Climate modelling and AI: opportunities, language, and ways forward.

Climate is what you expect, weather is what you get.

The IPCC identifies an important distinction between prediction and projection.

What does those distinctions mean for the use of AI/ML in climate?

Bryan Lawrence

NCAS and University of Reading

With input from Pier Luigi Vidale, Len Shastry, Reinhard Schiemann,
and more to some of the content ! Usual caveats, mistakes are mine etc

Outline

1. Some language for talking about AI in “modelling science”.
2. The abstract art of modelling and “fitness for purpose”.
3. What is climate modelling and how does it differ from weather modelling?
4. Opportunities for AI in climate science.

Language:

So you are using AI for climate modelling/science. What do you really mean?

- AI-Before:
  - Using AI to precondition and/or setup inputs
- AI-Inside:
  - Using AI to replace some parts of a “traditional” model.
- AI-Outside (or AI-On-Top)
  - Using AI in the workflow following traditional model simulation.
- AI-Instead:
  - It’s data-driven all the way down baby!

There are huge opportunities for Before, Inside, and Outside, but Instead is quite problematic, for some definition of problematic.

Stevens, B., Adami, S., Ali, T., Anzt, H., Aslan, Z., Attinger, S., Bäck, J., Baehr, J., Bauer, P., Bernier, N., Bishop, B., Bockelmann, H., Bony, S., Brasseur, G., Bresch, D. N., Breyer, S., Brunet, G., Buttigieg, P. L., Cao, J., … Ziemen, F. (2024). Earth Virtualization Engines (EVE). *Earth System Science Data*, *16*(4), 2113–2122. https://doi.org/10.5194/essd-16-2113-2024

Language:

So you are using AI for climate modelling/science. What do you really mean?

- AI-Before:
  - Using AI to precondition and/or setup inputs
- AI-Inside:
  - Using AI to replace some parts of a “traditional” model.
- AI-Outside (or AI-On-Top)
  - Using AI in the workflow following traditional model simu
- AI-Instead:
  - It’s data-driven all the way down baby!

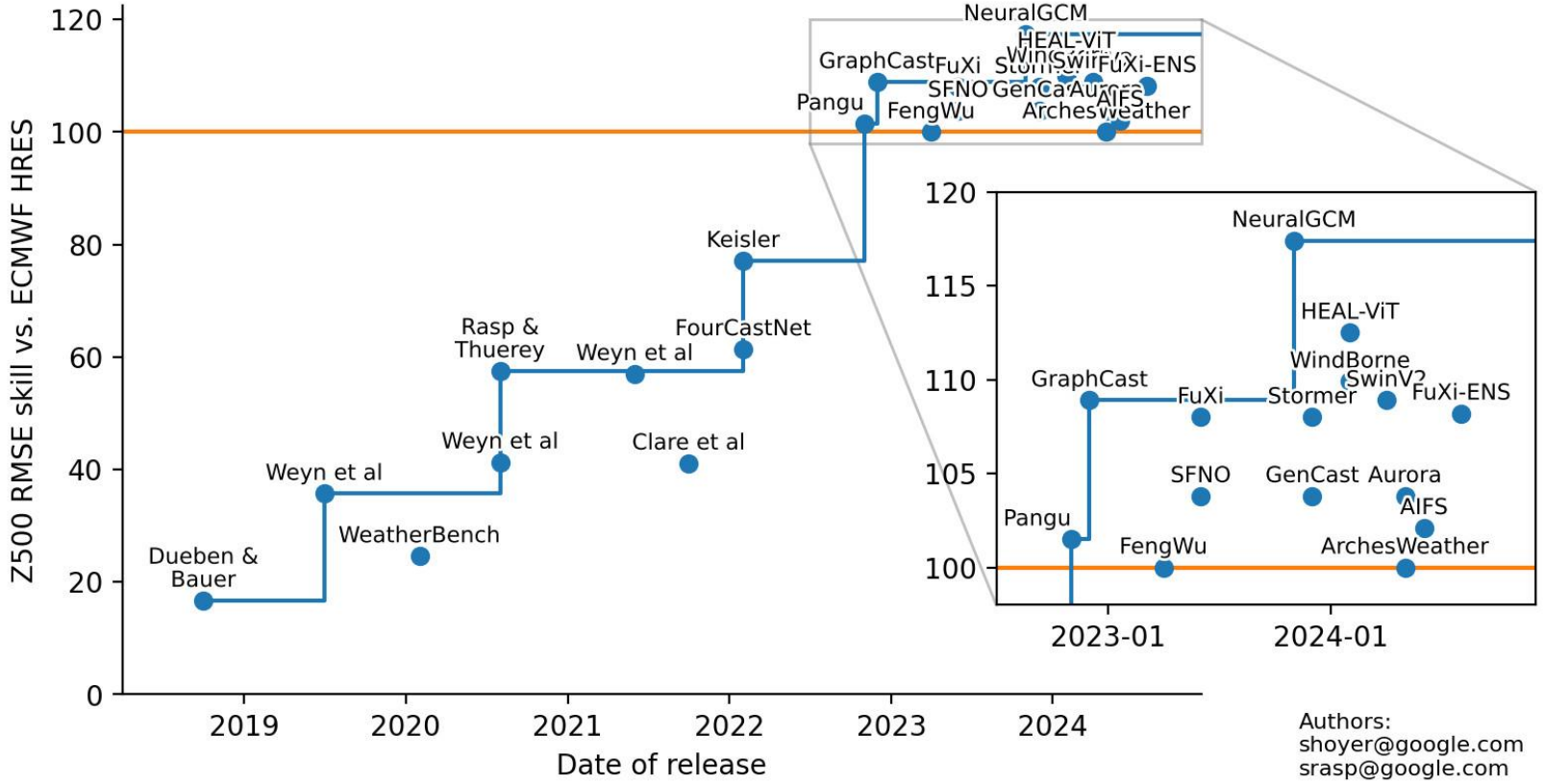


For a different audience, I’d give examples, but the rest of today’s agenda is all about Before, Inside and Outside.

Today I’ll concentrate on Instead.

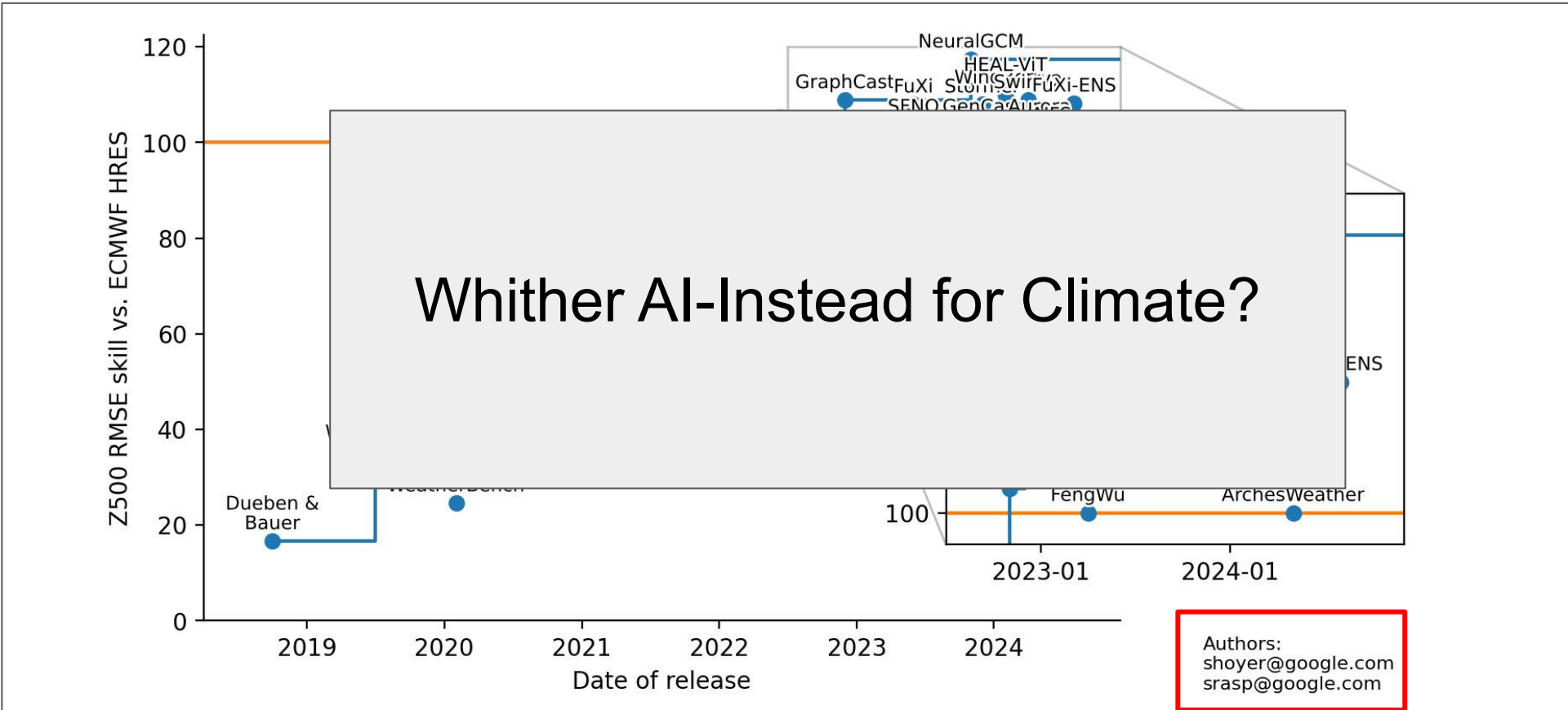
There are huge opportunities for Before, Inside, and O
problematic, for some definition of problematic.

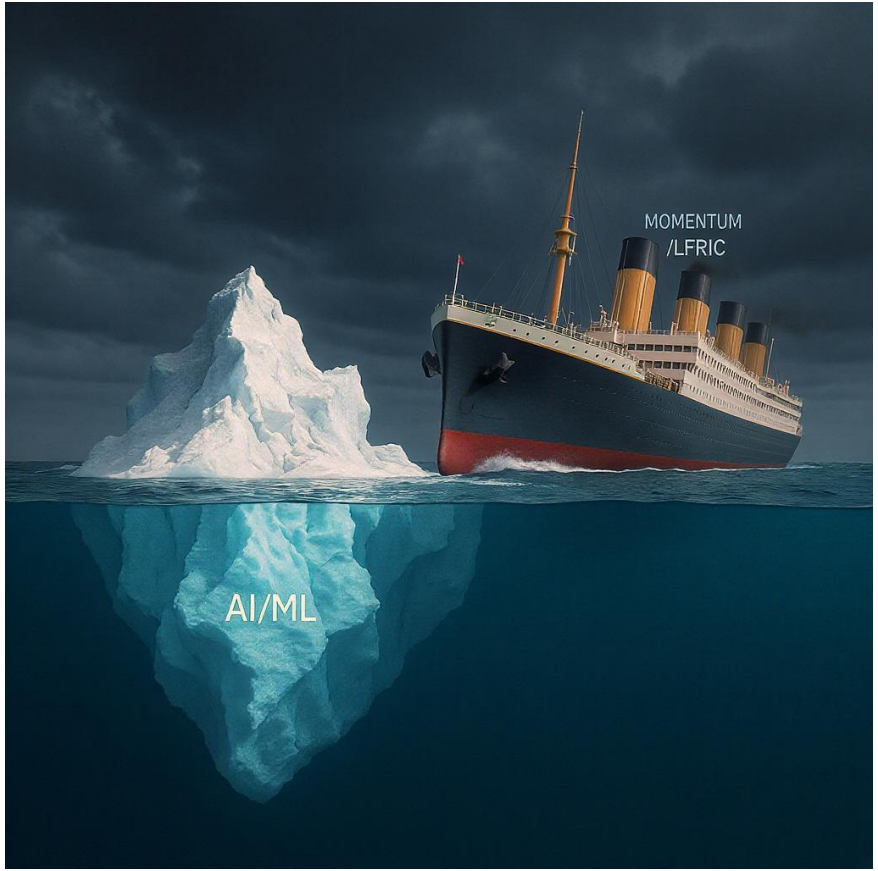
S



Authors:
shoyer@google.com
srasp@google.com

State of the Art in AI Weather Prediction (Nov, 2024)

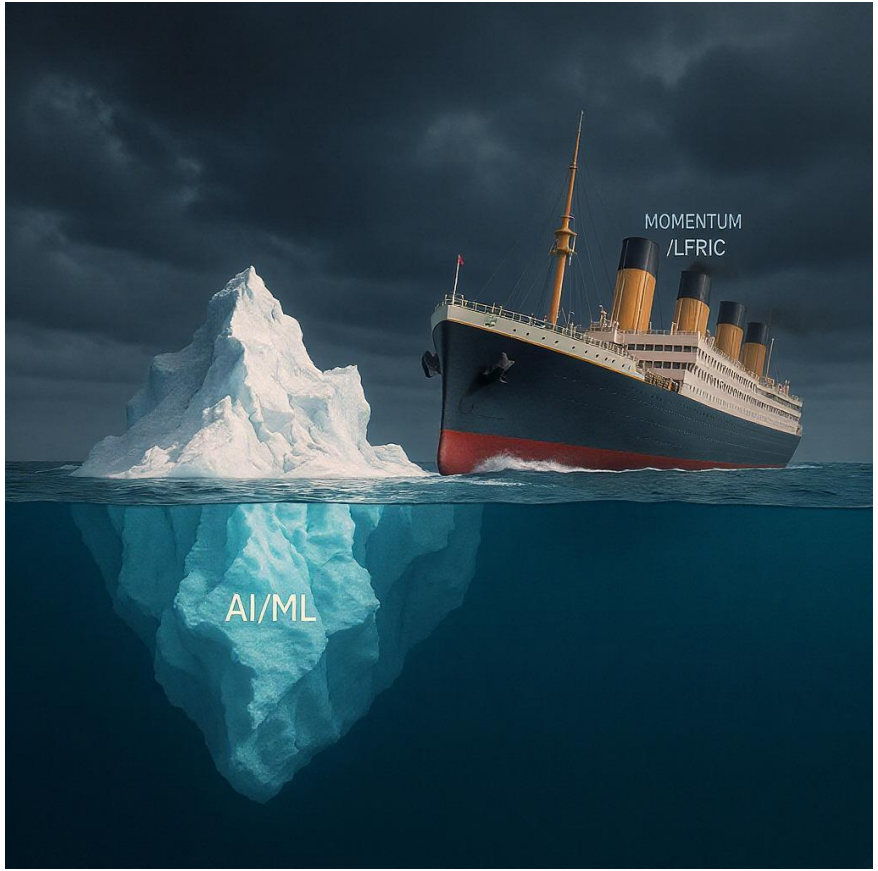




National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

A couple of different takes on the existential risks for ESM modelling right now!

University of Reading



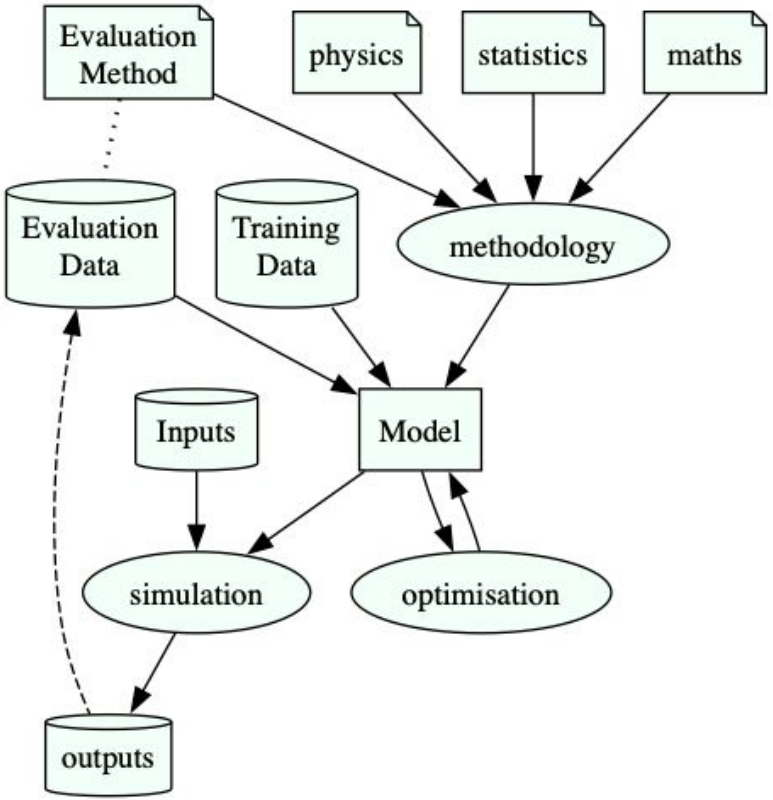


National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

A couple of different takes on the existential risks for ESM modelling (right now! AI, the risk, or the saviour?

University of Reading

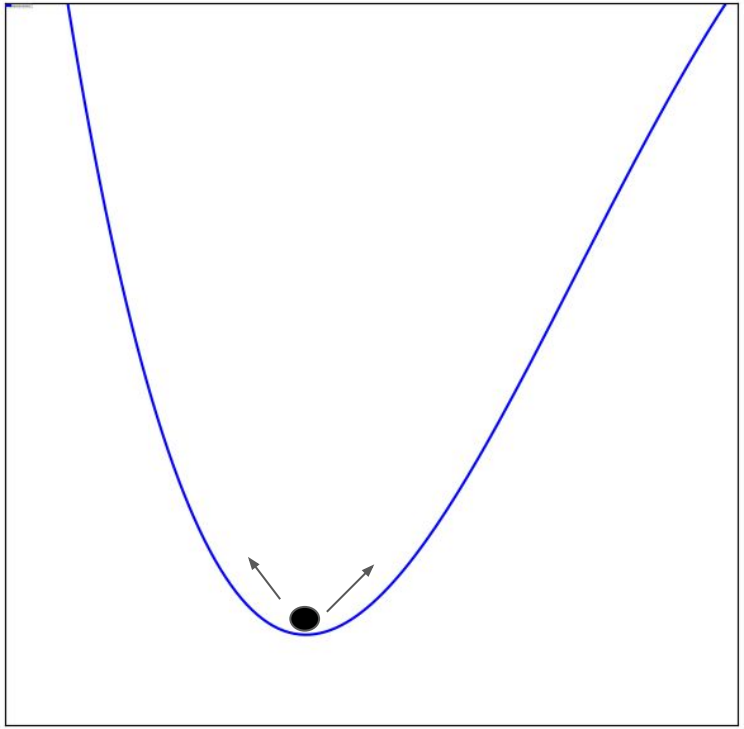
All modelling is “Data Driven Modelling”



The interesting questions are:

- What is the balance of physics and statistics in the methodology, and
- What is the evaluation *method*, used during the building of the model?
- What is the evaluation *data*, and what is the relationship between the evaluation in building the model and evaluating it’s outputs?
  - (The sub-question here is what is the difference between evaluation and training, and how do they interact?)

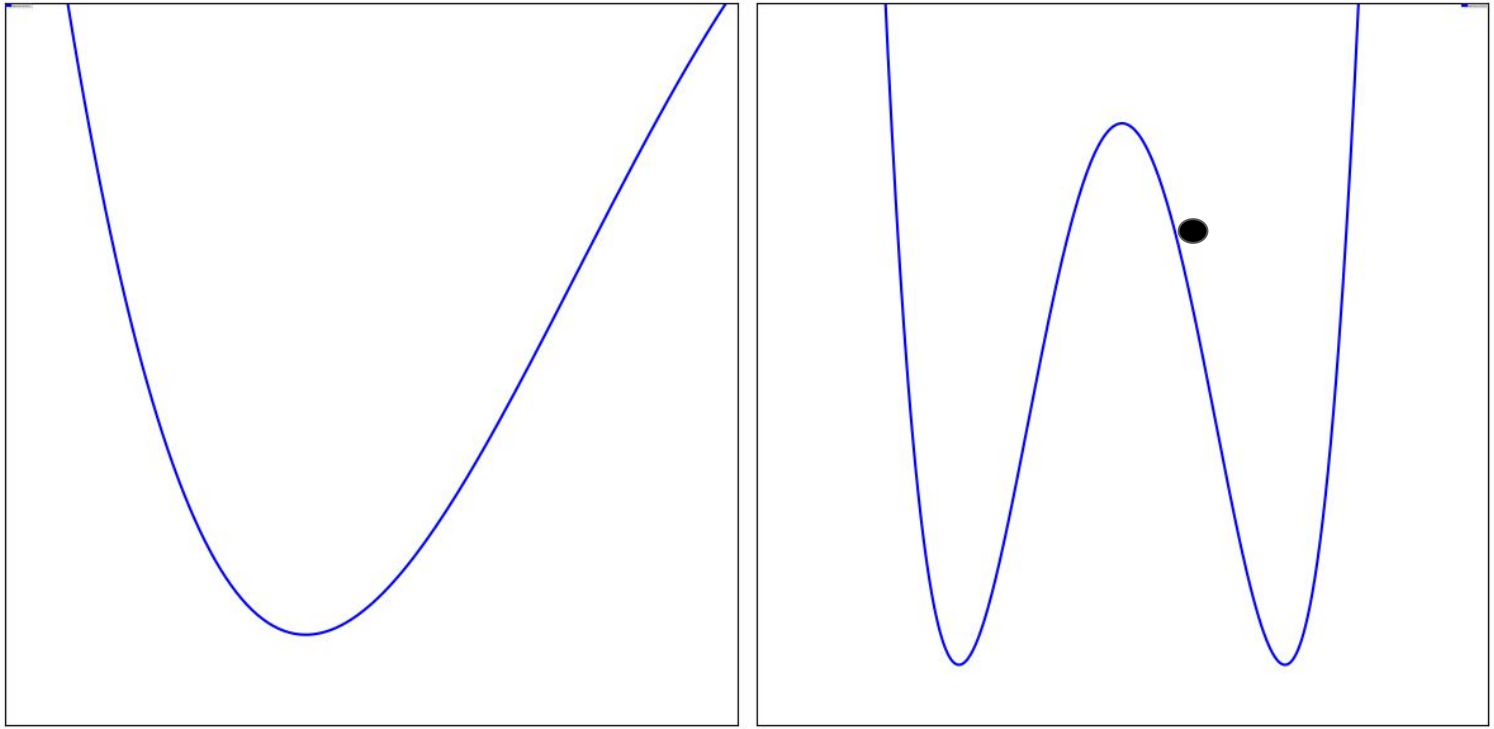
A simple system



Consider the climate as a simple system where the state is represented by a ball which rattles round the available set of states depicted by the blue curve.

This is the world which we use to inform our physical model development and from where we gather observations to train our AI based models.

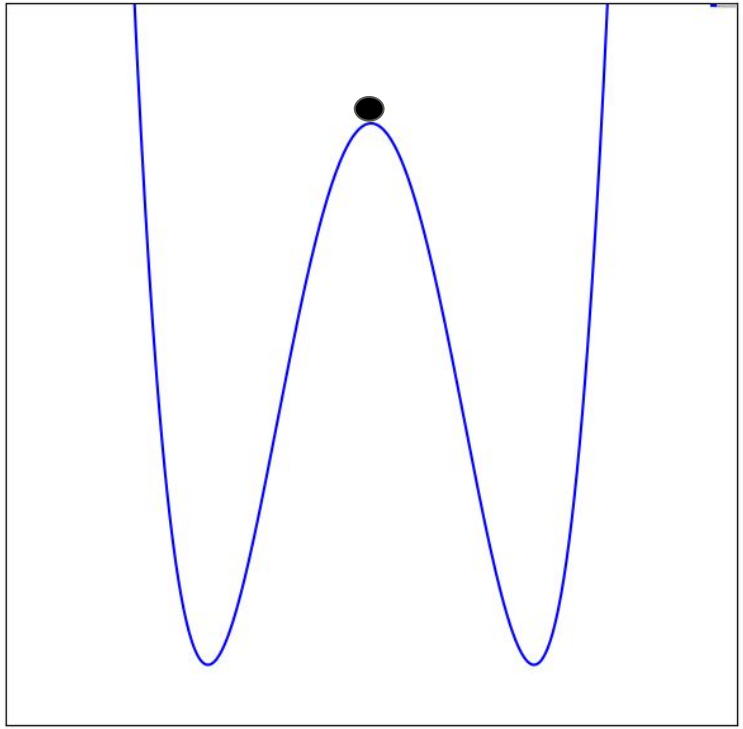
But we might be sampling only one part of a complex system



But we might be sampling only one part of a complex system

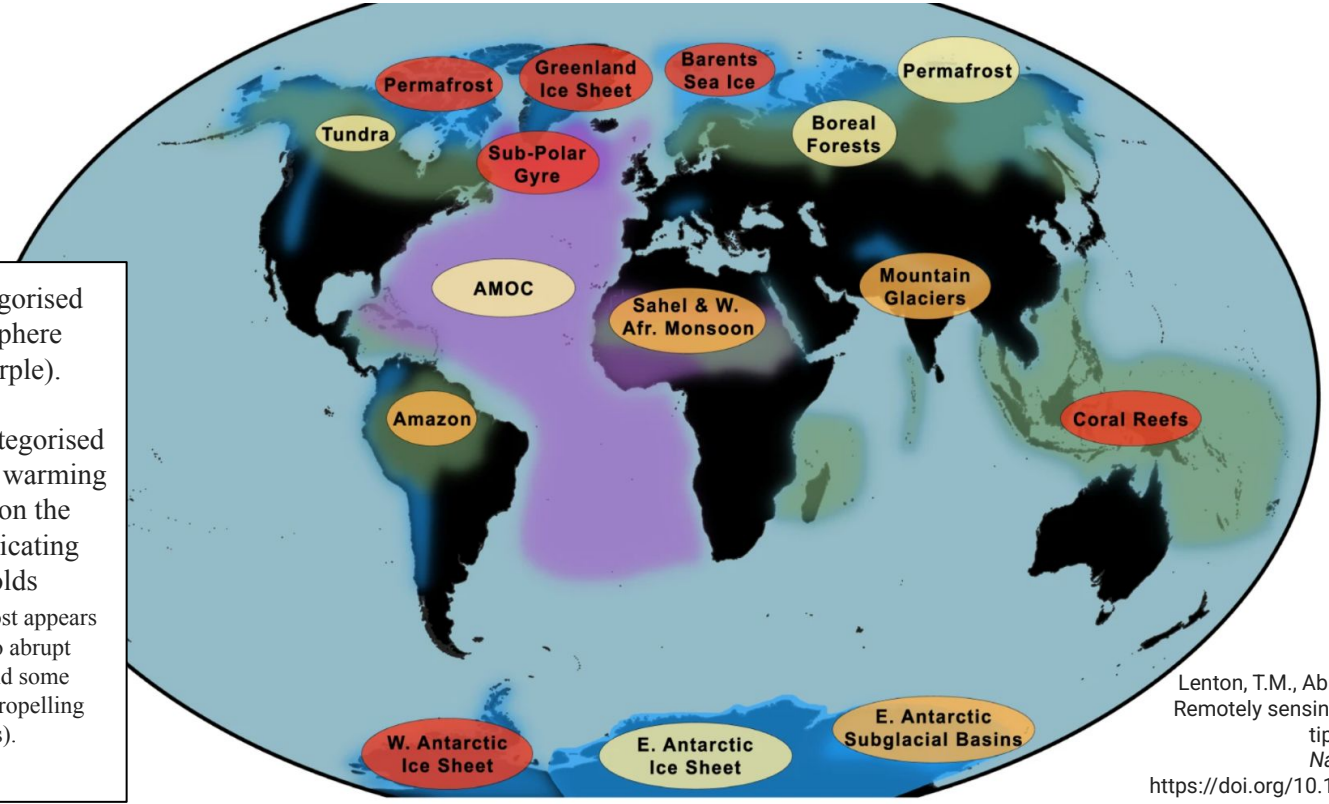
We can frame this lack of visibility of the future as a conversation round tipping points, as a conversation around how we build confidence in our models, and/or as being explicit about for what purposes our models are useful.

But we can't proceed with climate science without addressing these issues for each and every purpose!



Climate Simulation - Tipping Points

TPs have not been observed in the historical period, which makes it impossible to train statistical and AI models directly.



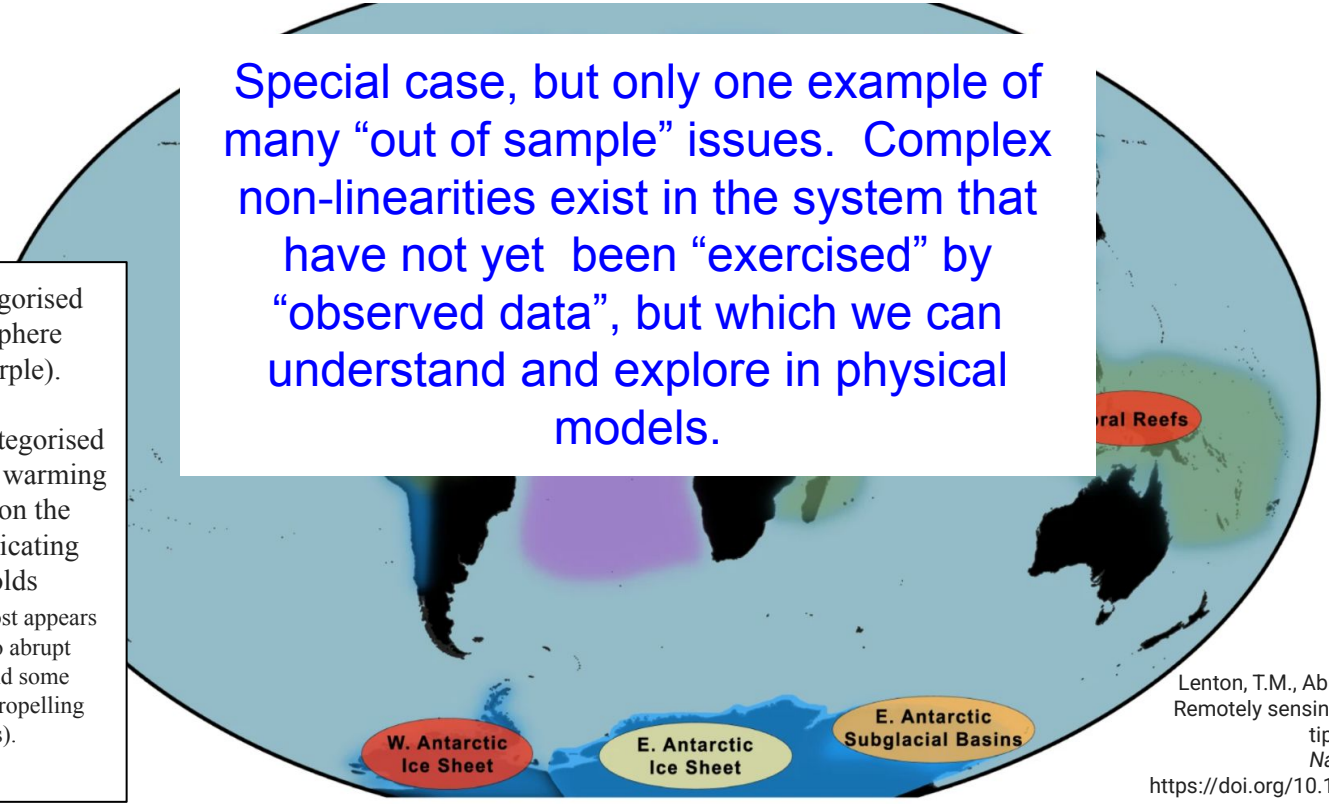
Tipping elements are categorised as cryosphere (blue), biosphere (green), or circulation (purple). Colours of labels denote temperature thresholds categorised into three levels of global warming above pre-industrial (key on the right), with darker red indicating lower temperature thresholds (greater urgency). Permafrost appears twice as some parts are prone to abrupt thaw (at lower temperatures) and some (organic-rich Yedoma) to self-propelling collapse (at higher temperatures).

Lenton, T.M., Abrams, J.F., Bartsch, A. *et al.* Remotely sensing potential climate change tipping points across scales. *Nat Commun* 15, 343 (2024). https://doi.org/10.1038/s41467-023-44609-w

Climate Simulation - Tipping Points

TPs have not been observed in the historical period, which makes it impossible to train statistical and AI models directly.

Special case, but only one example of many “out of sample” issues. Complex non-linearities exist in the system that have not yet been “exercised” by “observed data”, but which we can understand and explore in physical models.



Tipping elements are categorised as cryosphere (blue), biosphere (green), or circulation (purple). Colours of labels denote temperature thresholds categorised into three levels of global warming above pre-industrial (key on the right), with darker red indicating lower temperature thresholds (greater urgency). Permafrost appears twice as some parts are prone to abrupt thaw (at lower temperatures) and some (organic-rich Yedoma) to self-propelling collapse (at higher temperatures).

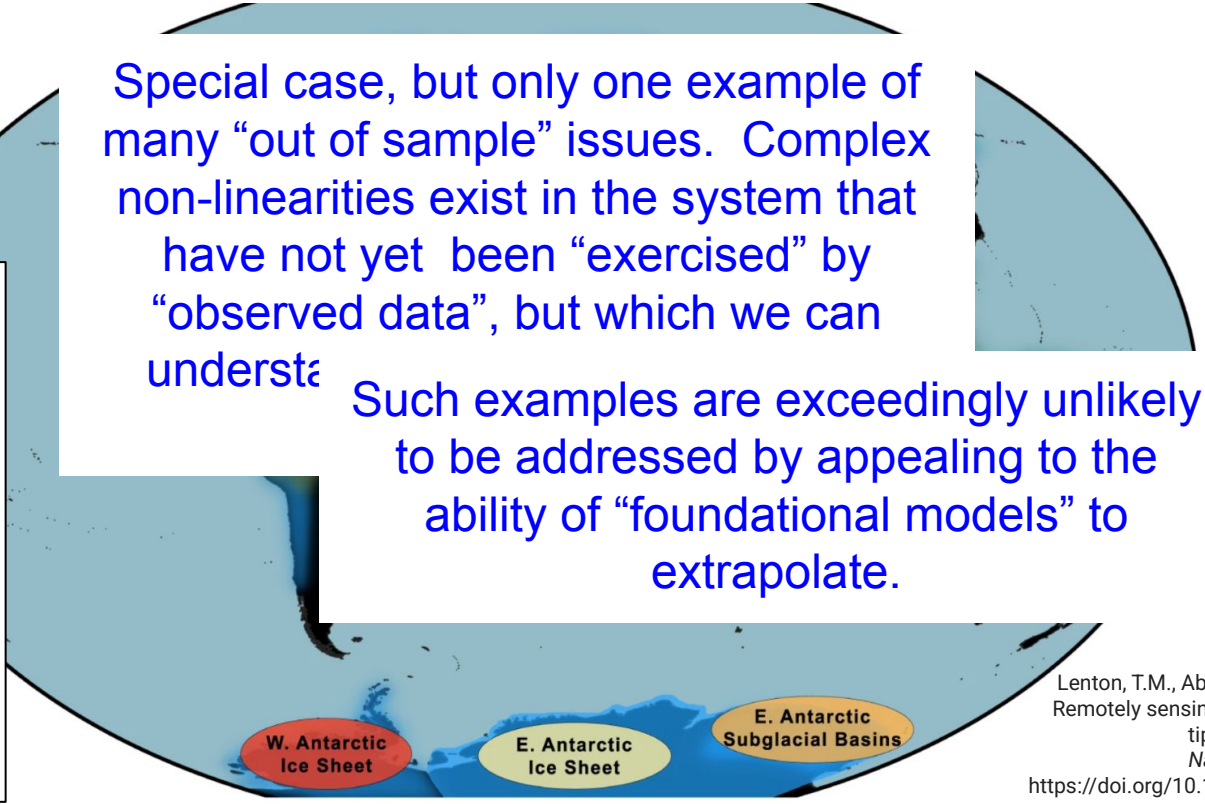
Lenton, T.M., Abrams, J.F., Bartsch, A. *et al.* Remotely sensing potential climate change tipping points across scales. *Nat Commun* 15, 343 (2024). https://doi.org/10.1038/s41467-023-44609-w

Climate Simulation - Tipping Points

TPs have not been observed in the historical period, which makes it impossible to train statistical and AI models directly.

Special case, but only one example of many “out of sample” issues. Complex non-linearities exist in the system that have not yet been “exercised” by “observed data”, but which we can understa

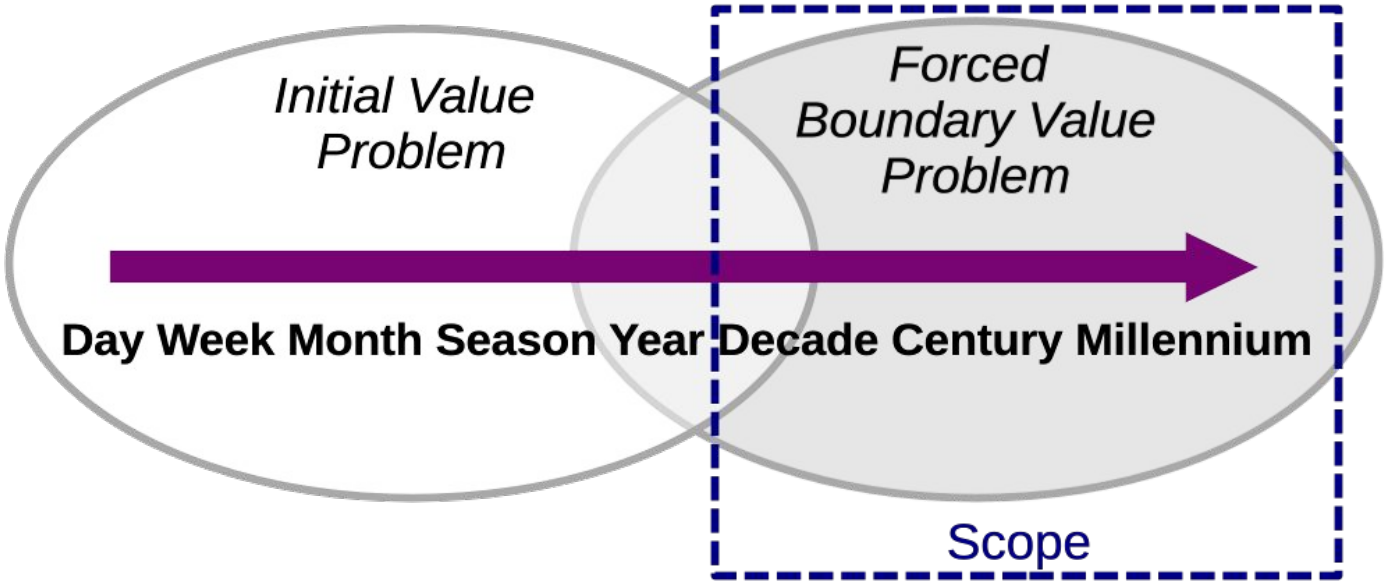
Such examples are exceedingly unlikely to be addressed by appealing to the ability of “foundational models” to extrapolate.



Tipping elements are categorised as cryosphere (blue), biosphere (green), or circulation (purple). Colours of labels denote temperature thresholds categorised into three levels of global warming above pre-industrial (key on the right), with darker red indicating lower temperature thresholds (greater urgency). Permafrost appears twice as some parts are prone to abrupt thaw (at lower temperatures) and some (organic-rich Yedoma) to self-propelling collapse (at higher temperatures).

Lenton, T.M., Abrams, J.F., Bartsch, A. *et al.* Remotely sensing potential climate change tipping points across scales. *Nat Commun* 15, 343 (2024). https://doi.org/10.1038/s41467-023-44609-w

Climate Simulation - A paucity of observations for machine learning!



The application of ML to the future, will probably most benefit from the application of ML to data from simulations of the future!

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

State of the Art in AI-Instead for Climate?

Hybrid (but more Instead than Inside)

- Machine Learn physics with a “simplified traditional” dynamical core. Trained on ERA data,.
- Considerations include whether or not the physics training can be extrapolated, but there are some interesting approaches, including “Climate-Invariant ML” (Beucler et.al. (2024), https://doi.org/10.1126/sciadv.adj7250)

Only one AFAIK (NeuralGCM) which purports not to drift when integrated for 40 years and to support ensembles.

- When used with “modest”(1-2K) modified SST for the future exhibits “some” of the robust features of climate warming, but for more “substantial” (4K) future warming, “response differs from expectations”

True AI-Instead

- ACE2 is the best example AFAIK.
- Use Spherical Fourier Neural Operators (ML tool to map between spherical functions.)
- Force conservation of key properties (moisture, mass) using a corrector.
- Algorithm includes both traditional loss function based on 12 hour forecast plus an additional weighting over a longer period of time and a training ensemble to get “climate accuracy”.

Stable for runs of up to 1000 years! (1 degree resolution at > 1000 SYPD!)

Reproduces lots of interesting phenomena (ENSO,tropical cyclones etc).

Trained on the output of a climate model to get the best results for climate!

PNAS PERSPECTIVE OPEN ACCESS Check for updates

Are general circulation models obsolete?

https://doi.org/10.1073/pnas.2202075119

V. Balaji[a,b,1], Fleur Couvreux[c], Julie Deshayes[d], Jacques Gautrais[e], Frédéric Hourdin[f], and Catherine Rio[c]

This paper is starting to be mis-referenced in the ML literature. I suspect the people who do that have not read it. This paper is WELL worth a careful read!

Emphasises the key need for counter-factuals!

"It is our assertion that parameterized GCMs can expect to get a new lease of life at this moment through sophisticated approaches to model calibration based on methods borrowed from ML."

"The future modeling landscape must rest on the principle of a traceable model hierarchy"

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

The Other Climate Crisis

- Predictions anchored solely by complexity (robustness or agreement across comprehensive climate model ensembles) lack a conceptual understanding.
- Predictions anchored solely by simplicity, such as tipping points, lack robustness when physical complexity is added.
- Some tipping points also involve small-scale processes that are not well constrained by observations.

… **a lack of consensus can often be traced to predictions that are anchored at only one or the other end of the hierarchy**.

The paper goes on to argue for a better understanding of the coupled atmosphere-ocean-land system, via a theoretical (and by implication, a practical) hierarchy of understanding, and **amongst other approaches "mechanism denial experiments".**

Shaw, T. A., & Stevens, B. (2025).*Nature*, *639*(8056), 877–887. https://doi.org/10.1038/s41586-025-08680-1

Climate Simulation - Diverse Range of Applications

Lawrence et al, in prep.

		Deep Time	Spinup, Control & Millenium	Scenarios (Past, Future, Sensitivity)	Single Model Large Ensembles (SMLEs)	Decadal Prediction (Hindcasts, Forecasts)	Equilibrium /Transient Timeslices	Process Studies	Storyline Timeslices
Duration		> 1000y	o(1000)y	o(100)y		o(10)y		o(10-100)d	
Domain	ESM	x	x	x	x				
	GCM		x	x	x	x	x	x	x
	RCM			x			x	x	x
Speed (SYPD)		> 50	> 4	> 4	> 1	> 1	> 0.5	> 0.1	> 0.1
Ensemble Size		1-1000	o(1)	o(10)	10-100	o(10) x NY	o(1)	o(1)	o(1)
Atmos XY Res (km)		o(500)	100-500	25-500	50-300	25-500	1-100	1-10	1-10
Ocean XY Res (km)		300+	100-250	25-250	25-250	25-250	10+	2.5+	2.5+
Exascale Status		Currently impossible		Not currently usable				Currently some capability	

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Climate Simulation - Diverse Range of Applications

Can training ML from simulations generate bigger and cheaper ensembles?

Lawrence et al, in prep.

		Deep Time	Spinup, Control & Millenium	Scenarios (Past, Future, Sensitivity)	Single Model Large Ensembles (SMLEs)	Decadal Prediction (Hindcasts, Forecasts)	Equilibrium /Transient Timeslices	Process Studies	Storyline Timeslices
Duration		> 1000y	o(1000)y	o(100)y		o(10)y		o(10-100)d	
Domain	ESM	x	x	x	x				
	GCM		x	x	x	x	x	x	x
	RCM			x			x	x	x
Speed (SYPD)		> 50	> 4	> 4	> 1	> 1	> 0.5	> 0.1	> 0.1
Ensemble Size		1-1000	o(1)	o(10)	10-100	o(10) x NY	o(1)	o(1)	o(1)
Atmos XY Res (km)		o(500)	100-500	25-500	50-300	25-500	1-100	1-10	1-10
Ocean XY Res (km)		300+	100-250	25-250	25-250	25-250	10+	2.5+	2.5+
Exascale Status		Currently impossible		Not currently usable				Currently some capability	

Can we ML learn improvements from HR models for LR models?

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

AI on top. AI Inside, AI Instead … but simulation underneath!

For climate, our use of AI depends on the quality of our "traditional simulations"

- The key AI inside opportunity is emulation for spin-up and emulation of physical processes from "high resolution" or "observations" (cf "parameterisation").
- AI on top opportunities include downscaling and complex analyses (including of ensemble diversity) - with all the usual caveats about GIGO.

Take home

The foundation of our understanding of the future is upon models in which we can investigate unobserved futures, that is, ones in which we can interrogate our assumptions, and develop plausible scenarios.

It is more than likely we will be utilising AI/ML heavily in much of what we do, but it is **unlikely that AI/ML will** ever **be** as **central in climate science** as it will be in disciplines where the phrase “data driven” rests upon “observational data” - for us, it will rest upon “simulated data”, and we will continue to need to invest in our capabilities to build, utilise, and evaluate such models.

Take home

The foundation of our understanding of the f
models in which we can investigate unobse
is, ones in which we can interrogate our ass
develop plausible scenarios.

It is more than likely we will be utilising AI/ML
what we do, but it is **unlikely that AI/ML will** ever **be** as
in climate science as it will be in disciplines where the phrase "data driven" rests upon "observational data" - for us, it will rest upon "simulated data", and we will continue to need to invest in our capabilities to build, utilise, and evaluate such models.

Nothing in this presentation should be taken as anyone suggesting that **we** don't need to **evaluate our physical models by confronting them with data,** it is simply that that while **such confrontation is necessary**, it is not sufficient!

Hybrid = AI-Inside

- A bit of both!
- But which bit is the most likely to suffer from the “unsampled” regimes?
- Can we work out whether or not the unsampled physics is more or less a problem for “future projection” in this model cf a traditional physics parameterisation?

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., & Hoyer, S. (2024). *Neural General Circulation Models for Weather and Climate* (No. arXiv:2311.07222). arXiv. http://arxiv.org/abs/2311.07222

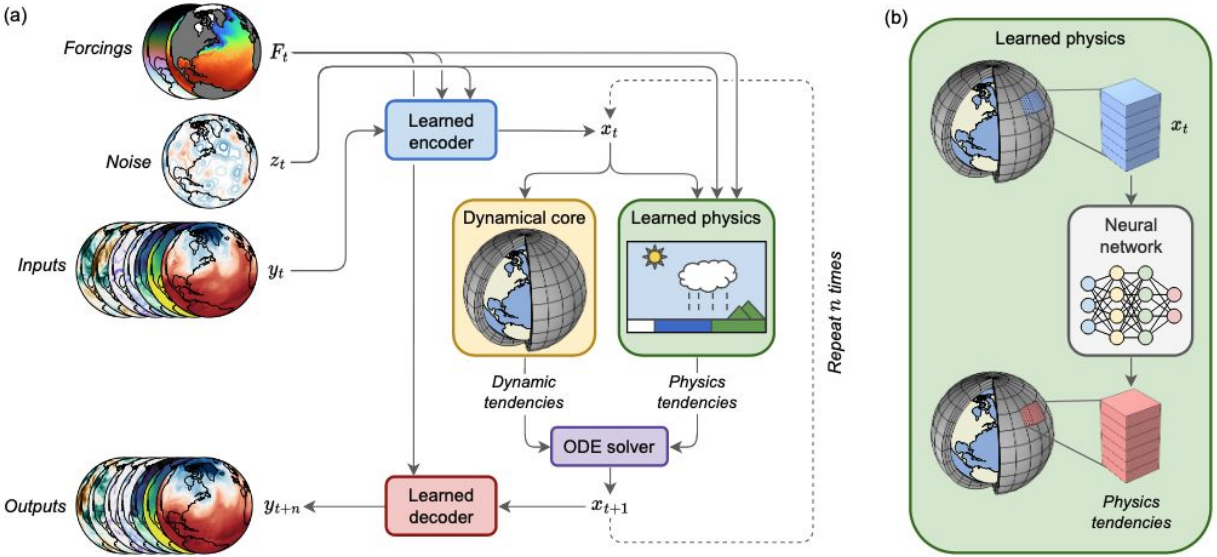


Fig. 1 Structure of the NeuralGCM model. (a) Overall model structure, showing how forcings F_t, noise z_t (for stochastic models), and inputs y_t are encoded into the model state x_t. Model state is fed into the dynamical core, and alongside forcings and noise into the learned physics module. This produces tendencies (rates of change) used by an implicit-explicit ODE solver to advance the state in time. The new model state x_{t+1} can then be fed back into another time step, or decoded into model predictions. (b) Inset of the learned physics module, which feeds data for individual columns of the atmosphere into a neural network used to produce physics tendencies in that vertical column.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Bibliography

1. Balaji, V., Couvreux, F., Deshayes, J., Gautrais, J., Hourdin, F., & Rio, C. (2022). Are general circulation models obsolete? *Proceedings of the National Academy of Sciences*, *119*(47), e2202075119. https://doi.org/10.1073/pnas.2202075119
2. Beucler, T., Gentine, P., Yuval, J., Gupta, A., Peng, L., Lin, J., Yu, S., Rasp, S., Ahmed, F., O'Gorman, P. A., Neelin, J. D., Lutsko, N. J., & Pritchard, M. (2024). Climate-invariant machine learning. *Science Advances*, *10*(6), eadj7250. https://doi.org/10.1126/sciadv.adj7250
3. Stephen Hoyer. (2024, November 20). *Curious how the latest AI weather models compare?* Bluesky Social. https://bsky.app/profile/stephanhoyer.com/post/3lbdxw24cs22l
4. Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., & Hoyer, S. (2024a). Neural general circulation models for weather and climate. *Nature*, *632*(8027), 1060–1066. https://doi.org/10.1038/s41586-024-07744-y
5. Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., & Hoyer, S. (2024b). *Neural General Circulation Models for Weather and Climate* (No. arXiv:2311.07222). arXiv. http://arxiv.org/abs/2311.07222
6. Shaw, T. A., & Stevens, B. (2025). The other climate crisis. Nature, 639(8056), 877–887. https://doi.org/10.1038/s41586-025-08680-1
7. Stevens, B., Adami, S., Ali, T., Anzt, H., Aslan, Z., Attinger, S., Bäck, J., Baehr, J., Bauer, P., Bernier, N., Bishop, B., Bockelmann, H., Bony, S., Brasseur, G., Bresch, D. N., Breyer, S., Brunet, G., Buttigieg, P. L., Cao, J., … Ziemen, F. (2024). Earth Virtualization Engines (EVE). *Earth System Science Data*, *16*(4), 2113–2122. https://doi.org/10.5194/essd-16-2113-2024
8. Watt-Meyer, O., Henn, B., McGibbon, J., Clark, S. K., Kwa, A., Perkins, W. A., Wu, E., Harris, L., & Bretherton, C. S. (2024). *ACE2: Accurately learning subseasonal to decadal atmospheric variability and forced responses* (No. arXiv:2411.11268; Version 1). https://doi.org/10.48550/arXiv.2411.11268