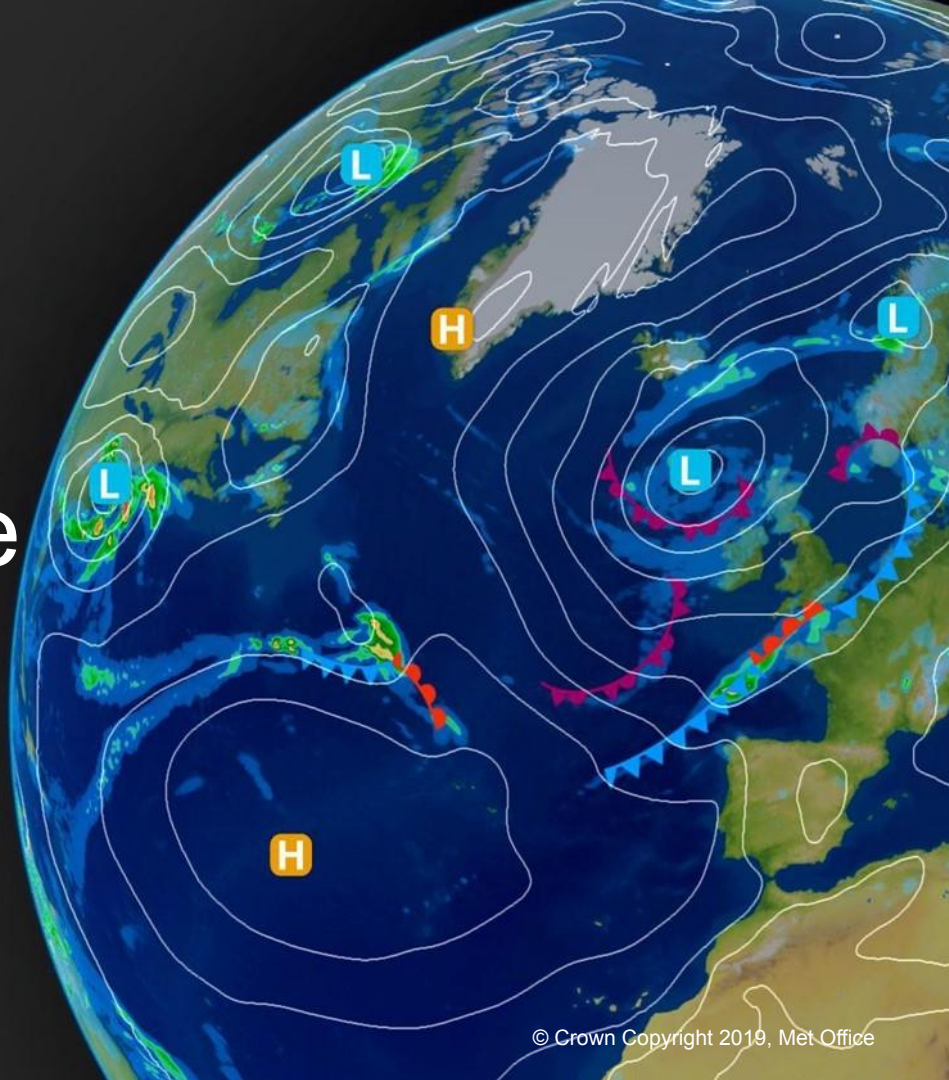
Challenges and opportunities for seamless model development for the 2030s: a climate modelling view

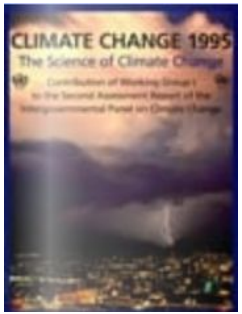
Richard Wood

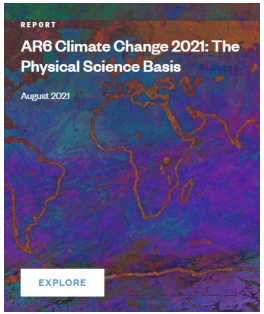
Met Office Hadley Centre, UK



The evolving questions for climate science

- UNFCCC, IPCC
  - Is the climate changing?
  - Is human activity responsible? How much?
  - How will that change continue in future? (conditional on various socio-economic / emissions scenarios)
  - What will be the likely impacts?

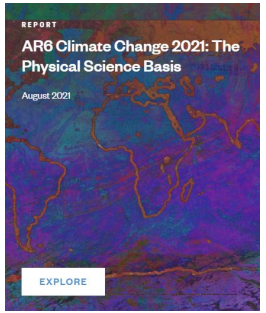




The evolving questions for climate science

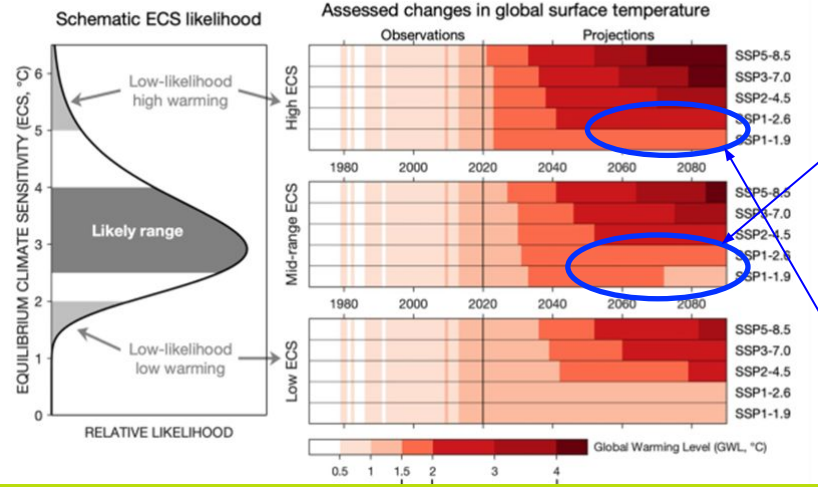
- UNFCCC, IPCC
  - Is the climate changing? YES
  - Is human activity responsible? YES How much? Quite a lot of it
  - How will that change continue in future? (conditional on various socio-economic / emissions scenarios) More than we've already seen. Long term commitment (e.g. sea level)
  - What will be the likely impacts? Far-reaching, with considerable uncertainties

- So what are we going to do about it?
  - Mitigation pathways (multiple forcings)
  - Adaptation and resilience (local/regional)



New questions for mitigation

- More precise evaluation of the effect of different *emissions* pathways
  - Multiple forcers
  - Earth system feedbacks (to convert emissions to concentrations)
  - Often long timescales for feedbacks (e.g. C cycle) and responses (e.g. ice sheets)
- Overshoot scenarios
- Full assessment of risk, including likely outcomes and High Impact-Low Likelihood risks



Low-emissions pathways keep warming below 1.5 or 2C, if the real- world's climate sensitivity has a mid-range value…

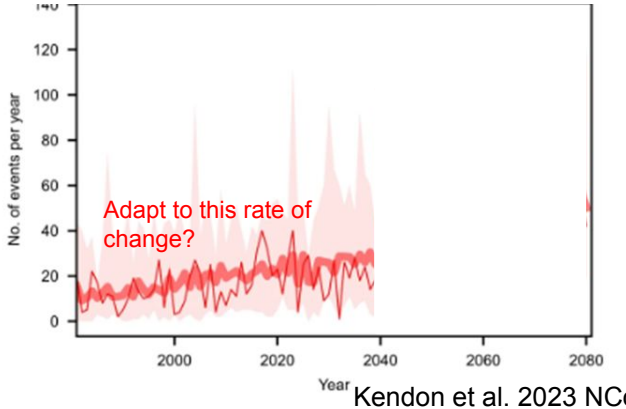
… but if the climate sensitivity turns out to be at the upper end of the plausible range, the same emissions pathway would lead to warming >2C. When could be know if this was the case? How can we address this risk?

(IPCC AR6 WG1 Figure TS.6)

New questions for adaptation

- More precision in time and space
  - Regional/local projections
  - Changing weather types / extremes
  - How might climate change impacts play out?

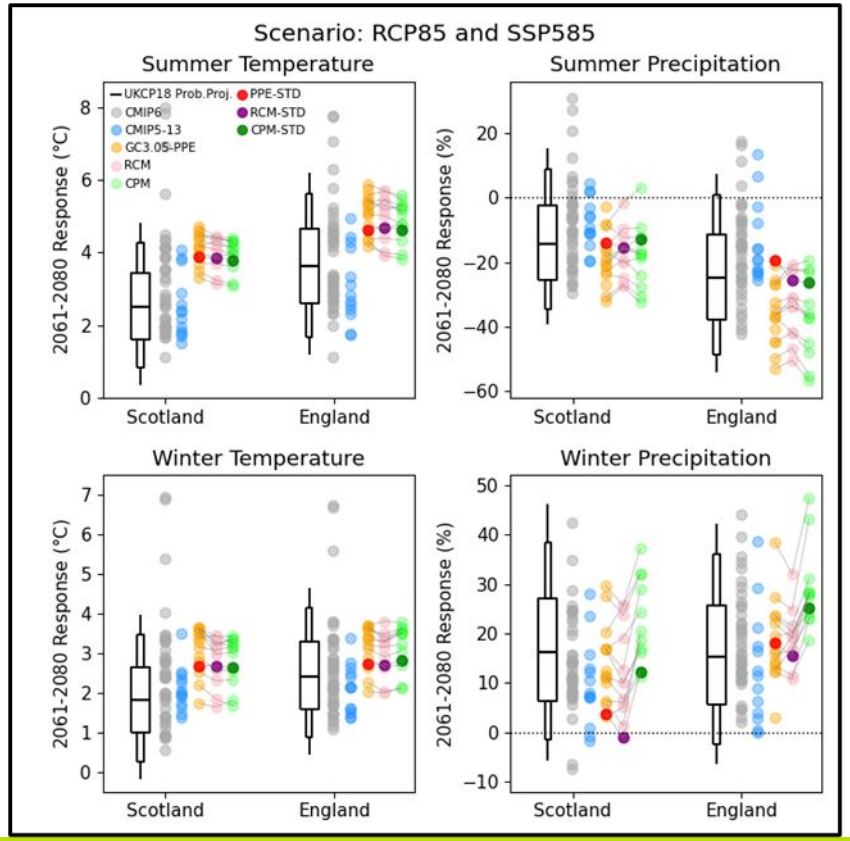
Number of UK extreme rainfall events.
Thin line shows one realisation.



Kendon et al. 2023 NComms
https://doi.org/10.1038/s41467-023-36499-9

- Full assessment of risk, including likely outcomes and High Impact-Low Likelihood risks

Ensembles and probability



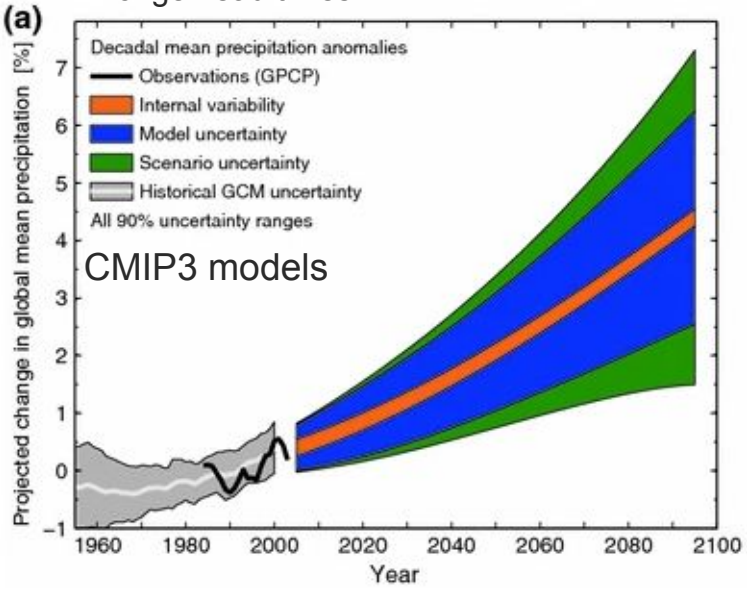
- Projected changes (2061-80 minus 1981-2000) (UKCP18)
- Perturbed parameter ensemble (downscaled) vs CMIP5 and CMIP6 ensembles
  - Are we sampling the full range of possibilities?
  - Does frequency imply probability?
  - Are all members equally credible?

Figure courtesy James Murphy, Met Office

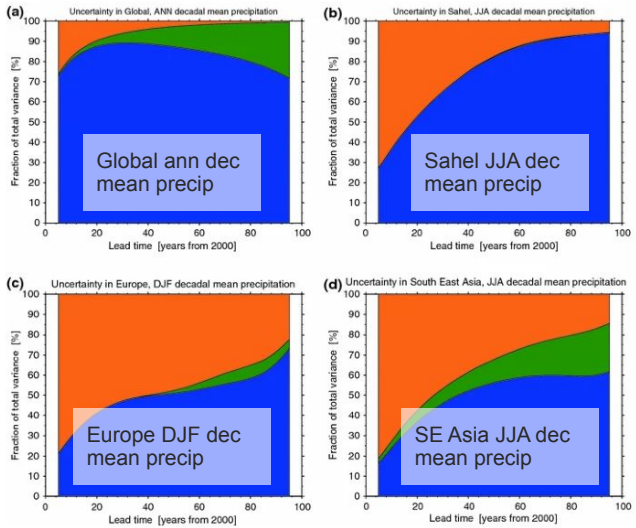
Have we got enough climate models?

- TAR (2001): 31
- AR4 (2007): 23
- AR5 (2013): 39
- AR6 (2021): 60

Model uncertainty can dominate initial condition or scenario uncertainty at longer lead times…

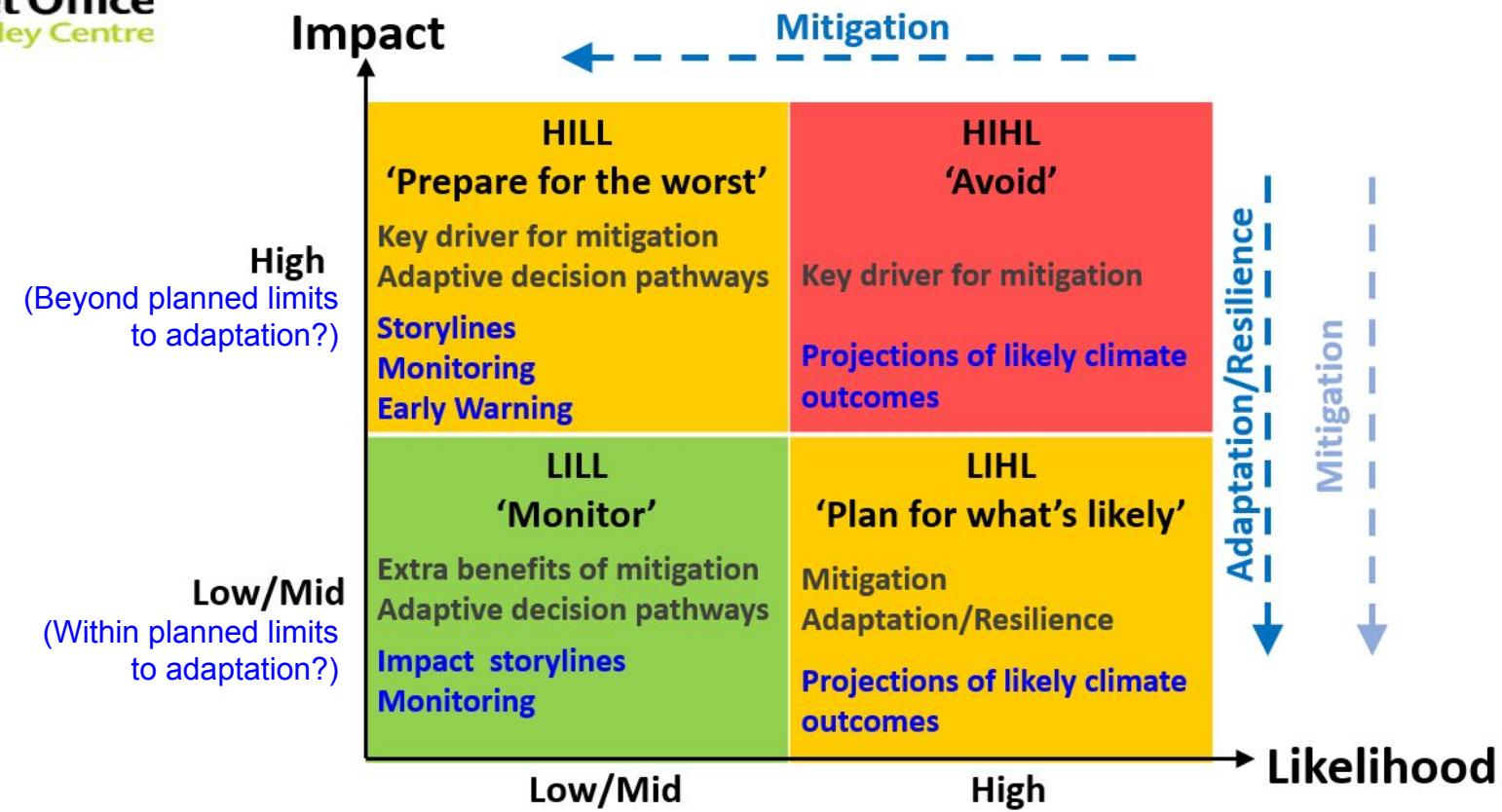


… but the balance depends on variable and period



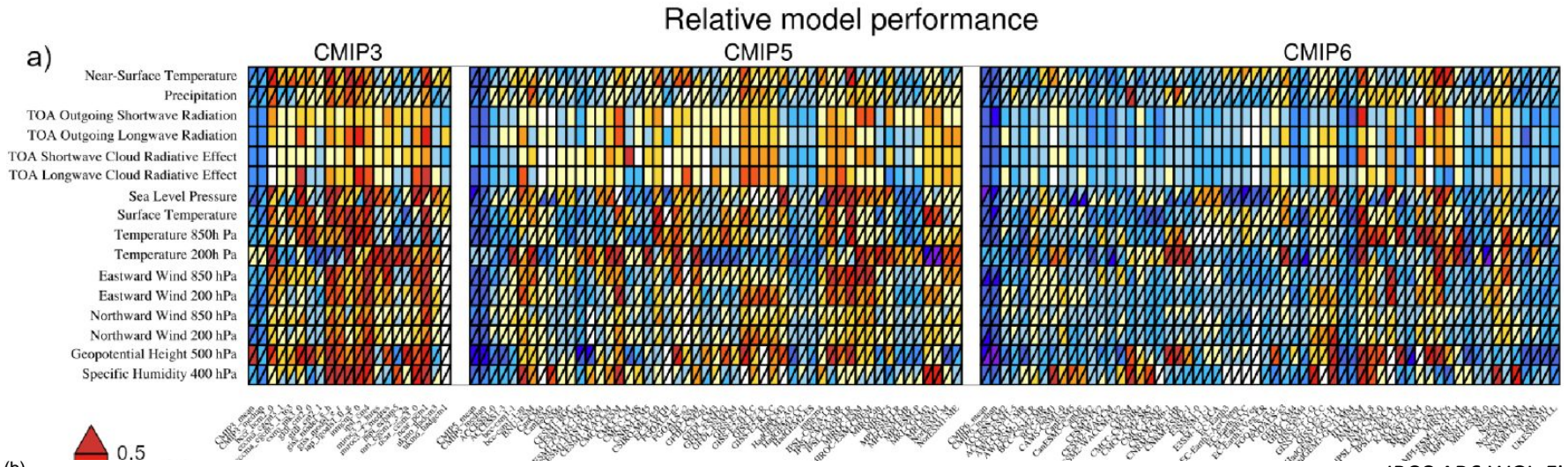
Hawkins & Sutton Cl. Dyn. 2010

‘Uncertainty’ and risk

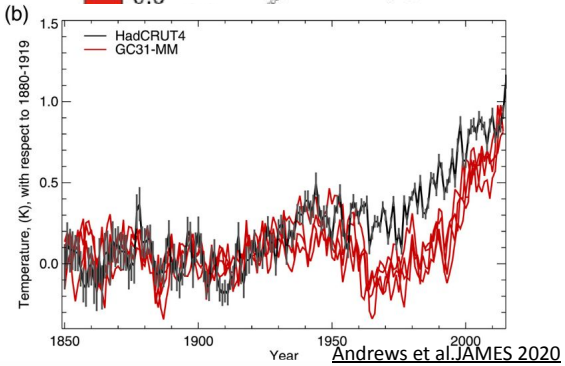


(Wood et al. 2023 Earth’s Future https://doi.org/10.1029/2022EF003369)

What makes a ‘good’ model?



IPCC AR6 WGI, Fig. 3.42

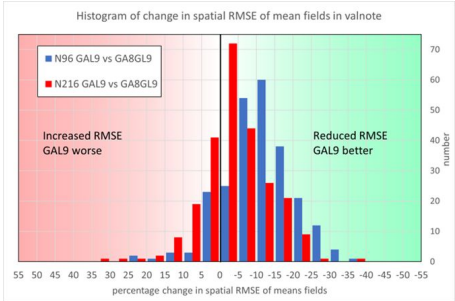


Andrews et al.JAMES 2020

- A model that’s good by one metric may be less good by another
- Does it make sense to choose different models for different questions?
- How do we weight the quality of the model’s processes?

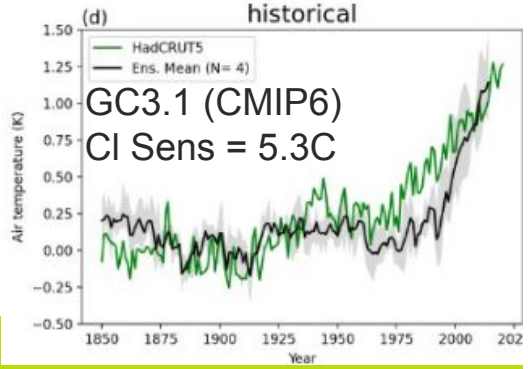
Example: HadGEM3-GC5 (2024)

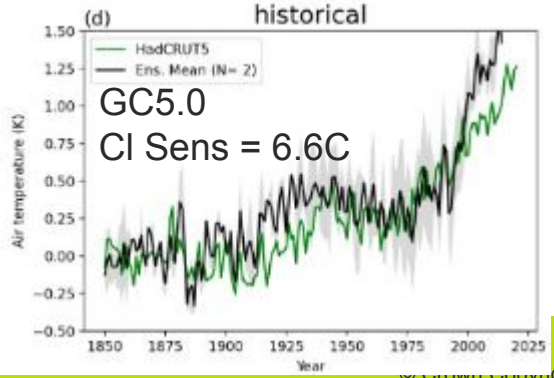
GC5 (2024): Developed through standard seamless model development process. Climate sensitivity allowed to emerge freely. Improved mean climate but high climate sensitivity.

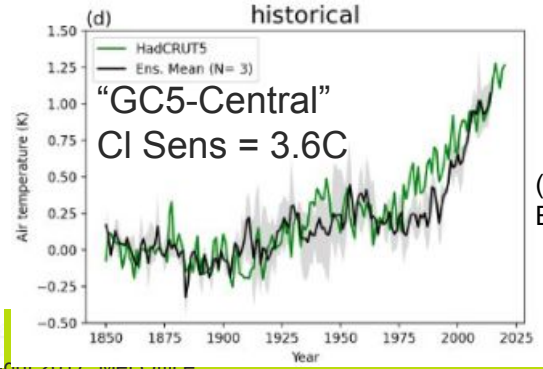


(Willett et al., in prep.)

GC5-Central (2025): Semi-automated parameter tuning to reduce Climate Sensitivity without ruining the mean climate simulation





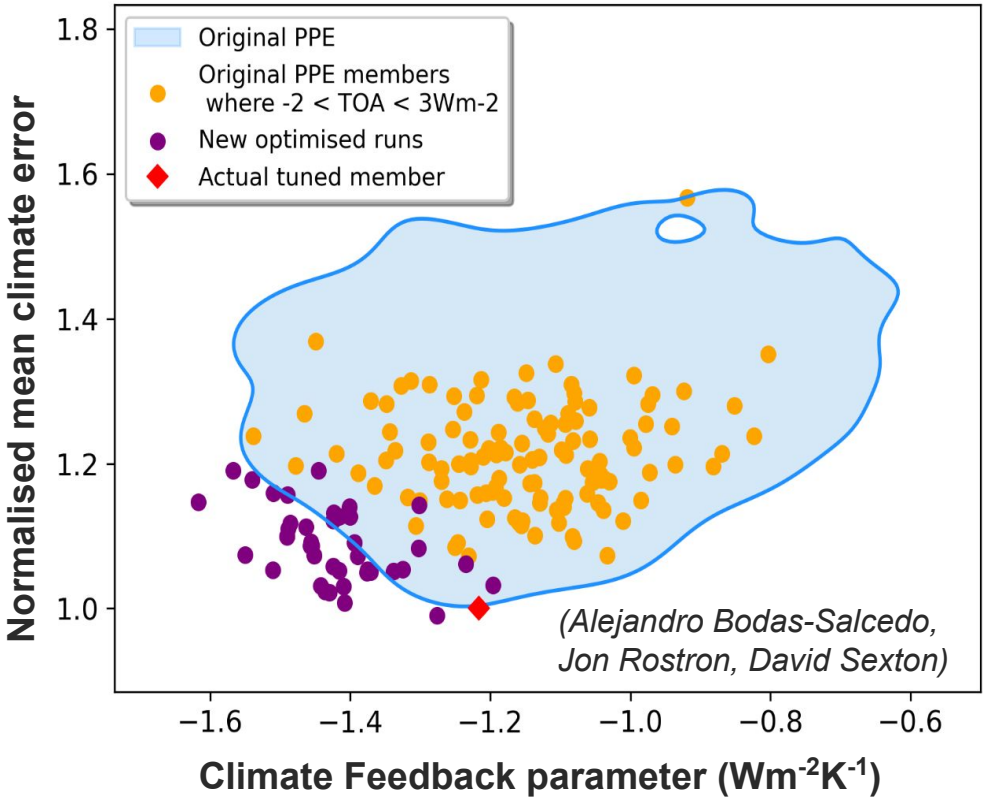


(Courtesy Alejandro Bodas-Salcedo et al.)

Use of PPE/emulator for GC5-central

- Based on Peatier et al., GRL, 2022.
- **PPE:** 5-yr runs of *amip* and *amipF[illegible]4K*.
- **Emulation + optimisation** to select alternatives with CFP < - 1.4 W m^{-2} K^{-1}.
- Coupled tests and manual TOA tuning with these parameters

- **It was really hard to maintain the mean climate performance of the baseline model!**
- **Could ML soon allow us to emulate and optimise to more target metrics?**

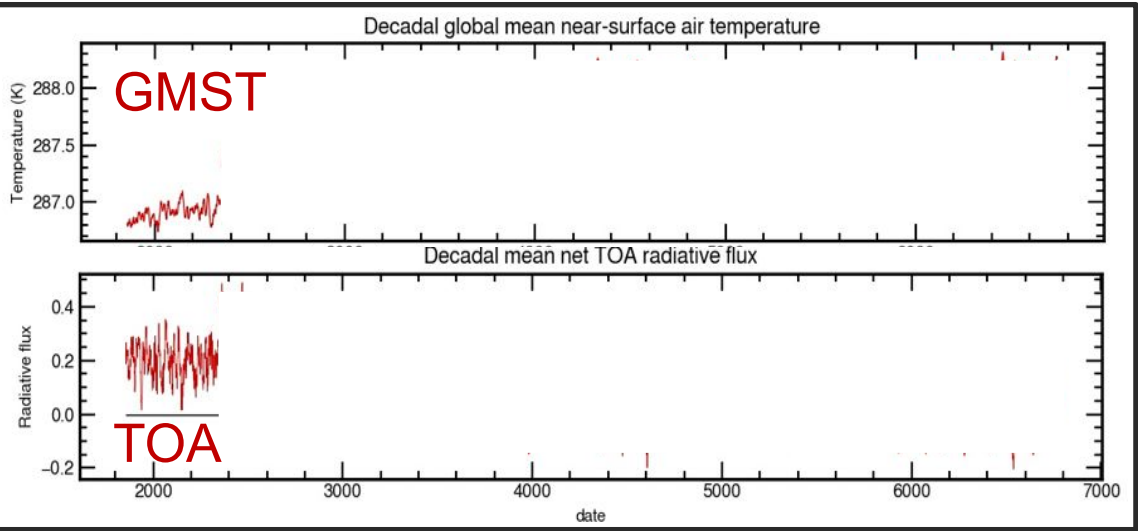


Model speed and spinup

- Climate properties of models only emerge after long runs / large ensembles
- How do we practically integrate evaluation/tuning of climate performance into seamless model development cycles?
  - Faster models? (may weaken seamlessness. ML?)
  - Identify key metrics for long-term behaviour (ML?)
  - Accelerated spinup methods? (ML?)

Adjustment toward equilibrium of HadGEM3-GC31-LL piControl

TOA adjusts towards zero through a succession of abrupt events (flushing) over 1000s of years



Courtesy Gareth Jones, Jeff Ridley
DOI:10.1029/2021GL097171

Machine Learning?

NWP may be a well-sampled problem amenable to ML (given a good reanalysis!)

Seasonal: do we have a large enough sample to train ML predictions?

Climate Change: fundamentally out-of-sample.

- "Turing test": Train a ML model on a GCM historical simulation that shows no AMOC collapse. Force the ML model with a forcing that collapses the AMOC in the GCM. What does the ML model predict?

Possible ML applications

- Downscaling
- Hybrid models (interpolation – beware extrapolation!)
- Ensemble enhancement (emulation/interpolation)
- Automated model tuning to multiple metrics
- Emergent constraints on long-term behaviour

Seamless model development into the 2030s

- Increased emphasis on changing weather types and extremes. It's the same (fast) physics!

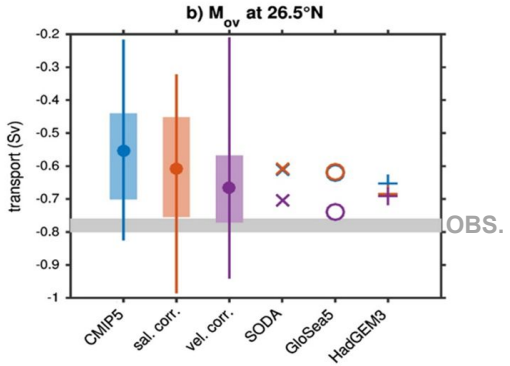
- Different priorities for what makes a good model (e.g. slow physics). Can't get everything right!
- Need for speed, for model development and application. E.g. do global climate models need to be non-hydrostatic?
- Need to span full range of uncertainty/risk. Model uncertainty a bigger factor for climate than weather? Is the concept of an ensemble of each centre's 'best' climate model outmoded? Storylines ~~rather than~~ as well as pdfs?
- Foundational ML may work for NWP, may not for climate. Pushing us to fundamentally different approaches?
- Traceability rather than seamlessness? Global ↔ Regional? Scale-aware parametrisations?
- ML may help us tackle several of these challenges!

- G

- G

Traceable model hierarchies

- By definition a large ensemble will contain few or no HILL events. How do we explore HILL storylines and target our limited GCM runs?
- (Even worse) GCMs as a class may be biased too stable. How do we get to places that the MME/PPE can't reach?



CMIP5. Bias in key fresh water transport controlling AMOC stability.
Mecking et al.Tellus A 2017