The other climate crisis

Bjorn Stevens

(bjorn.stevens@mpimet.mpg.de)

MAX-PLANCK-INSTITUT
FÜR METEOROLOGIE



A Kuhnian Crisis

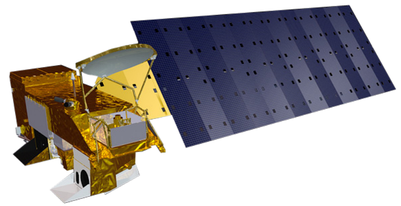
A period characterized by:

- An accumulation of anomalies (discrepancies between the standard model/explanation, and observations)
- The emergence of new theories vying to replace the standard model

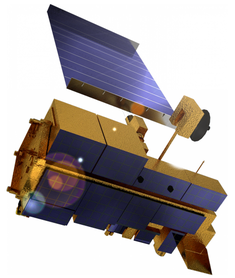
… and often scientifically fertile

Shaw and Stevens, *Nature* (2025)

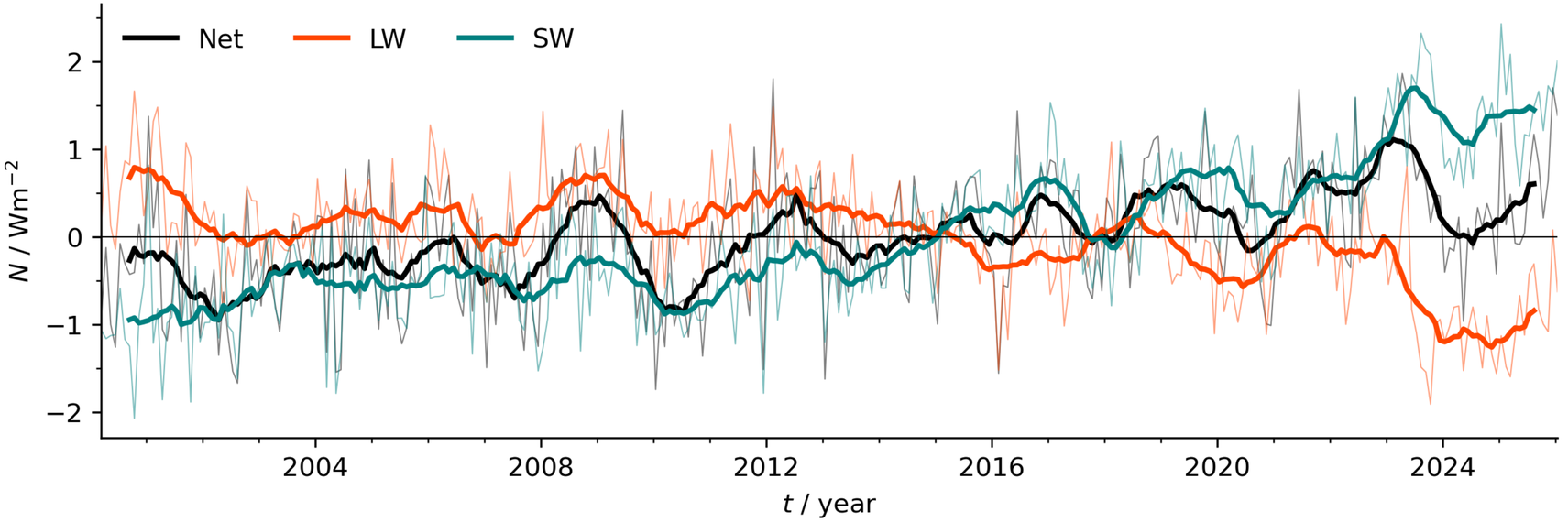
We can now measure the net flow of energy into the Earth system, and its trend over time



Artist's rendering of the *Aqua* satellite



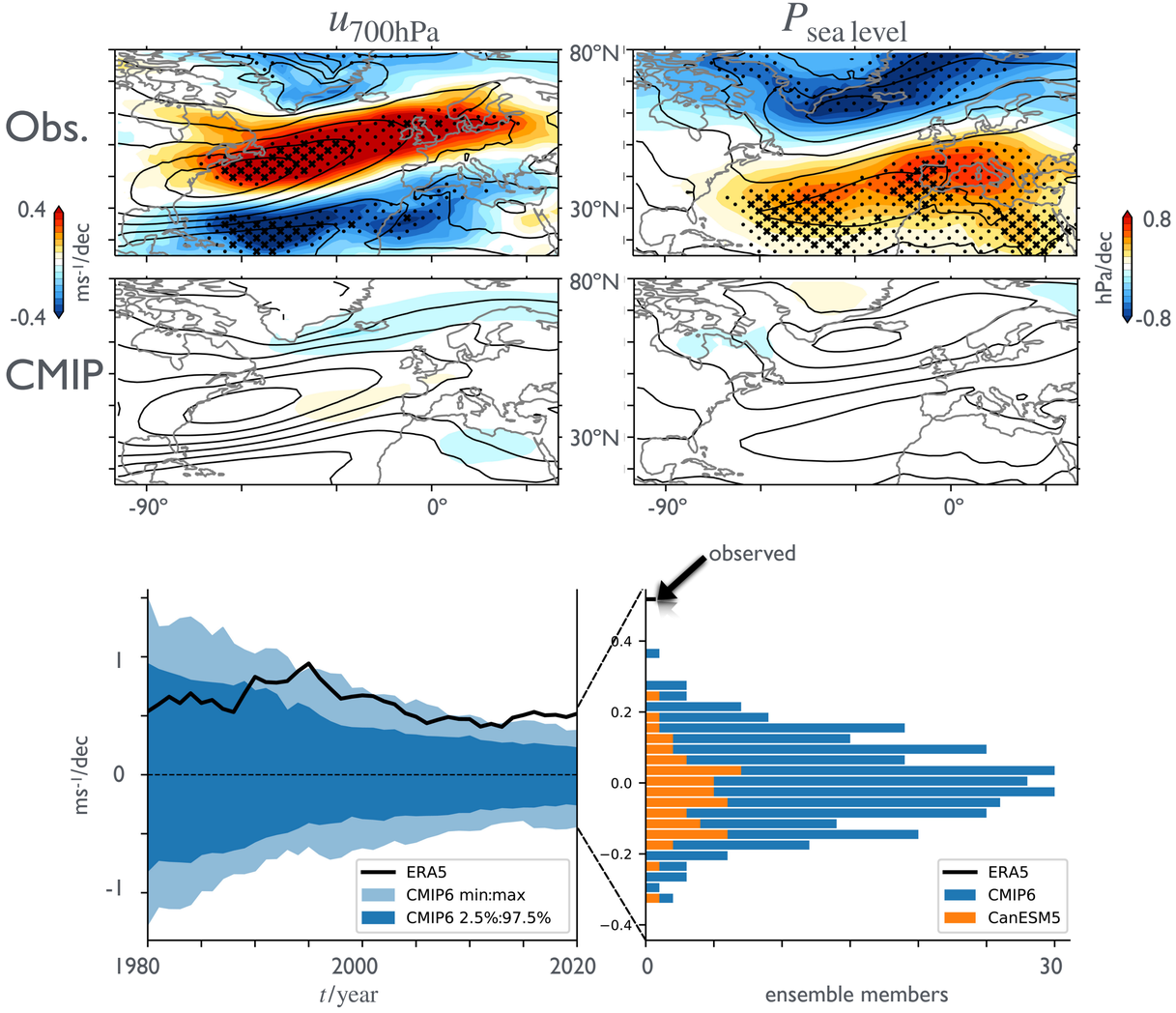
Artist's rendering of the *Terra* satellite

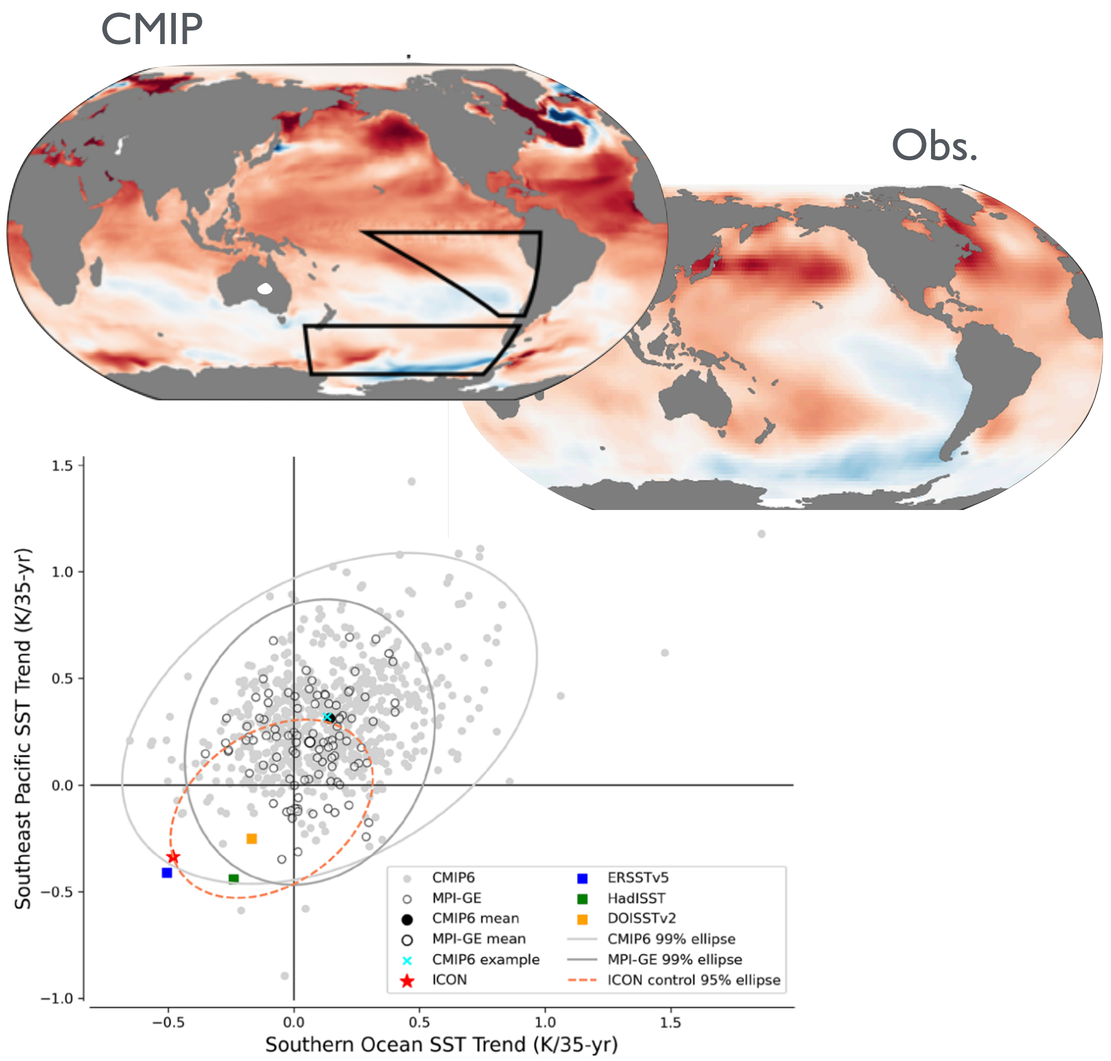


It is increasing at a rate, with a pattern, and following a spectral decomposition that we don't understand

Earth's net incoming energy-flux imbalance (and its implications for understanding climate change), Stevens et al., 2026

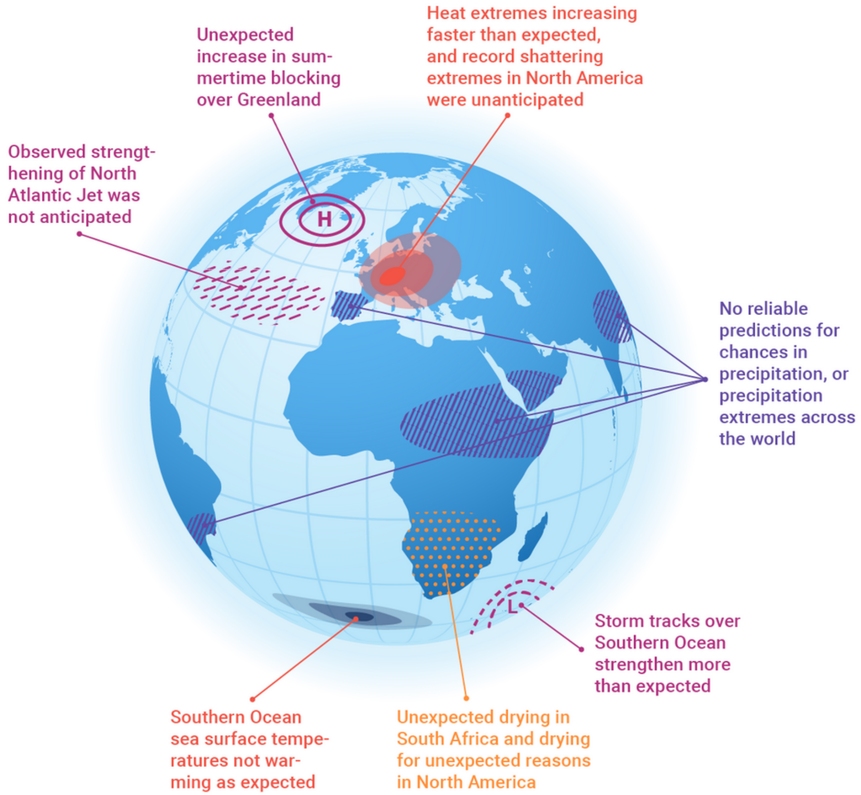
The accumulation of anomalies — (planetary scale)





Blackport and Fyfe, *Sci.Adv* (2022), Kang et al, *PNAS* (2026)

The accumulation of anomalies — regionally



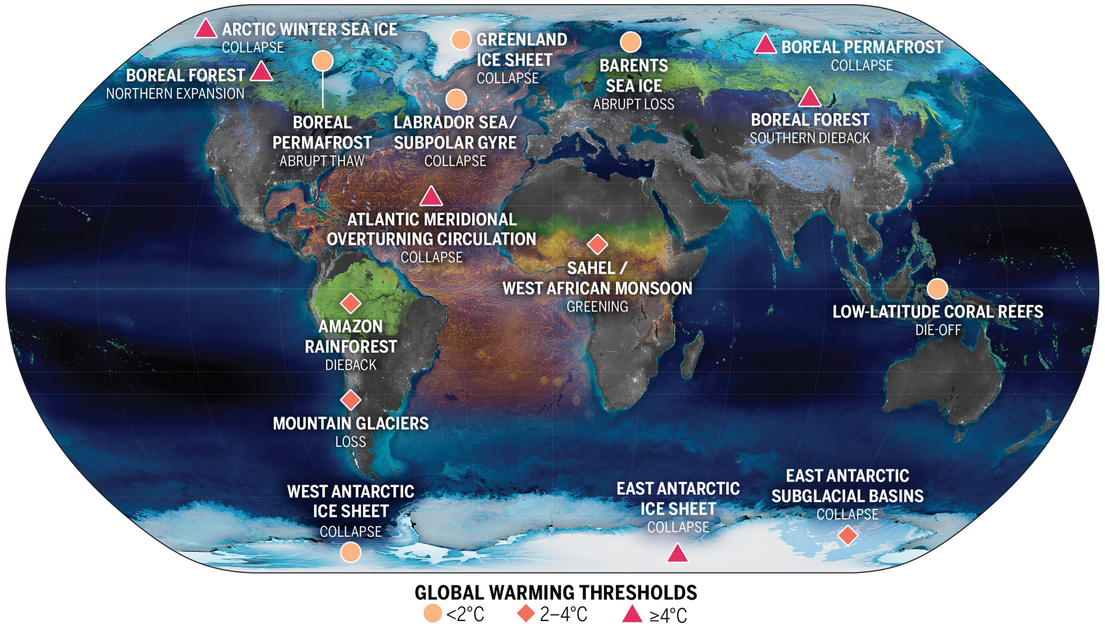
Quantity	Summary	References	
Tropical SST pattern	Most model ensemble members fail to capture the observed strengthening of the tropical Pacific SST gradient and instead predict a weakening (Fig. 1B). This is true both in a narrow band at the equator and for the broader tropical pattern.	(*53*, *73–75*, *126*)	D
Wintertime North Atlantic jet	Models fail to capture the observed strengthening of the North Atlantic jet and associated impacts on European precipitation since 1951.	(*88*)	D
JJA Greenland blocking	The recent increase in Greenland blocking events seen in observations is not captured in model simulations.	(*89*, *90*)	D
Exacerbated summer warming in western-central Europe	Western-central Europe has seen exacerbated warming and drying compared to the global mean and a substantial rise in heat extremes that is not well captured by models.	(*91–94*)	D
Arid region near-surface specific humidity	Models suggest that near-surface specific humidity in arid regions should have risen over the historical record. A rise has not been observed.	(*85–87*)	D
Southern Ocean SSTs and sea ice extent	It is rare for model ensemble members to reproduce the observed slight decline in Southern Ocean SSTs and increase in Southern Ocean sea ice extent since 1979, although rapid declines in Southern Ocean sea ice have been	(*53*, *79*)	D
Arctic warming	The observed warming of the Arctic during the satellite era lies within the modeled distribution of trends.	(*20*)	S

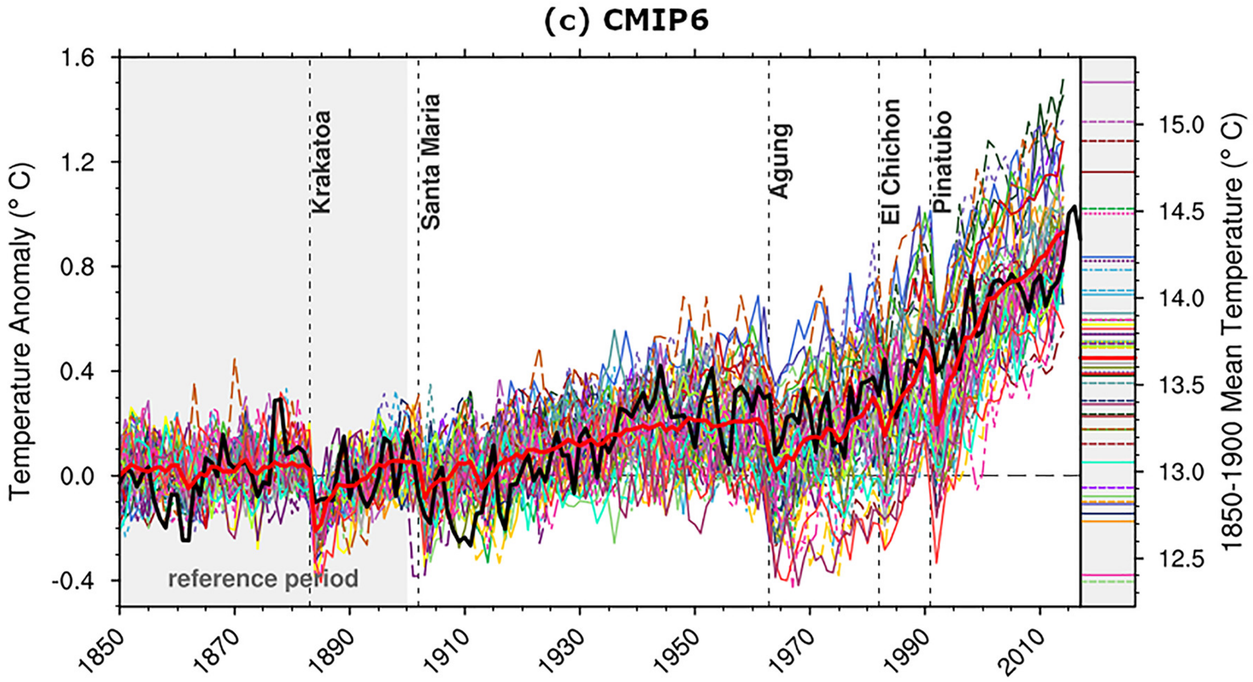
Models trends 10 times more likely to be outside of the 3σ range of observed trends … "but compared to uncertainty in the observations this number does not allow the conclusion that models are systematically inconsistent with the observations" … "we show that observed and simulated trends in sea surface temperature (SST) patterns are globally consistent when accounting for internal variability" †

… this despite the bar for success not being set particularly high.

Shaw and Stevens, *Nature* (2025), Simpson et al., *Sci. Adv.* (2025), † Olonscheck et al., (2020)

Tipping points





"where is everybody" — Fermi.

Lenton et al.., *Science* (2022), Bock et al., *J. Geophys. Res. Atmos.*(2023)

Accumulating anomalies



- Despite the setting of a high-bar, anomalies are accumulating.
  - models aren't consistent with observations
  - models aren't consistent with expectations
- How to rule out the hypothesis that improved alignment with observations is the result of either curve fitting
  - in the forcing, or
  - in the response.

Of course the ensemble mean is better, all are throwing darts at the same target

Image created by Gemini with prompts: "Can you draw me a picture of a dart board with many darts scattered about the target, but with large errors, some not even hitting the board?"
Followed by "Can you replace the sign with CMIP modelers only and put it in a more modern esthetic"

Competing narratives

- The standard model is fine, we just haven't sampled it adequately (large ensembles)
- The standard model is fine, we just haven't found the right closure (AI)
- The standard model is flawed, we need a new approach (km-scale)

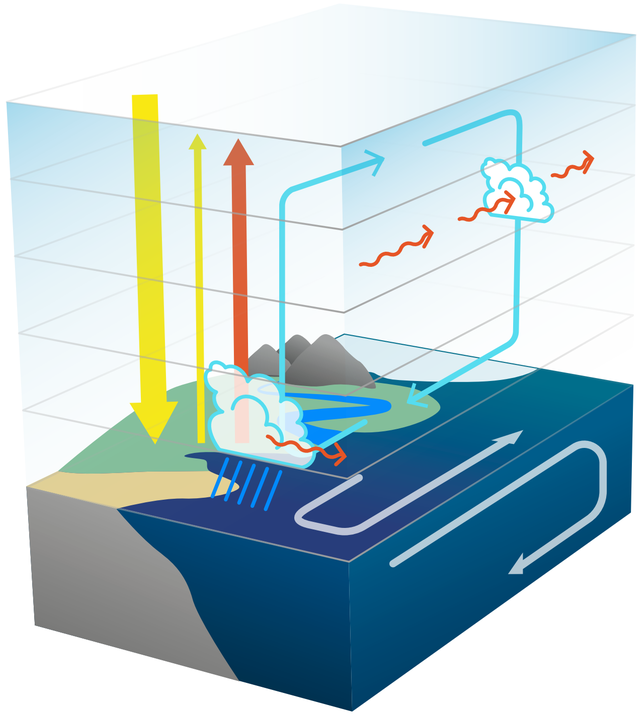
… how could these be wrong?

Shaw and Stevens, Nature (2025)

But first — what is the "standard model"?

Quasi-static equations of motion with Large-scale determinism: It states that small-scale physical processes can be parameterized by quasi-equilibrium statistics that depend on the state of large scales. Small scales are important, but have no agency, and energy and momentum flow in one direction. Thus, LSD assumes a scale separation and is similar to the assumption underlying an effective field theory.

GCM



- Radiant energy exchange
- Instabilities of a zonal current
- Large-scale waves response to heating, orography
- Quasi-static
- Laminar ocean

… statistically fit (tuned) to what we know about the present climate

Shaw and Stevens, Nature (2025)

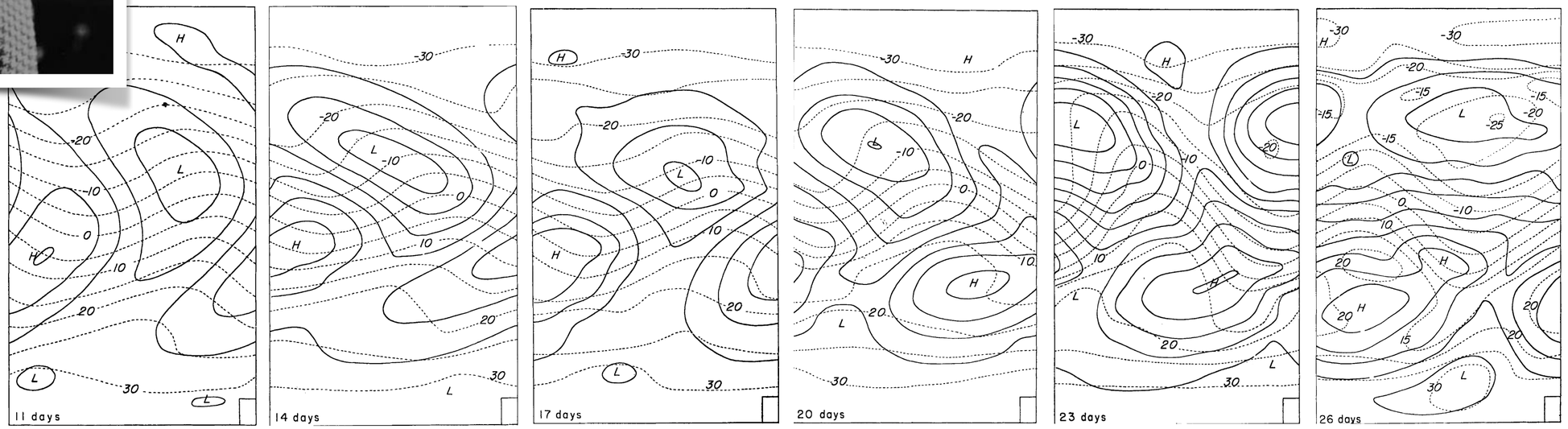


$$\partial u/\partial t + \mathbf{v}\cdot\nabla u - fv = \partial\phi/\partial x + A_v\nabla^2 u + g\partial\tau_x/\partial p$$

$$\partial v/\partial t + \mathbf{v}\cdot\nabla v + fu = \partial\phi/\partial y + A_v\nabla^2 v + g\partial\tau_y/\partial p$$

$$\nabla\cdot\mathbf{v} + \partial\omega/\partial p = 0$$

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)\ln\theta = \frac{1}{c_p T}Q$$



Norman Phillips (1956)

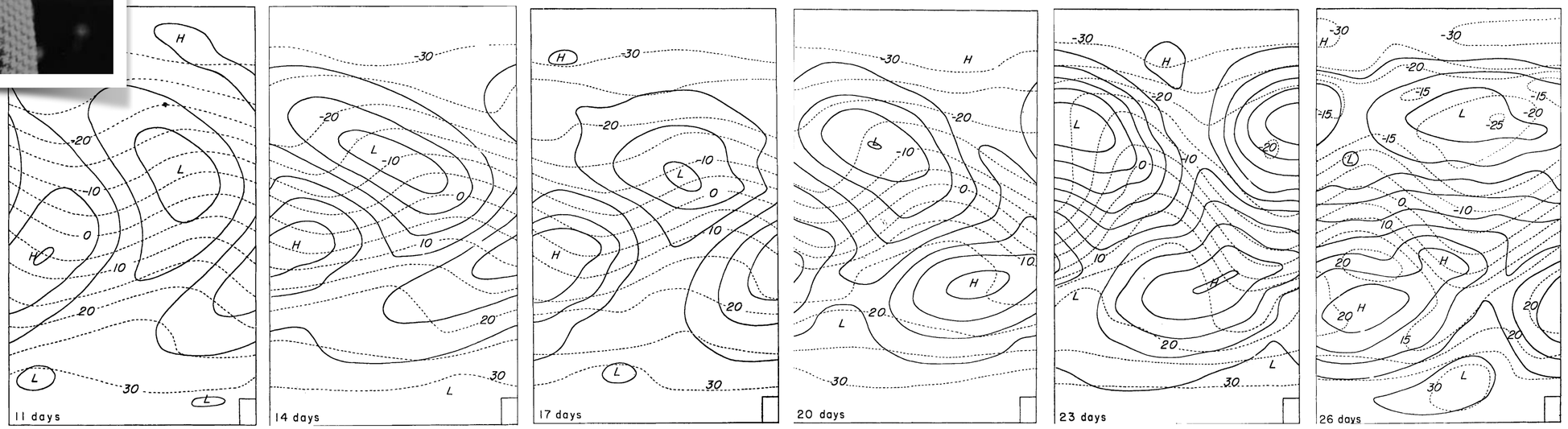


$$\partial u/\partial t + \mathbf{v}\cdot\nabla u - fv = \partial\phi/\partial x + A_v\nabla^2 u + g\partial\tau_x/\partial p$$

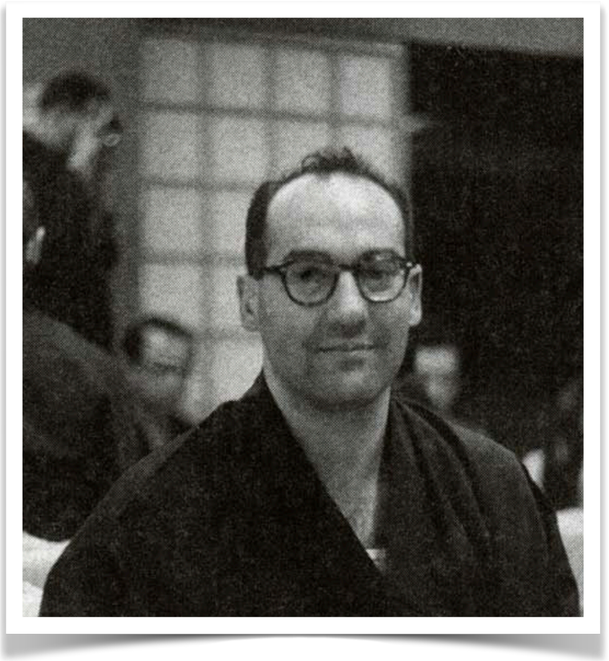
$$\partial v/\partial t + \mathbf{v}\cdot\nabla v + fu = \partial\phi/\partial y + A_v\nabla^2 v + g\partial\tau_y/\partial p$$

$$\nabla\cdot\mathbf{v} + \partial\omega/\partial p = 0$$

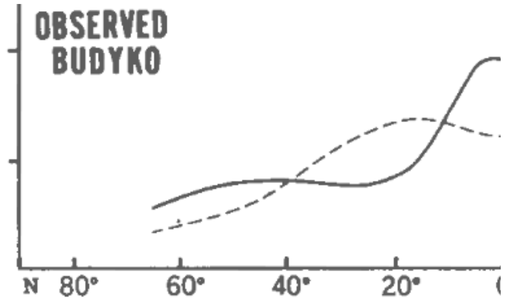
$$\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)\ln\theta = \frac{1}{c_p T}Q$$

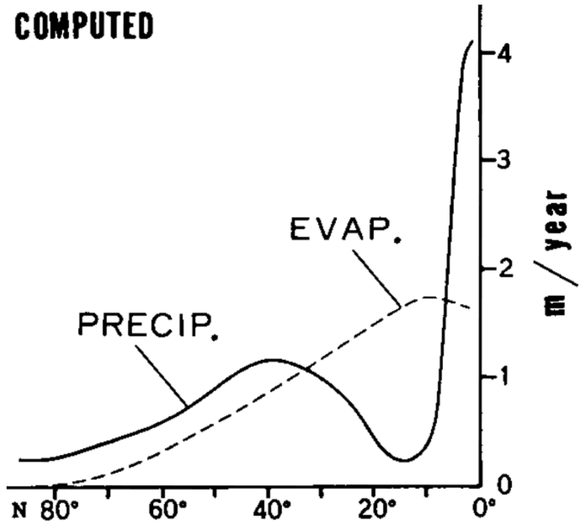


Norman Phillips (1956)



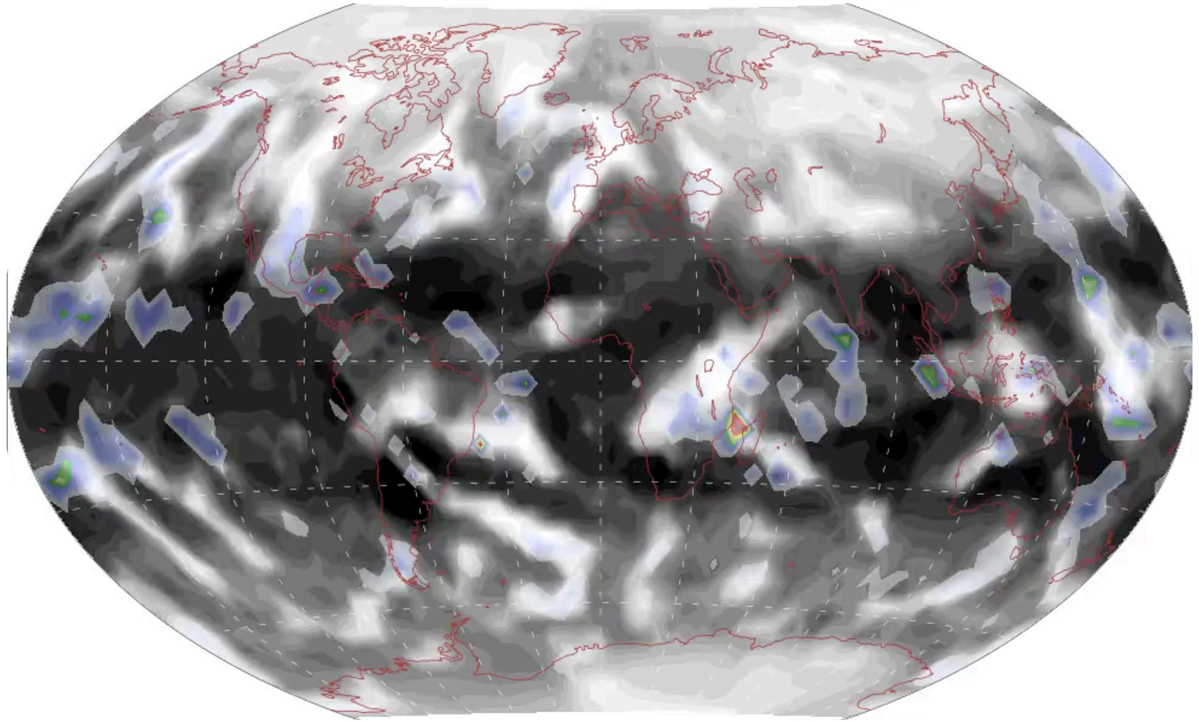
- Added the vertical advection: $\omega \partial \mathbf{X} / \partial p$,
- Spherical geometry
- Moisture equation: $\mathrm{D}q/\mathrm{D}t = \mathscr{C}$
- Rationalized sub-grid terms and boundary conditions

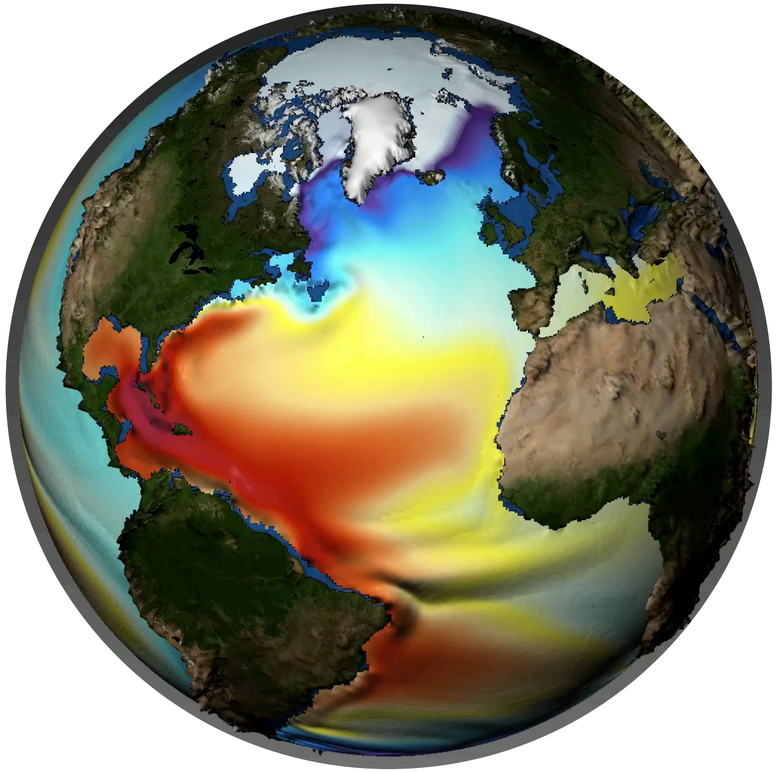




Joseph Smagorinsky (1963, 1965a, 1965b)

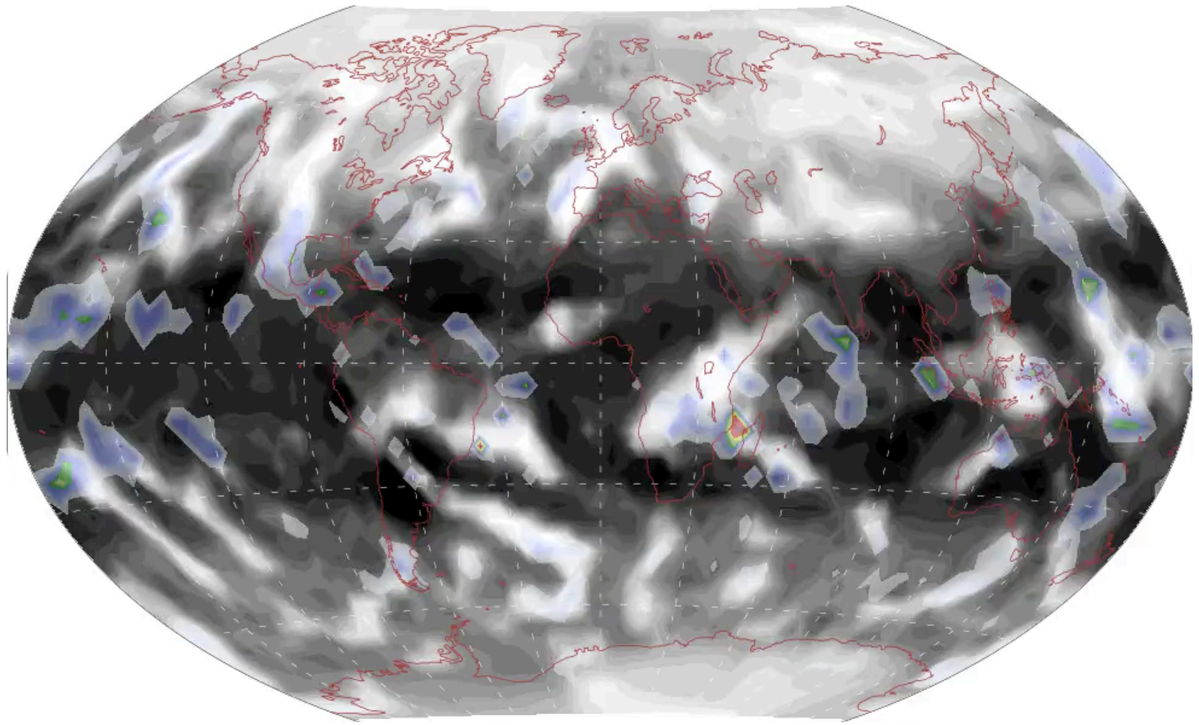
Simulating 2D geophysical fluid dynamics (layerwise)

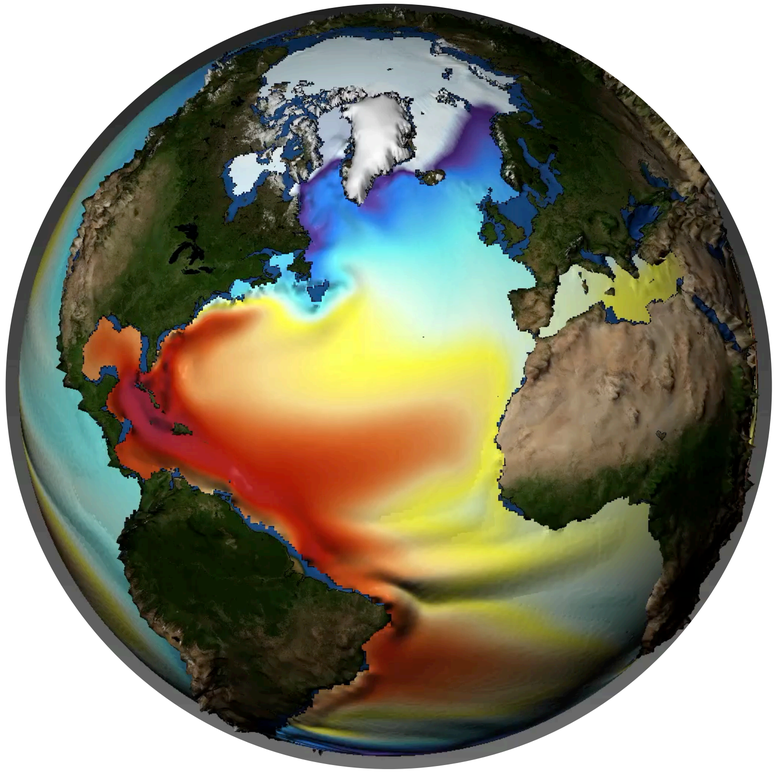




Phillips (1956)

Simulating 2D geophysical fluid dynamics (layerwise)

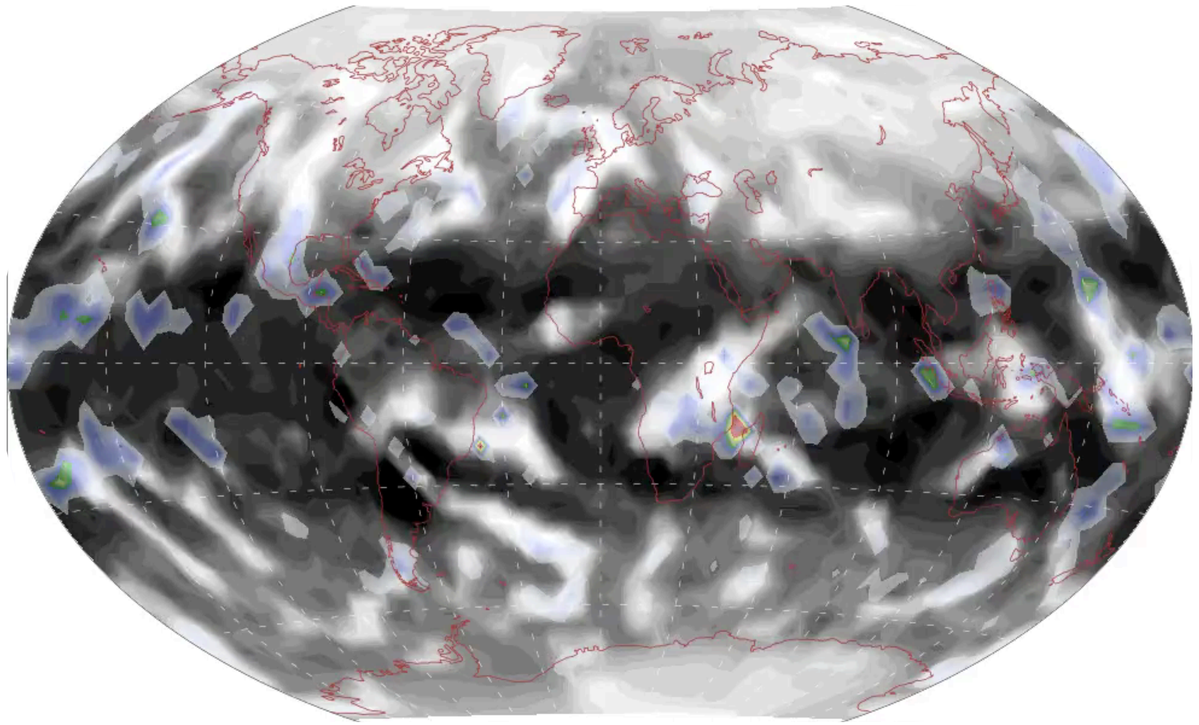


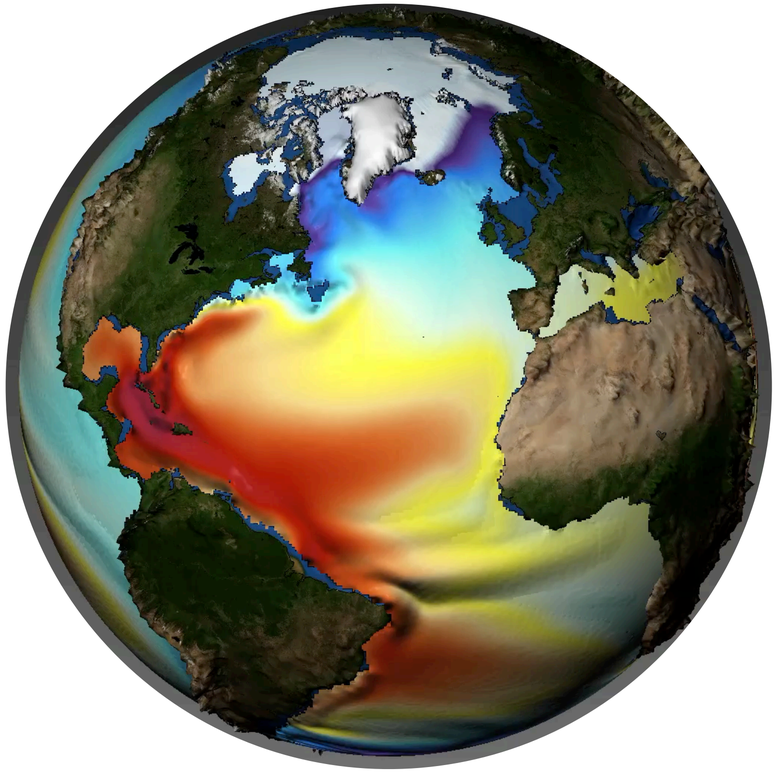


the DNA of a "modern climate model"

Phillips (1956)

Simulating 2D geophysical fluid dynamics (layerwise)





the DNA of a "modern climate model"

Phillips (1956)

Competing narratives





- The standard model is fine, we just haven't sampled it adequately (large ensembles)
- The standard model is fine, we just haven't found the right closure (AI)
- The standard model is flawed, we need a new approach (km-scale)

… and the scientific criticism

Shaw and Stevens, *Nature* (2025), images from Gemini following earlier prompts

Large Ensembles (SMILES … MMEs and PPEs)

One idea is that a better understanding of climate will come from better accounting for variability through the more routine use of large ensembles..

Scientific Criticism

- Increasing the spread about the wrong answer doesn't make it right.
- Why do we believe the models variability?

AI Climate Models

AI can solve the parameterization problem where humans failed.

Scientific Criticism

- The problem it tries to solve is ill posed
- It is not causal … how would we know if it is right?

km and finer scale simulation

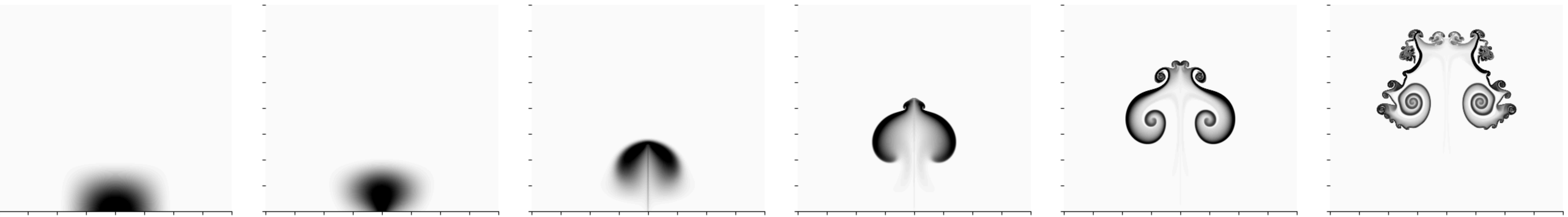
Using more physical models will better approximate the Earth system

Scientific Criticism

- Irreducible imprecision
- throughput is limited to small ensembles and centennial or finer time-scales

… how could we test this?

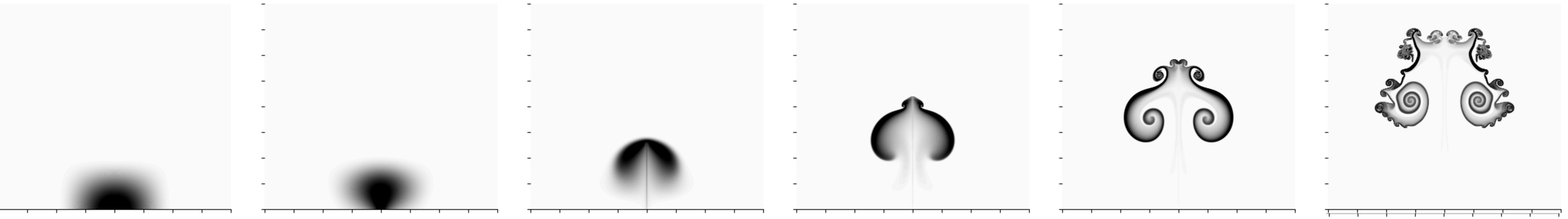
What is meant by “more physical” models

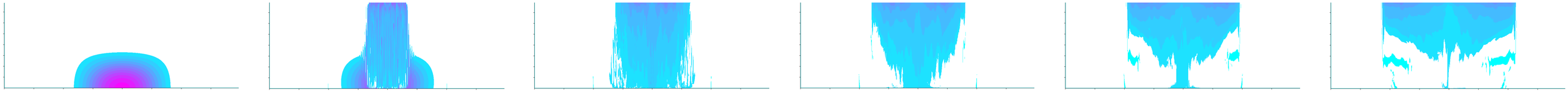


Simulation of aline thermal (a slab symmetric bubble of warm air rising) with DNS

Joint work with Juan Pedro Mellado (DNS) 2026

As represented by the standard model (blue)

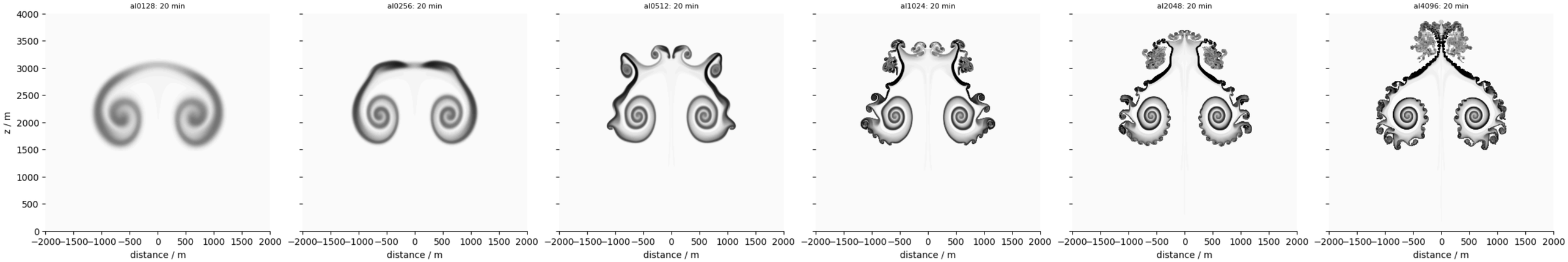




… the equations used by the standard model (quasi-static) don't generalize, they can't even simulate a warm bubble

Joint work with Juan Pedro Mellado (upper), Leonidas Linardakis and Peter Korn (lower) 2026

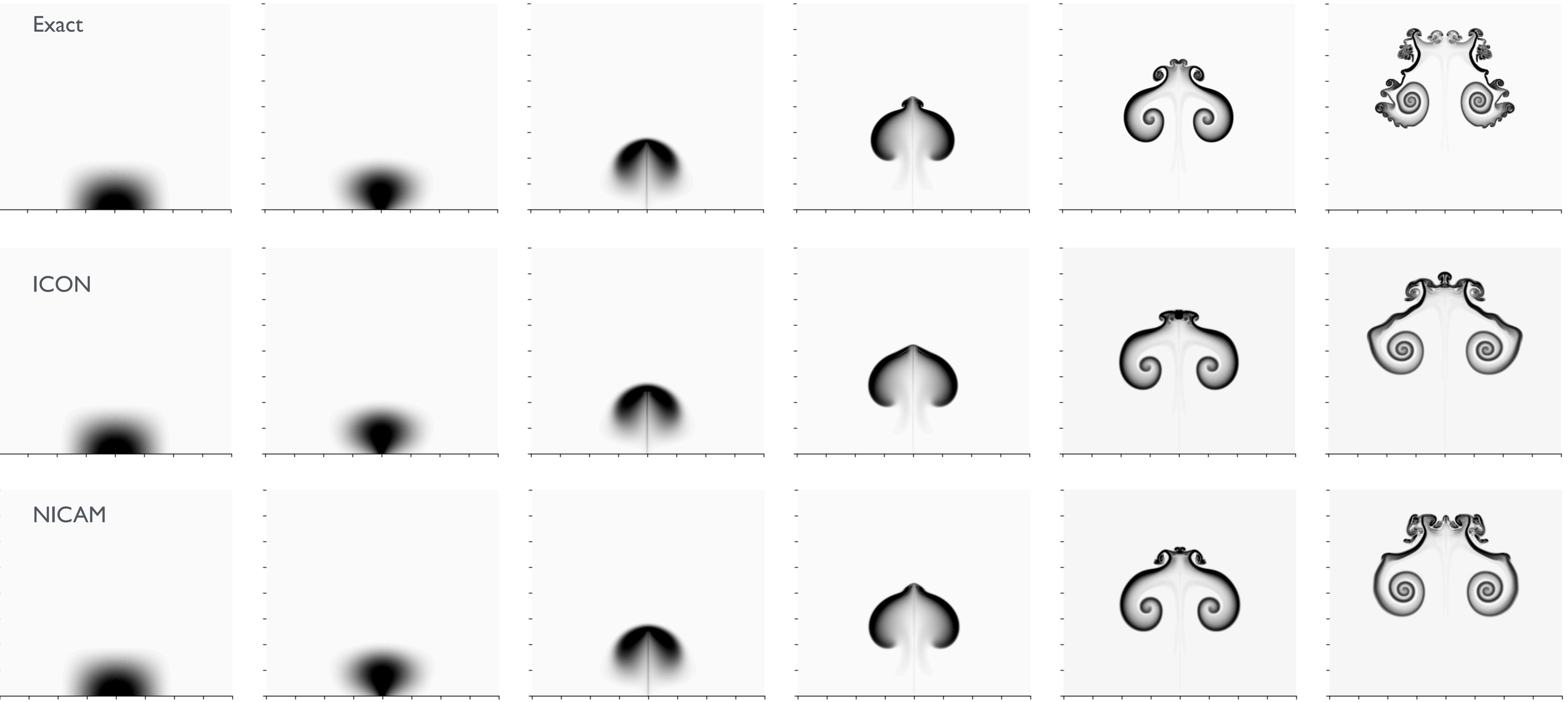
With simulation, reducing resolution distorts the bubble



… but large scale features remain similar

Joint work with Juan Pedro Mellado (DNS) 2026

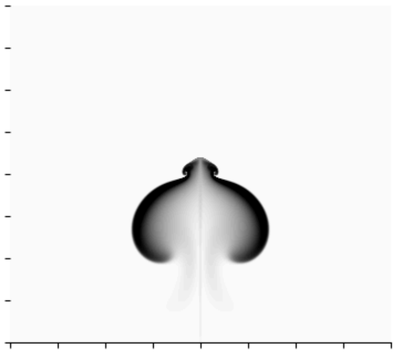
km-scale models can simulate a warm thermal



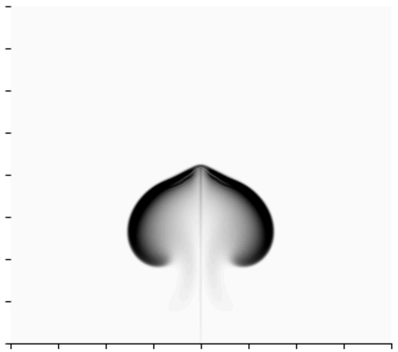
Credit: Tomoki Ohno (NICAM), Juan Pedro Mellado (DNS)

… the standard model can't (simulate a warm thermal)

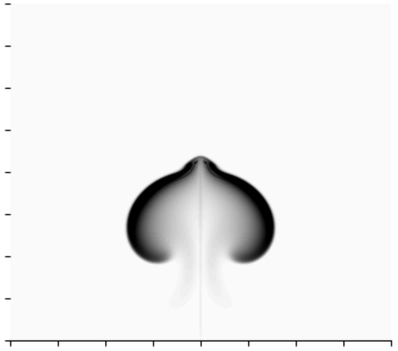
Exact



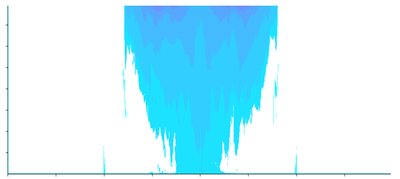
ICON



NICAM



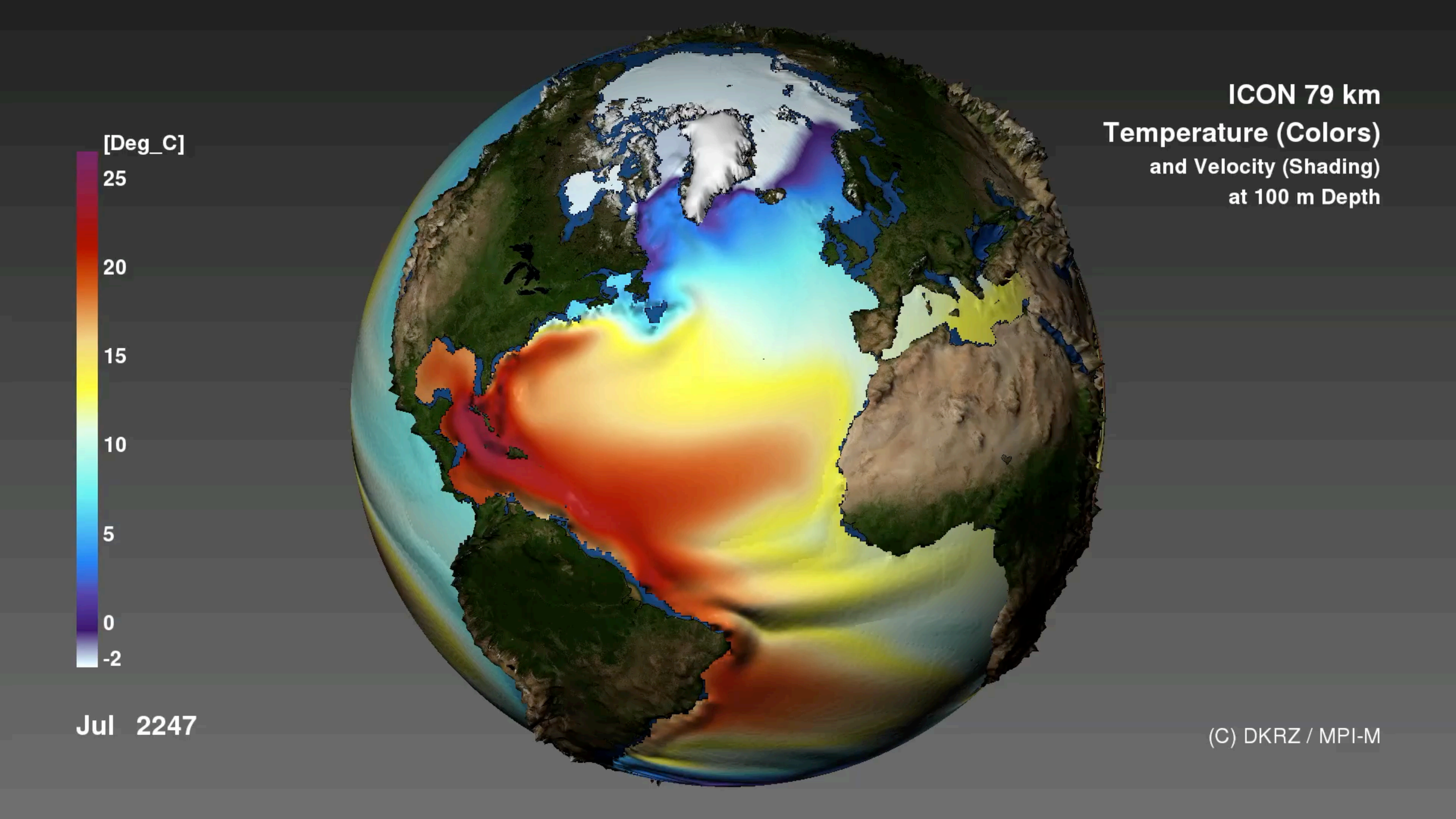
Standard model

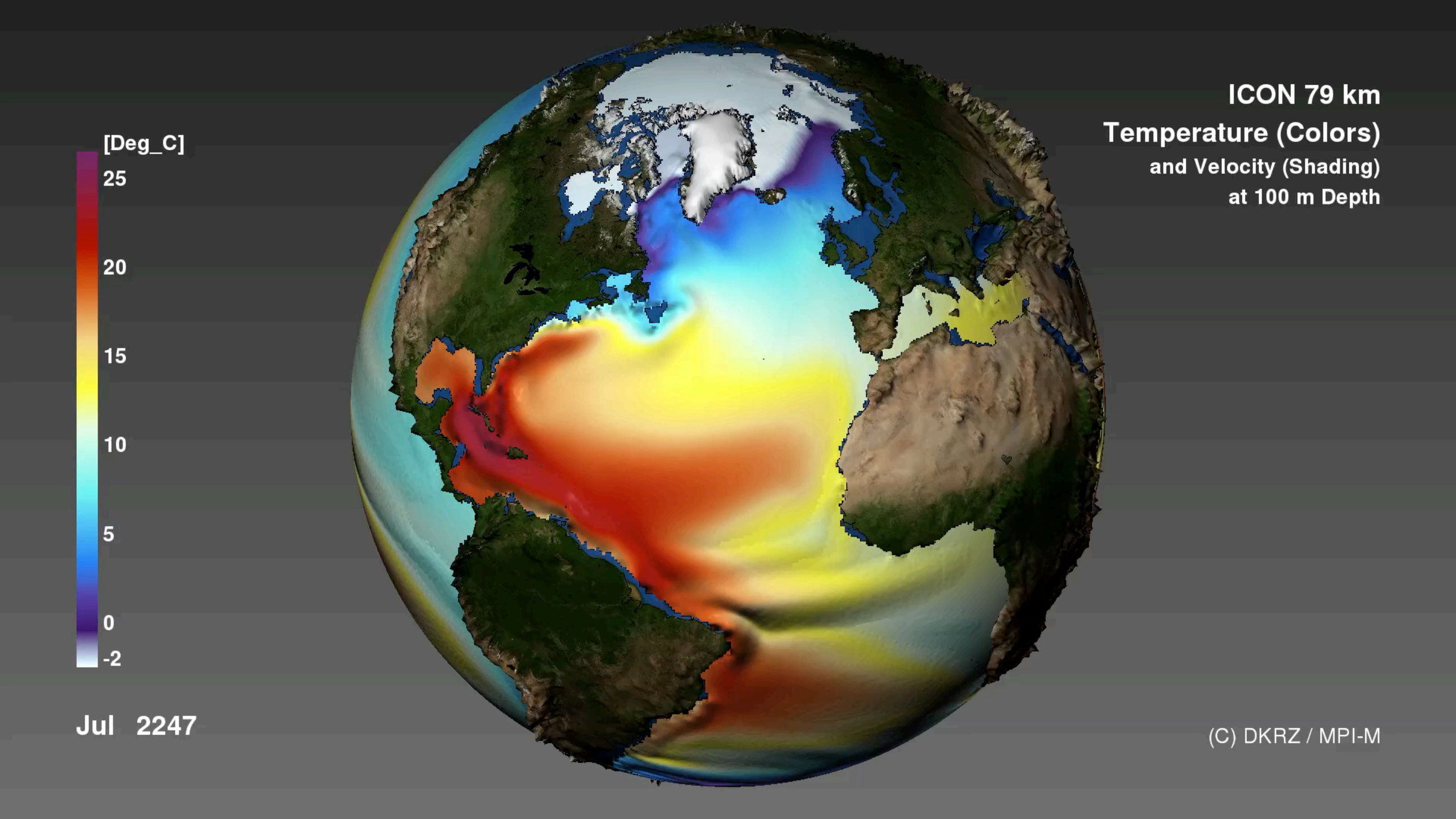


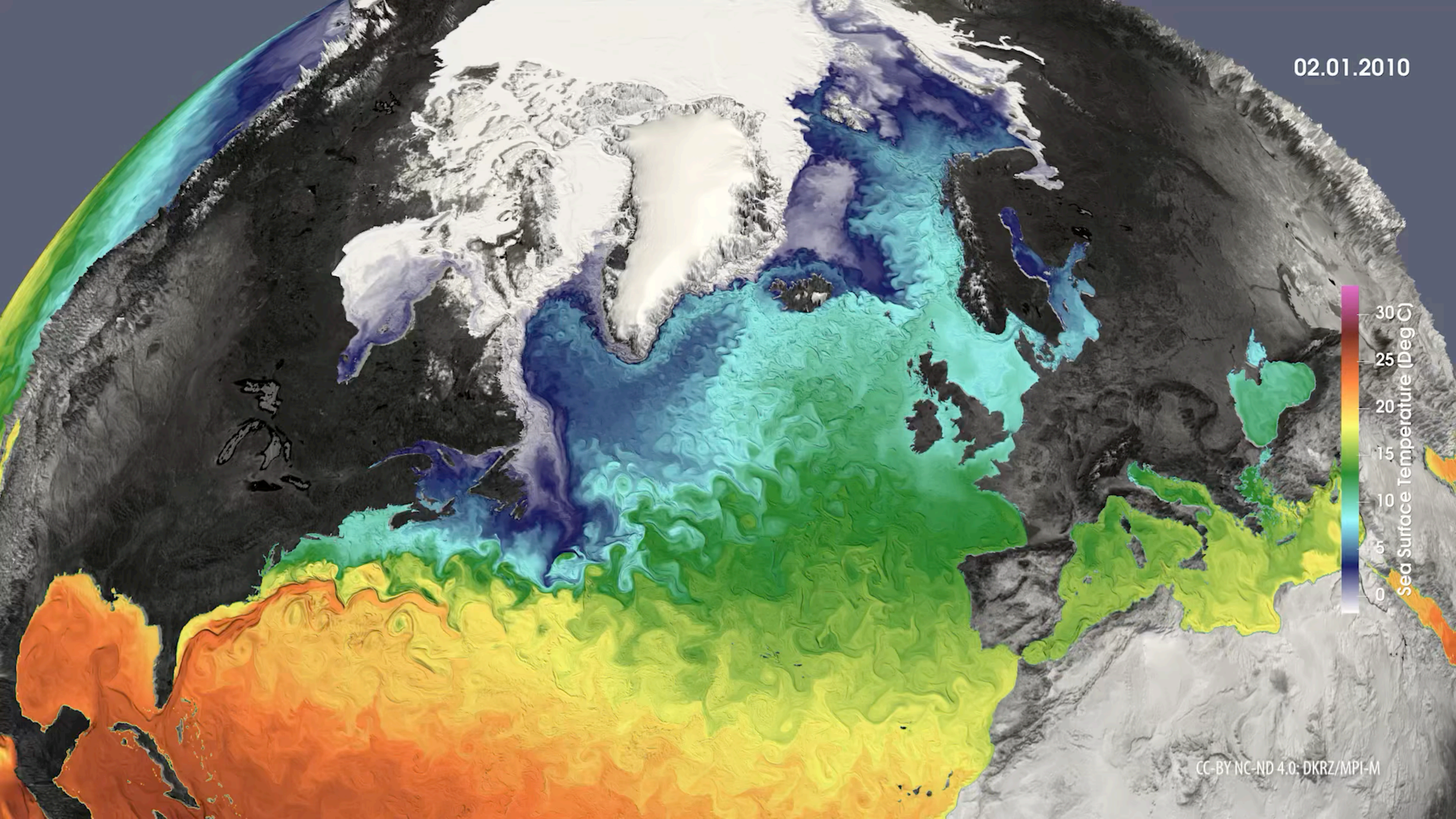
… worse yet the standard model:

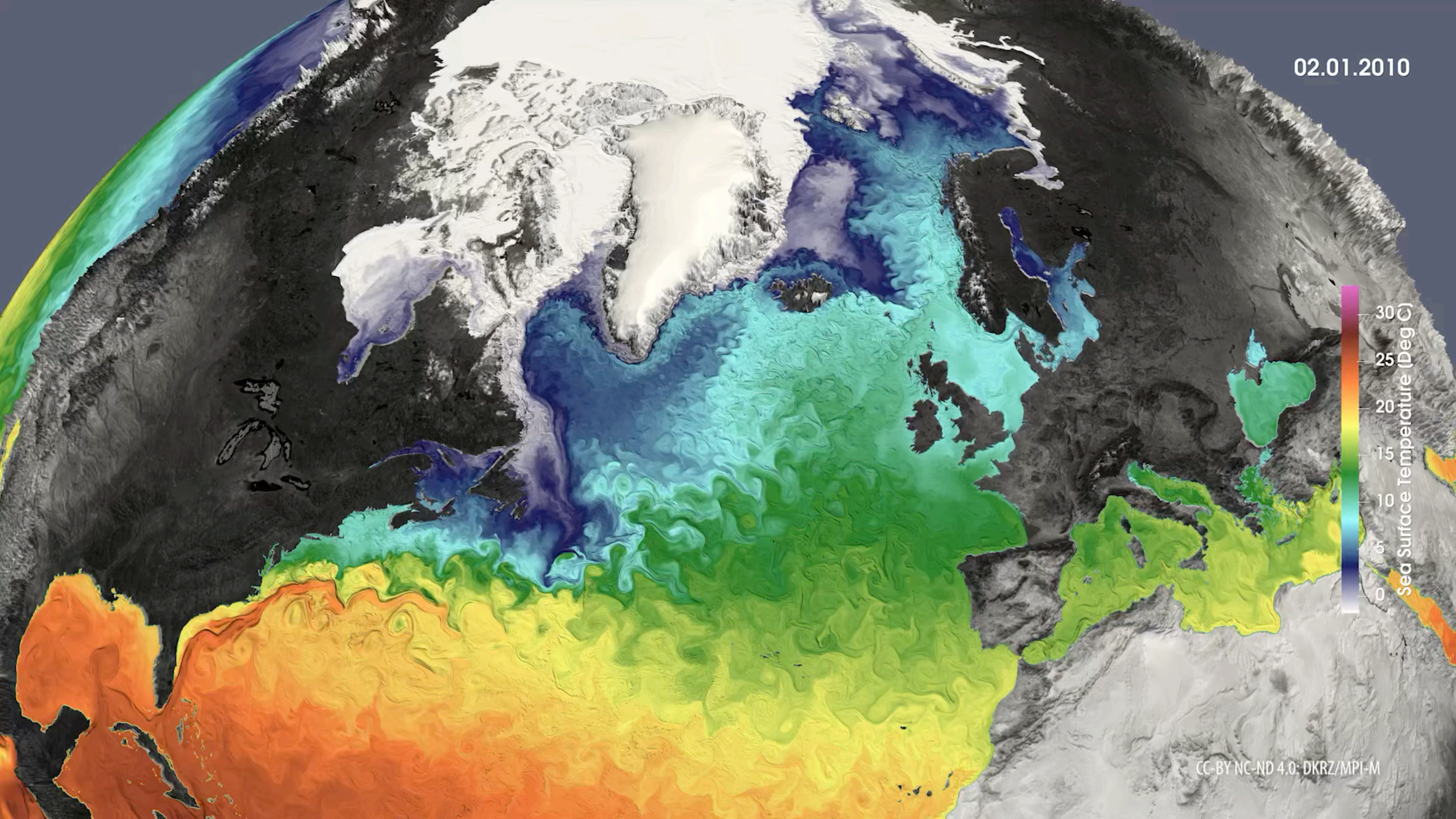
- *is designed to prevent the bubble, and damp extremes;*
- *becomes even more unrealistic at higher resolution.*

Credit: Tomoki Ohno (NICAM), Juan Pedro Mellado (DNS)


[Deg_C]
25
20
15
10
5
0
-2
Jul 2247
ICON 79 km
Temperature (Colors)
and Velocity (Shading)
at 100 m Depth
(C) DKRZ / MPI-M

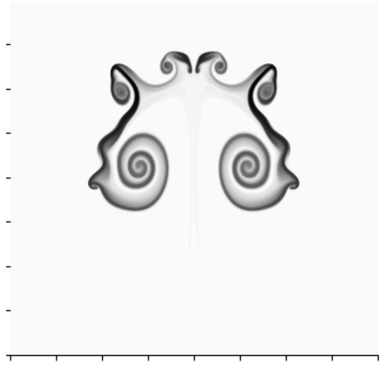

[Deg_C]
25
20
15
10
5
0
-2
Jul 2247
ICON 79 km
Temperature (Colors)
and Velocity (Shading)
at 100 m Depth
(C) DKRZ / MPI-M


02.01.2010
30
25
20
15
10
5
0
Sea Surface Temperature (Deg C)
CC-BY NC-ND 4.0: DKRZ/MPI-M


02.01.2010
30
25
20
15
10
5
0
Sea Surface Temperature (Deg C)
CC-BY NC-ND 4.0: DKRZ/MPI-M

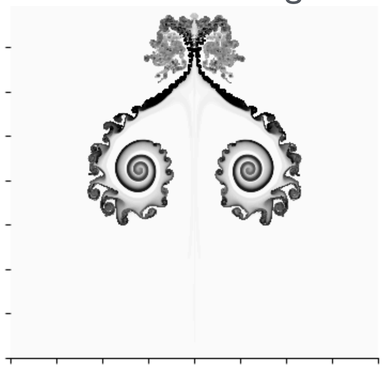
Irreducible imprecision

Exact

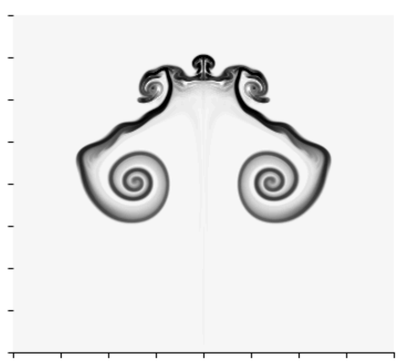


low Re

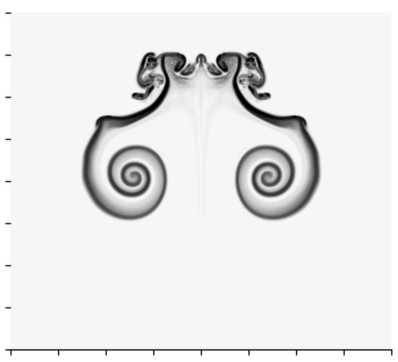
high Re



ICON



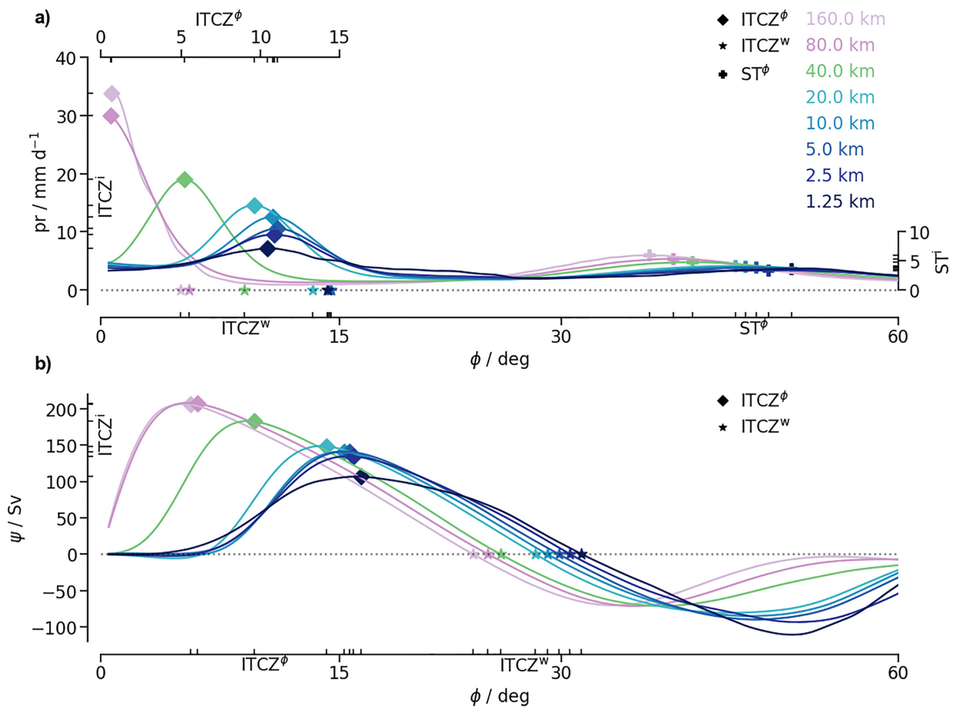
NICAM

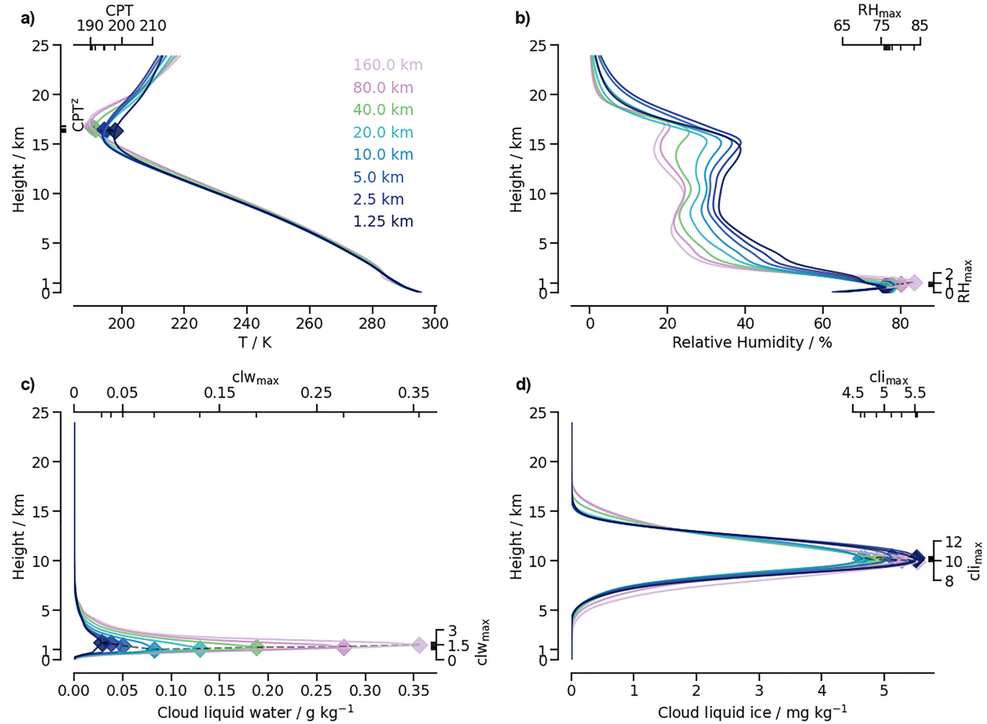


… do the details (difference between ICON and NICAM), equivalently the emergence of finer scale features at high resolutions (disproportionately) matter?

Joint work with Tomoki Ohno (NICAM), Juan Pedro Mellado (DNS)

Evidence the small details don't matter for large scales — rough limits

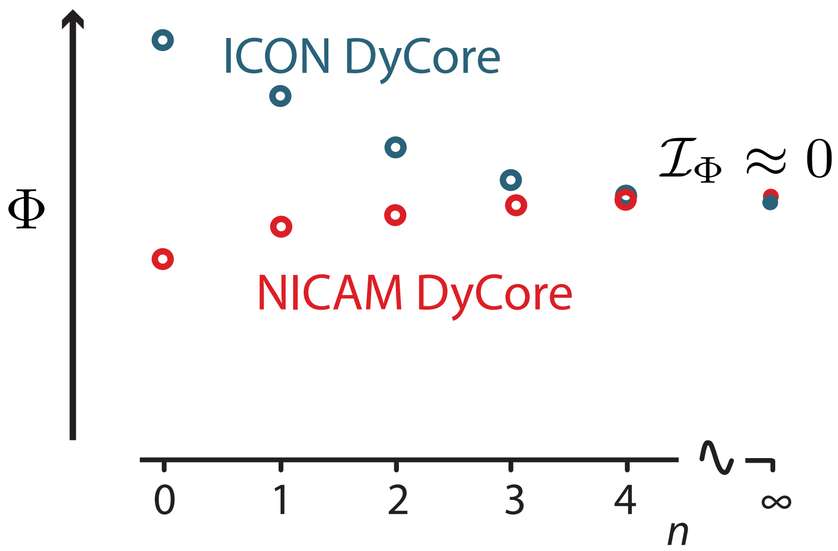


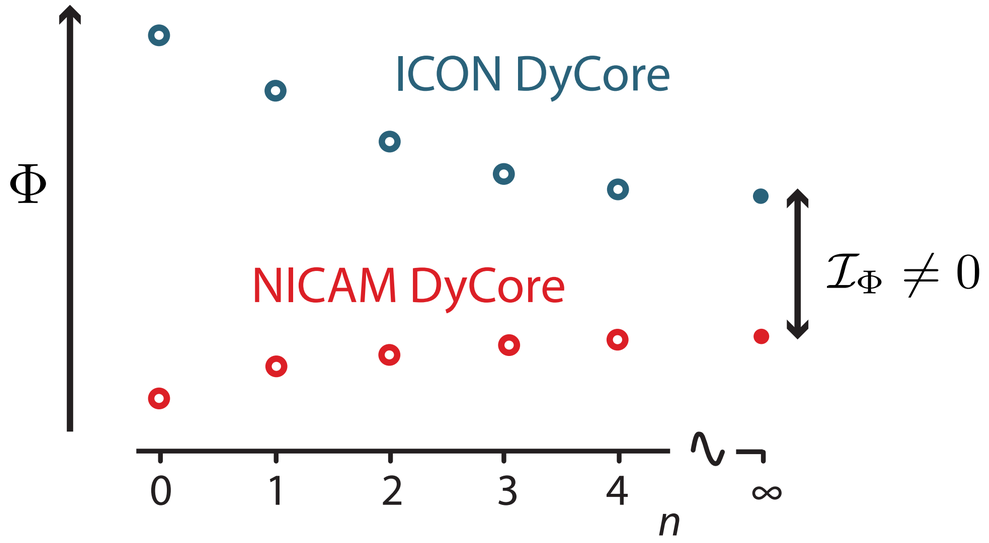


… the convergence to a particular value with increasing resolution is what we call a "rough limit"

Peinado Bravo et al., *J. Adv. Model. Earth. Syst.* (2026)

But do different models have different rough limits?





$\mathcal{I}_\Phi$ is measure of irreducible imprecision.

Sources of uncertainty

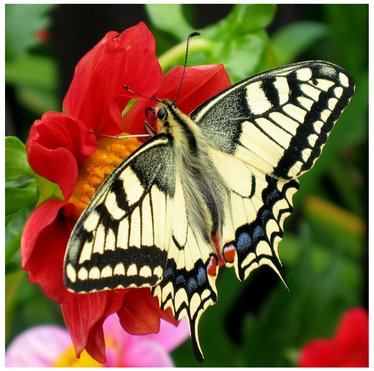
- $\mathcal{I}_\Phi$ measures irreducible imprecision
- $\mathcal{U}^s_\Phi$ measures sample noise — the topic of this summer school
- $\mathcal{U}^e_\Phi$ measures epistemic uncertainty

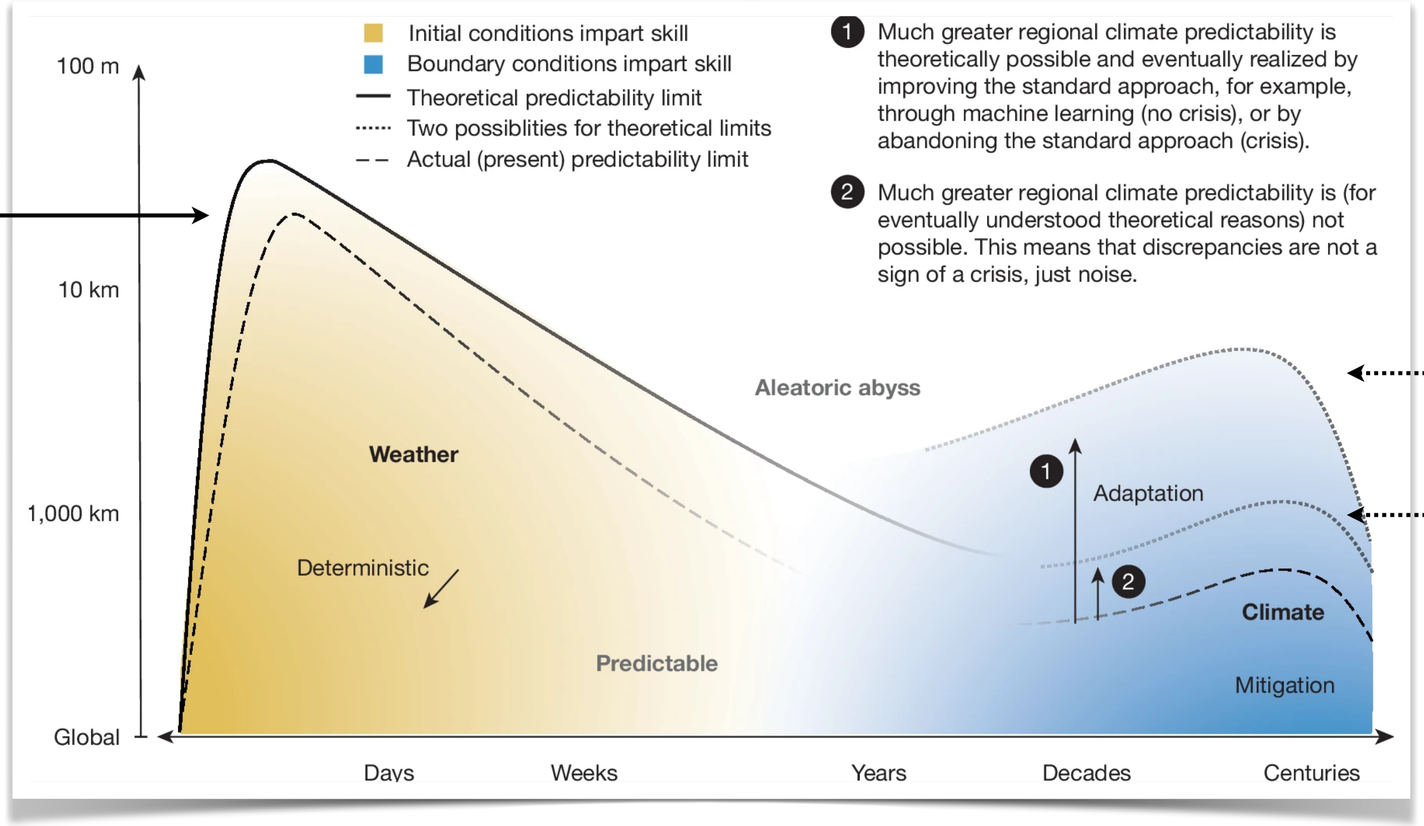
Modelling efforts to reduce uncertainty focus on $\mathcal{U}^e_\Phi$, they will only be successful if $\mathcal{U}^e_\Phi \gg \max(\mathcal{U}^s_\Phi, \mathcal{I}_\Phi)$

Tremendous attention is given to estimating $\mathcal{U}_\Phi$ almost none is given to $\mathcal{I}_\Phi$

Credit: Tomoki Ohno (NICAM), Juan Pedro Mellado (DNS)

But $\mathcal{I}_\Phi$ shapes what is knowable (Butterflies and Hawkmoths)





?



$\mathcal{I}_\Phi$ limits the projection skill of the best models, where it is large, better models won't make better predictions.

Edward Lorenz, *Tellus* (1982) Erica Lucy Thompson, PhD Thesis (2013)

Summary

What I hope you learned:

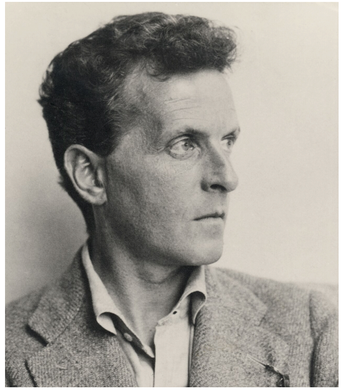
- Even with a very forgiving standard of agreement, anomalies are accumulating
- Different approaches are vying for the scientific narrative (to be the dominant theory)
- This suggests there is a crisis (with the standard mode)… The vying approaches to reconcile it with experience are to claim that
  - The standard model is fine, just not properly sampled (Ensembles)
  - The standard model could be fine, we just haven't settled on the right closure (AI)
  - The flaws of the standard model can be avoided by using more physical models
- New ideas: Rough limits and irreducible imprecision (butterflies and hawkmoth) and why better models might not make better predictions.

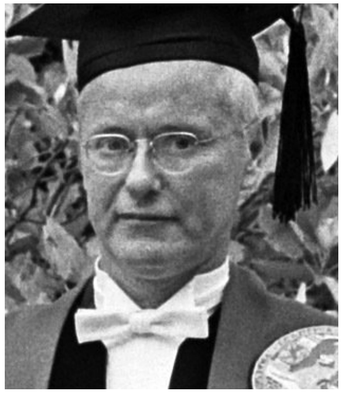




Some implications of the above:

- If the standard model agrees better with observations than a more physical model, we should not trust the standard model.
- Ensembles might be better used for understanding what controls variability, but here there is a notable absence of theoretical work.
- Climate science has entered a new (more scientific) era. Instead of speculating about the future we are increasingly having to explain the past and present.
- Approaching these questions in an uncritical, self-satisfied, manner is neither scientific†, nor likely to advance the frontiers of knowledge





† If you are interested in the last implication you will have to ask me about the two gentlemen shown in the lower pair of panels.