The Equilibrium Response of Atmospheric Machine-Learning Models to Uniform Sea Surface Temperature Warming

Zhang and Merlis, 2025

ML for climate-timescale atmospheric modeling

Potential benefits of ML climate modeling:

- Significantly reduced computational cost → large ensembles
- Capturing processes that traditional models can't
- Make better use of existing observations

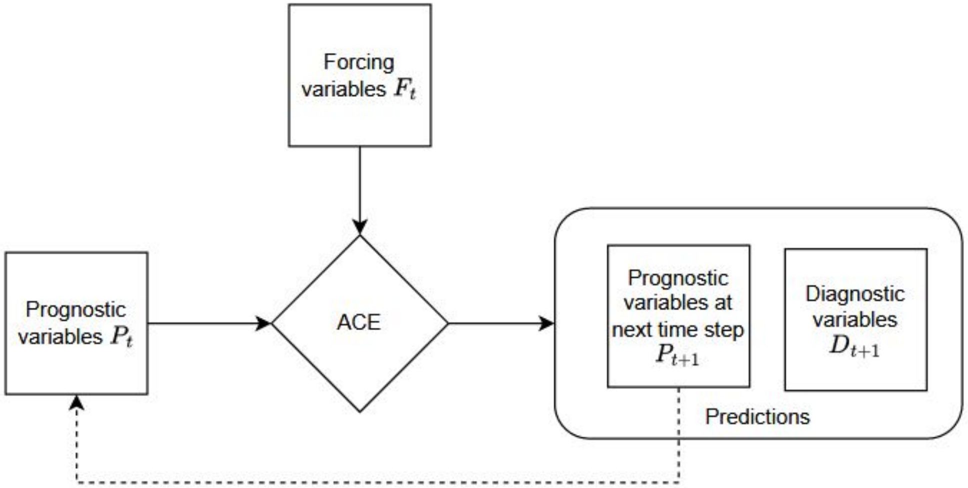
Two large challenges compared to weather modeling:

- Decadal to centennial stability
- **Generalizing to unseen climate states**

Types of ML climate models - pure ML

Examples - ACE, DLESyM

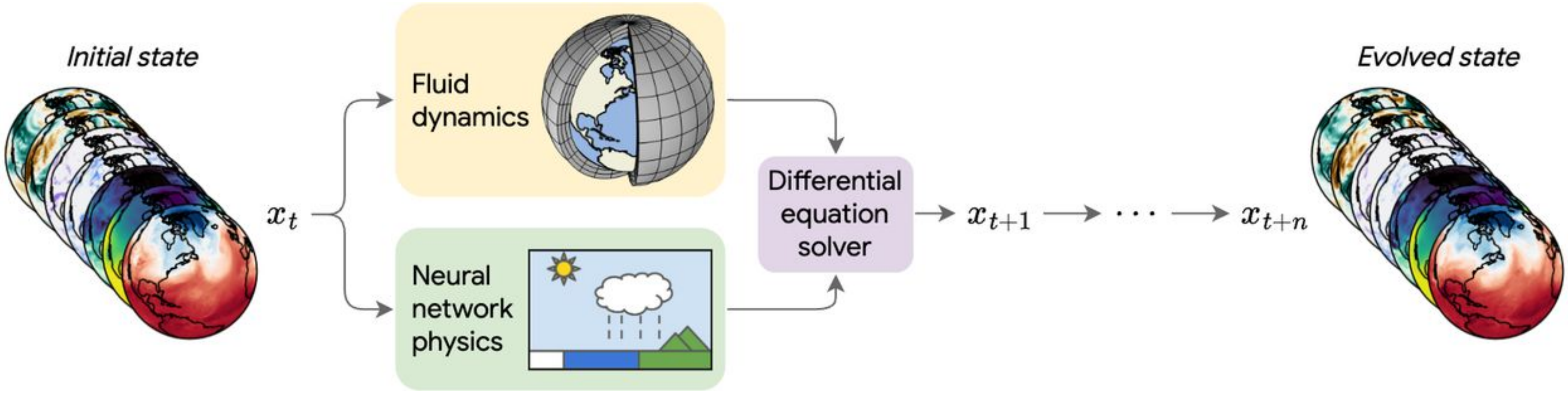
Key feature: autoregressive prediction



Types of ML climate models - hybrid physics/ML

Example - NeuralGCM

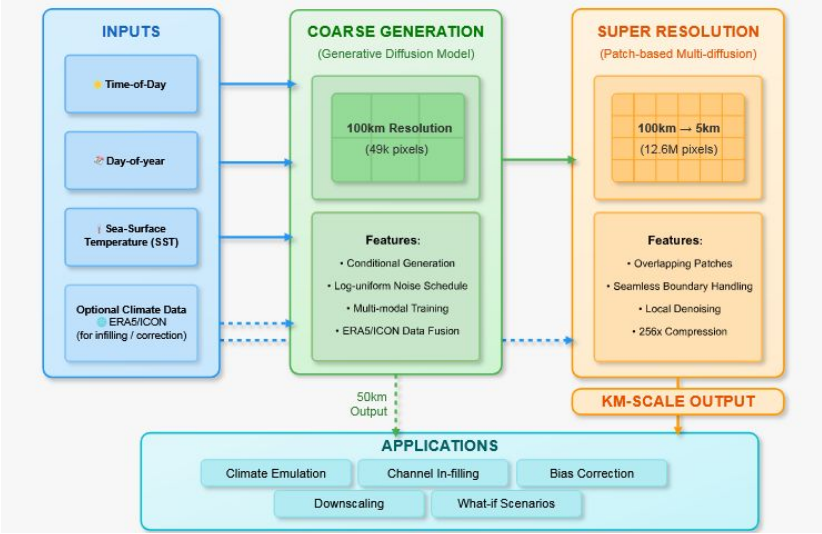
Key feature: retaining dry dynamical core



Types of ML climate models - generative

Example - cBottle

Key feature: no autoregression, directly predicts outputs from boundary conditions (SST, time of year, etc.)



Uniform SST warming experiments

- Standard benchmark for climate models
- Evaluates a model’s climate sensitivity and response to warming
- Easy to set up, rapid equilibration

Experimental setup

ML models tested: ACE, NeuralGCM, cBottle

Traditional reference model: AM4 (GFDL)

Control simulation: forced with monthly climatological SST and SIC from 1981-2014

Warming simulation: +2 K SST applied at every grid point

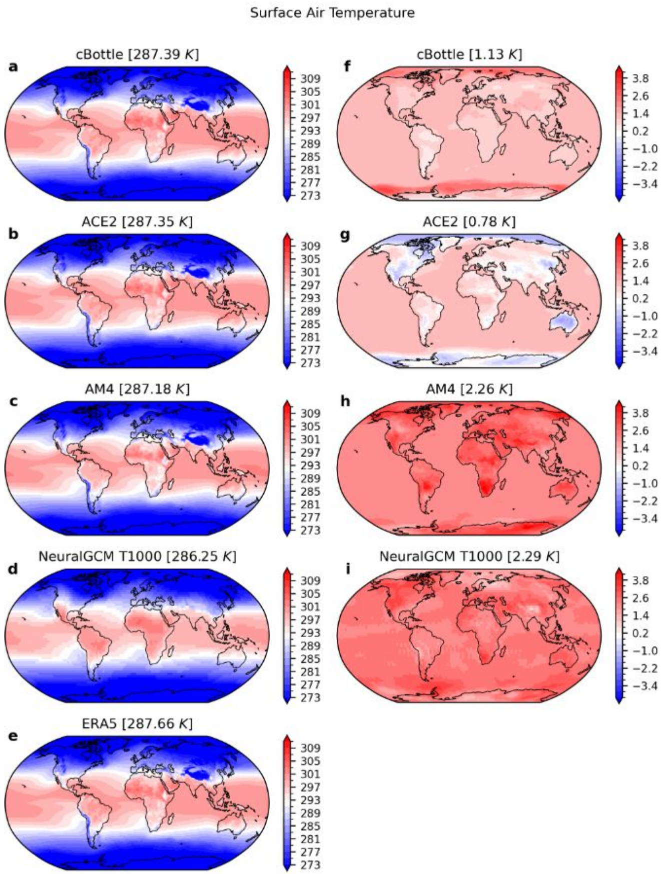
Surface temp. response

We expect

1. Polar amplification
2. Enhanced land warming

Control

+2K minus control



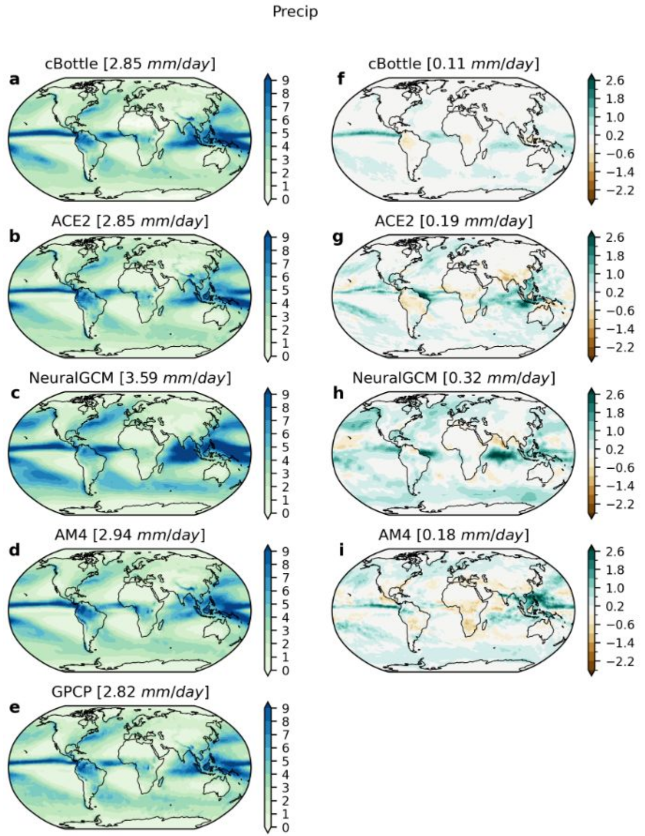
Precipitation

We expect

1. ~3%/K increase in precip
2. Strongest changes at tropics

Control

+2K minus control

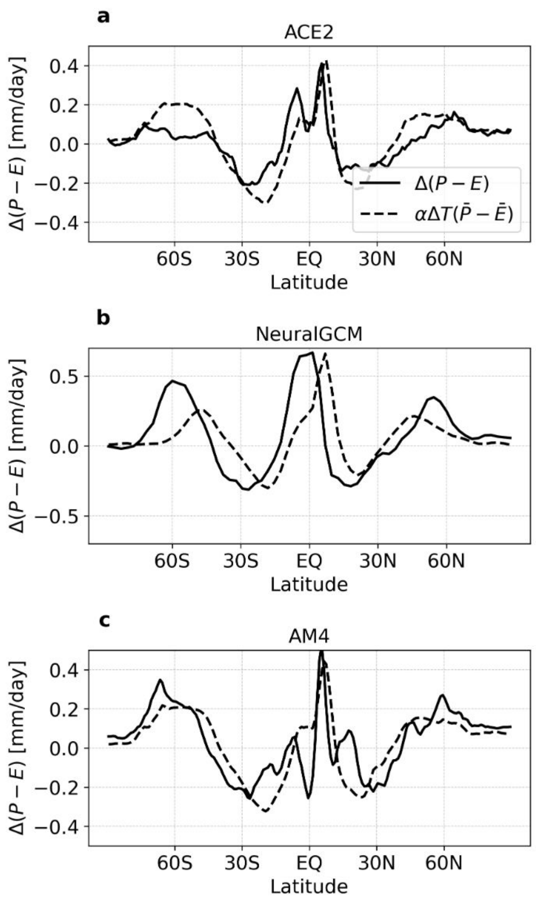


Precipitation

We expect

- "Wet-gets-wetter, dry-gets-drier"

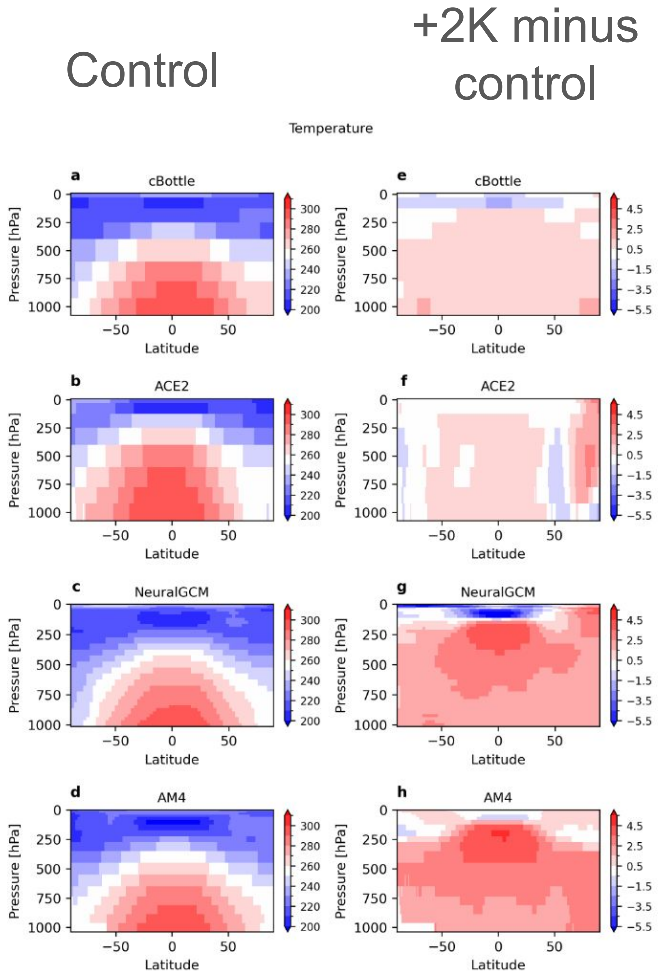
+2K minus control



Upper level temp. response

We expect

- Upper tropospheric warming amplification



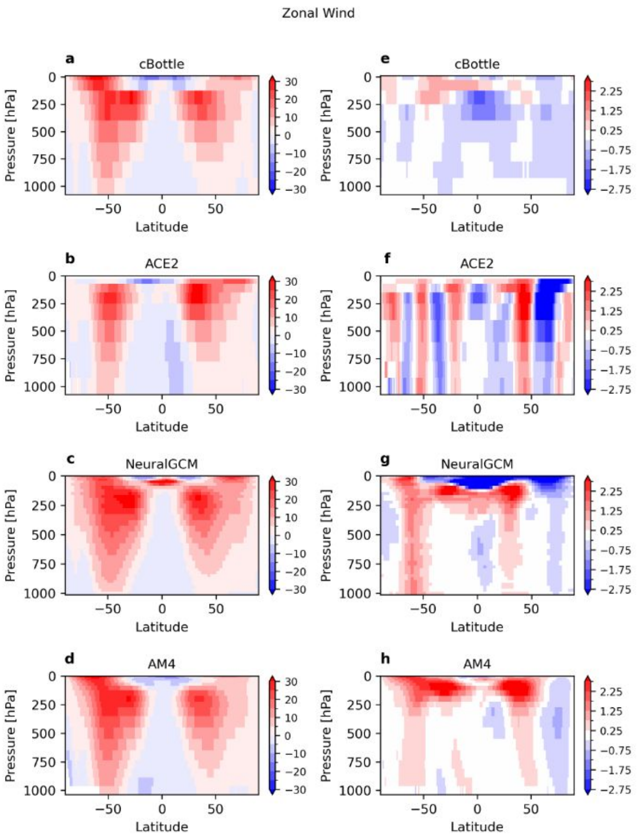
Upper level temp. response

We expect

- Amplified upper tropospheric westerlies (matching upper tropospheric warming)
- Poleward shift of surface easterlies

Control

+2K minus control



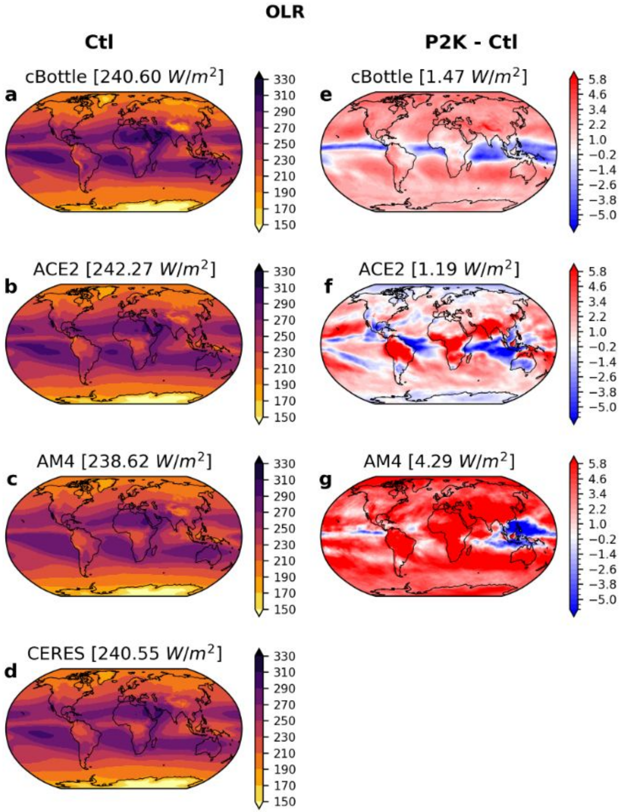
Radiation

We expect

- Global mean increase in outgoing longwave radiation (OLR), indicating enhanced longwave cooling

Control

+2K minus control



Conclusion and takeaways

- Uniform SST experiments remain a tractable benchmark for ML and traditional GCMs alike
- ML models reproduce some key physical responses, particularly in precip., but struggle with others like land warming and radiative response

Conclusion and takeaways

- Uniform SST experiments remain a tractable benchmark for ML and traditional GCMs alike
- ML models reproduce some key physical responses, particularly in precip., but struggle with others like land warming and radiative response

Big question for the field: Can we train a generalizable ML climate model from exclusively historical data?

- Distilling physical laws from historical data
- In distribution vs. out of distribution

How should we test our ML models?

Group discussion