ***metrics***), possibly with priorities/weights.
- Need for high-quality ***references/observations***, with well-quantified uncertainties
- +1 W m$^{-2}$ at TOA ~ +0.5–1.5 K of global mean near-surface temperature (*Hourdin et al. 2017*).
    - ➢ *Given current uncertainties, present-day global-mean temperature in a climate model is mostly a result of tuning.*

# Climate model calibration

***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***

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

## *Calibration (or tuning)*

- Common to most modelling frameworks
- Can be seen as an optimisation procedure under constraints (or ***metrics***), possibly with priorities/weights.
- Need for high-quality ***references/observations***, with well-quantified uncertainties
- +1 W $m^{-2}$ at TOA ~ +0.5–1.5 K of global mean near-surface temperature (*Hourdin et al. 2017*).
  - ➢ *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***

# Climate model calibration

***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***

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

## *Calibration (or tuning)*

- Common to most modelling frameworks
- Can be seen as an optimisation procedure under constraints (or ***metrics***), possibly with priorities/weights.
- Need for high-quality ***references/observations***, with well-quantified uncertainties
- +1 W $m^{-2}$ at TOA ~ +0.5–1.5 K of global mean near-surface temperature (*Hourdin et al. 2017*).
  - ➢ *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?

# A rationale for addressing the calibration bottleneck

***History matching***

- Determine the plausible sets of model parameter values rather than optimize
- Or equivalently rule out the ***implausible*** sets of model parameter values
- For a given set of ***quantitative metrics***
- Considering ***reference/observations uncertainties***
- And introducing priors for ***model structural errors*** (interpreted here as tolerances to error)

***Machine learning***

- ***Emulate*** the model behaviour (i.e. the dependence of metrics to model parameters), to explore at very weak cost the parameter space.
- Consider also the ***emulators' uncertainty***
- ***Gaussian processes*** nicely provide predictions with an uncertainty estimate

***Iterative refocussing***

- Be as ***parsimonious*** as possible in terms of true simulations
- Start with a 'few' simulations and progressively add new ones to improve the emulator quality, but only where it is needed
- Possibly add new metrics along the way, based on more expansive simulations (pre-conditioning with cheaper configurations)

➢ *A **formalized calibration procedure**, transparent and reproducible.*
➢ *A semi-automatic and efficient framework, that allows us to **quantify the true benefit of a new development** more rapidly*
➢ *More **rigorous comparison** between parameterizations or model physics.*
➢ *Possibly not a single acceptable configuration but several: being able to explore the **model parametric uncertainty** of its emergent properties.*

*Bower et al. (2010), Williamson et al. (2015, 2017)*

# The technical framework

1. Define targeted (scalar) ***metrics*** $f$, their ***reference values*** $r_f$ and associated ***uncertainties*** $\sigma_{r,f}$
2. Identify the relevant model ***parameters*** $\boldsymbol{\lambda}$, and their "acceptable" range >> ***input parameter space*** $\Lambda$
3. Define a simulation strategy, build an experimental design, run simulations >> ***learning dataset***
4. ***Emulate*** $f(\boldsymbol{\lambda})$ for each metric (Gaussian Processes)
5. Identify the sub-space of $\Lambda$ 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*** $T$

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

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

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

*Based on HTExplo (Couvreux et al. 2020)*

*Bower et al. (2010), Williamson et al. (2015, 2017)*

# The technical framework

1. Define targeted (scalar) ***metrics*** $f$, their ***reference values*** $r_f$ and associated ***uncertainties*** $\sigma_{r,f}$
2. Identify the relevant model ***parameters*** $\boldsymbol{\lambda}$, and their "acceptable" range >> ***input parameter space*** $\Lambda$
3. Define a simulation strategy, build an experimental design, run simulations >> ***learning dataset***
4. ***Emulate*** $f(\boldsymbol{\lambda})$ for each metric (Gaussian Processes)
5. Identify the sub-space of $\Lambda$ 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*** $T$

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

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

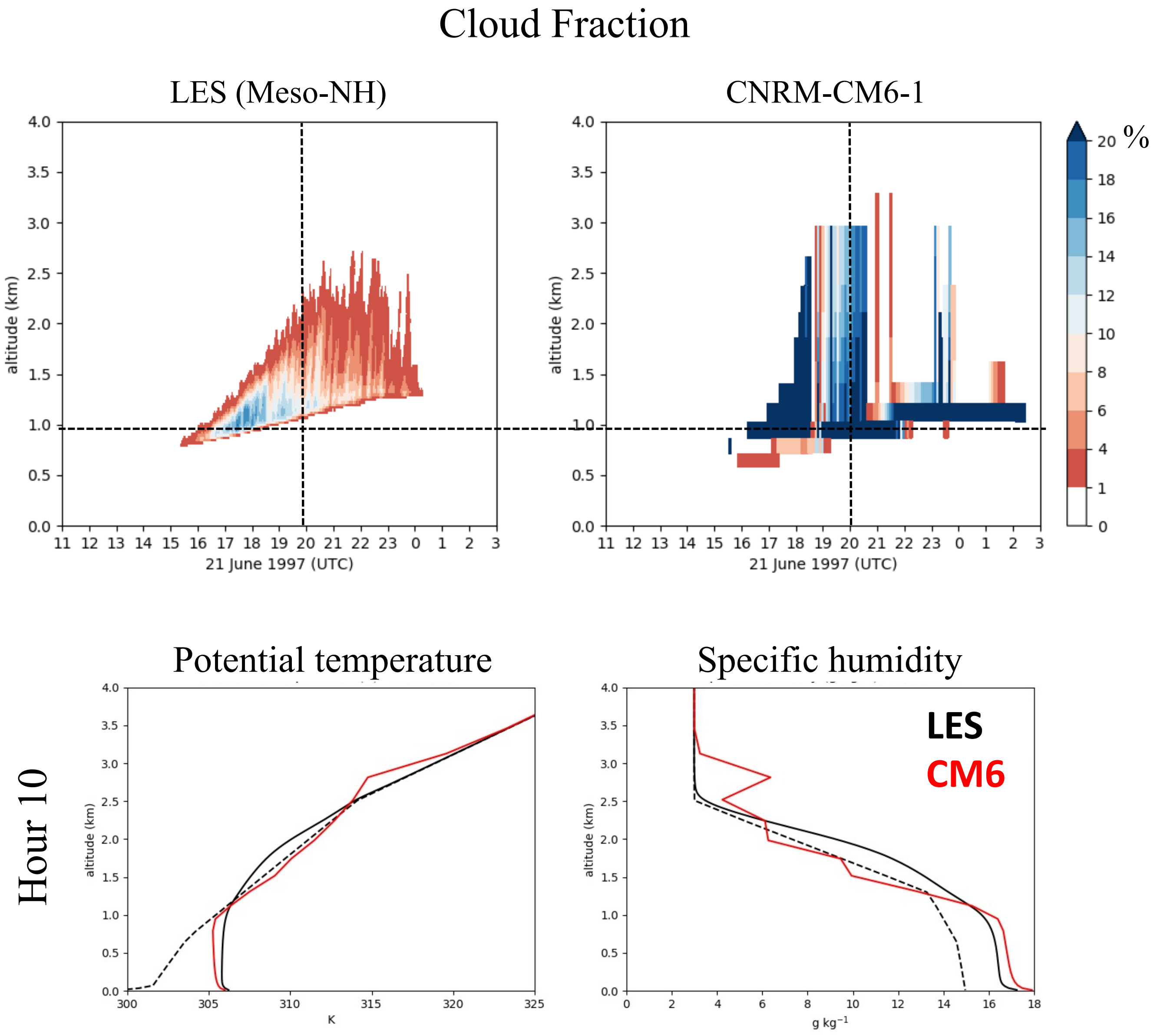
$$\mathrm{NROY}^1 = \bigcap_f \mathrm{NROY}_f^1 = \{\lambda \mid I_f(\lambda) < T, \text{for all } f\}$$

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

*Based on HTExplo (Couvreux et al. 2020)*

*Bower et al. (2010), Williamson et al. (2015, 2017)*

# First application: 1D configurations of CNRM-CM6-1




***1D (Single-column model) vs Large-Eddy Simulation (LES):***

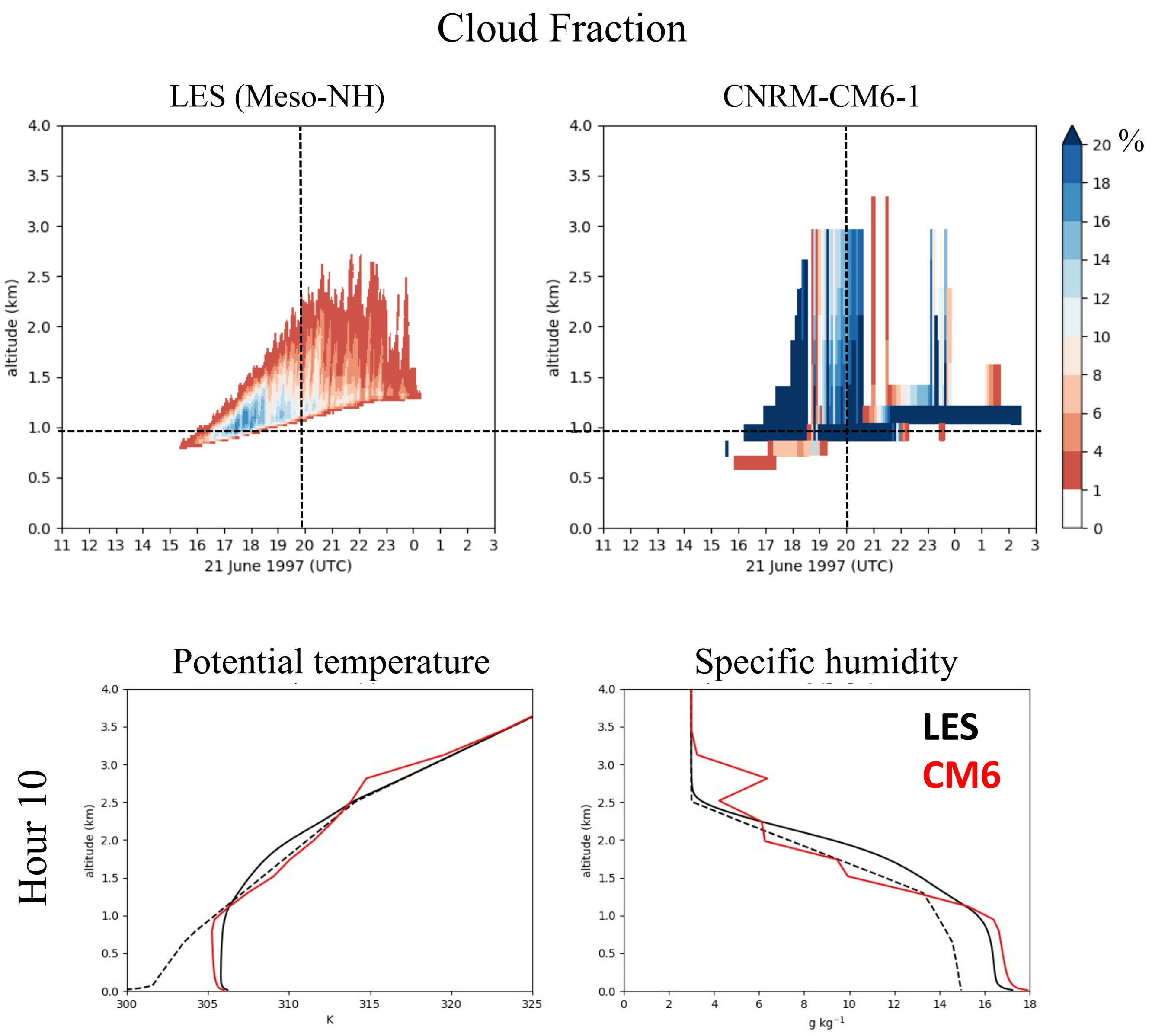
➢ ***inspire the development of atmospheric parameterizations, validate and calibrate them***

ARM-Cumulus: diurnal cycle of continental cumulus

In CNRM-CM6-1:

- Overestimate of cloud fraction
- Cloud base too low
- Boundary layer not deep enough, mixing too weak
- Intermittency and numerical issues

# First application: 1D configurations of CNRM-CM6-1




***1D (Single-column model) vs Large-Eddy Simulation (LES):***

➢ ***inspire the development of atmospheric parameterizations, validate and calibrate them***

ARM-Cumulus: diurnal cycle of continental cumulus

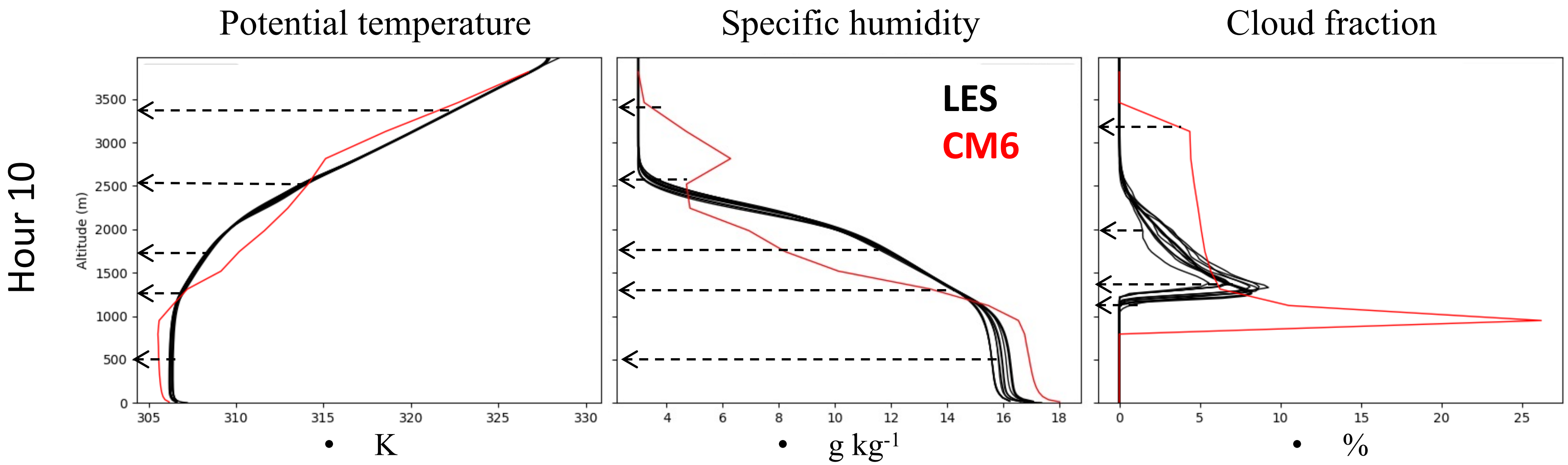
In CNRM-CM6-1:

- Overestimate of cloud fraction
- Cloud base too low
- Boundary layer not deep enough, mixing too weak
- Intermittency and numerical issues

➢ ***Poor calibration, or intrinsic limits?***

# Calibration setup

- **14 metrics**: temperature, specific humidity and cloud fraction at a few levels (Hour 10)
- **Reference**: LES simulation with Meso-NH, uncertainty from an ensemble of LES simulations

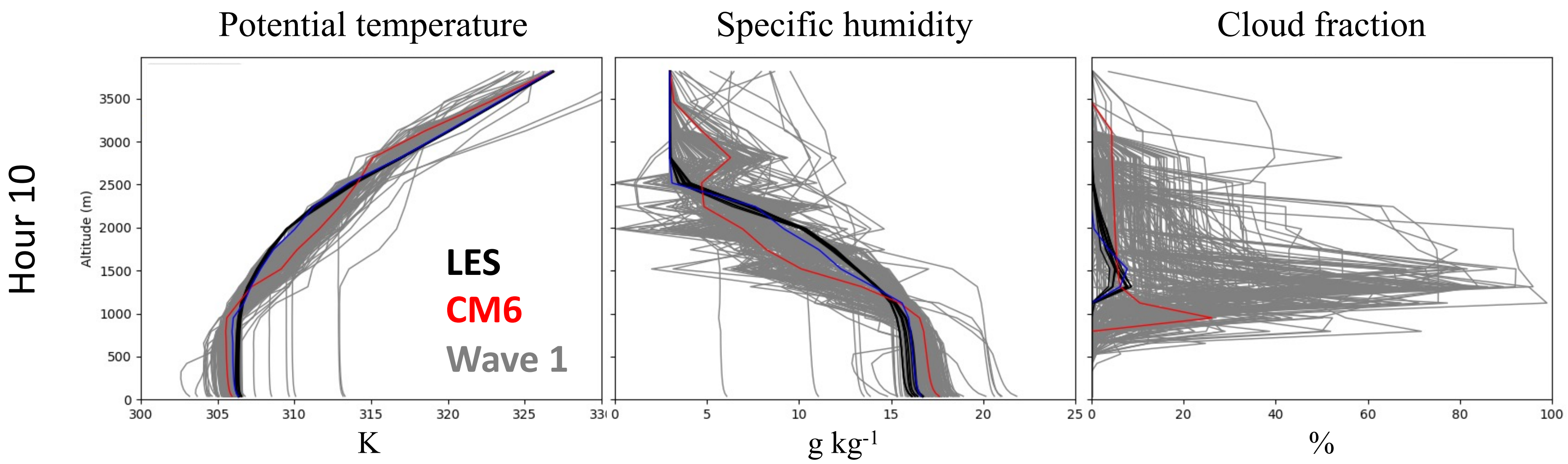



- **26 parameters**:
  - 13 related to convective transport (entrainment/detrainment, drag, buoyancy parameter)
  - 2 related with convective closure
  - 1 related to convective cloudiness
  - 6 related with liquid water microphysics
  - 4 related to turbulence
- Wave 1: 200 simulations, sampling based on a Latin hypercube
- Next waves: 200 simulations (within NROY$^{N-1}$)

# Wave 1

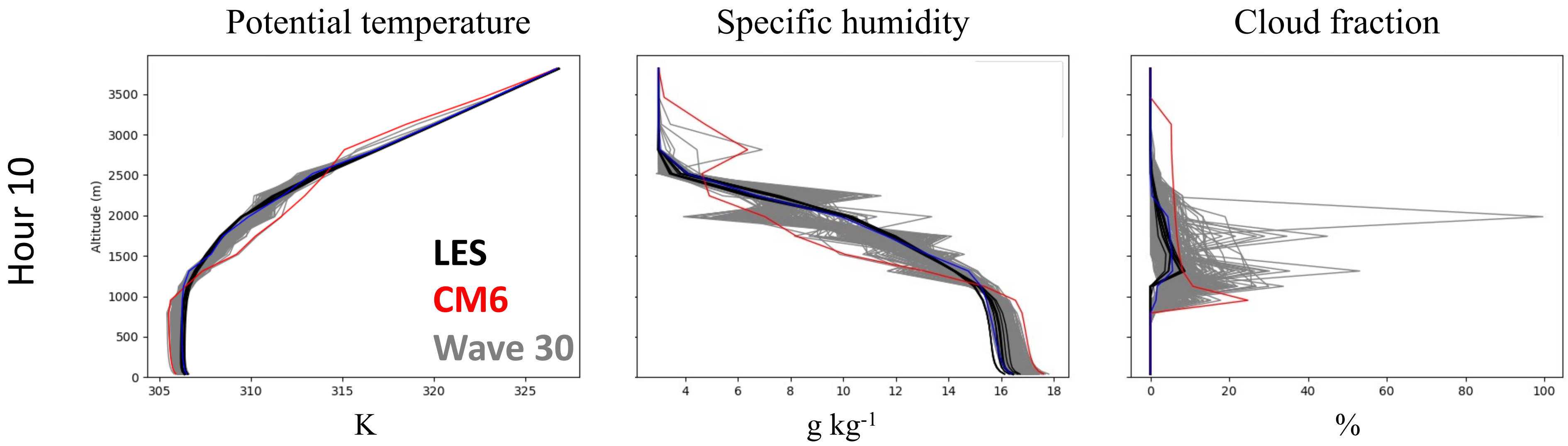
### *Tolerance to error*

- Potential temperature: 0.5 K except at 3400 m (0.1 K)
- Specific humidity: 0.5 g kg$^{-1}$
- Cloud fraction: 5 %



# Wave 33

- Several configurations much better than default.
- Convergence not fully achieved (esp. for cloud fraction).
- ***Intrinsic limits of the model physics***:
  - ➢ Some irregularities in the profiles seem intrinsic (esp. $q_v$)

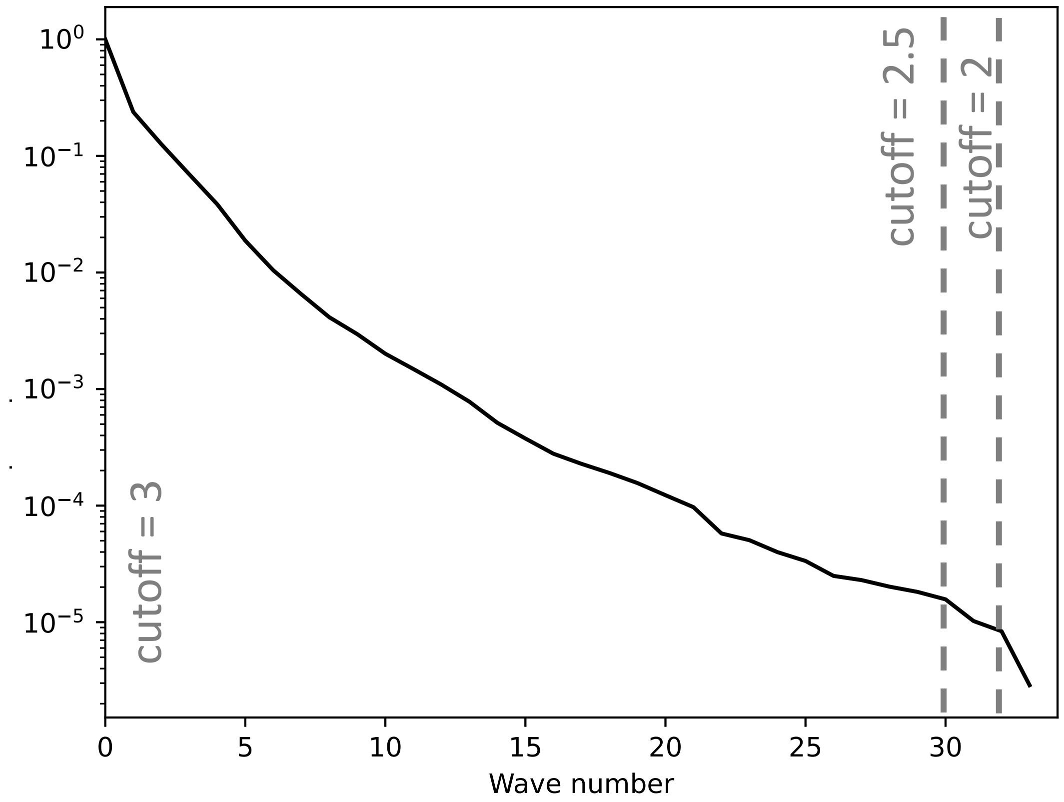


# NROY space evolution and convergence

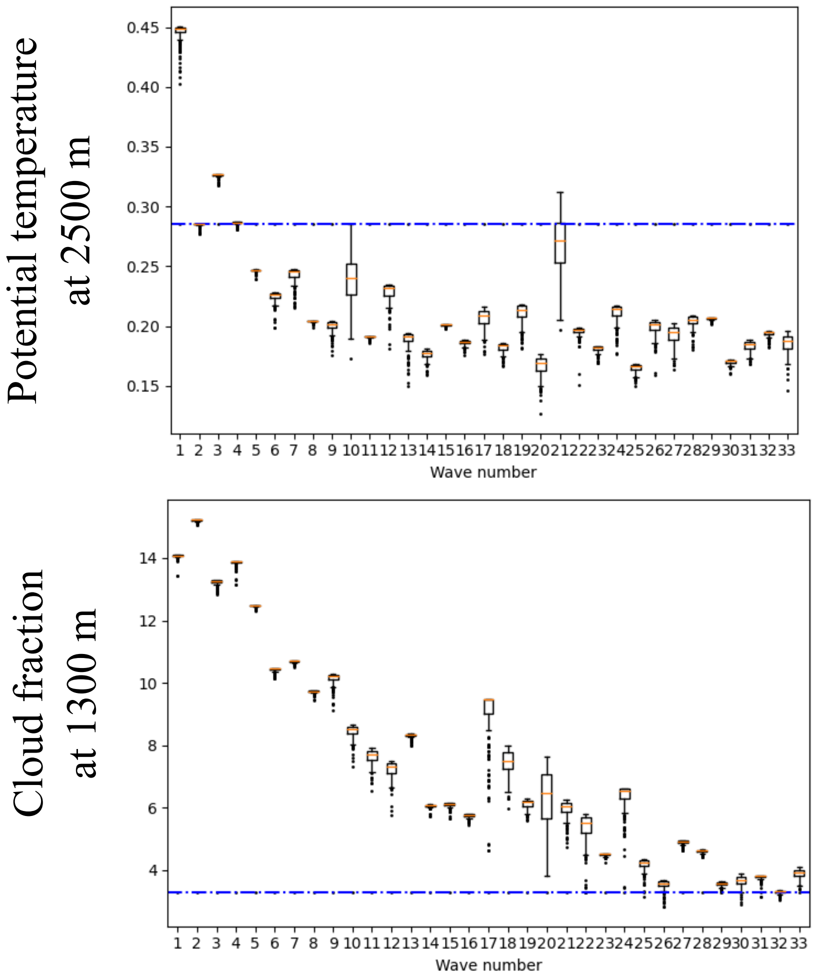
- Decrease of cutoff
  - when emulator uncertainty comparable/lower than reference uncertainty/tolerance-to-error
  - when NROY space size approximately stable
- After 33 waves, NROY $\sim 2.9\times10^{-6}$ of the input parameter space

  >> $\sim$290 over $10^8$ configurations

NROY space fraction




GP standard deviation (Leave-One-Out)

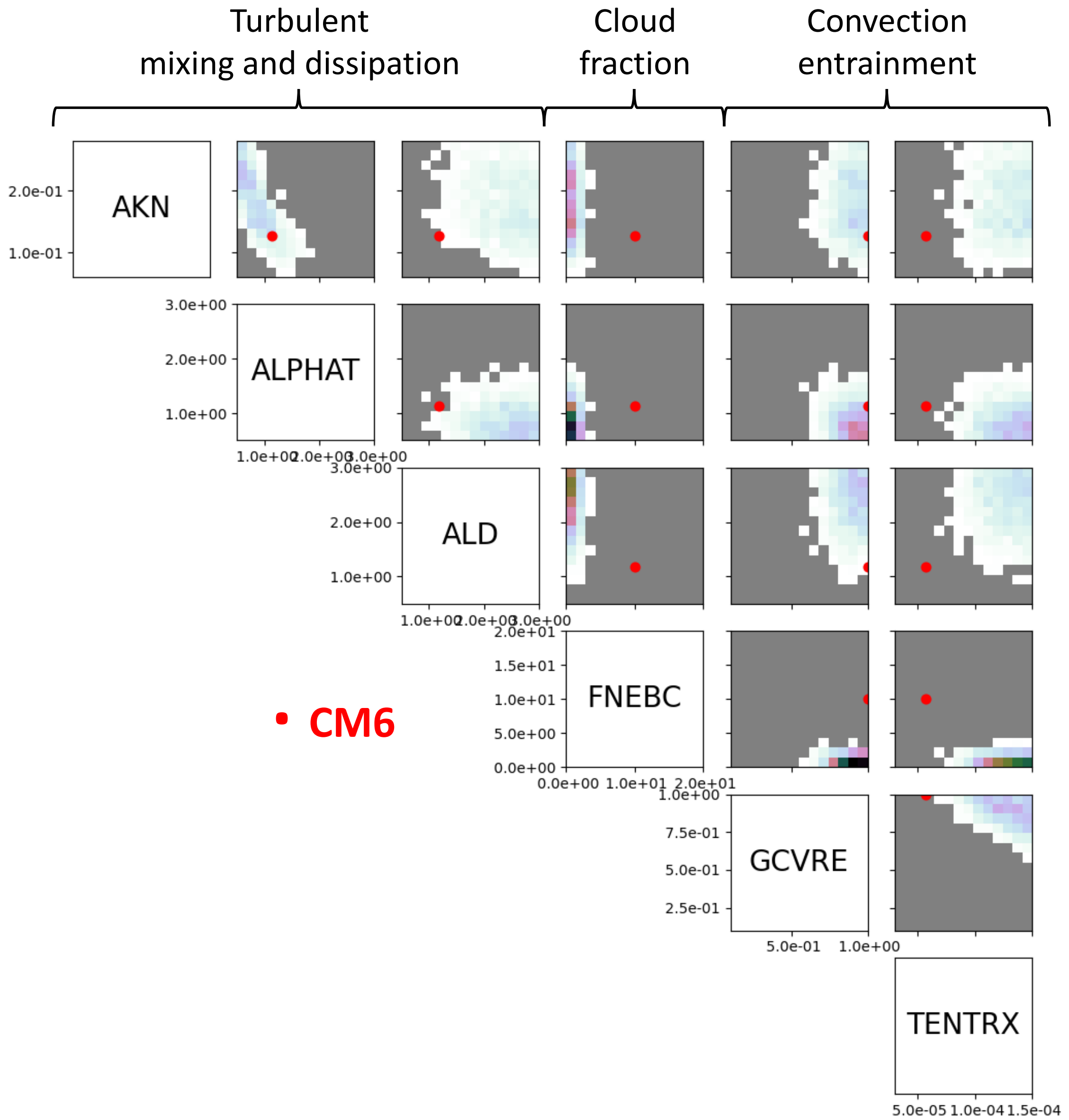


# Representing the NROY space

***Numerical characterization*** of the NROY space:

- (very) large sampling (LHS) of the input parameter space
- Use emulators to compute associated implausibilities
- Compute densities of points within NROY space as a function of parameters
- 1D or 2D representations

# Dominant parameters



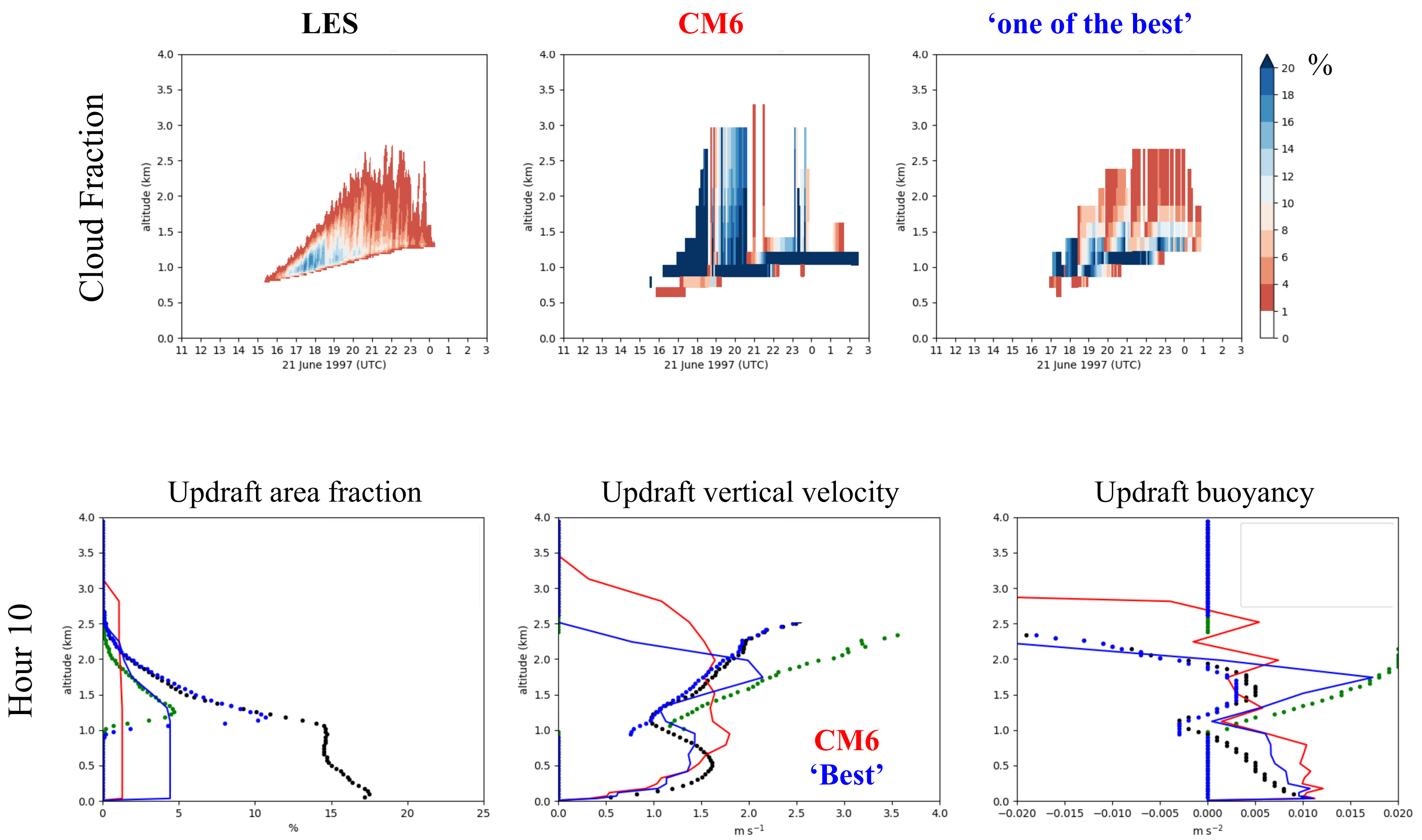

***Dominant parameters***

- Turbulent mixing (AKN and ALPHAT)
- TKE dissipation (ALD)
- Convective cloud fraction (FNEBC)
- Organized entrainement modulation factor (GCVRE)
- Maximum turbulent entrainment (TENTRX)

***Strong constraints on turbulent mixing and convective cloud fraction***

***High entrainment rates*** required for convection

# One configuration within the NROY space



# Calibration vs. structural limits

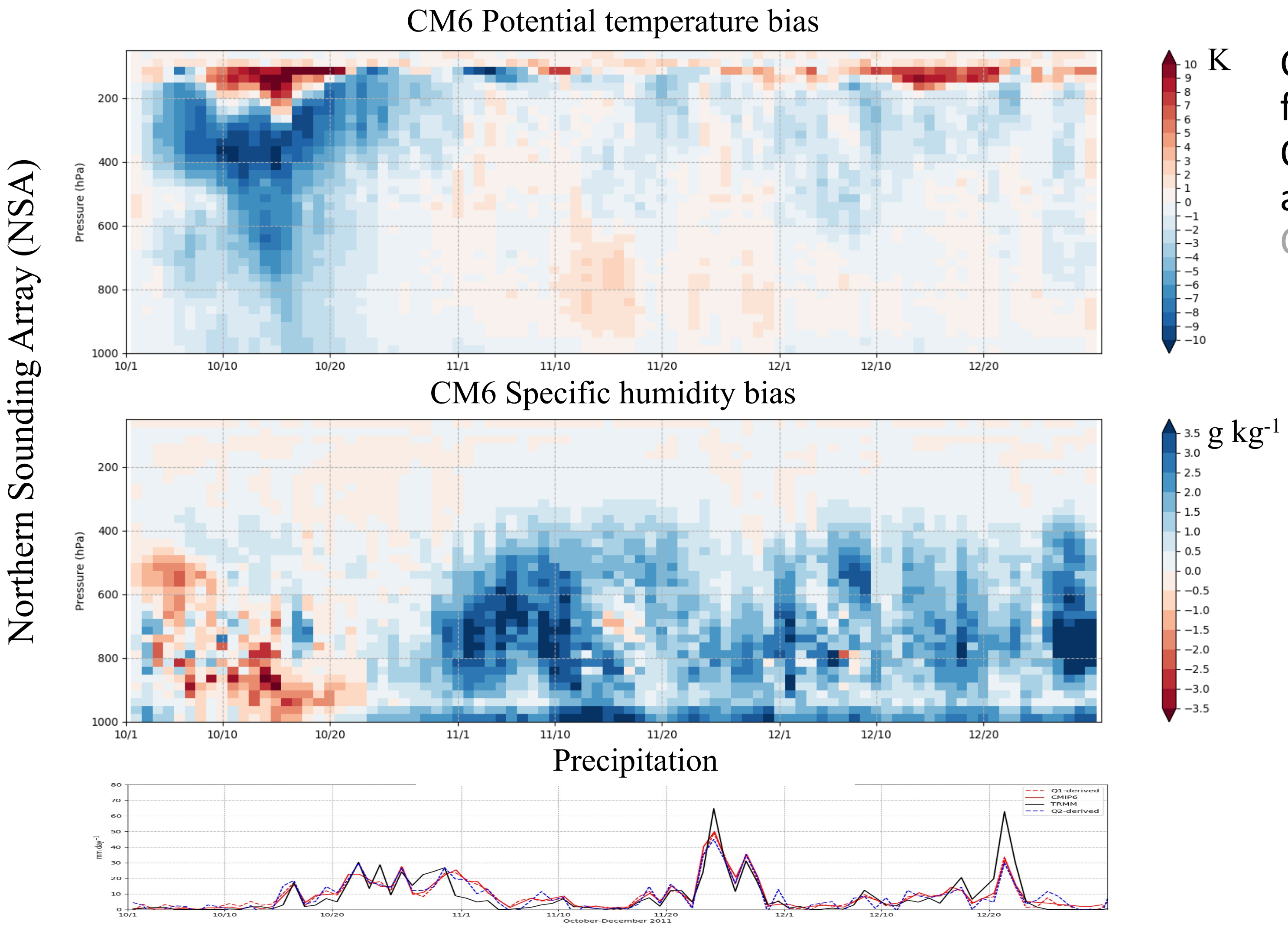
During the development and tuning of CNRM-CM6-1, hard to capture both shallow and deep convection (unified approach)

- ***Is the CNRM-CM6-1 unified approach for convection parameterization appropriate?***

# Calibration vs. structural limits

During the development and tuning of CNRM-CM6-1, hard to capture both shallow and deep convection (unified approach)

➢ ***Is the CNRM-CM6-1 unified approach for convection parameterization appropriate?***

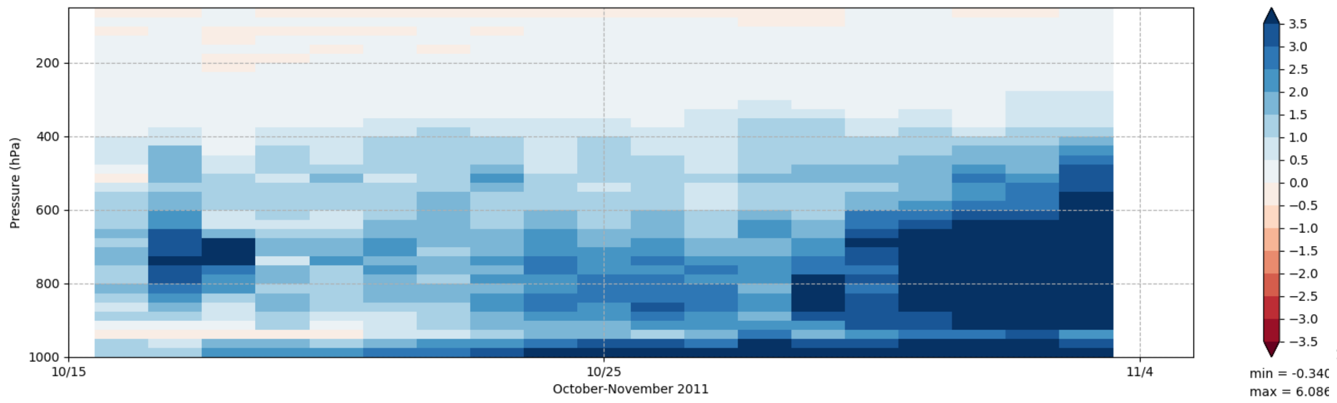



Consider the DYNAMO 1D case, based on a field campaign in OND 2011 in the Indian Ocean, and with alternating convectively-active and suppressed phases (Johnson and Ciesielski, 2013)

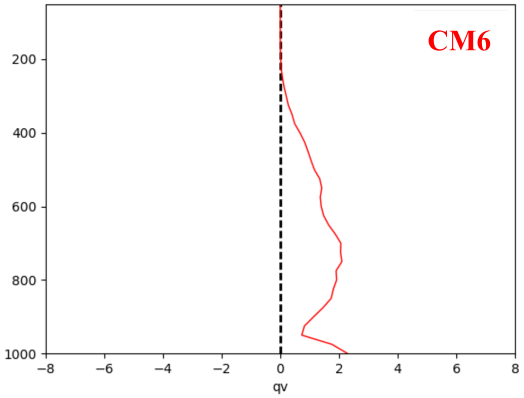
# Metrics, parameters and experimental design

- Focus on the first MJO event (15 October to 4 November 2011)
- **Metrics**: temperature/humidity average profiles
  - 3 levels for θ: 925, 600 and 200 hPa
  - 2 for $q_v$: 925 and 700 hPa
  - 4 for RH: 850, 700, 500 and 300 hPa
- **Reference** (field campaign observations):
  - temperature uncertainty ~0.1 K; tolerance 1 to 5 K
  - $q_v$ uncertainty ~0.1 g kg$^{-1}$; tolerance 0.2/0.1 g kg$^{-1}$
  - RH uncertainty ~2 %; tolerance 5 to 20%
- **Parameters**: same 26 parameters as for ARM-Cumulus
  + 13 for ice microphysics
  + 4 for cloud radiative properties
- Wave 1 and following: 200 simulations

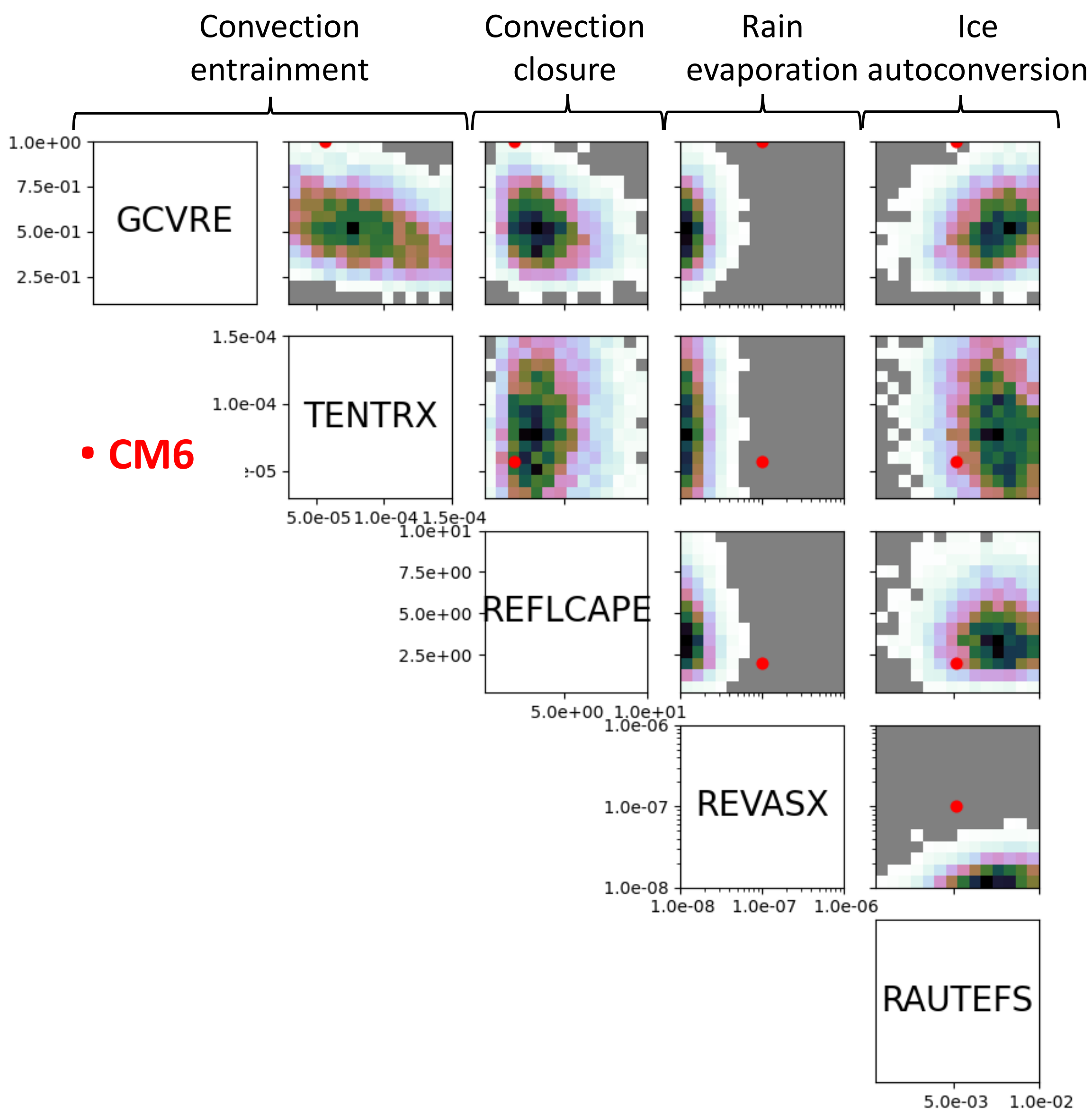
Specific humidity bias




20-day average bias



# Wave 29



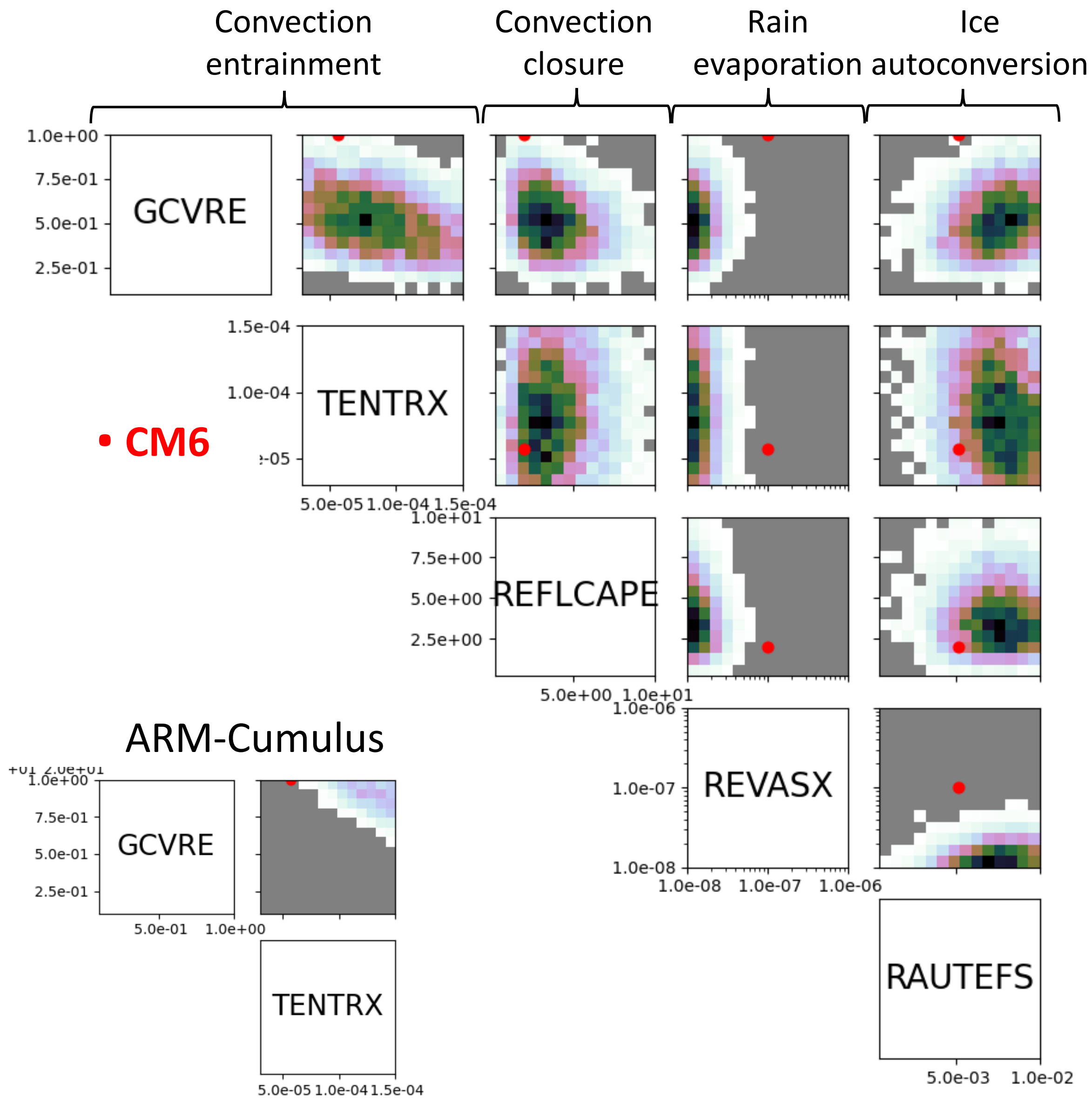

***Wave convergence not yet fully achieved.***

NROY$^{29}$ $\sim 6.9\times10^{-5}$ of the input space

***Dominant parameters*:**

- ***Rain evaporation*** rate key to get the moisture vertical profile right
- Reduced ice autoconversion required >> ***longer life cycle of (high-level) ice clouds***. Feedback through LW heating?
- ***Convection intensity*** (closure) is important
- ***Convective entrainment*** still key, towards more moderate values compared to shallow convection.

# Deep vs. shallow convection




***Wave convergence not yet fully achieved.***

NROY[29] $\sim 6.9\times10^{-5}$ of the input space

***Dominant parameters*:**

- ***Rain evaporation*** rate key to get the moisture vertical profile right
- Reduced ice autoconversion required >> ***longer life cycle of (high-level) ice clouds***. Feedback through LW heating?
- ***Convection intensity*** (closure) is important
- ***Convective entrainment*** still key, towards more moderate values compared to shallow convection.
- ***Weak overlap between the entrainment parameter NROY spaces for shallow and deep cases***

# What did we learn from 1D configurations?

- ***History Matching with Iterative refocusing:*** framework to rigorously explore the ***parametric calibration*** of a parameterization/model physics, and identify its ***structural limits***

- Better tuning of the CNRM convective parameterization (in fact the CNRM physics) can be achieved for each of the two 1D cases addressed here.

- The combination of a shallow convection case and a deep convection case emphasizes the difficulty to make the CNRM “unified” convection parameterization work for both.
  - ***The CNRM-CM6-1 single-plume mass-flux approach is structurally limited.***

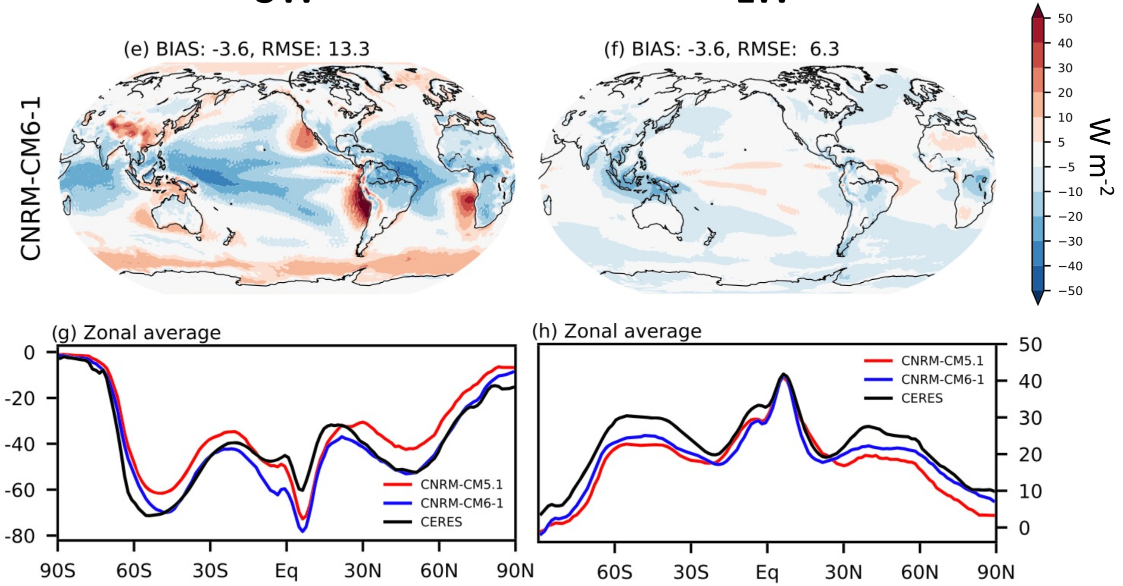
- Shallow and deep convection regimes require significantly different entrainment rates.

➢ ***What about more ‘realistic’ (3D) configurations?***

# Calibrating the 3D (atmospheric) configuration of CNRM-CM6-1

CRE at TOA

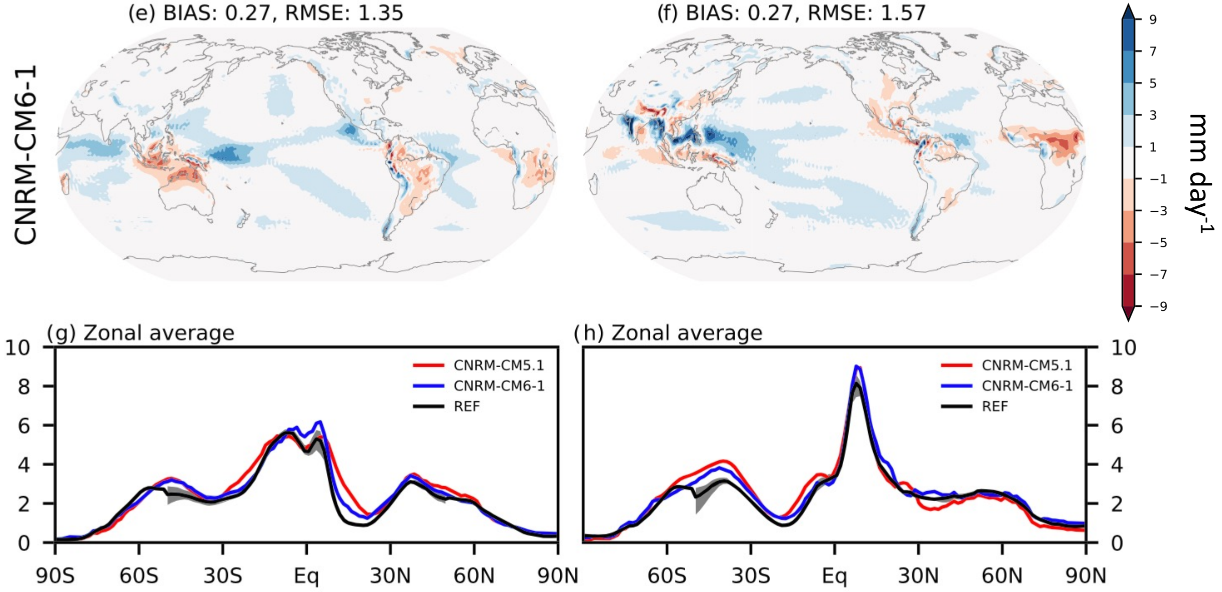
SW LW




Reference: *CERES-EBAF Ed. 4.1*

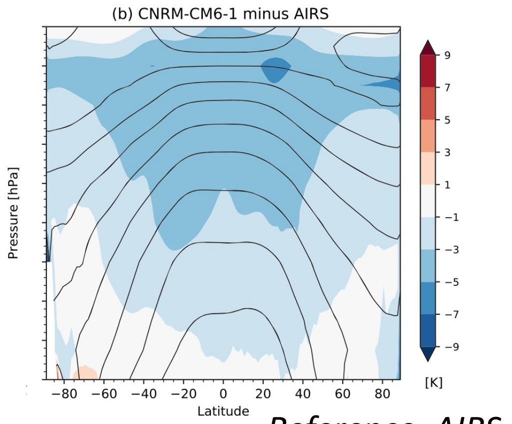
Precipitation

DJFM JJAS




*Reference: MSWEP v1.2 + GPCP v2.3 + TRMM 3B42 v7*

Annual Temperature



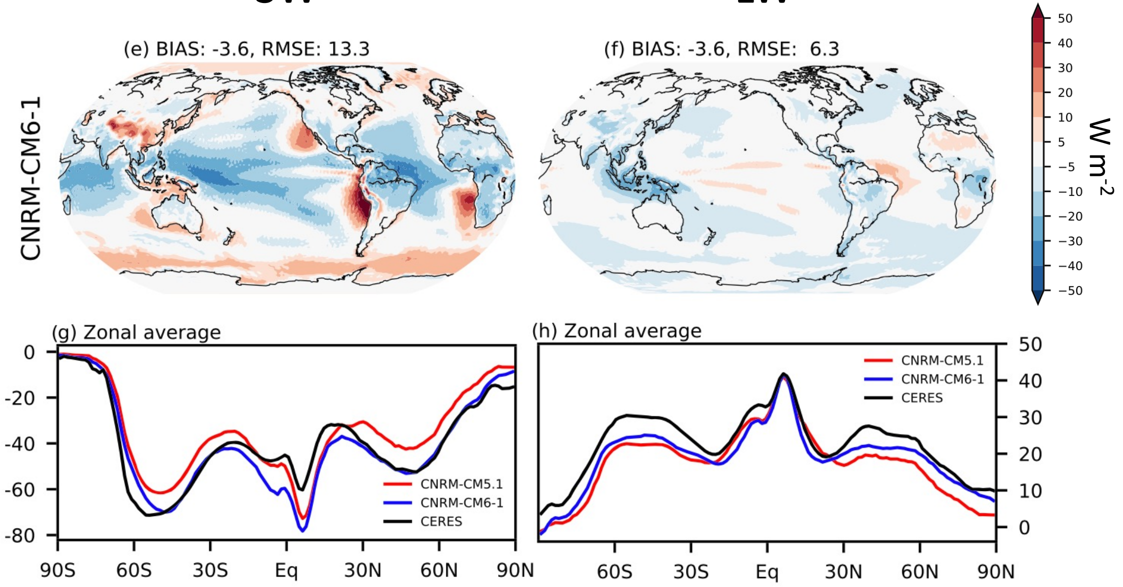

*Reference: AIRS*

*Roehrig et al. (2020)*

# Calibrating the 3D (atmospheric) configuration of CNRM-CM6-1

CRE at TOA

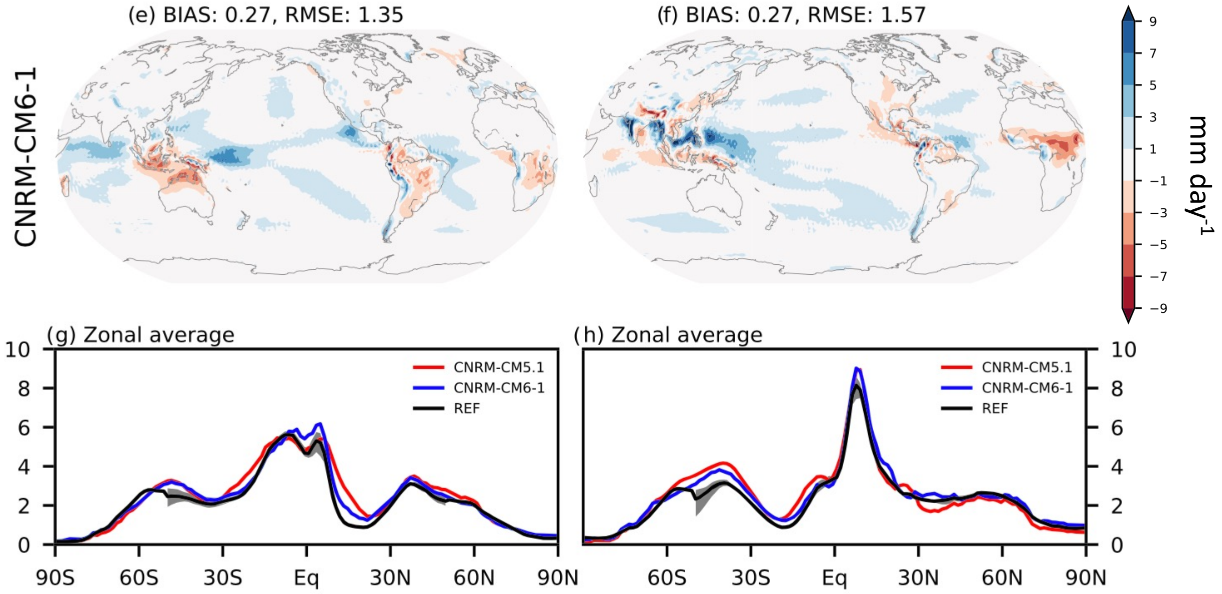
SW LW




*Reference: CERES-EBAF Ed. 4.1*

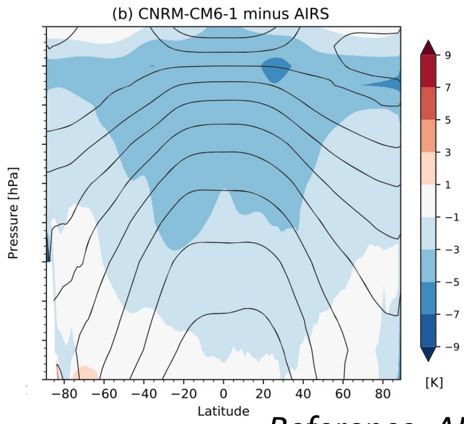
Precipitation

DJFM JJAS




*Reference: MSWEP v1.2 + GPCP v2.3 + TRMM 3B42 v7*

Annual Temperature




*Reference: AIRS*

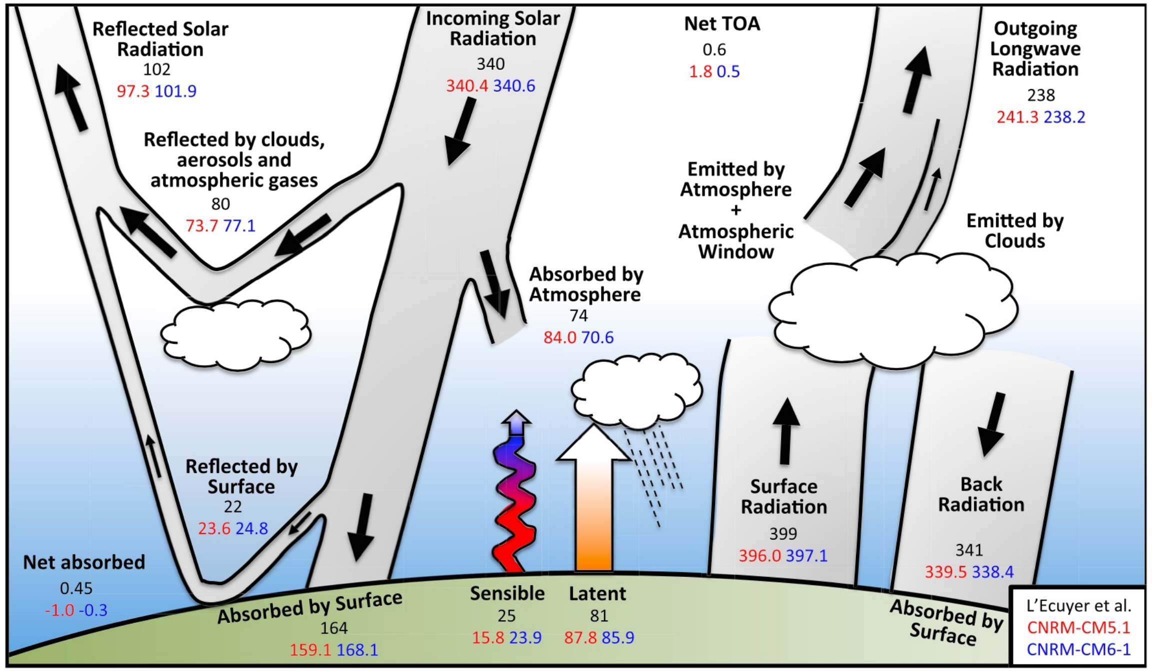
➢ ***Can we reduce CNRM-CM6-1 biases through better calibration?***

*Roehrig et al. (2020)*

# Metrics, references and uncertainty estimates

Following the CNRM-CM6-1 tuning strategy, 3 classes of metrics:

1. ***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 $m^{-2}$

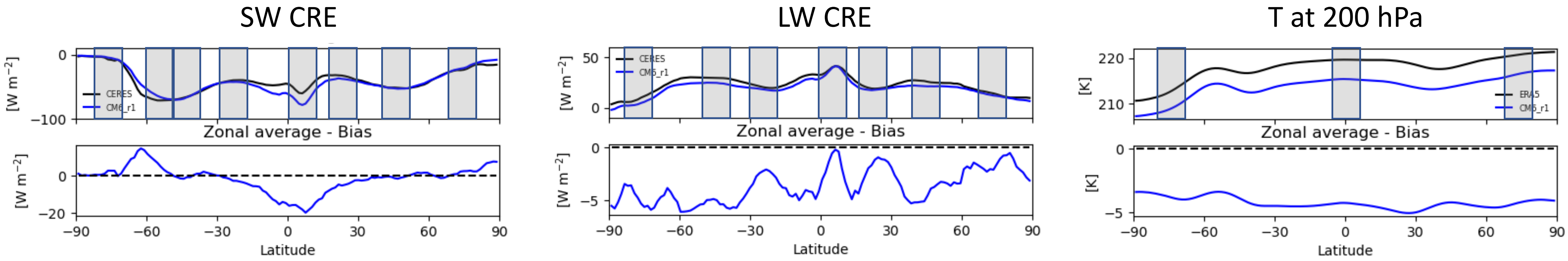



Roehrig et al. (2020)

# Metrics, references and uncertainty estimates

Following the CNRM-CM6-1 tuning strategy, 3 classes of metrics:

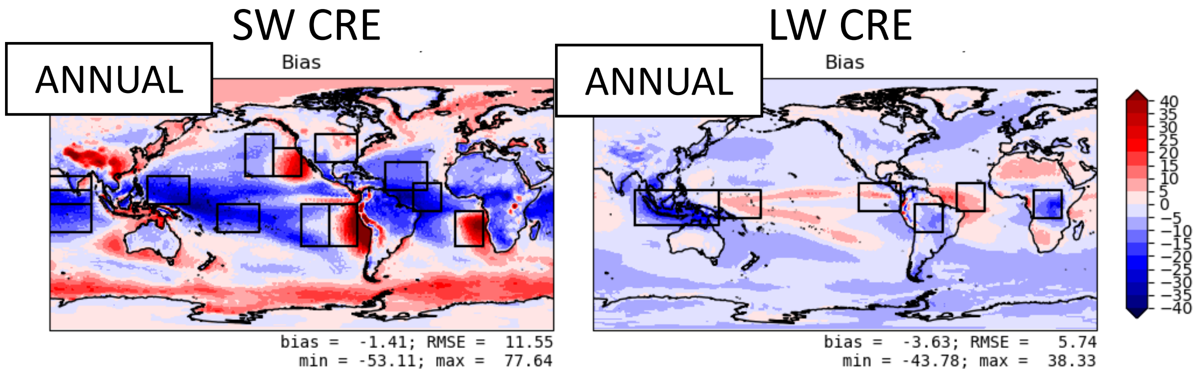
1. Global averages of the energy budget components
2. ***Zonally average profiles*** of SW/LW CRE + Temperature at 200 hPa
   - SW/LW CRE: CERES-EBAF with uncertainty of 2 W $m^{-2}$ + tolerance of 1 W $m^{-2}$
   - T200: based on ERA5/JRA55/MERRA/CFSR ensemble mean and std, tolerance 1.5 K

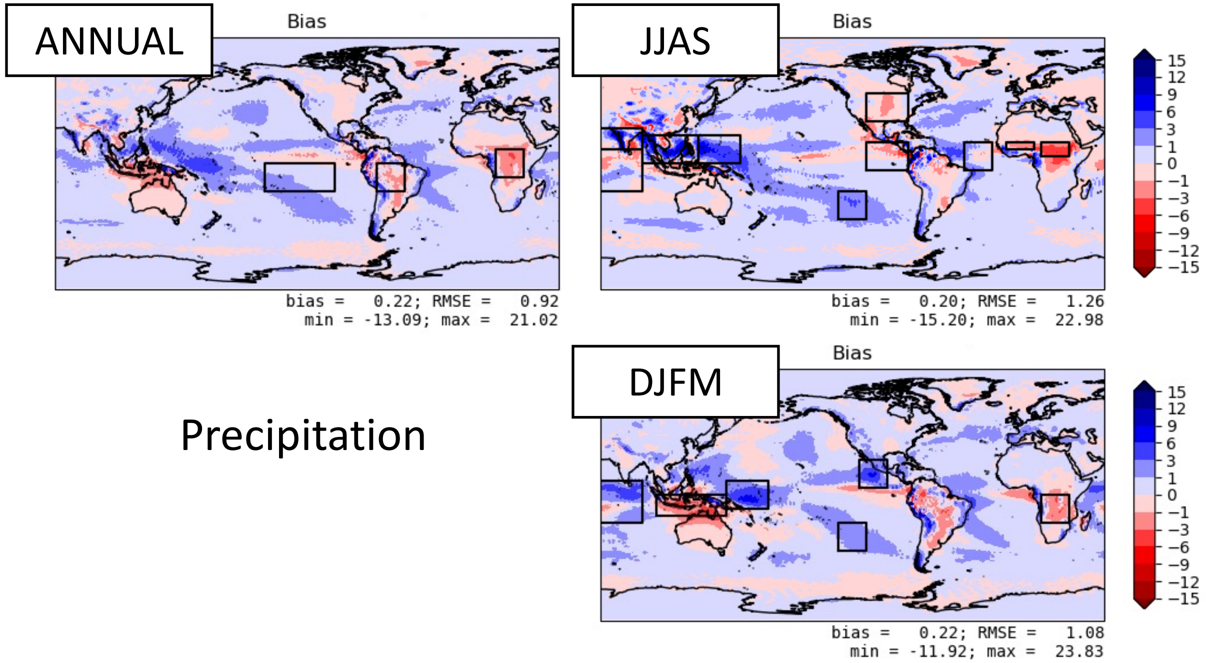


# Metrics, references and uncertainty estimates

Following the CNRM-CM6-1 tuning strategy, 3 classes of metrics:

1. Global averages of the energy budget components
2. Zonally average profiles of SW/LW CRE + Temperature at 200 hPa
3. ***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, tolerance between 0.5 and 1 mm $day^{-1}$



# Metrics, references and uncertainty estimates

Following the CNRM-CM6-1 tuning strategy, 3 classes of metrics:

1. Global averages of the energy budget components
2. Zonally average profiles of SW/LW CRE + Temperature at 200 hPa
3. Regional and seasonal averages of SW/LW CRE and precipitation

- ***63 (scalar) metrics***

# Model parameters and simulation strategy

***46 model 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)

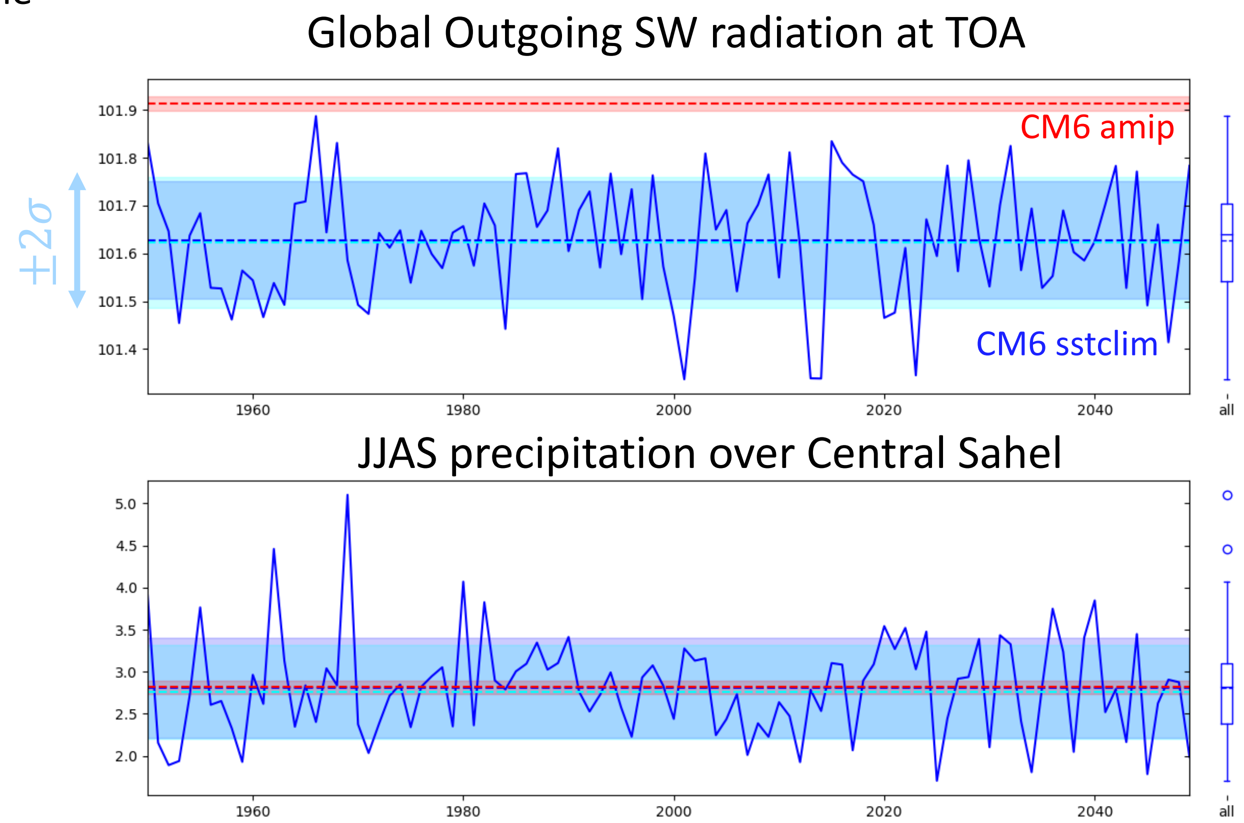
# Model parameters and simulation strategy

### ***46 model 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)

### ***Waves of 400 simulations***

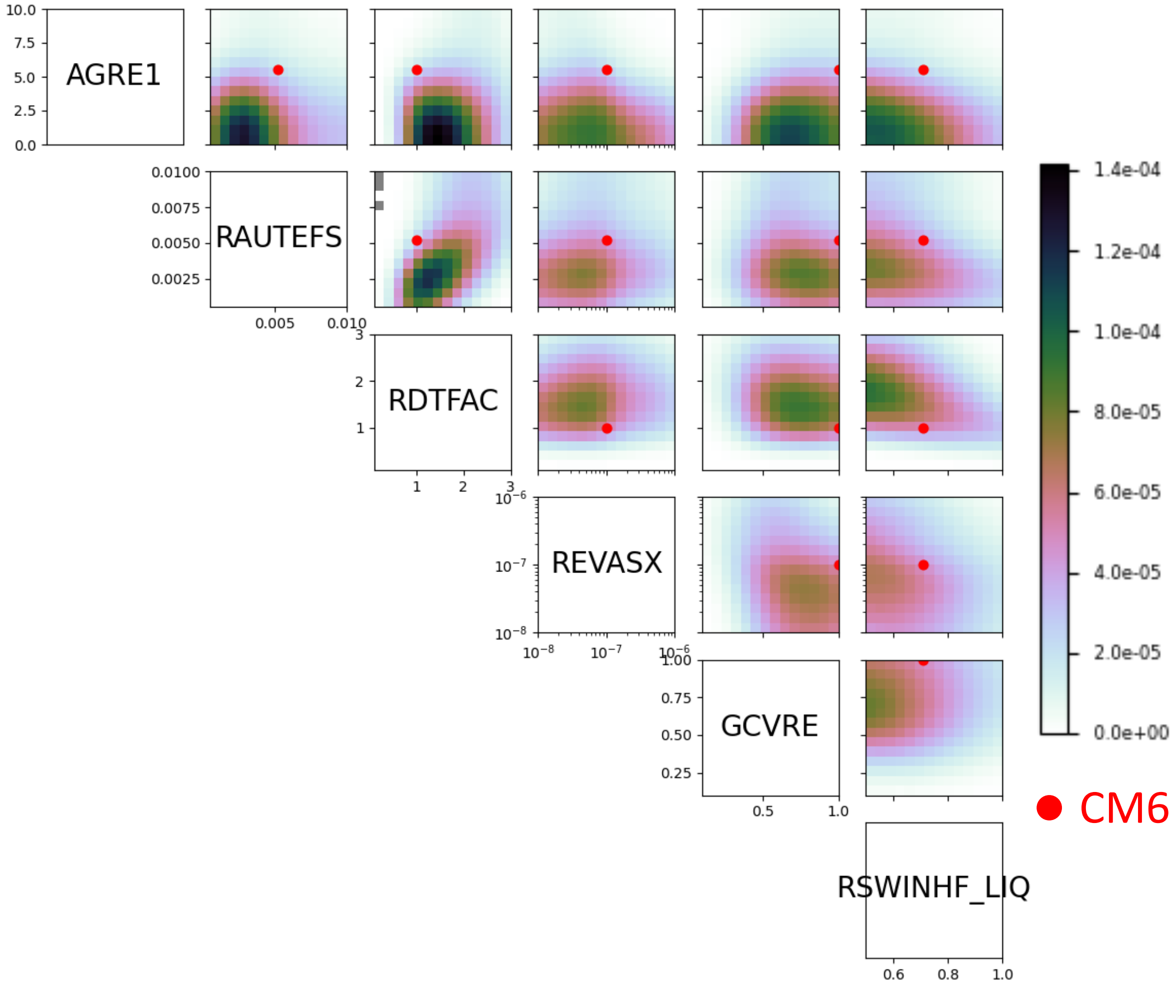
- 1-year ***sstclim*** simulations + 3-month spin-up
- sstclim vs amip correction of the reference target
- Consideration of ***internal variability uncertainty*** based on a 100-year sstclim simulation with CNRM-CM6-1.
- Latin Hypercube sampling for 1st wave



# Wave 1 results

- NROY[1] space: 0.66% of the input space
- Cloud (esp. ice) and convection parameters are critical

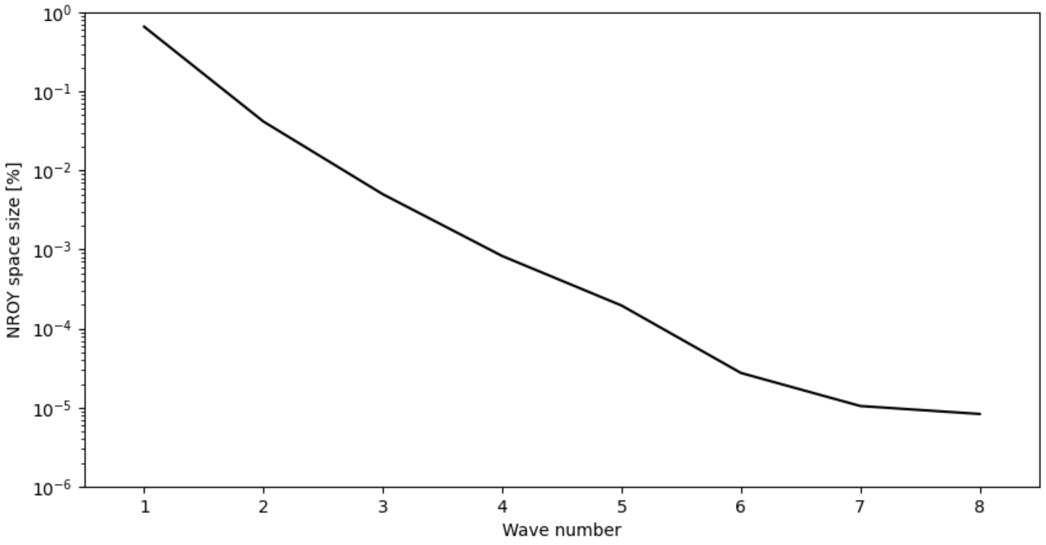
NROY[1] density within input parameter space
For some of the dominant parameters



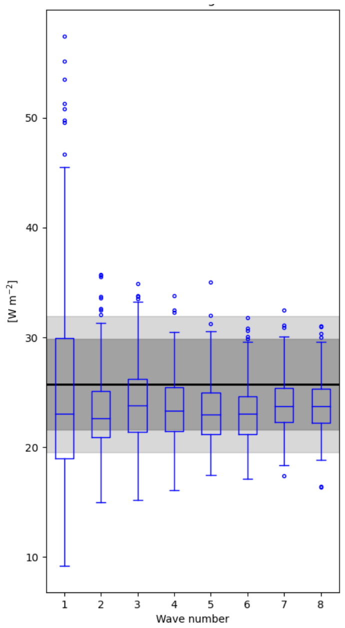
# From Wave 1 to Wave 8

- Large shrinking of the NROY space size (8 orders of magnitude)
- Some metrics have converged, some are more demanding

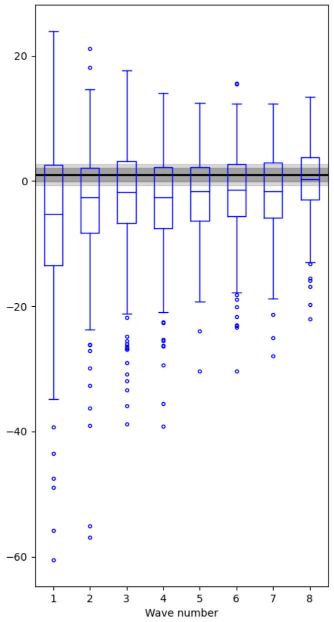
NROY space fraction of the input space




Global LW CRE at TOA

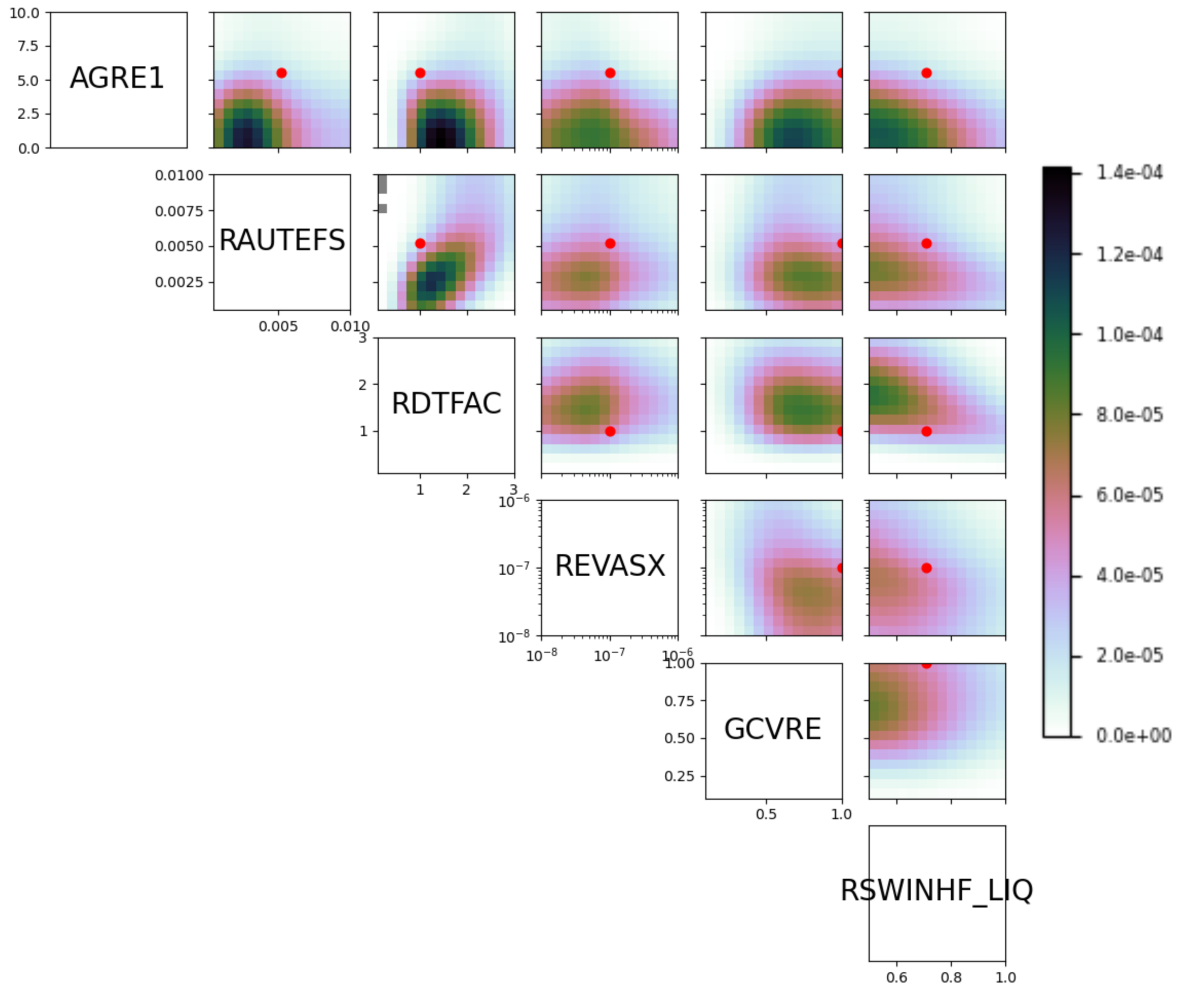



Ocean net energy flux

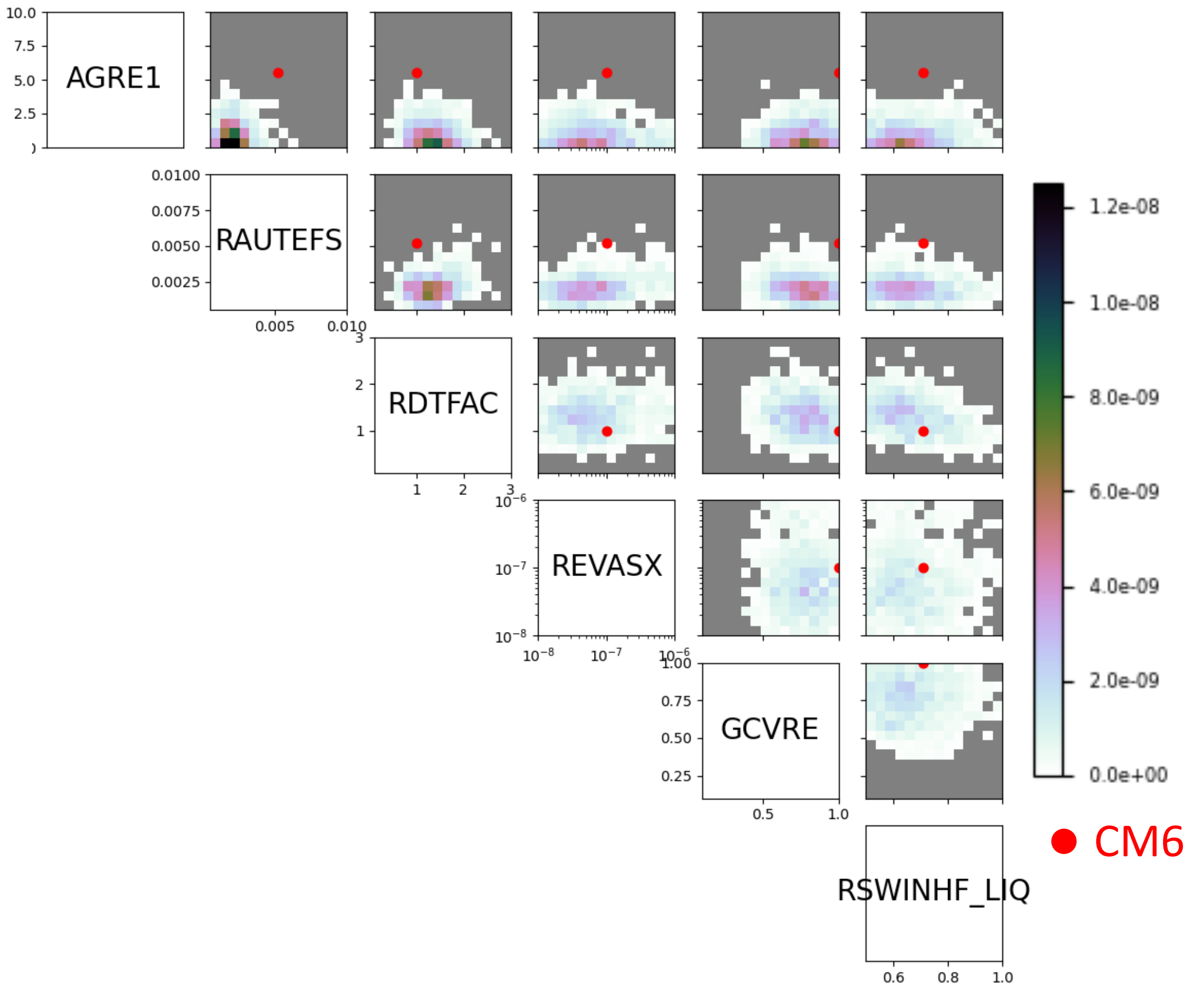


# From Wave 1 to Wave 8

NROY[1] density within input parameter space
For some of the dominant parameters




NROY[8] density within input parameter space
For some of the dominant parameters




+ some new dominant parameters have emerged

# Choosing configurations of interest

***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***

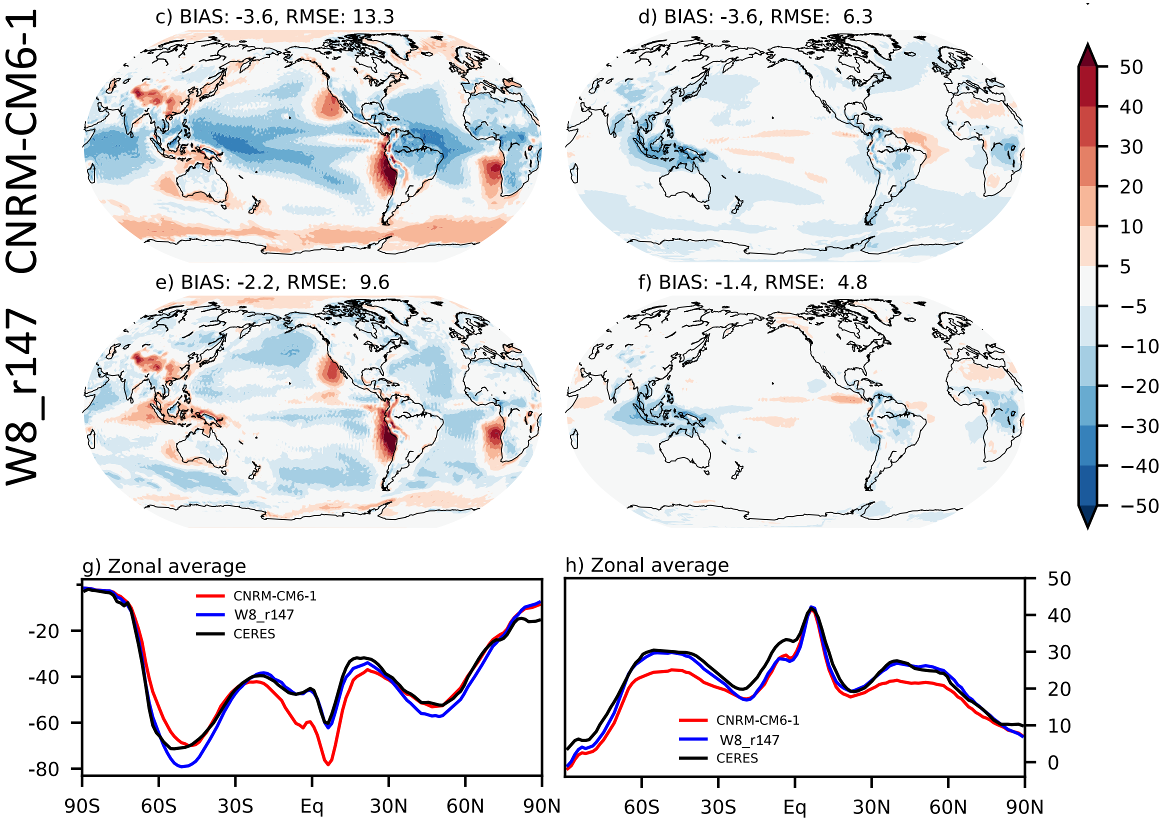
- A few simulations fulfil all the metrics but one (for a cutoff of 3)
- A few have interestingly low RMSEs for targeted variables

➢ A selection of these simulations is further analysed with 10-year amip-style simulations.

# Improved performance?

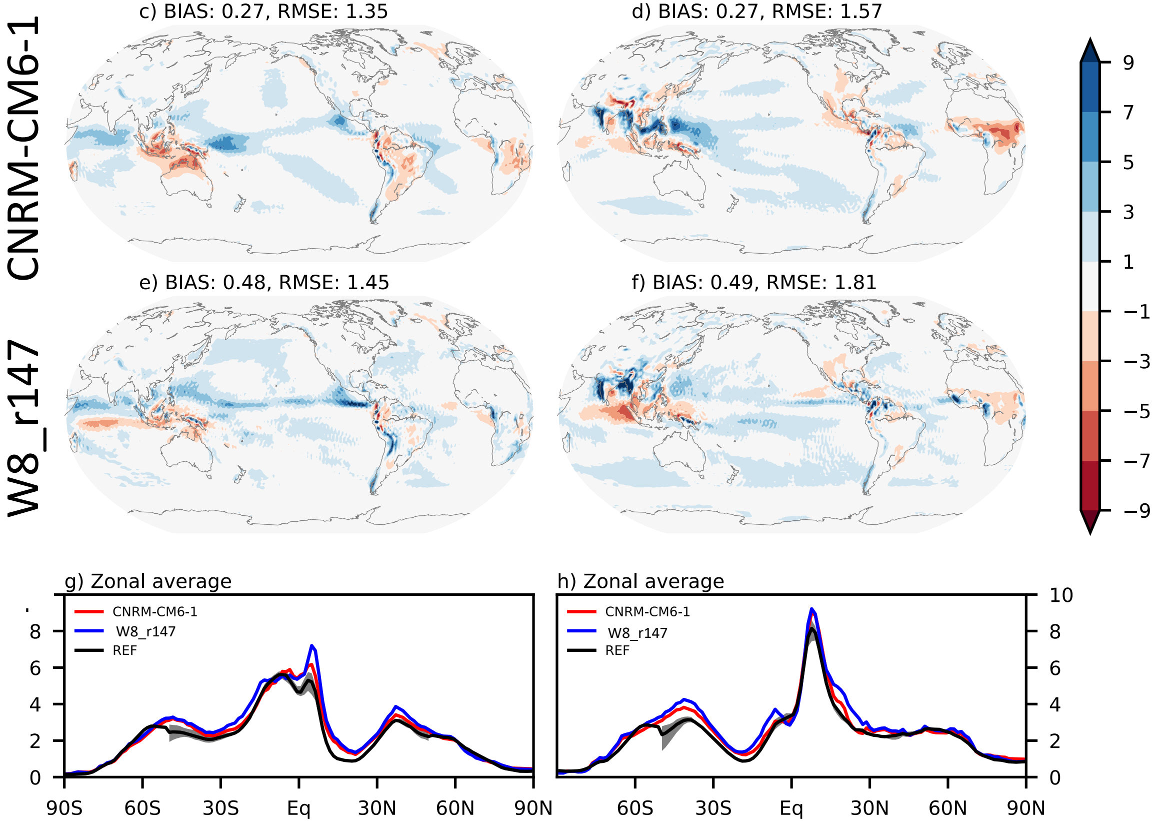
## CRE at TOA

SW LW




## Precipitation

DJFM JJAS



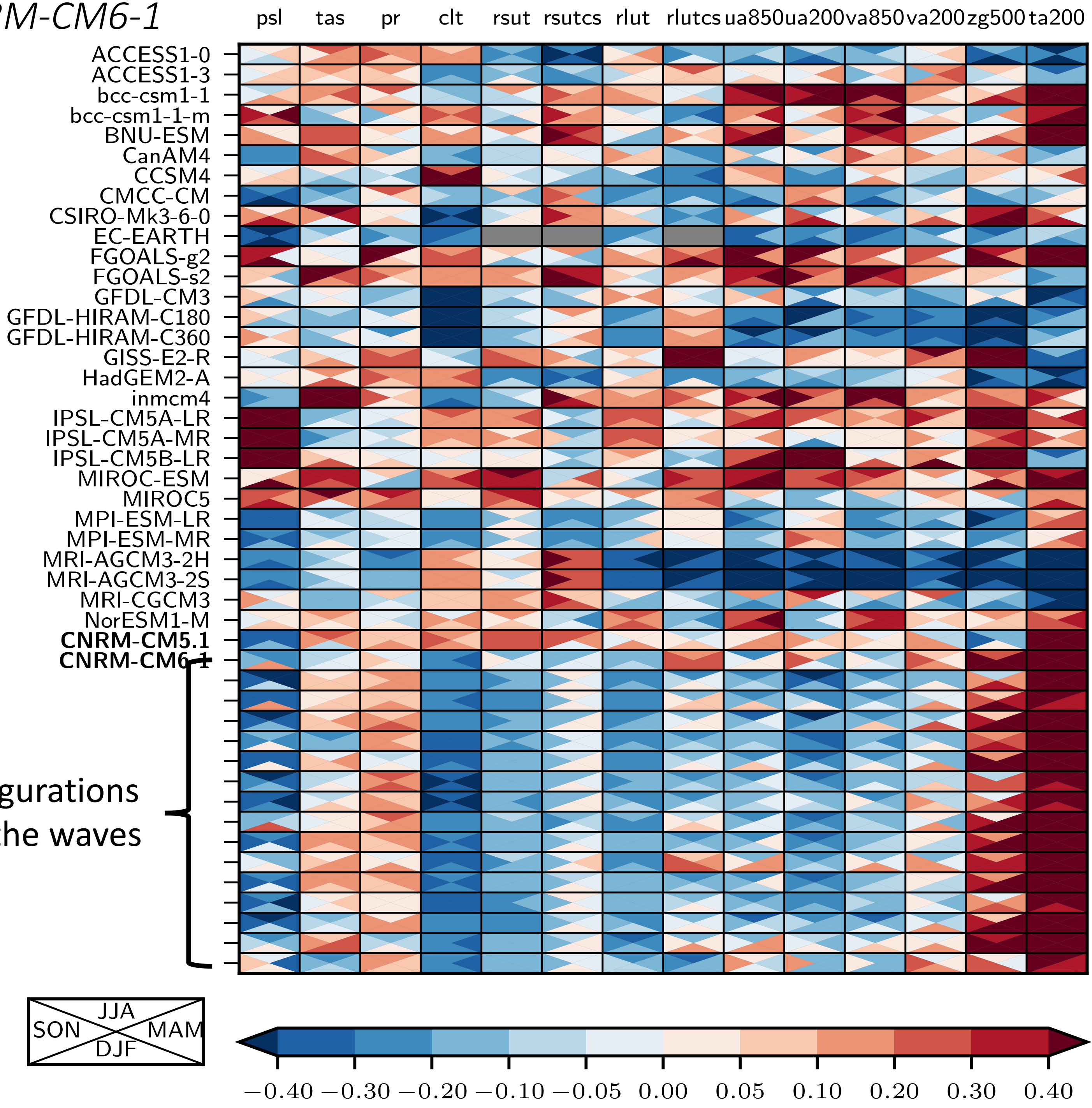

- ***Improved or similar performance*** on several mean state features
- Some errors seems truly structural: clouds/radiation over eastern part of ocean basins

# Improved performance?

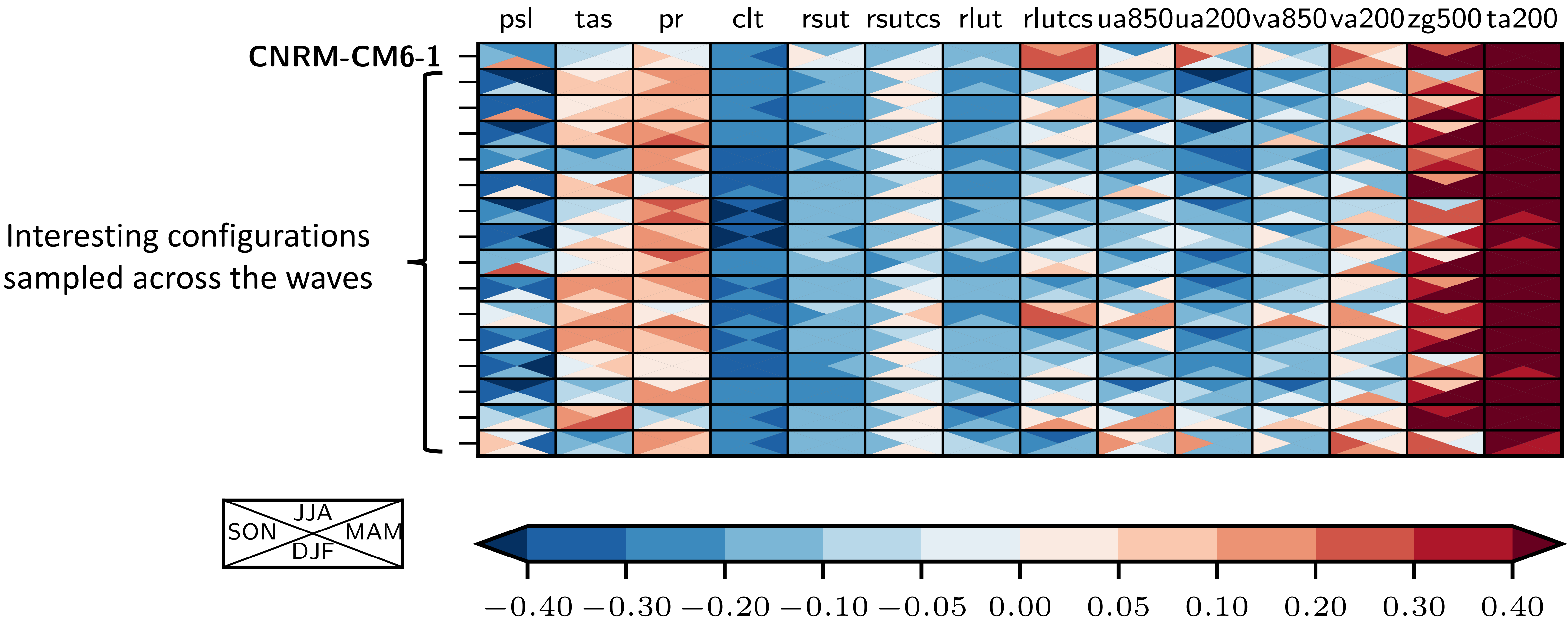
Relative score within the CMIP5 ensemble
*(Gleckler et al. 2016)*

$$\text{score} = \frac{\text{RMSE} - \text{RMSE}_{\text{median}}}{\text{RMSE}_{\text{median}}}$$

Interesting configurations sampled across the waves



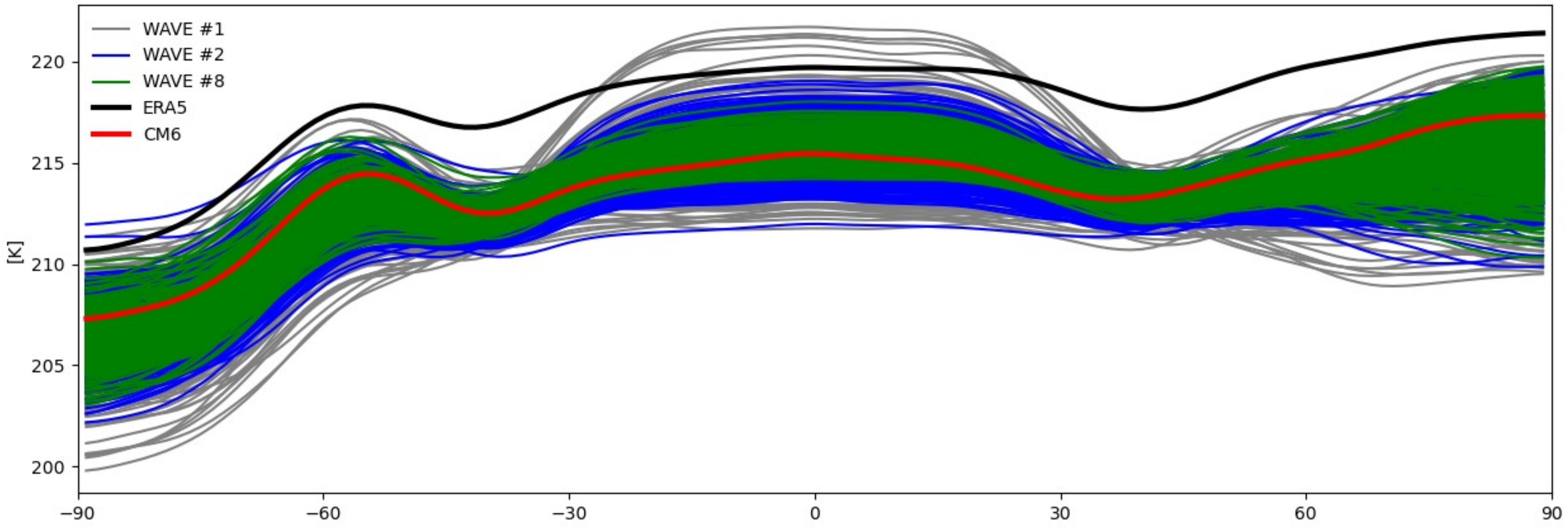
# Improved performance?




- ***Improved or similar performance*** on several mean state features
- Some errors seems truly structural: clouds/radiation over eastern part of ocean basins, upper-tropospheric temperature
- Some ***trade-offs*** are required

# Upper-tropospheric temperature bias: structural limit?

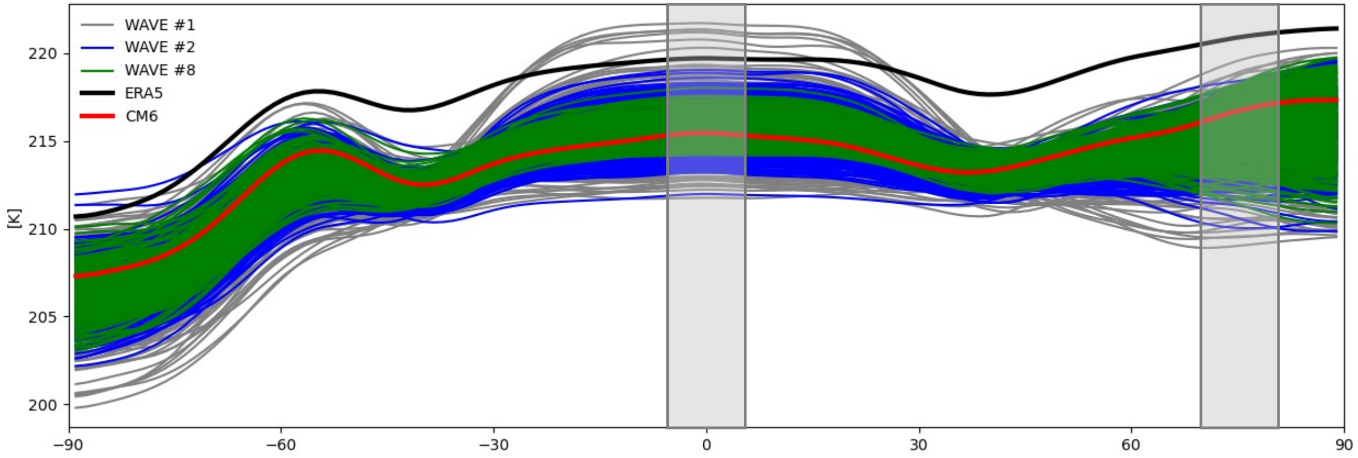
Zonally-average 200-hPa Temperature

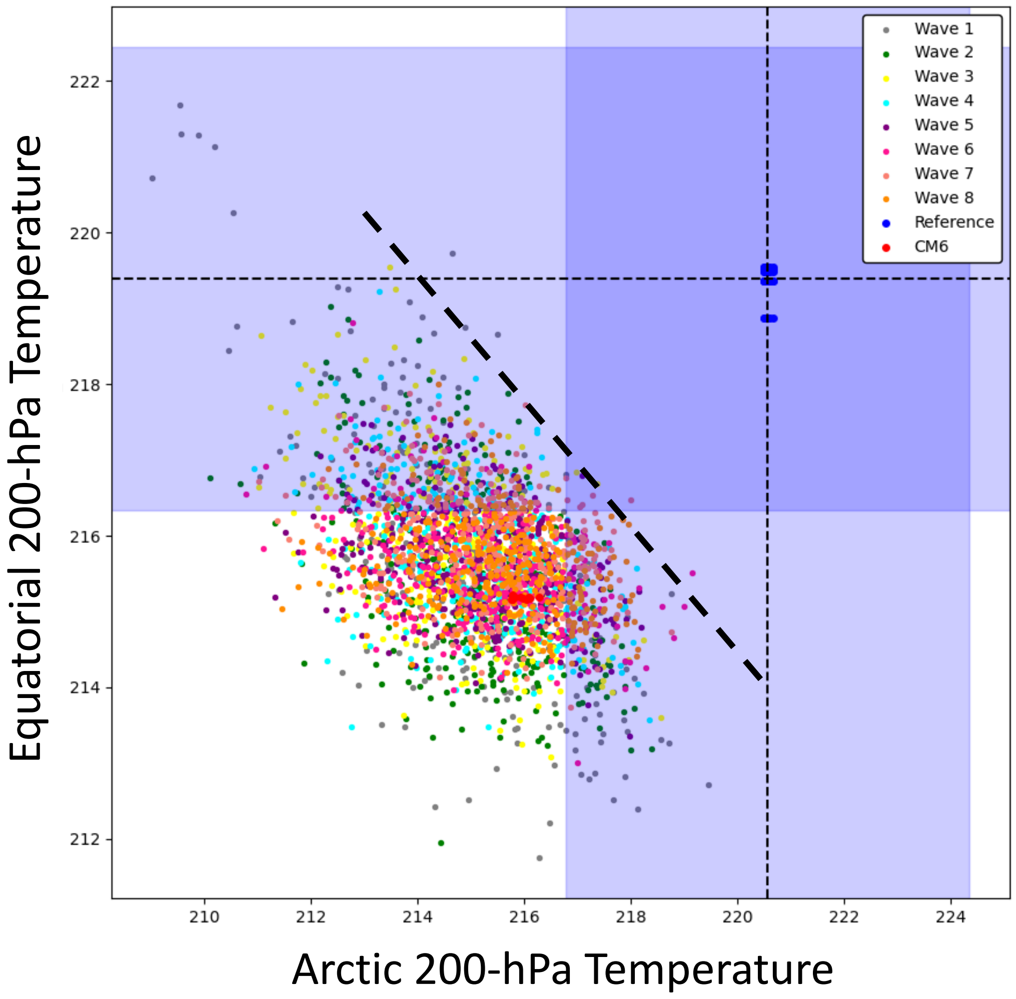



- Simulations performing well (beyond the chosen tolerance to error) in the equatorial regions disappear with successive waves
  - ➢ Incompatible with other metrics

# Upper-tropospheric temperature bias: structural limit?

Zonally-average 200-hPa Temperature




- Simulations performing well (beyond the chosen tolerance to error) in the equatorial regions disappear with successive waves
    - ➢ Incompatible with other metrics
- The model can most likely not capture both equatorial and arctic upper-tropospheric temperatures within uncertainty ranges, while remaining compatible with other metrics.

# From 1D and 3D: the High-Tune philosophy

- ***Start calibration at the process level*** (1D vs. LES) to ***pre-condition*** calibration of more realistic configurations
  - Keep the representation of processes right

- Requirement: have ***high-quality references*** (observations, LES or CPM) for the targeted processes.
  - Some are available (LES for boundary-layer convection/clouds, stable boundary layers)
  - Some are currently being developed (radiation for boundary-layer clouds)
  - Some are not fully appropriate or missing (e.g., microphysics)

# Proof of concept with LMDZ

### *Multi 1D case approach (boundary-layer clouds) :*

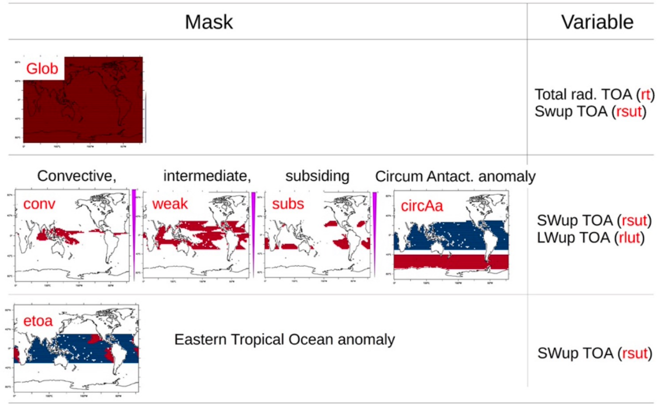
- Dry convective boundary layer (IHOP, Couvreux et al. 1996)
- Continental cumulus (ARMCU, Brown et al. 2002)
- Marine cumulus (RICO, van Zanten et al. 2011)
- Stratocumulus-cumulus transition (SANDU, Sandu et al. 2011)
- Reference LES: Meso-NH ou UCLA

*Metrics Retained for the SCM/LES Tuning*

| Case | IHOP | ARMCU | RICO | SANDU | SANDU | SANDU |
|---|---|---|---|---|---|---|
| Subcase | REF | REF | REF | REF | SLOW | FAST |
| Time | 7–9 | 7–9 | 19–25 | 50–60 | 50–60 | 50–60 |
| $\theta_{400-600\ \mathrm{m}}$ | X | X | - | - | - | - |
| $q_{v,400-600\ \mathrm{m}}$ | - | X | - | - | - | - |
| $\alpha_{cld,\mathrm{max}}$ | - | X | X | - | - | - |
| $z_{cld,\mathrm{ave}}$ | - | X | - | X | - | - |
| $z_{cld,\mathrm{max}}$ | - | X | - | X | X | X |

### *Metrics*

- 11 1D metrics: potential temperature, humidity averaged in the boundary layer, cloud fraction
- 11 3D metrics: TOA radiative budget components, global/regional averages




| Metrics | target $\mathrm{W\ m^{-2}}$ | error $\mathrm{W\ m^{-2}}$ |
|---|---|---|
| glob.rt | 2.5 | 0.2 |
| glob.rsut | 99.6 | 5 |
| circAa.rsut | 24.0 | 5 |
| circAa.rlut | -48.6 | 5 |
| subs.rsut | 84.9 | 5 |
| weak.rsut | 81.8 | 5 |
| conv.rsut | 103.2 | 5 |
| subs.rlut | 274.6 | 5 |
| weak.rlut | 264.3 | 5 |
| conv.rlut | 235.8 | 5 |
| etoa.rsut | 11.0 | 5 |

### *LMDZ model parameters*

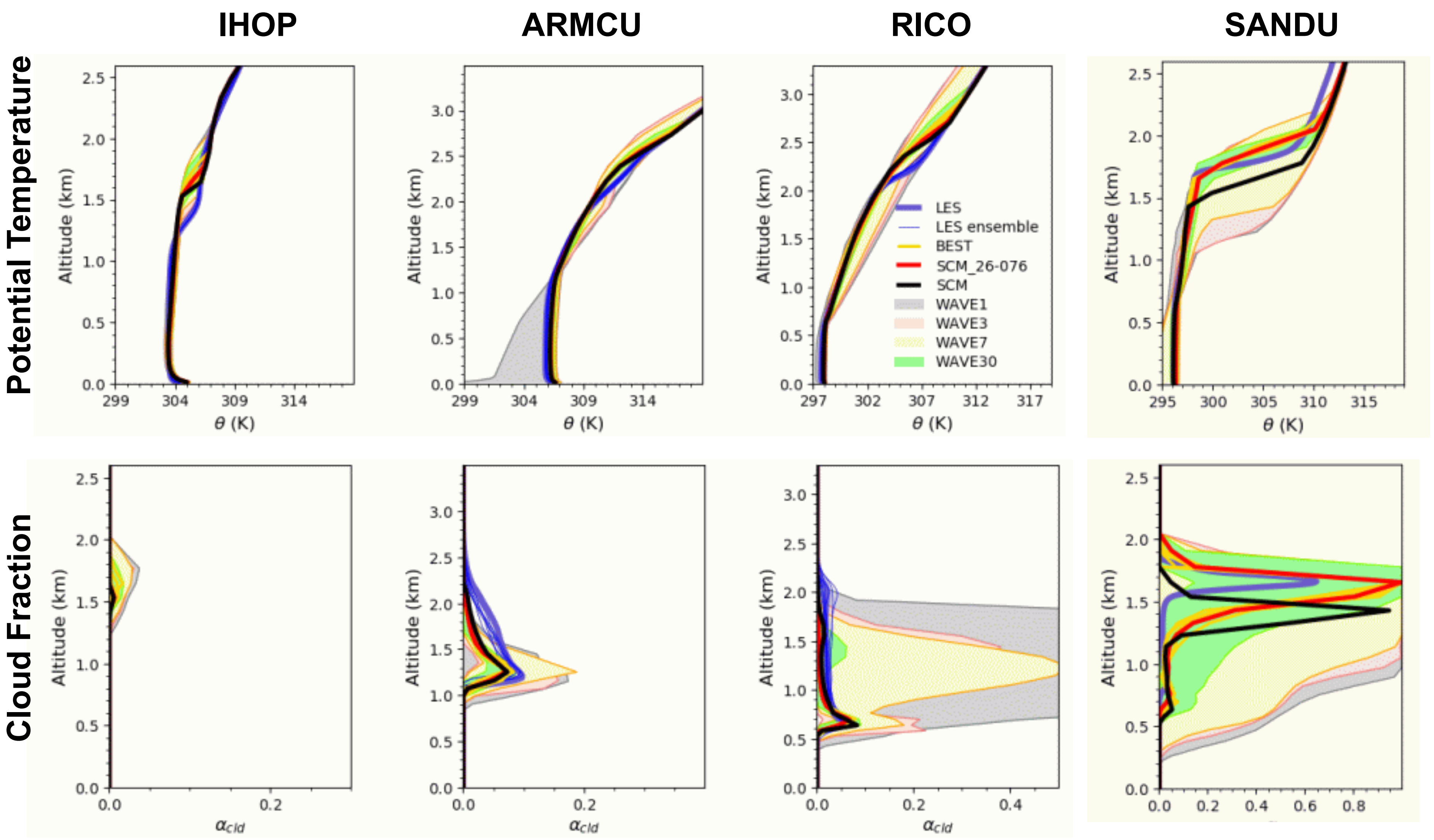
- 9 parameters from cloud, shallow convection and microphysics parameterizations

### *Waves*

- 30 waves with 1D configuration, progressive reduction of the implausibility cutoff (from 3 to 2)
- 2 waves with 3D configuration

*Hourdin et al. (2020, 10.1029/2020MS002225)*

# 30 Waves in 1D

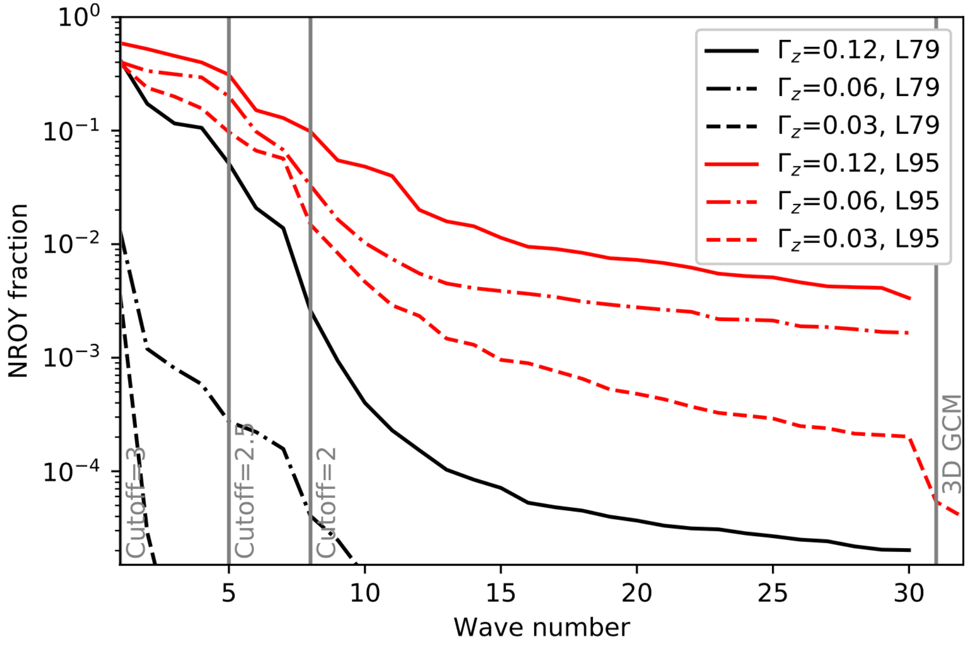


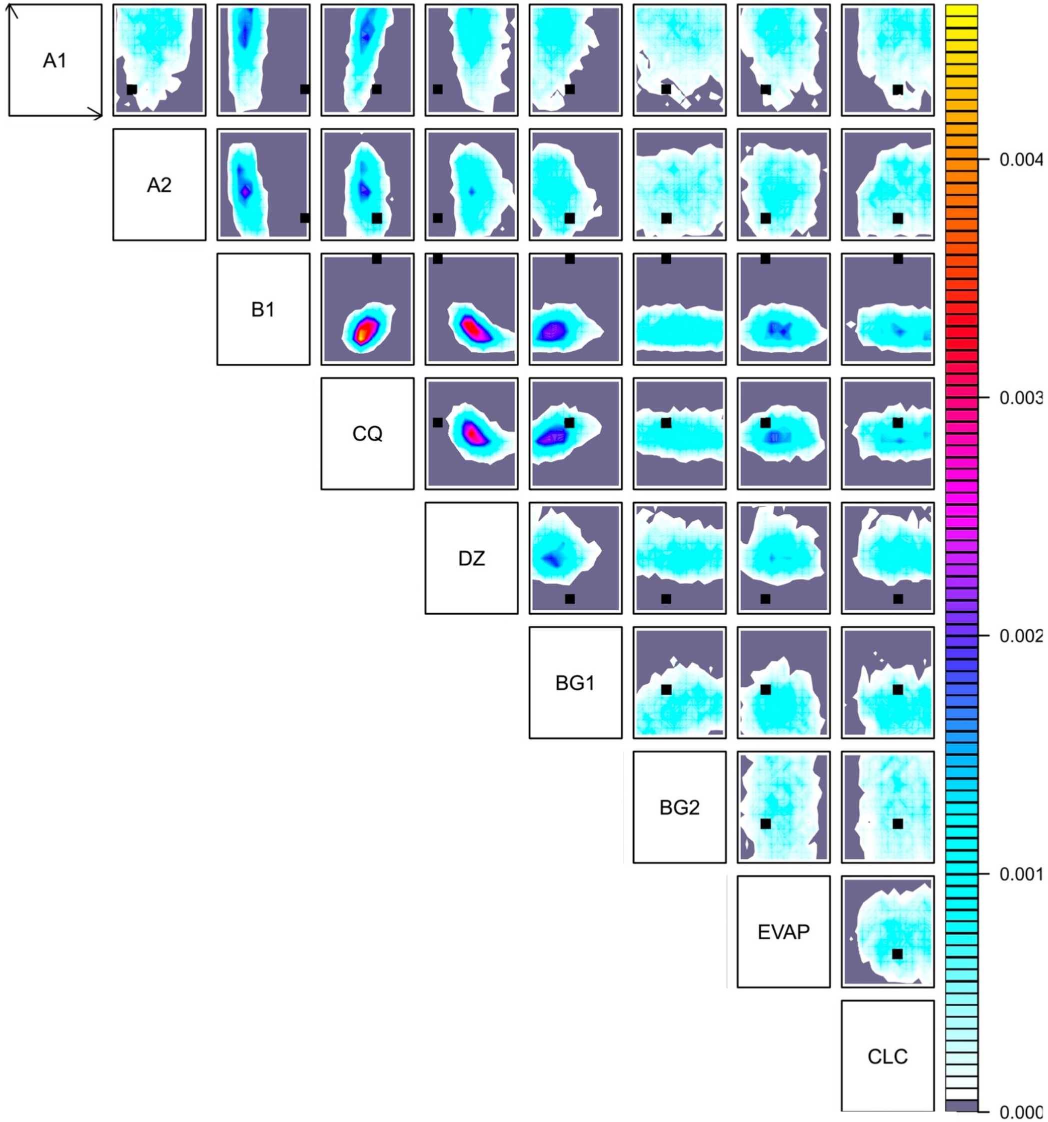

*Hourdin et al. (2020, 10.1029/2020MS002225)*

# 30 Waves in 1D

***NROY space***

- Approximate convergence
- Non-empty space (0.02% of the *input space*)
- LMDZ6 calibration non optimal



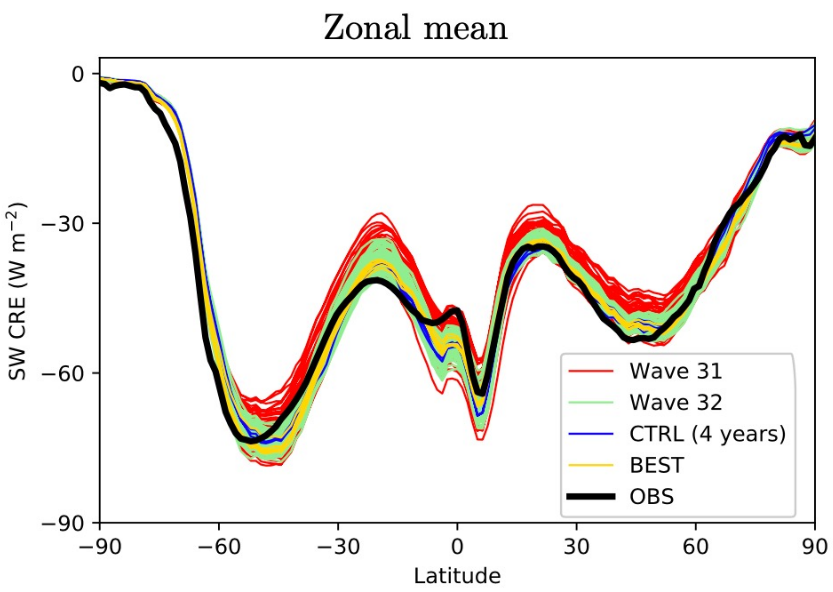



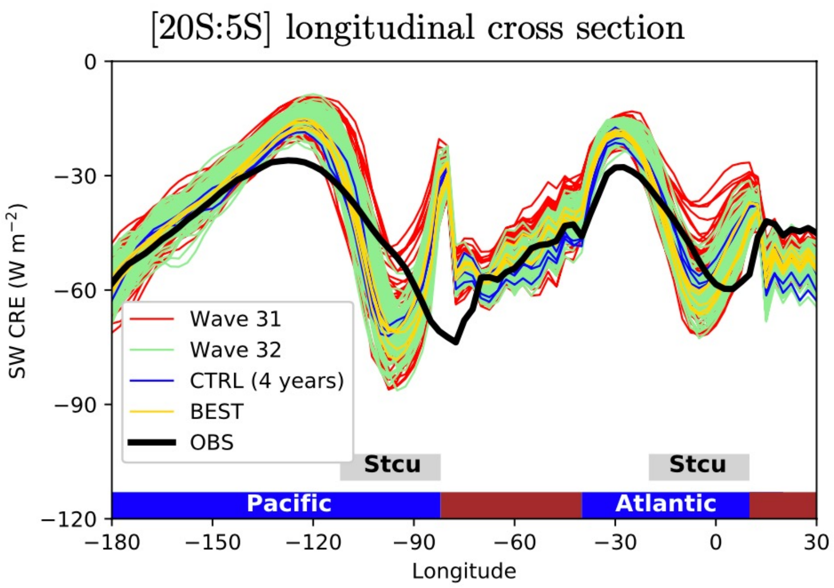

Hourdin et al. (2020, 10.1029/2020MS002225)

# 2 waves in 3D

***Continuing calibration in 3D, starting from 1D NROY space***

- Non-empty space (0.004% of the input space)
- 1D calibration resulted in an efficient pre-conditioning of the 3D model
- LMDZ6 calibrated 'by hand' relatively skilfull, though non optimal
- Probably necessary to add a few more waves (but dependent on computing resources)



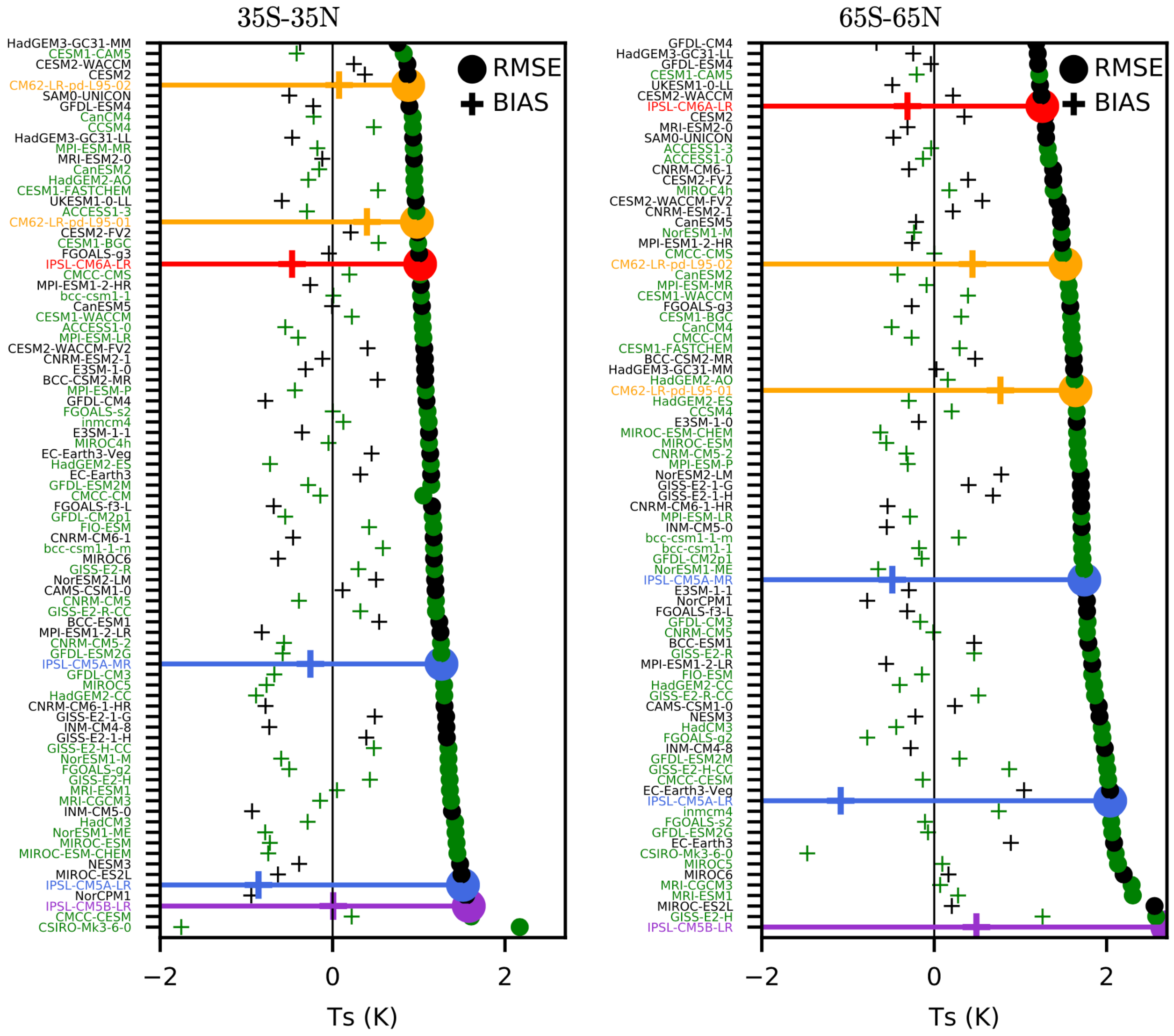




*Hourdin et al. (2020, 10.1029/2020MS002225)*

# Testing in ocean-atmosphere coupled configurations

***From atmosphere-only to OA coupled configration***

- Calibration by hand of the TOA energy budget (1 parameter)
- Behaviour similar to the version calibrated by hand over 2-3 years.




*Hourdin et al. (2020, 10.1029/2020MS002225)*

# Conclusions, Perspectives

## *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 to be made
- 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:

  accelerating spin-up, use of intermediate resolutions, fast/slow processes…
- Develop physical interpretations of what is happening in the calibration process.

## *More on the technical/statistical aspects*

- Develop strategies to better sample particularly small NROY spaces.
- Going beyond scalar metrics, emulate directly vectors/maps (e.g., using EOFs, *Salter et al. 2019*)
- Emulate the expert judgment (‘subjective metrics’, Xu et al. 2023)