Are GCMs obsolete?

Climeri Webinar

V. Balaji

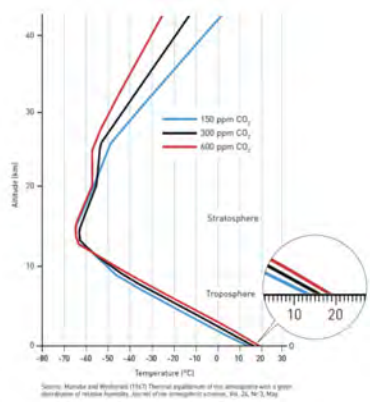
Laboratoire de Science de Climat et Environnement, Saclay Paris
et Schmidt Futures

11 juillet 2022

Outline

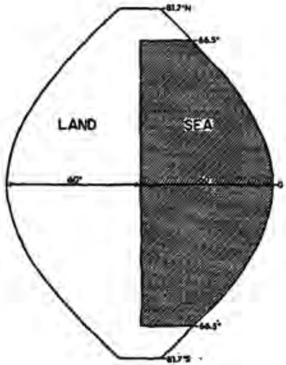
Outline

Manabe and Wetherald (1967): 1D model response to CO_2 doubling



“Radiative convective equilibrium of the atmosphere with a given distribution of relative humidity is computed as the asymptotic state of an initial value problem.”. Syukuro Manabe won the Nobel Prize in Physics, 2021.

- Recognized as a “milestone in scientific computing”, Nature (2006).
- Sector model of 120°
- 1 atmospheric year coupled to 100 ocean years
- 1200h for 1 simulated year (0.02 SYPD) on Univac 1108



Atmospheric response to doubled CO_2

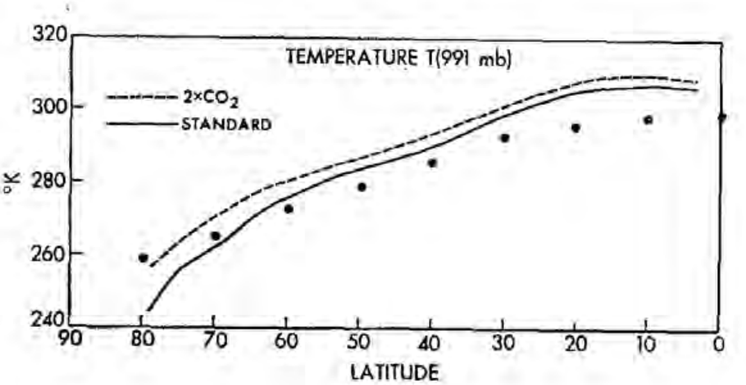


Fig 5 from Manabe and Wetherald (1975), equilibrium response to doubled CO_2.

Atmospheric response to doubled CO_2

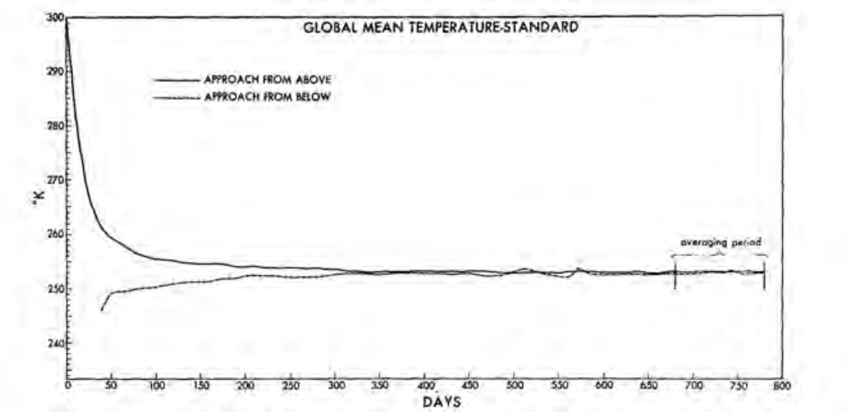
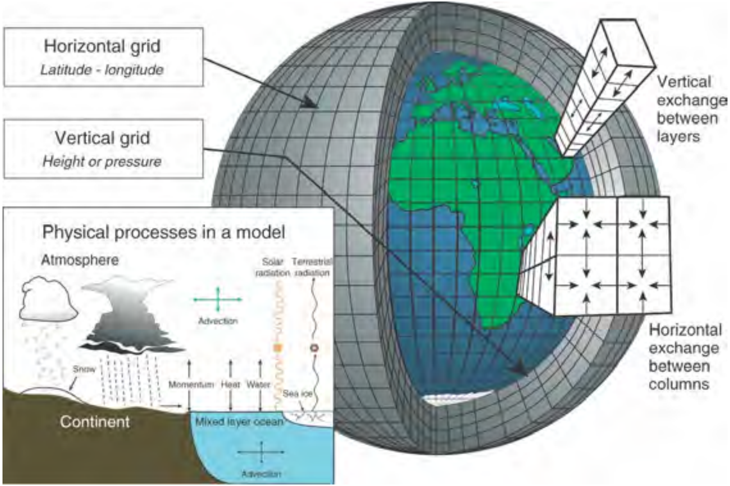


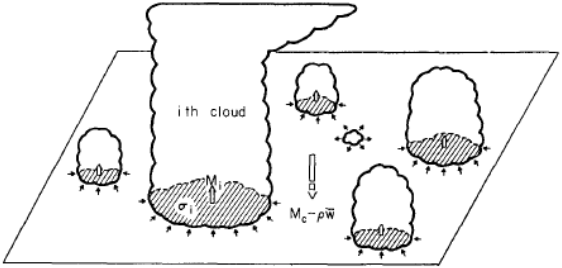
Fig 3 from Manabe and Wetherald (1975), equilibrium response to doubled CO_2. Spinup times in modern GCMs can be $\mathcal{O}$(1000 years).

The structure of a GCM, from Manabe to present day



From Edwards (2011). $\mathcal{O}$(10X) increase in resolution from Manabe and Bryan to CMIP6.

Parameterizing convection: slow(?) progress over 50 years



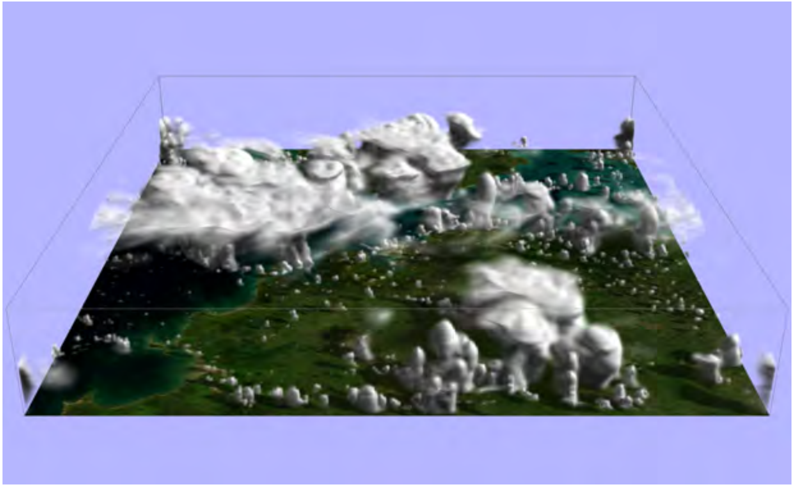
Arakawa and Schubert (1974): *Interaction of a cumulus cloud ensemble with the large-scale environment*.

Is the column broken?

Can the redistribution of heat, moisture, momentum by clouds be written as a function of a column state? In other words, are clouds parameterizable at all? The answer has often been no. “A problem that refuses to die”. Randall (2003)

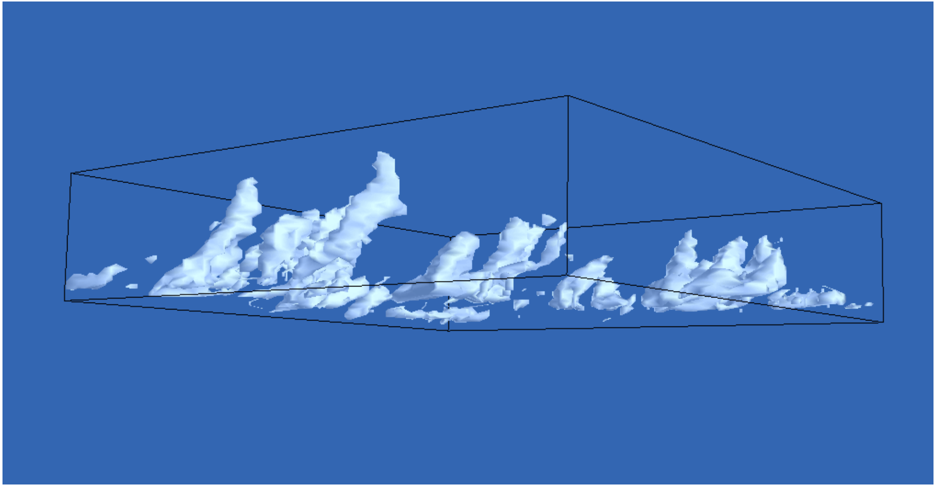
- Too many cloud habits
- Organized deep convection: squall lines, tropical cyclones, mesoscale convective systems: non-local physics from a column perspective.
- Sensitivity to small-scale dynamics: entraining updrafts, cold pools.
- Sensitivity to details of microphysics.
- Schemes tend to begin simple and end Ptolemaic.
- Systematic biases across GCMs.

Resolving atmospheric deep convection: CRMs



Courtesy W.-K. Tao, NASA. We can begin to resolve deep convection at km-scale.

Resolving boundary layer clouds: LES models

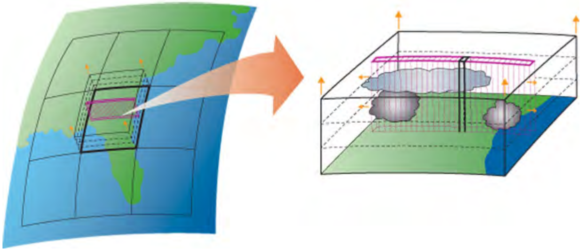


Courtesy UKMO GASS project. Typical resolution, $\mathcal{O}(10\text{ m})$.

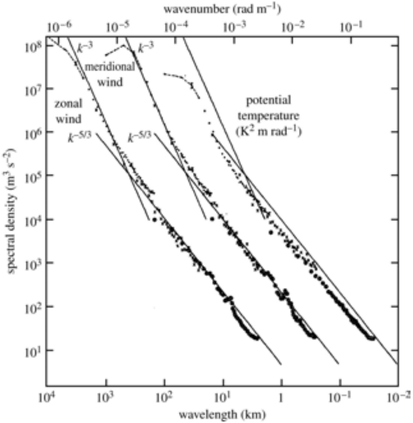
Superparameterization: embedded CRMs

From Randall (2003). Parishani et al (2017) attempt the same for boundary layer clouds.

- Assumes scale separation between GCM gridscale and embedded model.
- Only 2D CRMs feasible for computational reasons (or 2 orthogonal ones).
- CRMs can retain memory of their state, and potentially communicate with neighboring CRMs to enable mesoscale organization.
- Some success in improving climate simulations, but too expensive for a workhorse.
- Early target for ML: Gentine et al (2018)

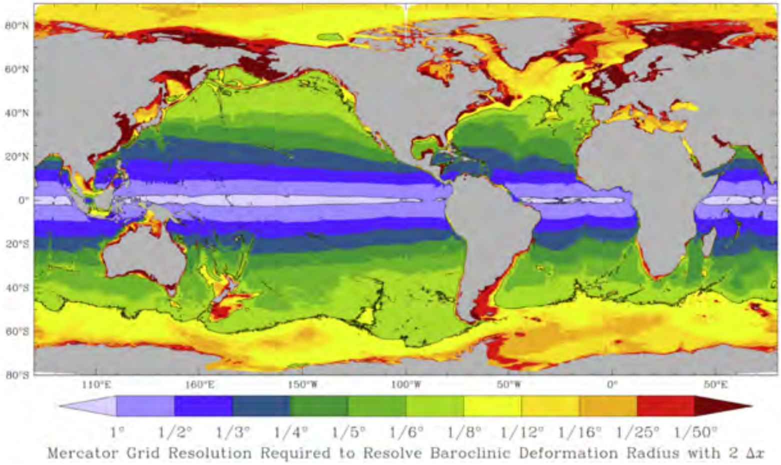


No separation of "large" and "small" scales



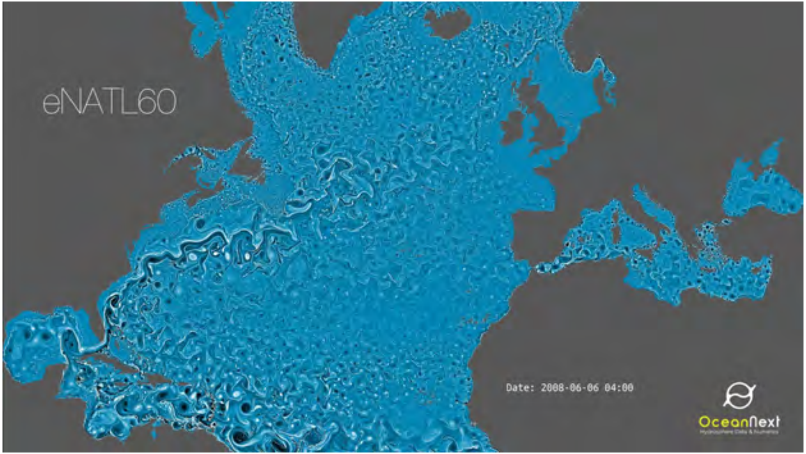
Nastrom and Gage (1985). More model fidelity, more complexity over time in small scales ("physics"). The backscatter idea (Jansen and Held 2014) provides an energetically consistent framework for SGS.

Eddy resolving scales in the ocean



From Hallberg (2013).

Coarse-graining without scale separation



eNATL60 dataset courtesy Julien le Sommer and collaborators. Can we assume a structure for learning. e.g “GM+E” Bachman 2019? See Sommer et al AGU 2019.

The climate Turing test: global CRMs

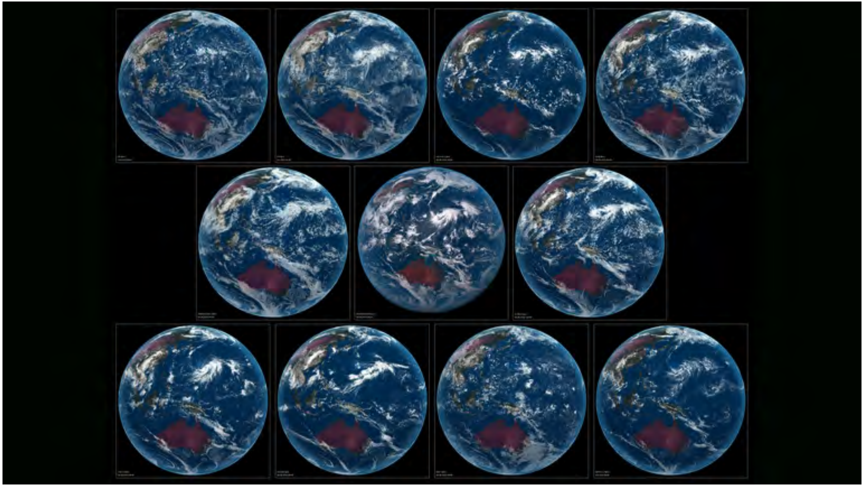
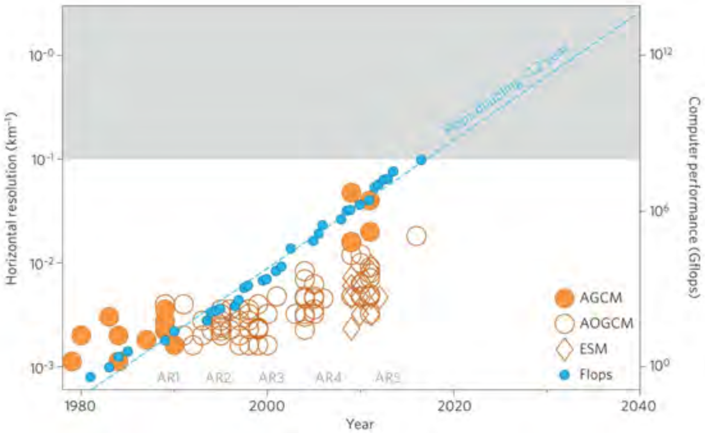


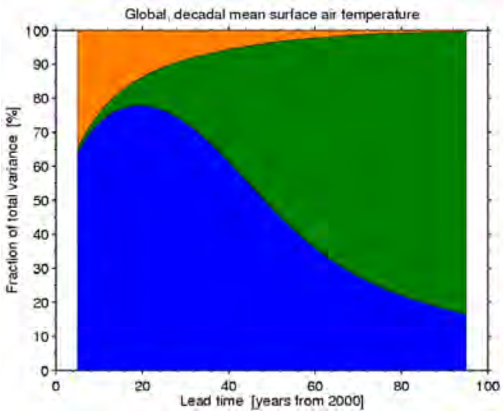
Figure courtesy the DYAMOND initiative.

Evolution of model resolution



From Schneider et al (2017). At GFDL: 10X from Manabe and Bryan (1969) to Held et al (2019).

Science requires going beyond observations

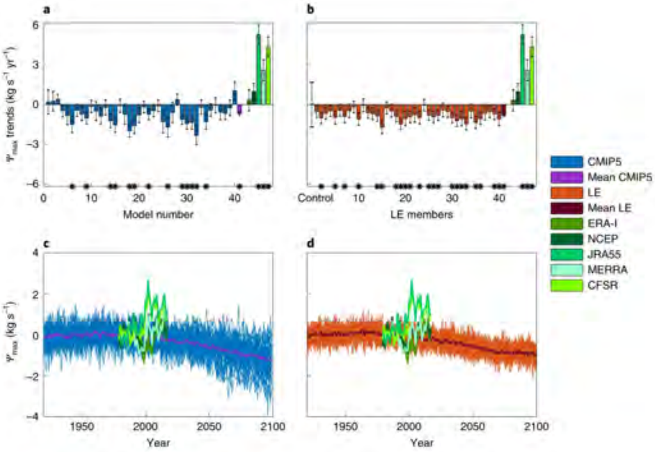


Sources of uncertainty in climate simulation:

- *chaotic uncertainty* or internal variability
- *scenario uncertainty* dependent on policy and human actions.
- *structural/epistemic uncertainty* or imperfect understanding.

Models must also generate counterfactual values! From Hawkins and Sutton (2009).
Baseline requirement: a climate model must be capable of 100 simulations of 100 SY each in 100 days.

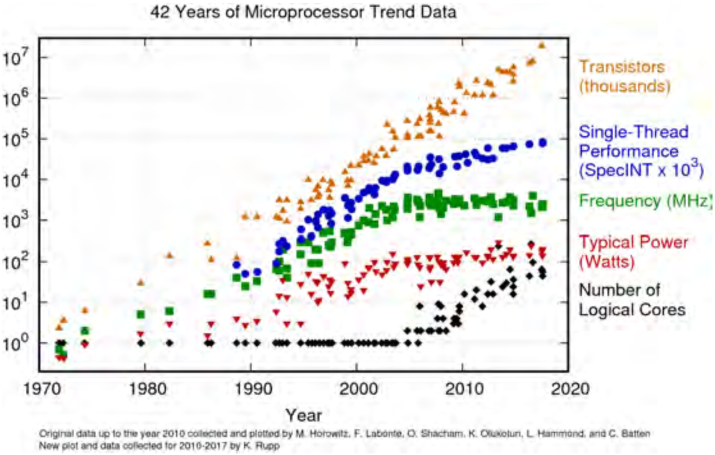
Overfitting to present day climate?



Hadley cell strength is likely correct in models and not in "observations"!
From Chemke and Polvani (2019).

Outline

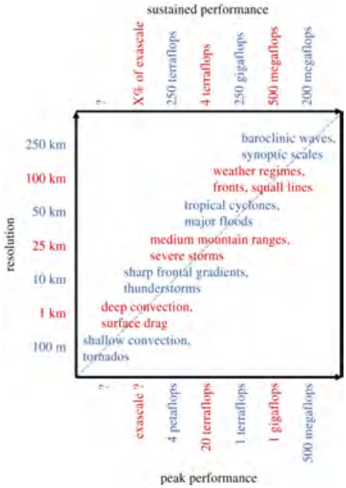
End of Dennard scaling: computers get bigger, not faster



From 42 Years of Microprocessor Trend Data, courtesy Karl Rupp. Weak scaling (bigger problems in the same time) works, strong scaling (same problem in less time) doesn't.

What can we expect at an exaflop?

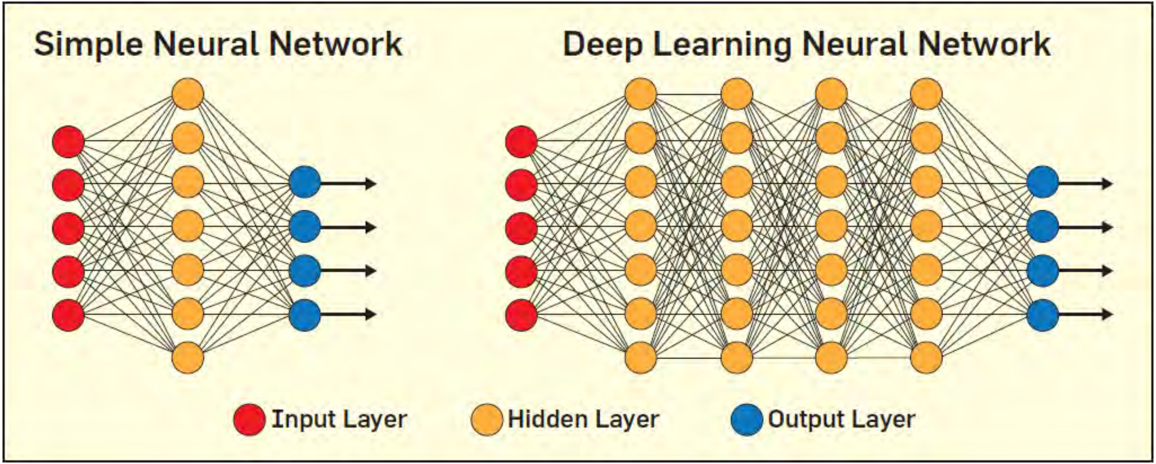
Will exascale be the rescue? Neumann et al (2019).



Hypothesis: vastly reduced uncertainty at ∼1 km (see "digital twins", DestinE, NextGEMS, ...)

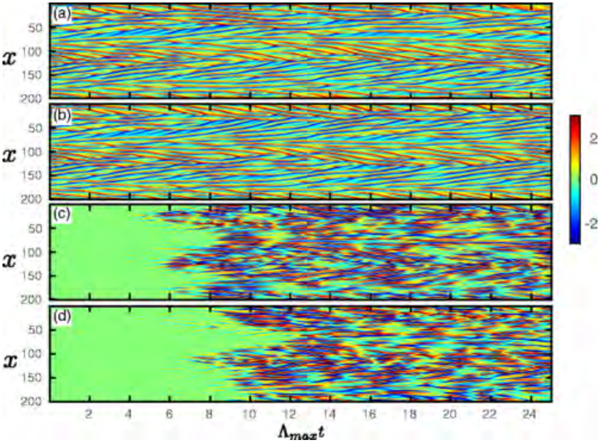
- ICON projects that a 1 km global model will run at 0.06 SYPD on "pre-exascale" technology: 17X improvement needed for 1 SYPD.
- Large nodecount, see e.g Caldwell et al (2021).
- DECK: 1000 SY.
- A full suite of hindcasts for seasonal forecasting: 10,000 SY.
- Ocean state needed for seasonal prediction and beyond as well!

Deep Learning



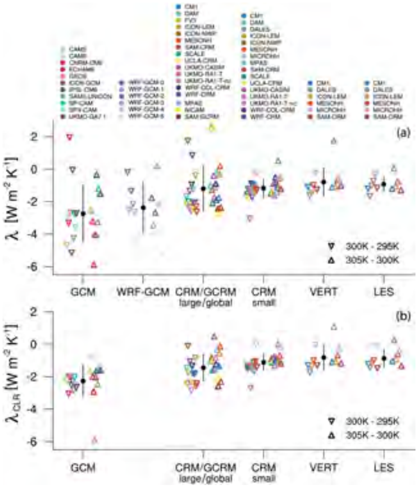
From Edwards (2018), ACM. Dense linear algebra with high operation intensity, data-intensive.

Model-free prediction for stationary problems



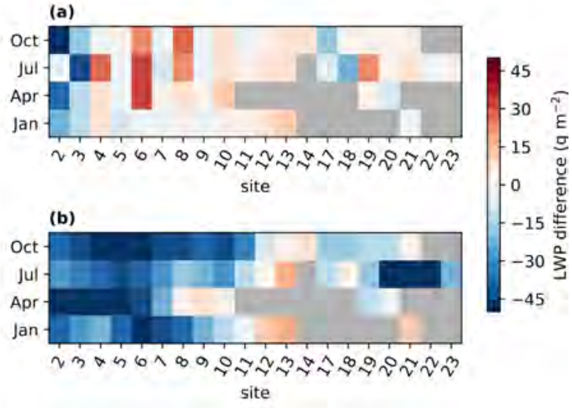
From Pathak et al, PRL (2018), *Model-Free Prediction of Chaotic Systems from Data*. See also Patel et al (2021). But climate is non-stationary, see O'Gorman and Dwyer (2018), Dixon et al (2016). Use models "up the ladder" for training.

Extreme spread in climate sensitivity in RCEMIP



From Becker and Wing (2020).

LES reduces GCM structural uncertainty, but has its own



From Shen et al 2021. Sensitive to LES details (numerics, closure), see Couvreux et al (2020), Beare et al (2006), Siebesma et al (2006), ...

Outline

ML for model calibration: the sales pitch!

- Models, even “seamless” ones, may be configured or calibrated differently for different problems (e.g forecast horizons).
- Each problem carries an implicit cost function by which a model configuration is declared suitable.
- Models do not converge cleanly with resolution: much unresolved physics is not yet “scale-aware”.
- Computation alone is not going to make the problem go away (not everyone agrees...)
- Important new constraints on models from observations (new generation of satellites, Argo...)
- While data science is a misnomer (what is non-data science?) the convergence of computation and statistics that we call ML provides paths forward toward seamlessness: traceable hierarchies of scale, *Charney’s ladder*

Model calibration

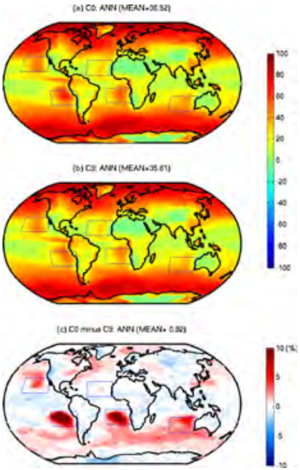
Model calibration or “tuning” consists of reducing overall model error (relative to some goal of modeling) by modifying parameters. In principle, minimizing some cost function:

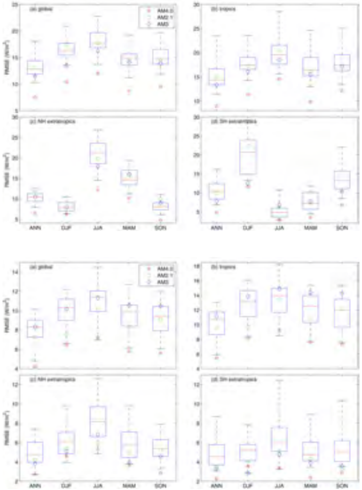
$$C(p_1, p_2, ...) = \sum_1^N \omega_i \|\phi_i - \phi_i^{obs}\|$$

- Usually the p must be chosen within some observed or theoretical range $p_{min} \leq p \leq p_{max}$.
- “Fudge factors” (applying known wrong values) generally frowned upon (see Shackley et al 1999 on “flux adjustments”.)
- The choice of ω_i is part of the lab's “culture”. Cost also plays a role.
- The choice of ϕ_i^{obs} is also troublesome:
  - overlap between “tuning” metrics and “evaluation” metrics.
  - “Over-tuning”: remember “reality” is but one ensemble member...

See for example, Hourdin et al (BAMS 2017)

Example: the tuning of GFDL's AM4/OM4/CM4 models





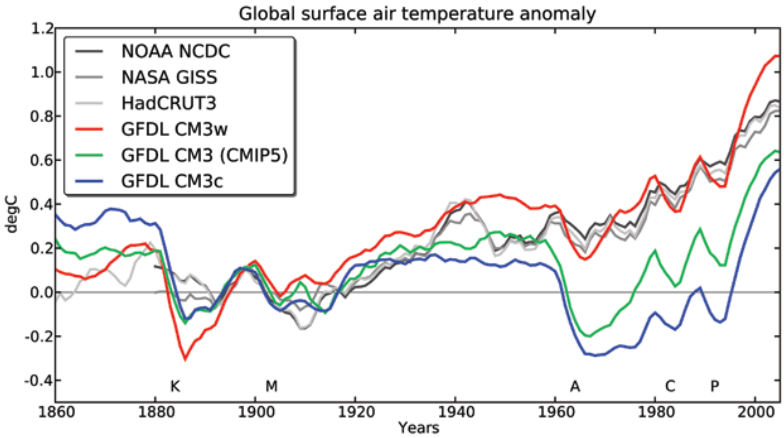
From Zhao et al (2018).

Example: the tuning of GFDL's AM4/OM4/CM4 models

- The GFDL Global Atmosphere and Land Model AM4.0/LM4.0: 2. Model Description, Sensitivity Studies, and Tuning Strategies
- The GFDL Global Ocean and Sea Ice Model OM4.0: Model Description and Simulation Features: "We hypothesize that the development of a climate model is optimized only with close coordination across component model development."
- Structure and Performance of GFDL's CM4.0 Climate Model: "CM4.0 is sensitive to a number of features [...] much less apparent in uncoupled atmosphere/land simulations"
- Climate Sensitivity of GFDL's CM4.0
- The GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1): Overall Coupled Model Description and Simulation Characteristics

The JAMES special issue on GFDL's "4 series" models. 50,000 SY of coupled models run during model development, 10,000 SY of "CMIP6 runs".

Should we tune to get the 20th century?



Tuning reduces model bias without violating process fidelity (but poses a problem for validation). From Golaz et al 2013.

Parameter optimization, elimination, uncertainty quantification

Goal: explore parameter space of model while minimizing the use of expensive forward models.

- Parametric uncertainty vs structural uncertainty.
- A two stage process: process fidelity followed by global constraints.
- The choice of cost function.
- Metric weights and normalization.
- Do observations sample the space sufficiently?
- If models "higher on the ladder" are used for calibration, are they representative of all possible states? What are the associated uncertainties?
- Internal feedbacks on multiple timescales, and compensating errors.

HighTune: Formulating the problem

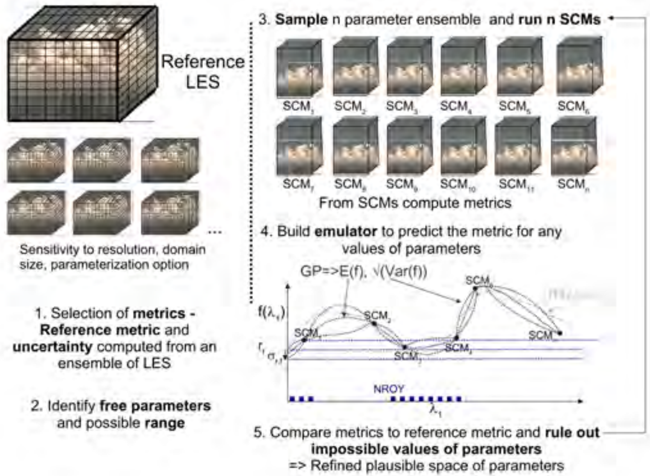
$$\frac{\partial \mathbf{x}}{\partial t} = D(\mathbf{x}) + \sum_n (\mathcal{P}_n(\mathbf{x}, \lambda_n))$$

- Structure is given by $\mathcal{P}$, we are trying to calibrate values of a vector of parameters λ
- Multiple metrics we wish to satisfy. For each metric f, define a distance given by:

$$I_f(\lambda) = \frac{\| r_f - E_f[\lambda] \|}{\sigma^2_{r,f} + \sigma^2_{d,f} + Var[f(\lambda)]}$$

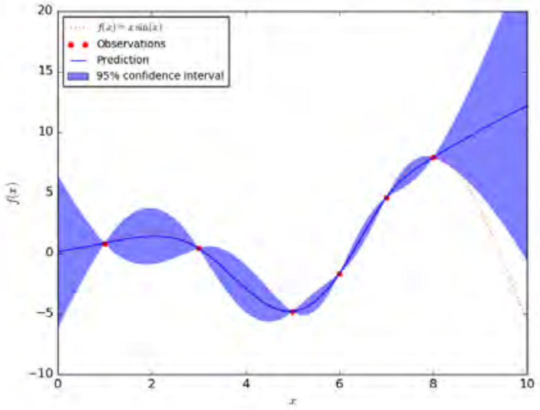
- Euclidean distance over history normalized by error (observational, structural, chaotic)
- Sample λ space as exhaustively as practical for $I < T$, the NROY space. Iterate in *waves*. Can use different metrics in subsequent waves.

$$\mathrm{NROY}^n = \cap_k \mathrm{NROY}_{f_k}$$



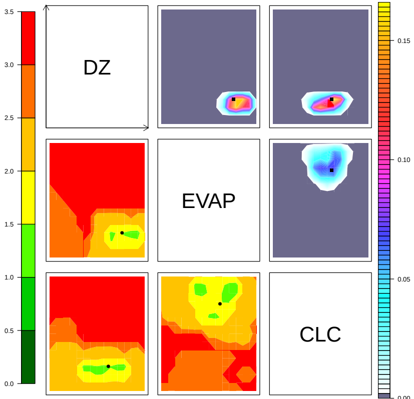
- LES as ground truth, multiple variants to get “observational error”.
- Emulate LES using SCMs encoding all the $\mathcal{P}$.
- Latin hypercube sampling of λ
- Fit Gaussian processes to SCMs to densely sample all values of λ

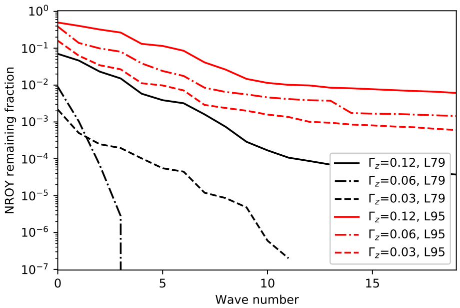
Gaussian processes



- Extremely standard emulator, widely available in python libraries
- Very poor at extrapolation, so training data must span phase space!

Hourdin et al (2020)





- Eliminate implausible parameter space comparing SCMs with LES.
- ... leaving irreducible (“structural”) model error.

Bibliography

- Couvreux et al 2020: Process-based climate model development harnessing machine learning: I. a calibration tool for parameterization improvement.
- Hourdin et al 2020: Process-based climate model development harnessing machine learning: II. model calibration from single column to global
- Hourdin et al 2017: The art and science of climate model tuning.
- Williamson et al 2013: History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble
- Williamson et al 2017: Tuning without over-tuning: parametric uncertainty quantification for the NEMO ocean model

CliMA: Calibrate, Emulate, Sample

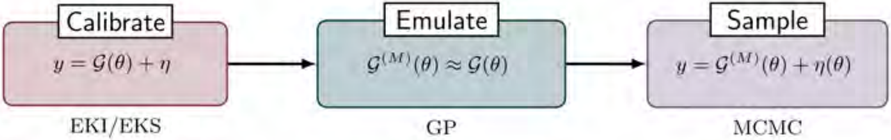


Fig. 1. Schematic of approximate Bayesian inversion method to find θ from y. EKI/EKS produce a small number of approximate (expensive) samples $\{\theta^{(m)}\}_{m=1}^{M}$. These are used to train a GP approximation $\mathcal{G}^{(M)}$ of $\mathcal{G}$, used within MCMC to produce a large number of approximate (cheap) samples $\{\theta^{(n)}\}_{n=1}^{N_s}$, $N_s \gg M$.

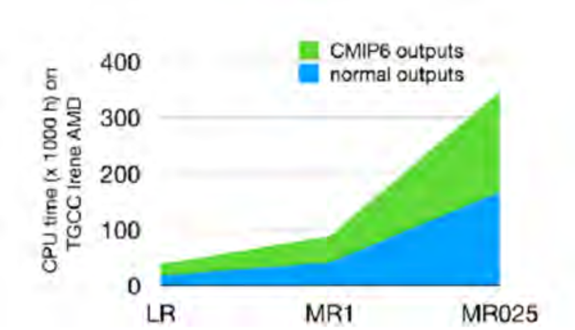
- *Calibrate*: approximately locate attractor using expensive forward model
- *Emulate*: cheap GP emulator to map parameter space near attractor
- *Sample*: MCMC sampling of parameter space for uncertainty quantification (parameter vector with error bounds)

Applied to boundary layer and shallow cloud (EDMF) parameterizations, Cleary et al 2020, Dunbar et al 2021.

Bibliography

- Schneider et al (2017): Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations
- Pressel et al 2017: Numerics and subgrid-scale modeling in large eddy simulations of stratocumulus clouds
- Cleary et al 2020: Calibrate, emulate, sample
- Dunbar et al 2021; Calibration and Uncertainty Quantification of Convective Parameters in an Idealized GCM
- Shen et al 2021: A Library of Large-eddy Simulations for Calibrating Cloud Parameterizations

Beyond LES: calibration of coupled models

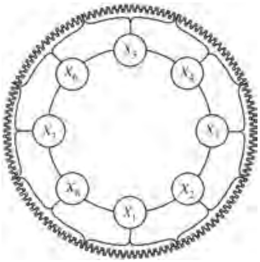


Post Hourdin et al automatic tuning:

- 5 new piCtrl coupled simulations, 250 SY each
- excessive cold biases and sea ice cover relative to baseline IPSL-CM6
- required extensive retuning of ocean and sea ice!

- GFDL experience is similar: about 50000 SY of coupled runs of CM4 and ESM4 during model calibration in addition to AMIP.

Lorenz 96, a nice abstraction

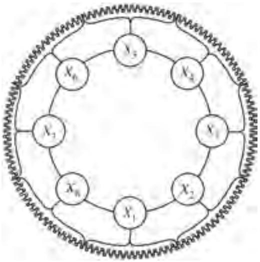


$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b}\sum_{j=1}^{J} Y_{j,k} + f \tag{1}$$

$$\frac{dY_{j,k}}{dt} = -cbY_{j+1,l}(Y_{j+2,k} - Y_{j-1,k}) - cY_{j,k} + \frac{hc}{b}X_k \tag{2}$$

- A simplified multiscale system (X and Y can stand for resolved/unresolved, slow/fast), where coupling strength can be varied... maybe too interesting? See metastability issues in Schneider et al (2017).
- Maybe too simple? (from Stephan Rasp's blog)

Lorenz 96 again: history matching for an "AOGCM"

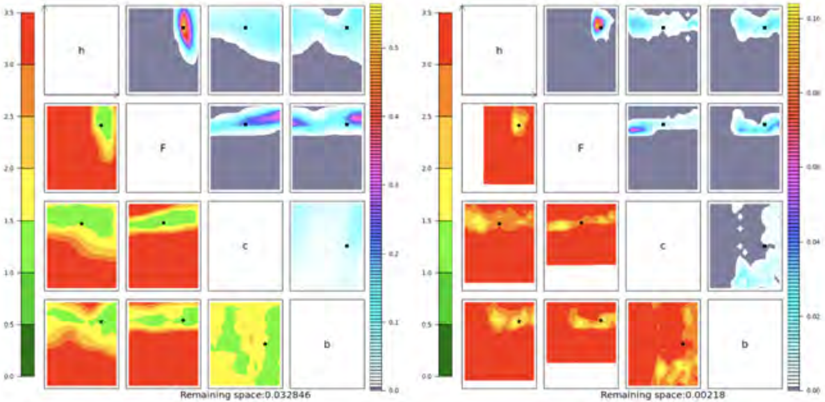


$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b}\sum_{j=1}^{J} Y_{j,k} + f \tag{1}$$

$$\frac{dY_{j,k}}{dt} = -cbY_{j+1,l}(Y_{j+2,k} - Y_{j-1,k}) - cY_{j,k} + \frac{hc}{b}X_k \tag{2}$$

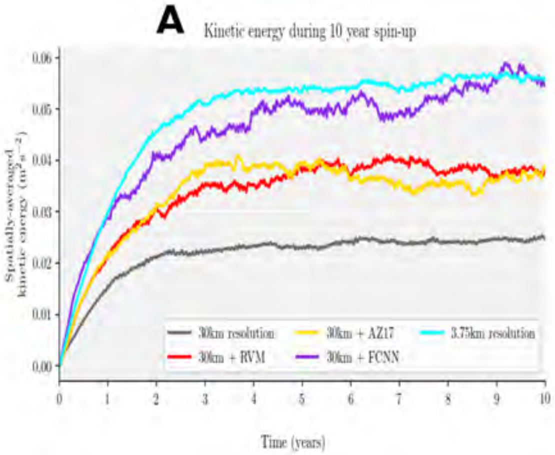
- Similar metrics to Schneider et al (2017) $f(X,Y) = (X, \overline{Y}, X^2, X\overline{Y}, \overline{Y}^2)$
- as usual try to recover $F, h, \log c, b$ from prior "truth" run.
- AMIP: apply only Y constraints; OMIP = apply only X constraints.
- Investigate length of sample needed for training.
- Lguensat, Balaji, Deshayes 2021, *in prep.*

History Matching on Lorenz96



- History matching efficiently reduces NROY space.
- “AMIP” and “OMIP” experiments underway.
- From Lguensat, Balaji, Deshayes, *in prep.*

Discovering subgrid momentum closures

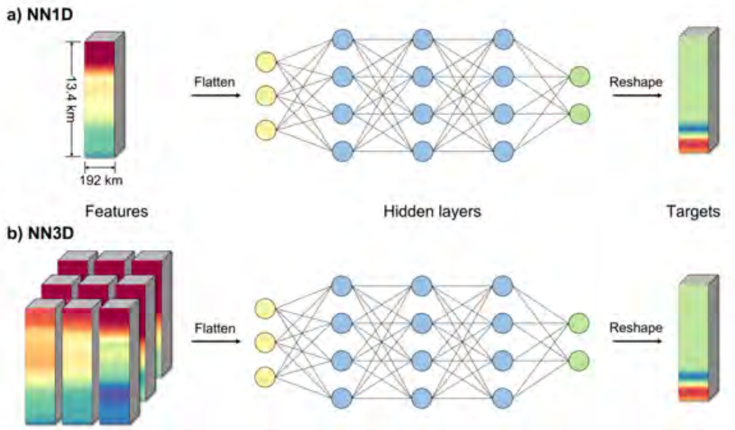


Zanna and Bolton 2020 returns a closed-form expression for subgrid momentum closures:

$$\mathbf{S_u} = (\overline{\mathbf{u}}.\nabla)\overline{\mathbf{u}} - \overline{\mathbf{u}.\nabla\mathbf{u}}$$

where *relevance vector machine* techniques yield a representation similar in form to Anstey and Zanna (2017).

Non-local parameterizations using ML

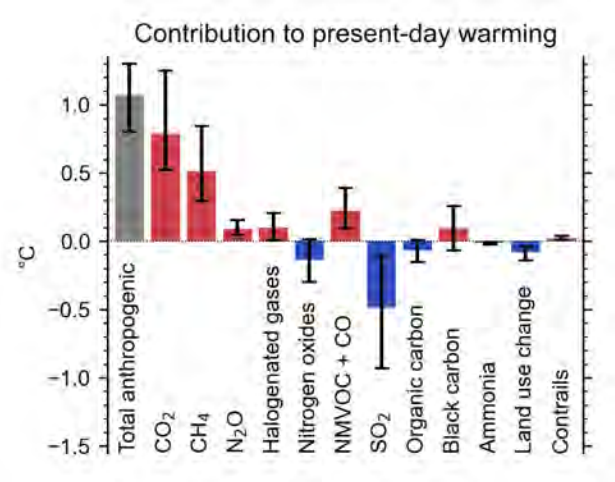


From Wang et al (2021). Can be non-local in (past) time as well! The DataWave project is attempting similar approaches for gravity wave parameterization.

Bibliography

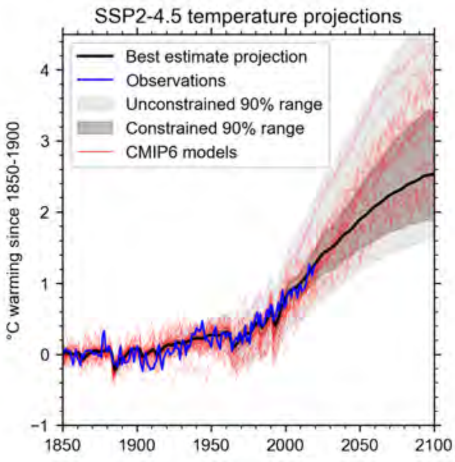
- Schmidt and Lipson, *Science*, 2009: Distilling Free-Form Natural Laws from Experimental Data.
- Gaitán, Balaji and Moore (2016): *Can we obtain viable alternatives to Manning's equation using genetic programming?*
- Rudy et al (2017): Data-driven discovery of partial differential equations
- Zanna and Bolton 2020: Data-driven Equation Discovery of Ocean Mesoscale Closures
- Balaji 2021: Climbing down Charney's ladder: machine learning and the post-Dennard era of computational climate science

Emulators in IPCC-AR6: sampling counterfactuals



From Chris Smith's *CarbonBrief* guest post, 28 Sep 2021.

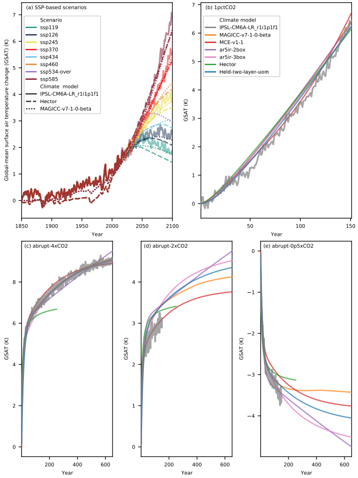
Emulators in IPCC-AR6: reducing CMIP6 spread in ECS



From Chris Smith's *CarbonBrief* guest post, 28 Sep 2021.

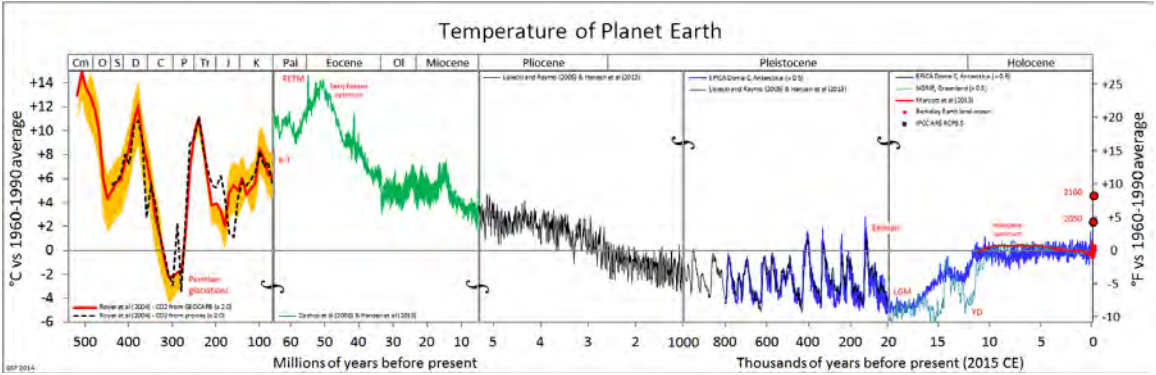
Structural uncertainty across reduced complexity models

- RCMs vary a lot in design: impulse-response models, single column models of varying complexity
- Millions of times faster than ESMs!
- Connection to climate physics can be tenuous!
- “The role of the ESM is increasingly as a target for robust emulation”: Ben Sanderson, WGCM24, Dec 2021.



From Nicholls et al (2020).

Paleoclimate constraints on sensitivity



- From Wikipedia.
- Paleoclimate provides stringent out-of-sample tests on novel climate models, and a strong separate line of evidence on ECS, see Sherwood et al (2020).

Past warm climates

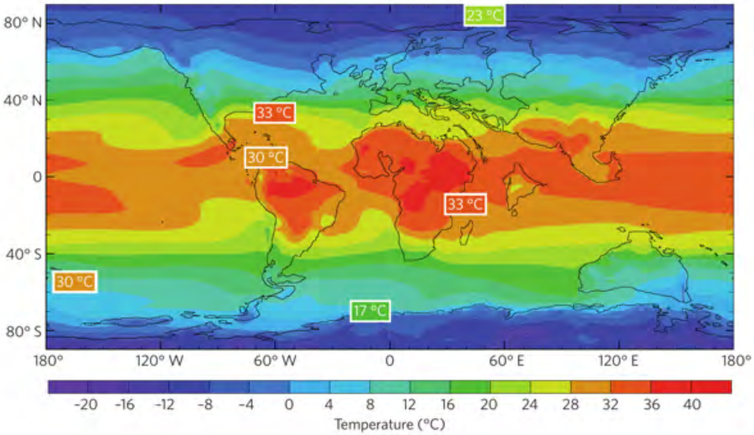


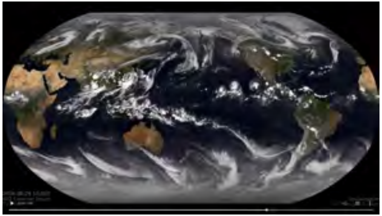
Fig 1 from Valdes (2011).

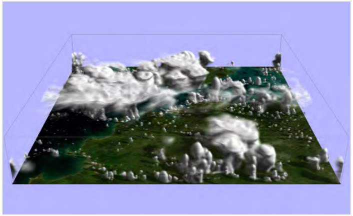
Outline

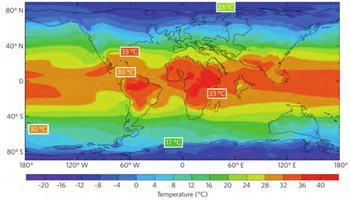
Some remarks

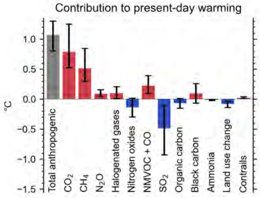
- Climate is not weather: in the ML era, they might further diverge, as model-free methods become successful in weather forecasting. See Guardian, 9 January 2022: Are we witnessing the dawn of post-theory science?
- Computers are getting bigger, not faster: which is ok for weak scaling problems (more degrees of freedom, same SYPD), but not for strong scaling (fewer degrees of freedom, more SYPD).
- There is always a cost function: any model has been calibrated to meet its requirements ("fit for purpose"). Can new methods yield model development in weeks, not years?
- Model calibration is needed at any resolution: we need methods of fast sampling of parametric and structural uncertainty.
- Decoupling of reduced-complexity models from climate models carries epistemic risk.
- Readings: Charney's ladder, Are GCMs obsolete? (submitted to PNAS), Saravanan's book The Climate Demon, Ben Sanderson's excellent talk at WGCM24.
- Workshops: Modeling Hierarchies 2022.

GCMS, not the end of the road, but the crossroads!









Extract from Manabe press conference, 5 October 2021

on Youtube. Other links: Annonce CEA, La Météorologie