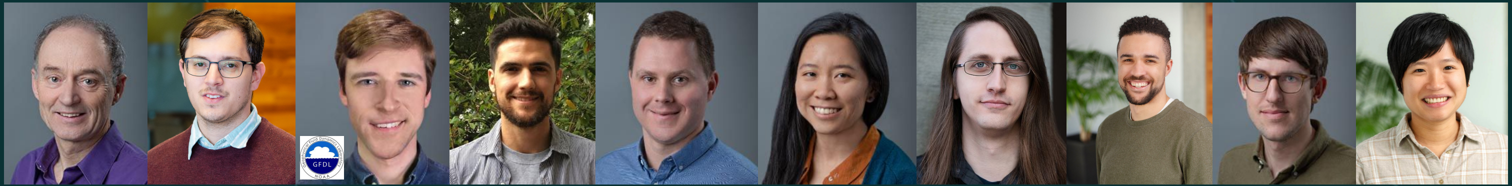
The Ai2 Climate Emulator: Capabilities, Challenges and Opportunities

Oliver Watt-Meyer
and the Ai2 Climate Modeling team



Chris Bretherton Troy Arcomano Spencer Clark James Duncan Brian Henn Anna Kwa Jeremy McGibbon Andre Perkins Oli Watt-Meyer Elynn Wu

Ai2

External partners

M²LInES GFDL LLNL NVIDIA

Who are we?

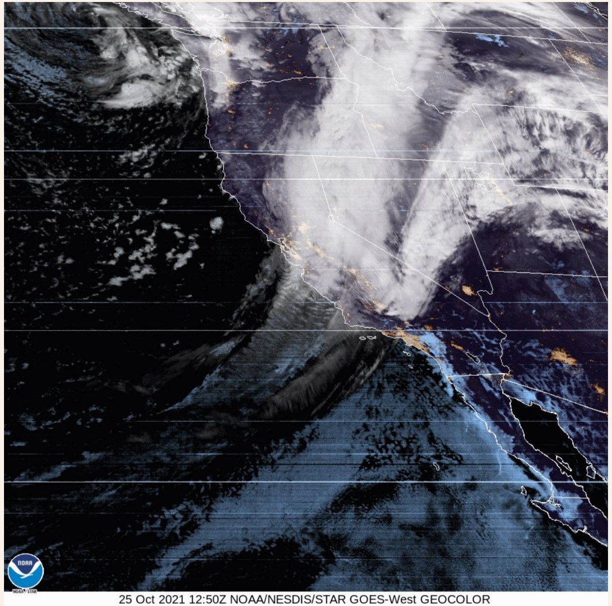
Allen Institute for AI (Ai2): a Seattle-based nonprofit funded by the Paul Allen estate

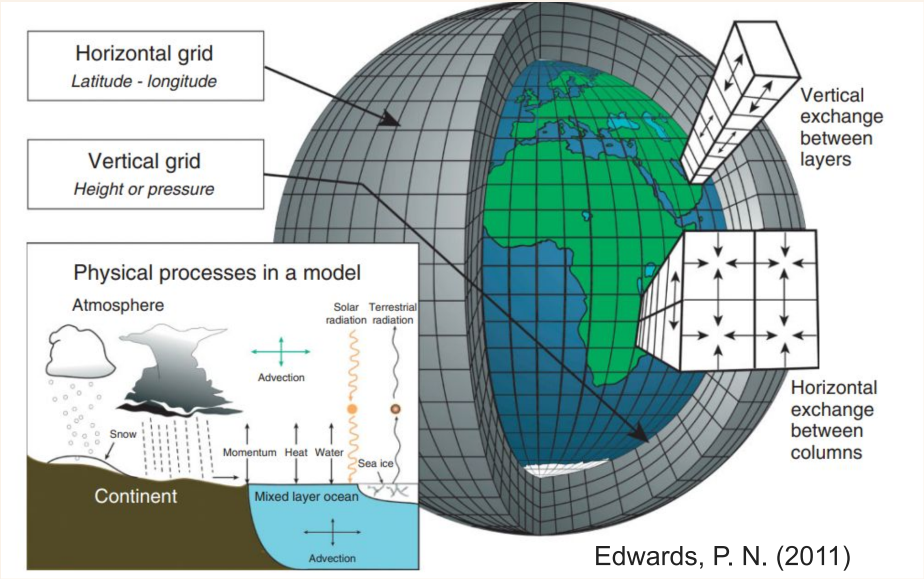
Building breakthrough AI to solve the world's biggest problems

Ai2 Climate Modeling:

- We sit within "AI for the Environment" group at Ai2
- **Goal: Use machine learning to make faster, better, and easier-to-use climate models**
- Partners: NOAA/GFDL and DOE/LLNL (km-scale atmosphere and climate modelers), M^2LInES and NVIDIA (fellow machine learners) & student interns

Physically based weather and climate modeling





Real system we'd like to represent in a… form compatible with computational prediction

How can ML help weather/climate models?

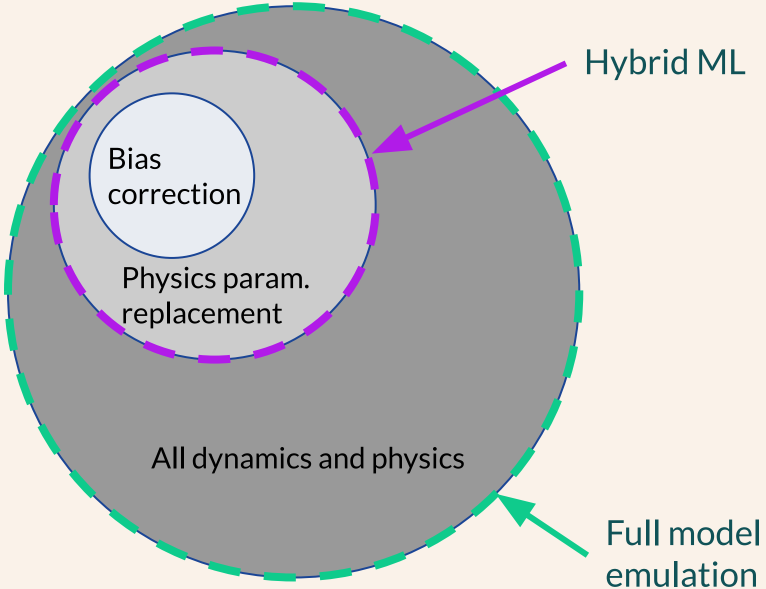
Make global weather and climate models:

- more accurate
- faster
- more affordable/accessible

Three general strategies (atmospheric focus):

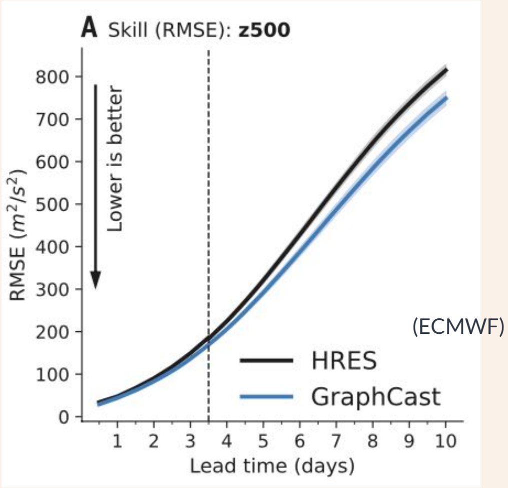
- ML for post-processing bias correction and/or downscaling (e.g. *McGibbon et al. 2024*)
- Hybrid: ML replaces/corrects parts of the atmospheric model (e.g. *Watt-Meyer et al. 2021; Kochkov et al. 2023*)
- **Full model emulation: machine learning of entire global atmospheric evolution**

What is the ML responsible for?



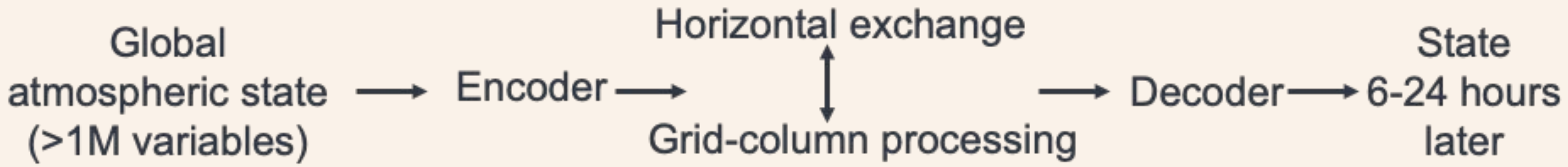
ML-based weather forecasts are state of the art

- **2019–2022:** pioneering work showing potential for ML-based global weather forecasting trained on reanalysis (e.g. *Dueben and Bauer*, *Weyn et al. DLWP*)
- **2022–2023:** Pangu, GraphCast, etc. demonstrate improved 3-10 day deterministic weather prediction over operational benchmarks
- **2024–:** development of probabilistic (e.g. GenCast, AIFS ENS) and coupled ocean-atmosphere models



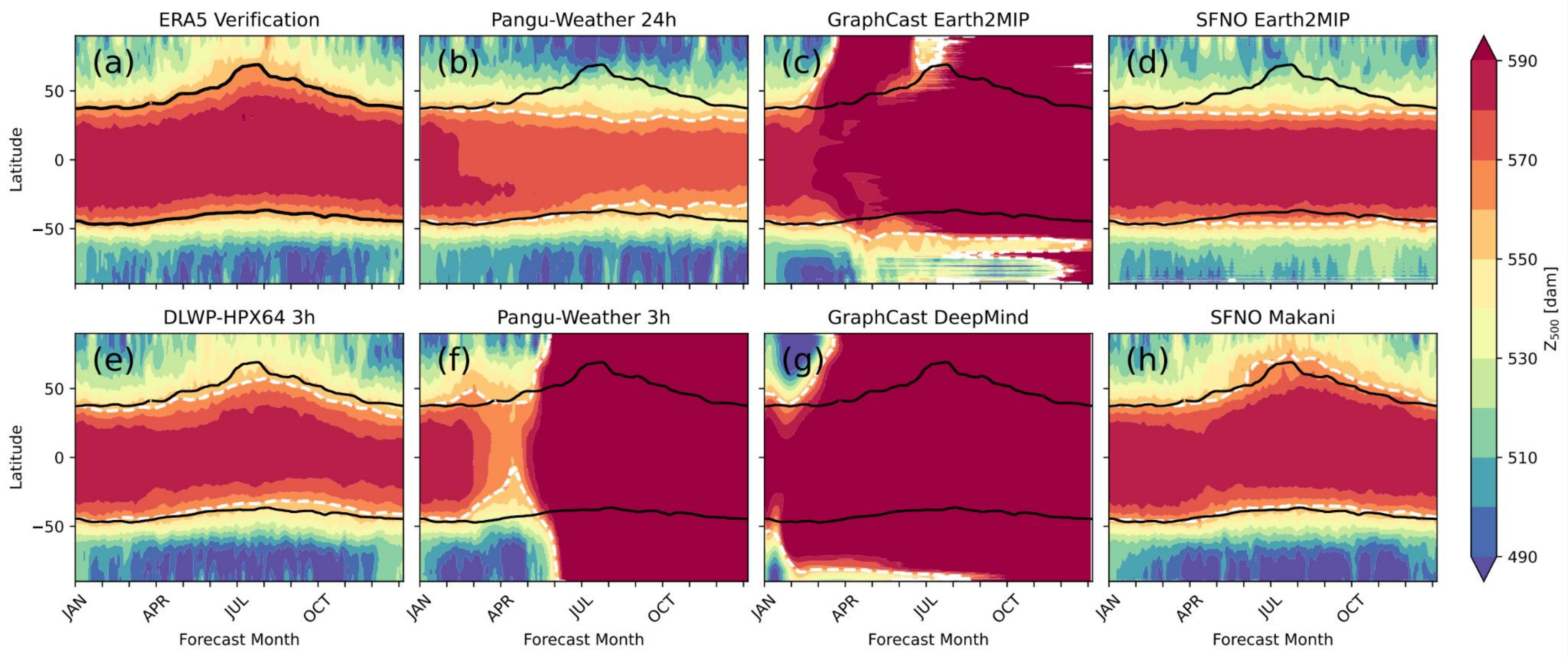
GraphCast: Lam et al., 2023 (*Science*)

Typical strategy for ML weather prediction



- U-Nets, Graph neural nets, neural operators, SWIN transformers, diffusion models
- 10^6 - 10^9 learnable parameters
- Train on ~40 years of ERA5 in days to weeks
- 10-day forecast on a single GPU in a few seconds
- Large efficiency gains from long time step and GPU-optimized matrix math

ML forecasts are often not stable and/or accurate beyond ~2 weeks



Karlbauer et al. 2024; JAMES

Weather vs. climate emulation

- ML weather emulators are trained on decades of global reanalysis, which are most reliable since 1980
- The goal of climate models is to predict past and future climates that may lie well outside this historical range.
- Data-driven ML models are generally unreliable for extrapolation, because they are not directly grounded in physical principles.
- Our goal: Train an ML emulator of an excellent physically-based global climate model for 100-10000x speedup of routine simulations, ensembles, downscaling…

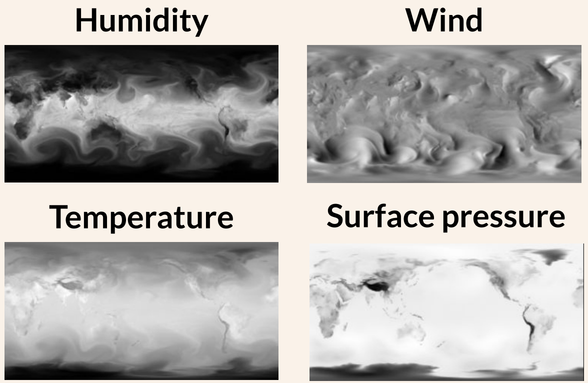
Our approach: Ai2 Climate Emulator (ACE)

- Derive climate as statistics of simulated weather, like physics-based GCMs
- Train on climate model output in addition to ERA5
  - Want to use data from diverse range of climates (global warming etc.)
- Start simple, then increase complexity:
  - Use relatively coarse 1° (100 km) grid and 8 finite volume terrain-following layers
  - Initially we used climatological SST and fixed external forcing
  - ACE2: varying external forcings (historical and increased CO_2)
- Focus on long-term stability and climate accuracy

Training Setup

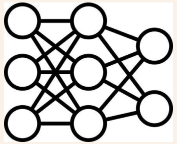
Loss function: mean-squared error of 6-hour forecast computed over all the output variables (for ACE2, accumulated over 2 forward steps)

Input at time *t*



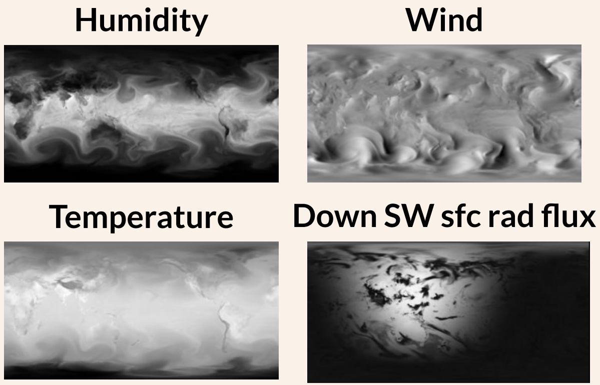
(**forcing** + **prognostic** vars) x 180 x 360

SFNO*



$$f(\mathbf{u};\theta)$$

Prediction of *t* + 6 hours



(**prognostic** + **diagnostic** vars) x 180 x 360

Showing subset of inputs/outputs

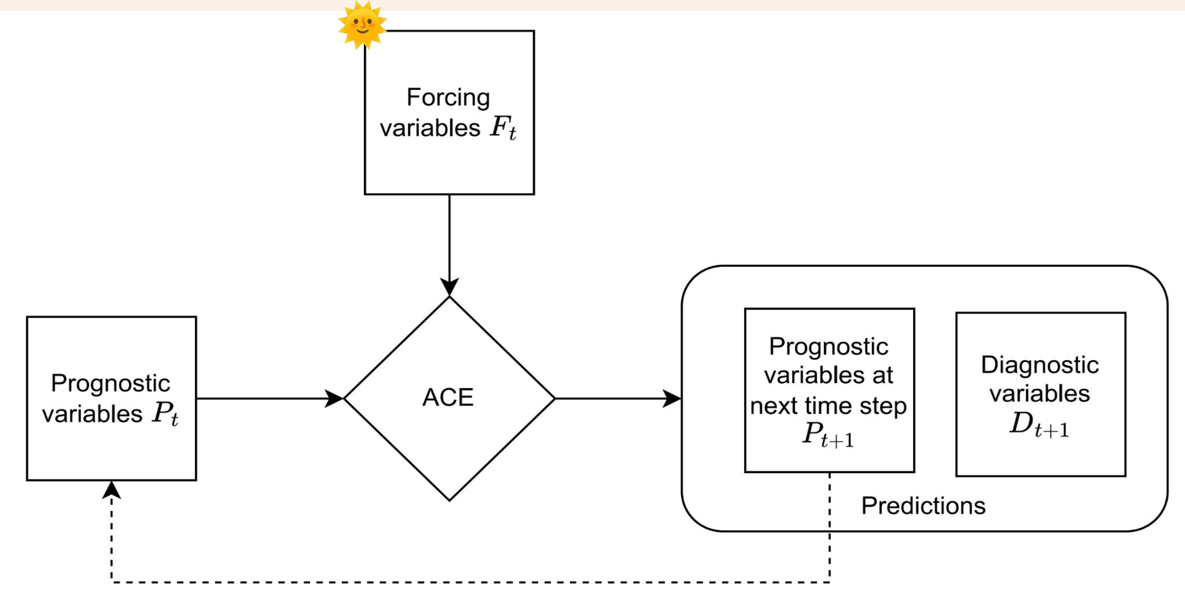
*__SFNO__: Spherical Fourier Neural Operator from Bonev et al., 2023 arxiv.org/abs/2306.03838

Our variable set

💨 **Prognostic**:
→ horizontal winds, temperature, specific total water, skin temperature of land/sea-ice, surface pressure

🌞 **Forcing**:
→ insolation, sea surface temperature, surface type fractions, elevation (and sometimes CO2)

💧 **Diagnostic**:
→ precipitation, TOA and surface radiative fluxes, surface turbulent heat/moisture fluxes



ACE training data

Train ACE on 6-hourly 3D data—either NOAA FV3GFS/SHiELD model or ERA5 reanalysis

Regrid to 8 vertical layers on a 1° Gaussian grid for SFNO compatibility, speed, less storage

Reference datasets:

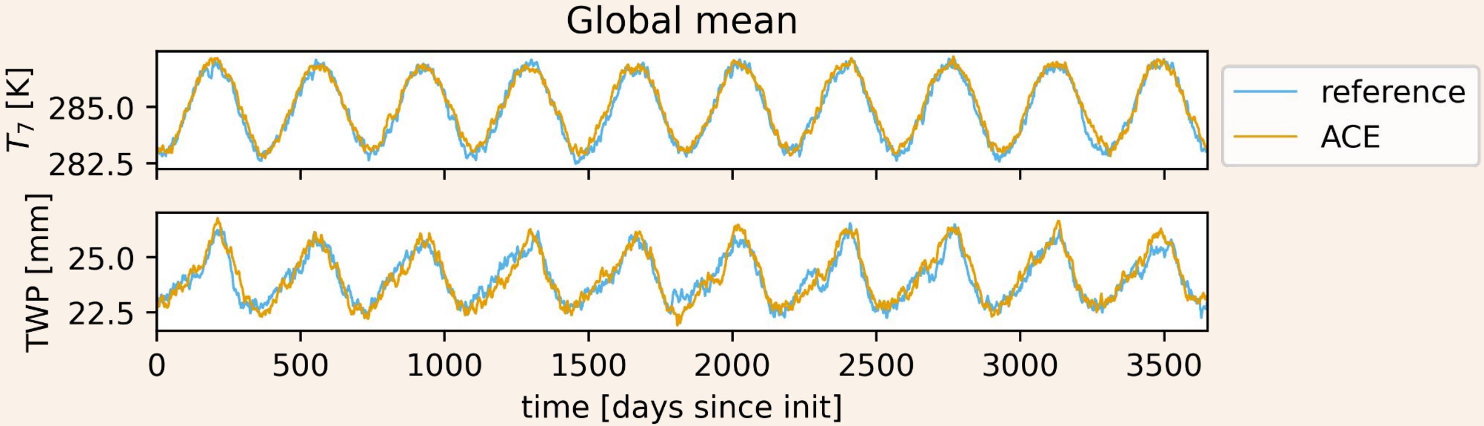
1. **ACE**: Annually-repeating SST forcing (climSST)
   - 100 years of training data, 10 years of validation
   - Each year is an independent sample of the same climate forcing
2. **ACE2**: Historical SSTs (AMIP)
   - ERA5 or an SHiELD AMIP Ensemble
3. **ACE2-SOM**: Slab ocean with CO_2 forcing
   - Present-day CO_2, 2xCO_2 and 4xCO_2, with a simple slab ocean

Ai2

ACE v1: Stable, accurate seasonal cycle!

Temperature
(boundary layer average)

Total Water Path
(vertically integrated moisture)



Indefinitely stable
800 simulated years/day on one A100 GPU

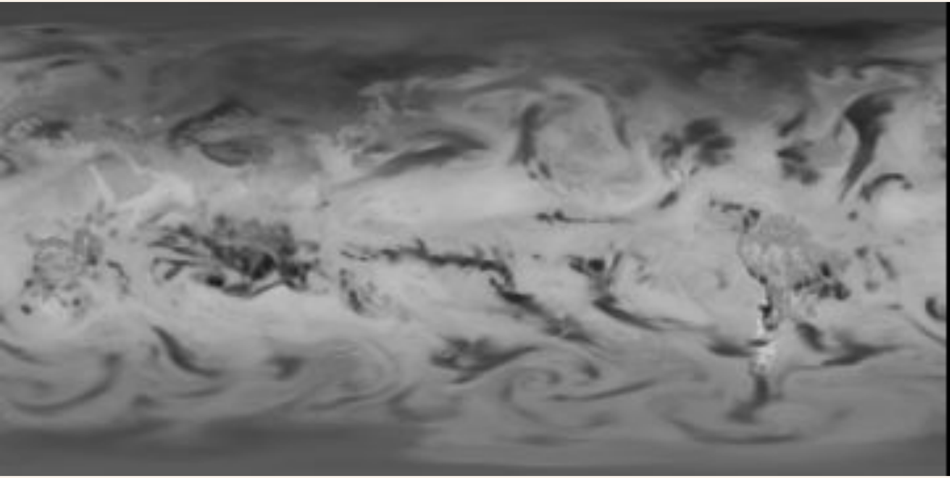
Ai2

Watt-Meyer et al. (2023)

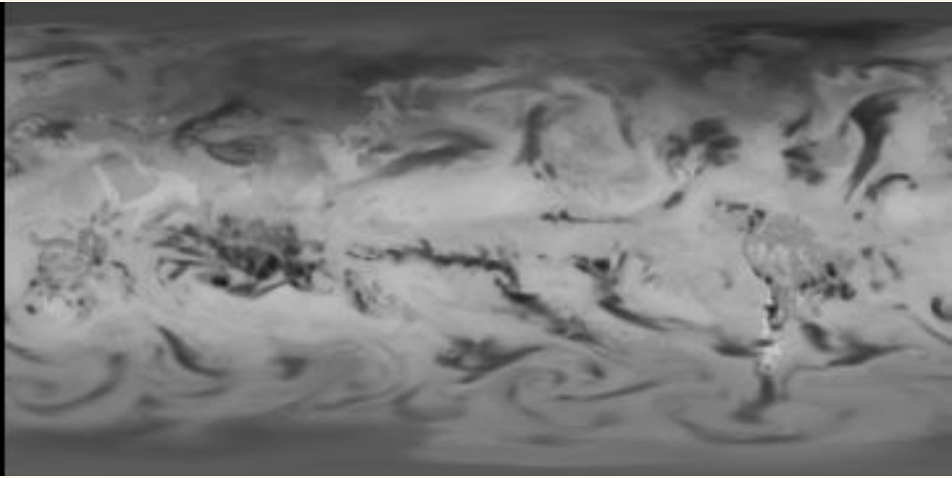
Realistic weather variability

Outgoing longwave radiation (OLR) for first 100 days of simulation

ACE (prediction)



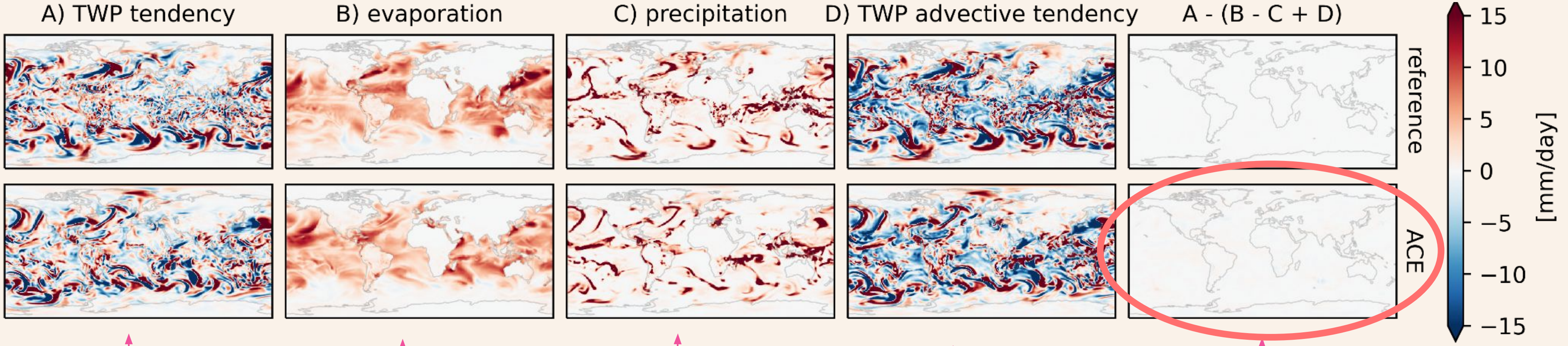
FV3GFS (target)



Mostly realistic OLR despite model not explicitly prognosing clouds!
But also see evidence of overly smooth prediction.

Physical consistency

$$\frac{\partial \mathrm{TWP}}{\partial t} = E - P + \left.\frac{\partial \mathrm{TWP}}{\partial t}\right|_{\mathrm{adv}} \qquad \mathrm{TWP} = \frac{1}{g}\sum_k q_k \Delta p_k$$



Change in Total Water Path

Surface evaporation rate

Surface precipitation rate

Horizontal moisture transport

Balance of all these terms

Should be zero!

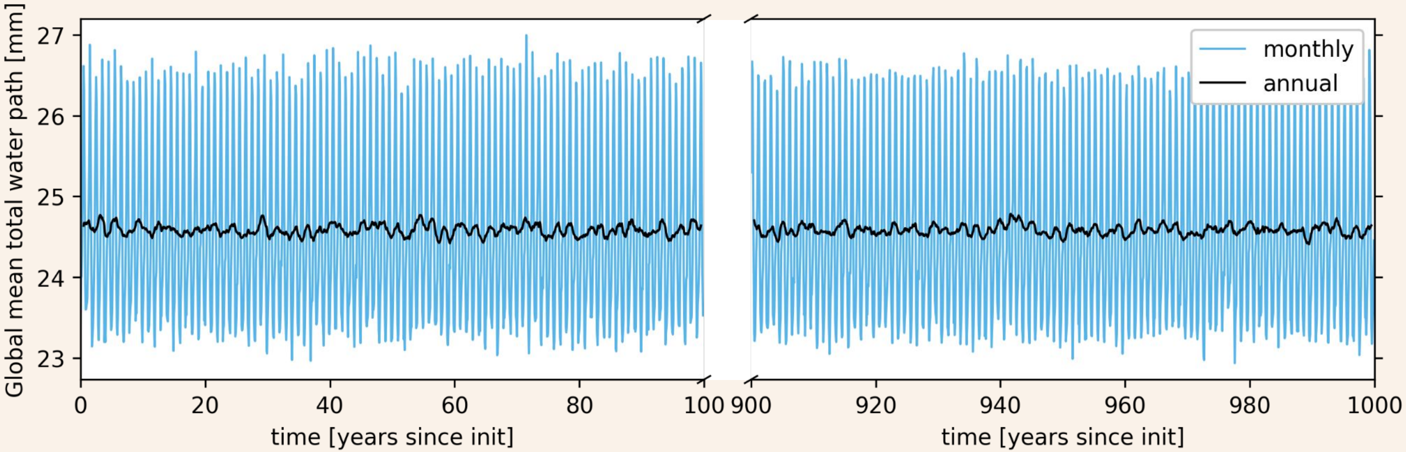
Plot shows a 6-hr time step 1 year into simulation

Ai2

Watt-Meyer et al. (2023)

Why is ACE more stable and accurate than weather emulators?

- Architecture makes a difference
  - SFNO is very stable
- Loss weighting of predicted fields
  - Accounts for different timescales of different variables
- Choice of variables, especially forcing inputs including SST
  - Weather emulators lock predicted ocean changes to weather, allowing error build-up
  - Insolation and topography are important and also used by some weather emulators



1000 year ACE2 simulation with repeating SST

Ai2

ACE2

- Built-in dry air and moisture conservation
- Architectural improvements in capacity, loss function, normalization
- Trained and tested on AMIP (historical SST variability)
- Add CO_2 as a forcing variable
- 1500 simulated years per day on an H100 GPU

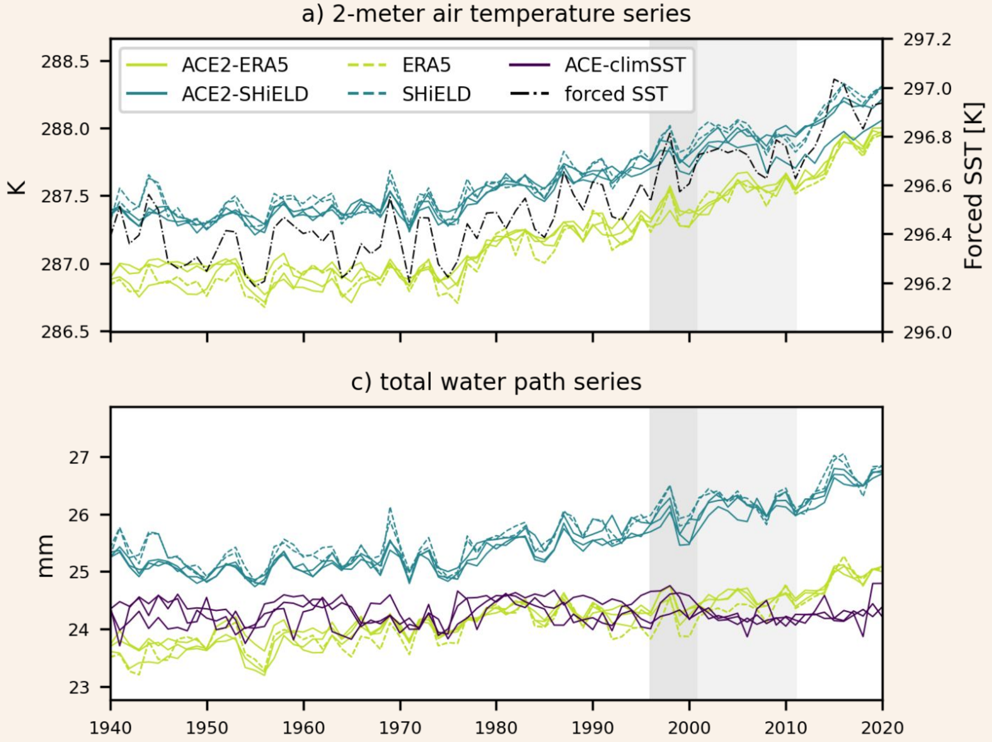
Watt-Meyer et al. 2024, http://arxiv.org/abs/2411.11268, submitted to *NPJ Climate & Atm Sci*

ACE2: historical SST warming

Two new training datasets spanning 1940-2020:

1. ERA5 → **ACE2-ERA5**
2. C96 SHiELD AMIP simulations → **ACE2-SHiELD**

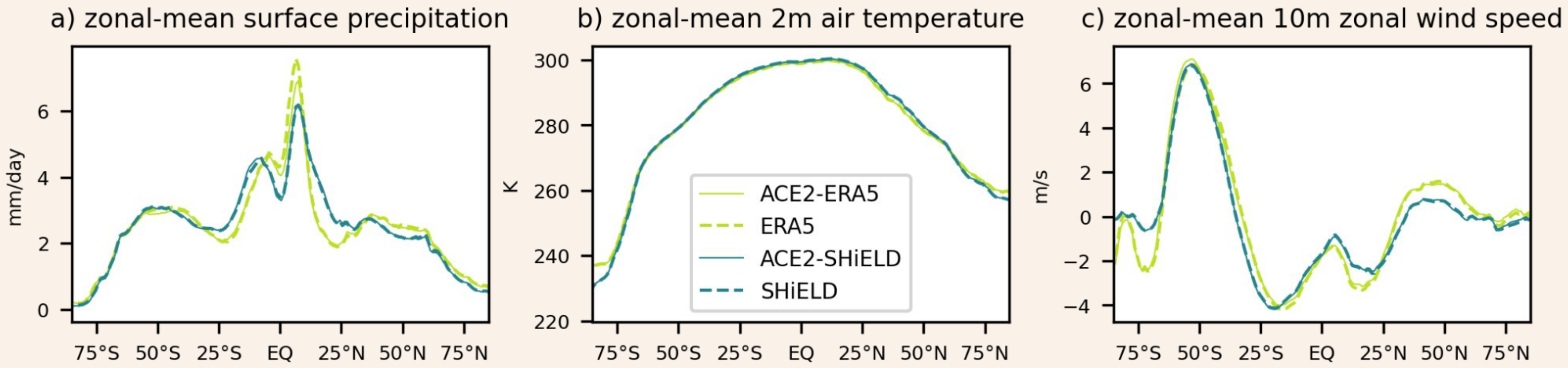
Training setup adds CO_2 as input and some new diagnostic outputs



ACE-climSST (the previous model trained only on climSST dataset) fails to get trend when forced with historical SST.

Ai2

ACE2 has low time-mean AMIP biases on ERA5/SHiELD

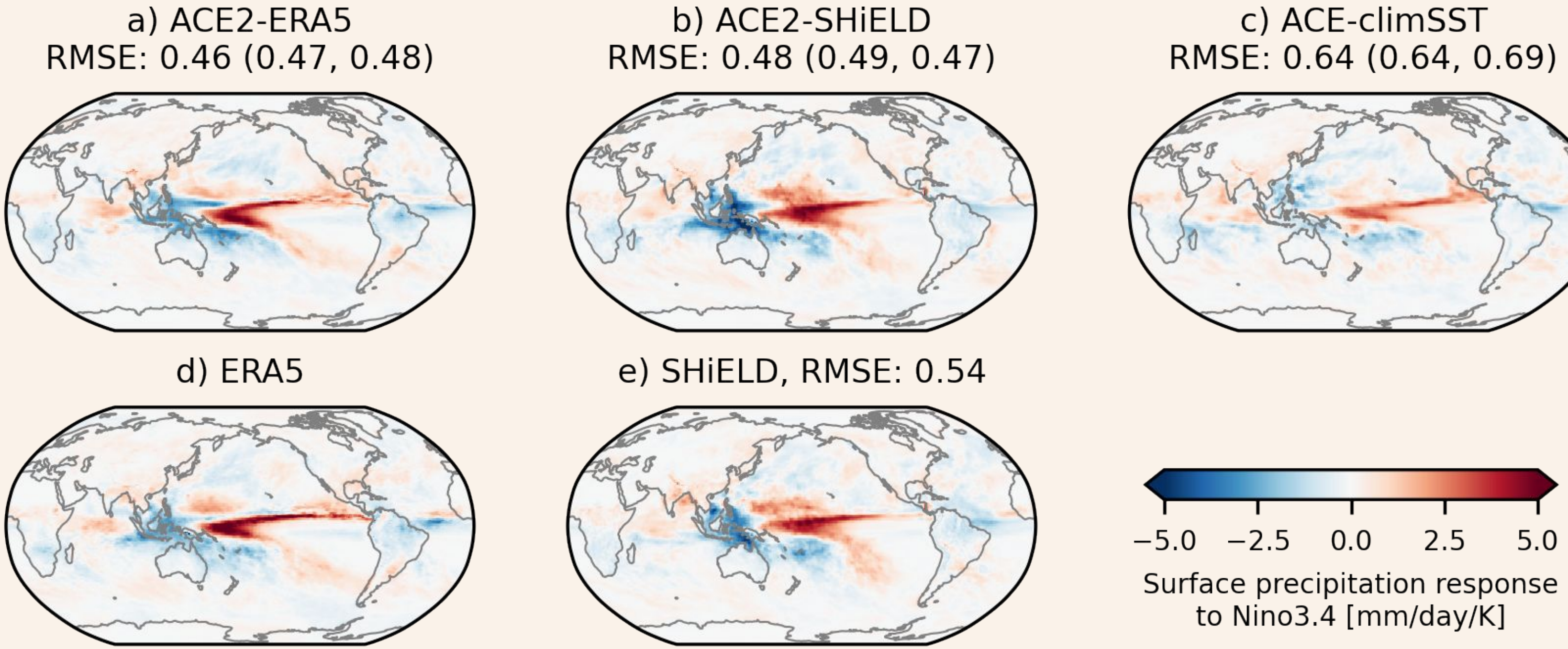


ACE2's biases relative to its target dataset are much smaller than the differences between SHiELD and ERA5.

Ai2

Held-out 2001-2010 test period

ACE2 responds accurately to El Niño SST variability



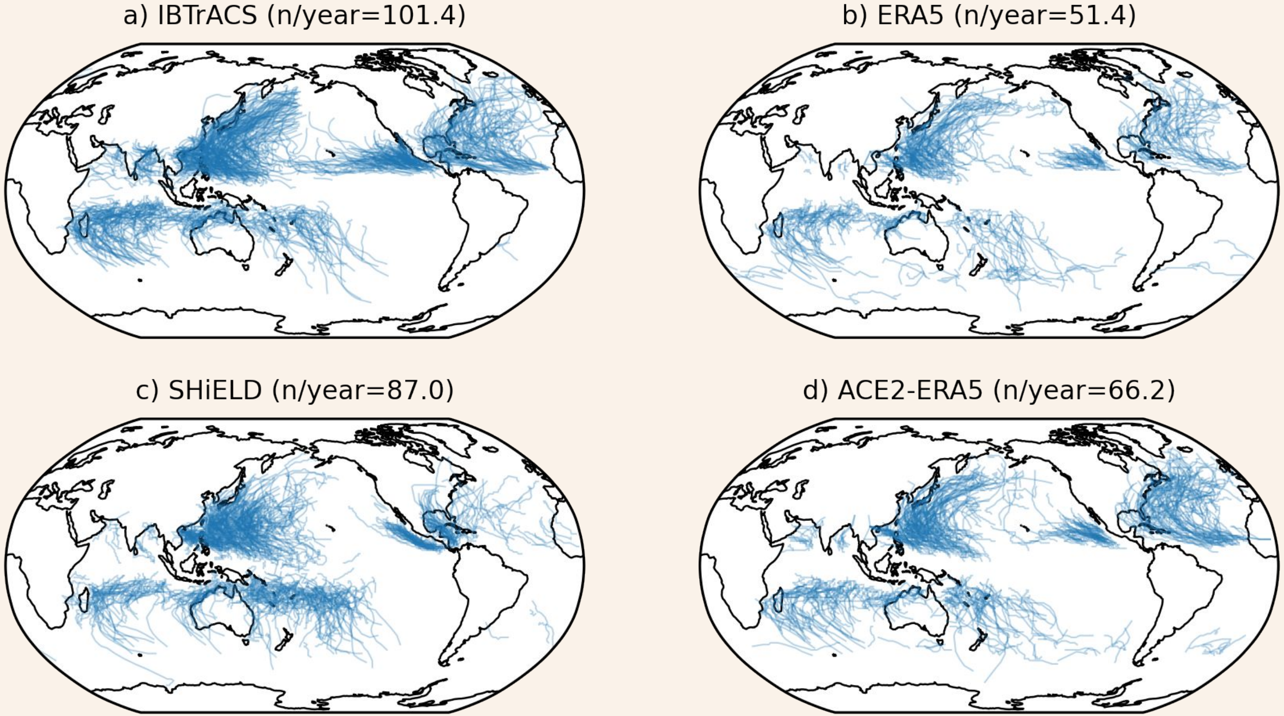
2001-2010 period

ACE2 “tropical cyclone” distribution

Caveat: for ERA5, SHiELD and ACE2-ERA5 we are using a cyclone tracking applied to 1° resolution data.

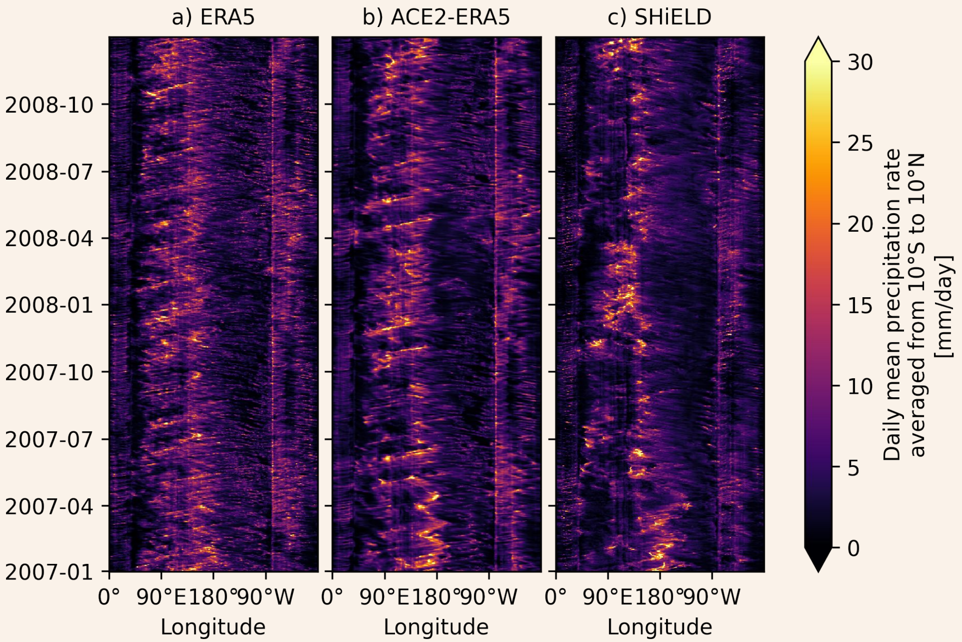
Global # of cyclones is highly tunable based on parameters used for cyclone tracking, so hard to compare directly to IBTrACS.

But differences between ERA5, SHiELD and ACE2-ERA5, as well as basin-to-basin differences, are robust to changes in tracking.

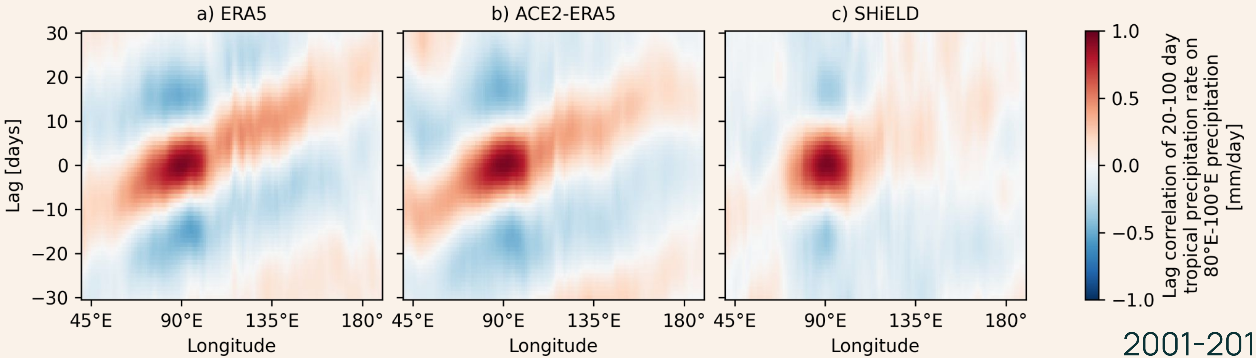


2001-2010 period

ACE2 tropical precipitation variability



Eastward propagation of MJO-like variability



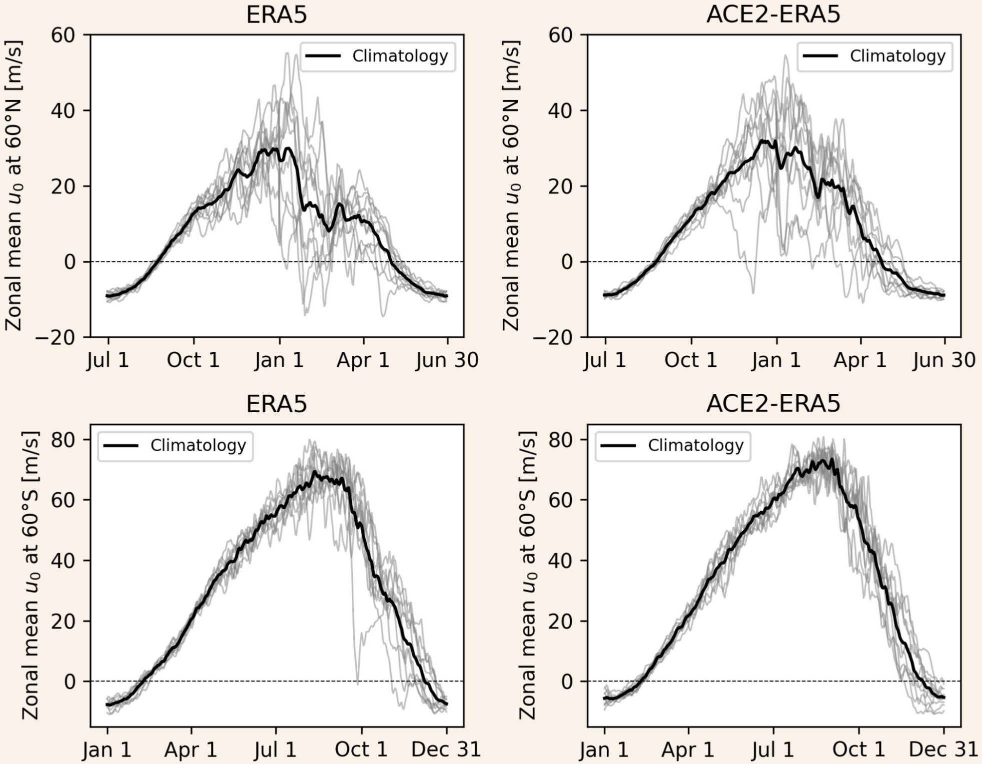
2001-2010 period

ACE2 polar stratospheric variability

Zonal mean zonal wind averaged from ~50hPa to TOA

Northern Hemisphere (60°N)

Southern Hemisphere (60°S)

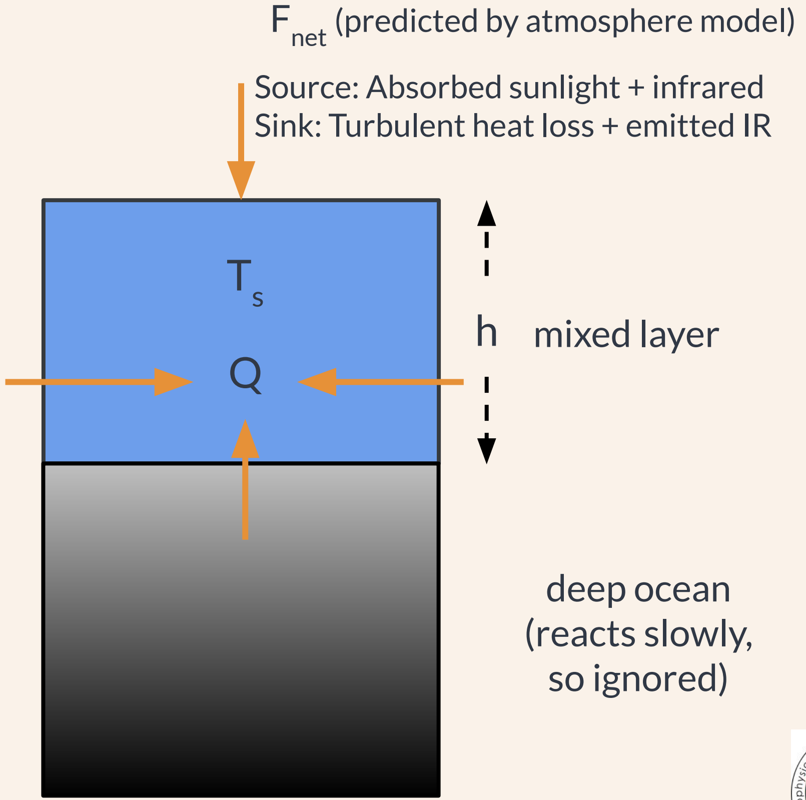


2001-2010 period

Ai2

ACE2 + interactive slab ocean (ACE2-SOM)

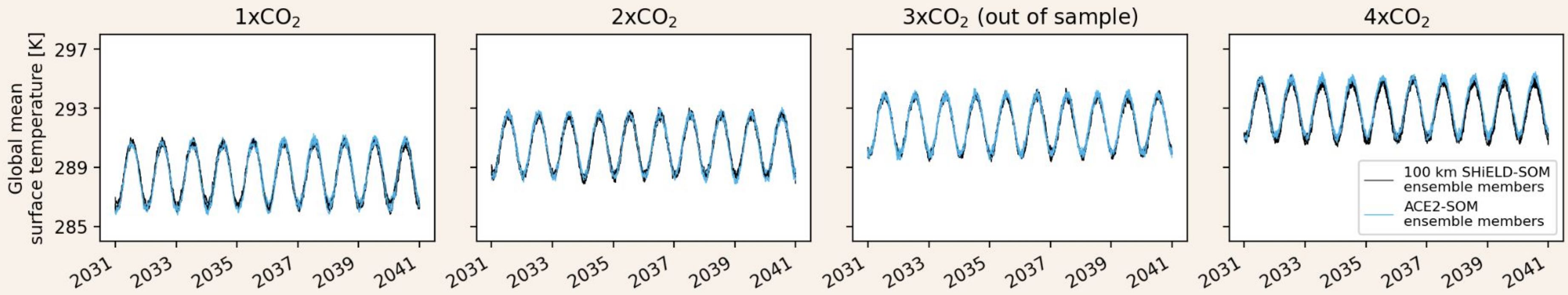
- Generate training/validation data from three 50-year simulations:
  - $1xCO_2$
  - $2xCO_2$
  - $4xCO_2$
- Train ACE-SOM with CO_2 as an input.



Clark et al. 2024, https://doi.org/10.48550/arXiv.2412.04418, submitted to *JGR-ML*

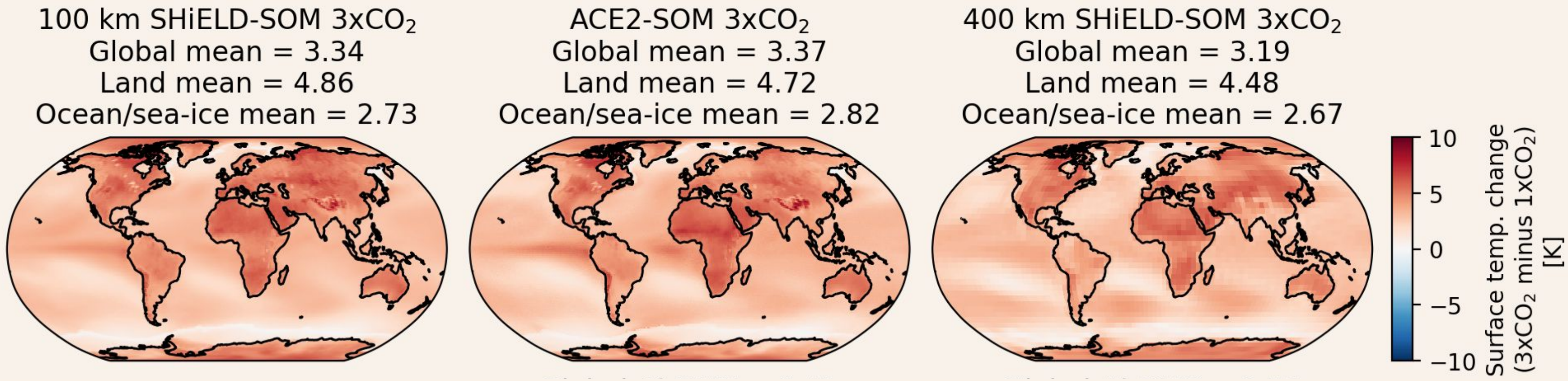
Ai2

ACE2-SOM is stable and accurate in multiple climates



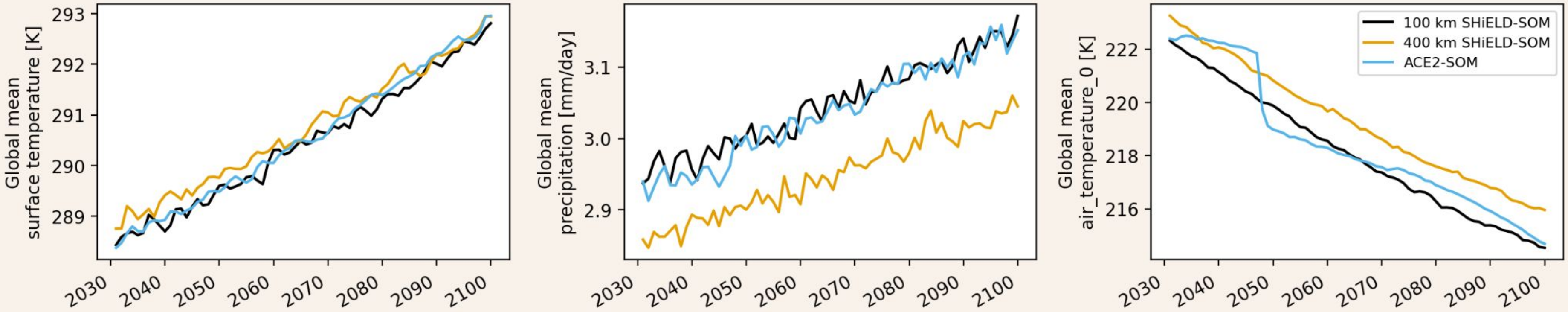
- Results are from five-member initial condition ensembles of 10-year 100 km SHiELD-SOM (target model) and ACE2-SOM simulations in each climate.
- ACE2-SOM did not see data from the 3xCO_2 climate during training.

ACE2-SOM is stable and accurate in multiple climates



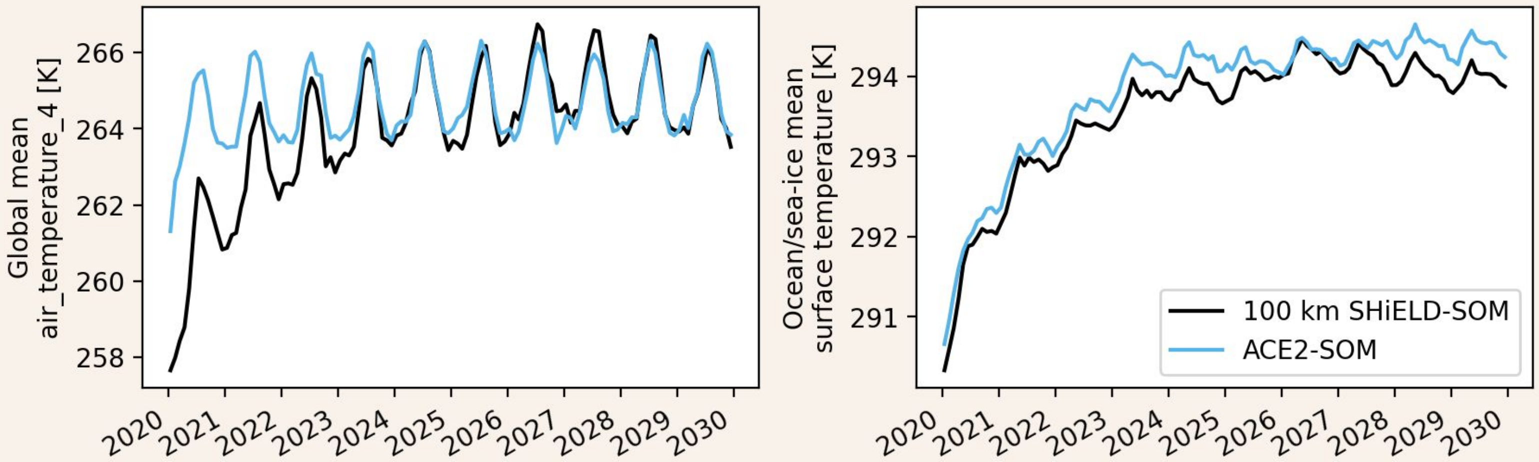
ACE2-SOM global warming pattern matches the physics-based model, capturing robust features like amplified warming over land.

Out of sample test: 2% CO_2 increase per year



- 70-year rollout ramping from 1xCO_2 to 4xCO_2 concentration
- Surface variables look good; stratosphere has a spurious climate jump
- Fixed in ACE3 by better learning CO_2 impacts on radiative heating

Out of sample test: abrupt 4xCO_2 increase



- ML-controlled fields (except ocean temperature) unrealistically jump to 4xCO_2 regime.
- Ocean temperature, aided by prescribed thermal inertia, warms realistically slowly.
- ACE2-SOM still finds the right steady state, helped by its equilibrium 4xCO_2 training.

Ai2

ACE2 speed and computational cost

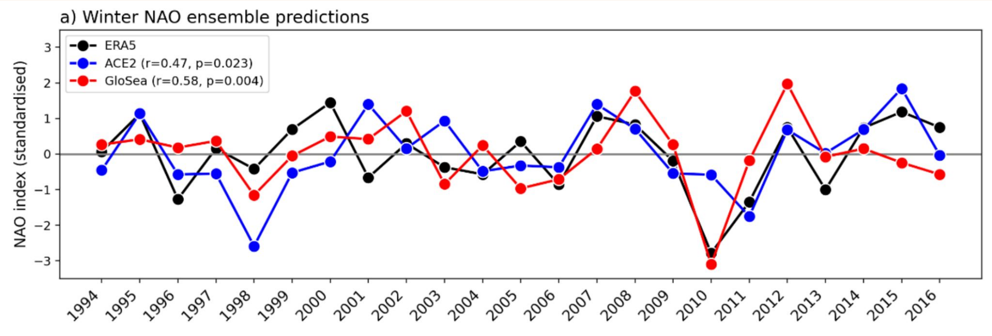
- Training time for each ACE2 model was about **4.5 days on eight 80GB NVIDIA H100s (~850 GPU-hours)**

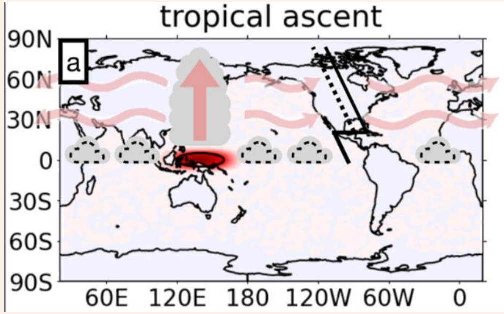
Inference throughput and energy cost

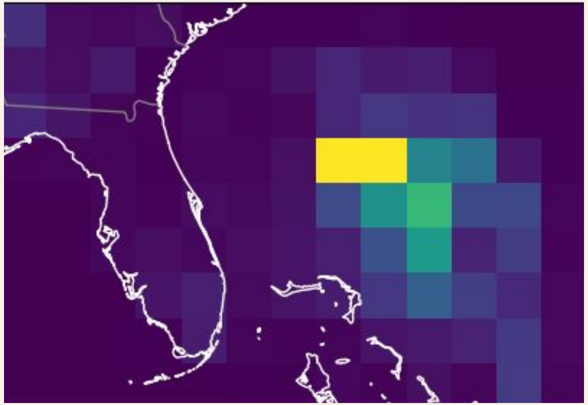
	C96 SHiELD	ACE2
Hardware	864 CPU cores (AMD EPYC 7H12)	1 80GB NVIDIA H100
Simulated years per wall clock day	~12	~1500
Energy cost per simulated year [Wh]	8250	11.2

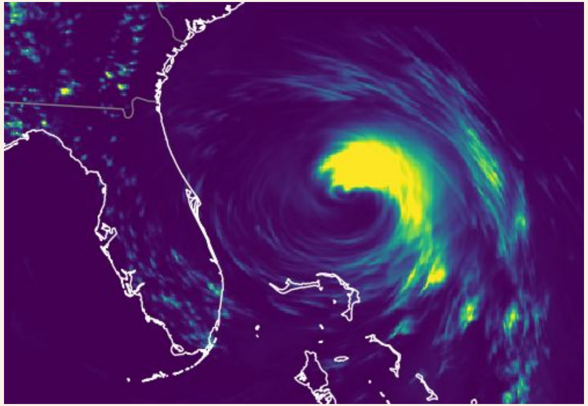
Other applications of fast climate model emulators

- Seasonal prediction
- Green's function experiments
- Local climate predictions on the fly



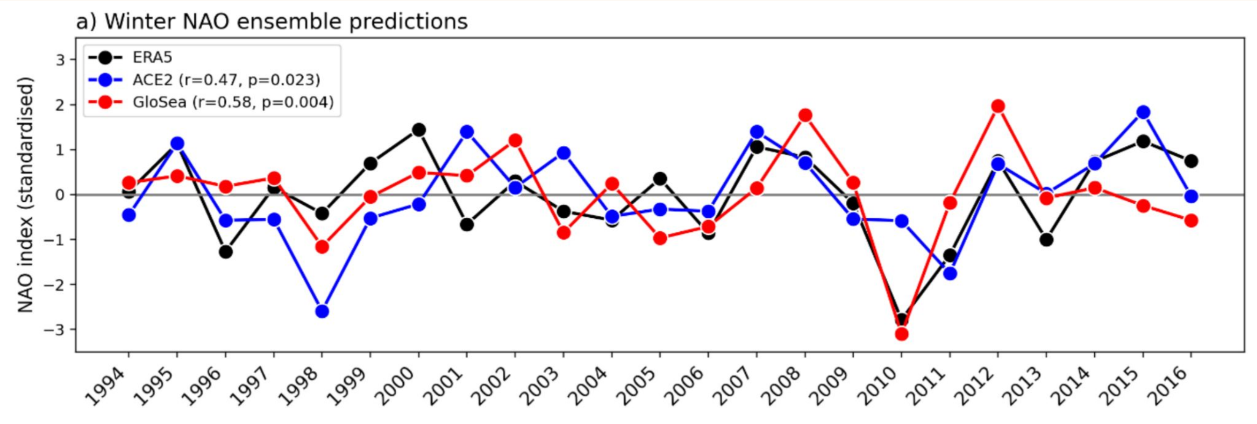






Seasonal prediction with ACE2-ERA5

- Work led by Chris Kent, Adam Scaife et al. at Met Office using our publicly available ACE2-ERA5 model
- Use persistence SSTs (since ACE doesn't have an ocean model yet...) and initialize ensemble forecasts on November 1 to predict DJF averages
- Skill in many respects is approaching that of Met Office's GloSea prediction model



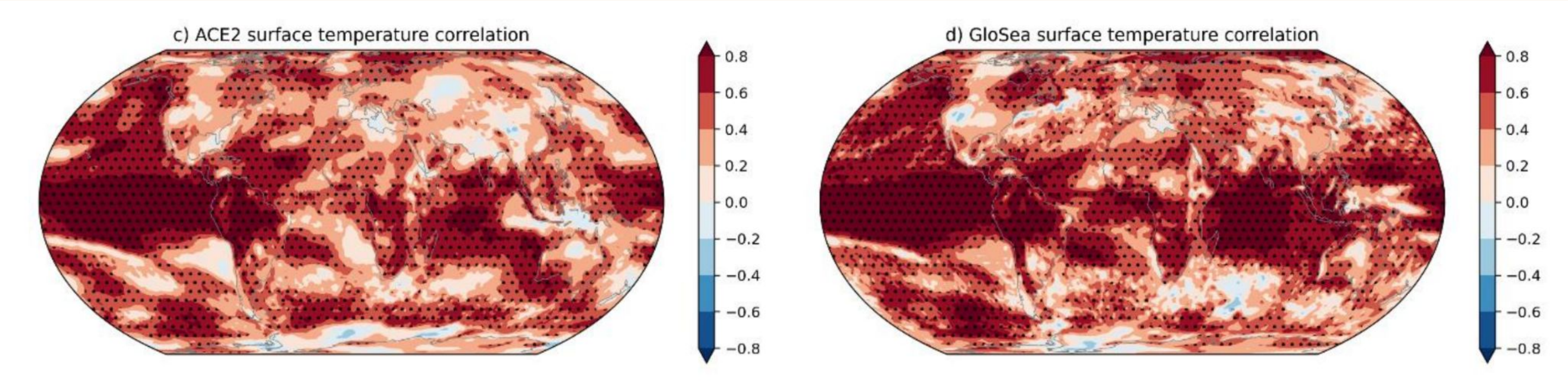
Caveat: only 2001-2010 is outside the training period of ACE!

Ai2

Kent et al. 2025 (arxiv)

Seasonal prediction with ACE2-ERA5

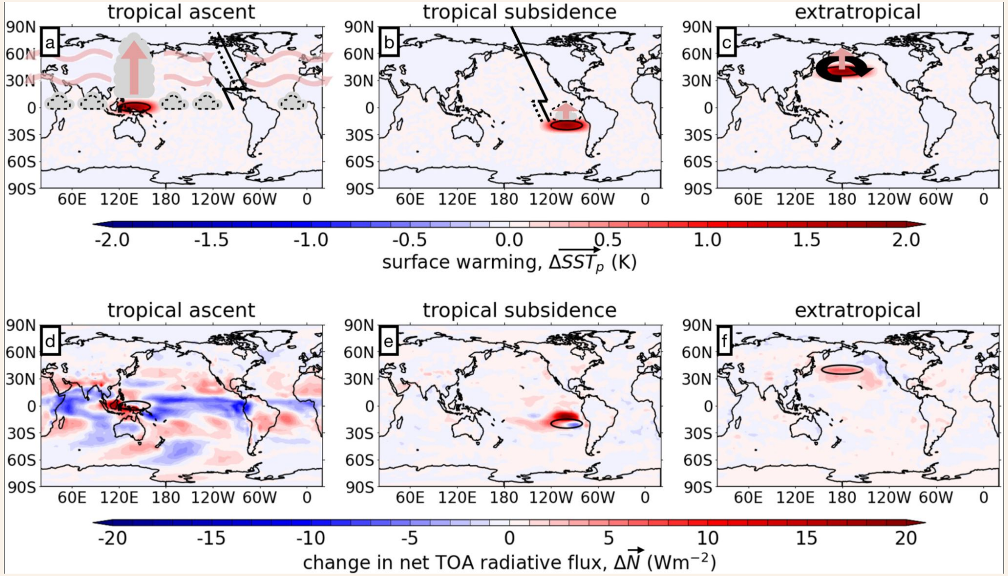
- Regions of globe with predictive skill are mostly similar between ACE2 and GloSea



Kent et al. 2025 (arxiv)

Assessing importance of SST warming pattern

- "Green's function" experiments have become popular to do with AGCMs
- Involves hundreds of multi-decade simulations, each perturbing SST in a small region of the globe and assessing TOA radiative response (Bloch-Johnson et al. 2024)
- **2000 to 4000 years of simulation required in total!**
- Can the massive cost-savings of AI emulators help us here?



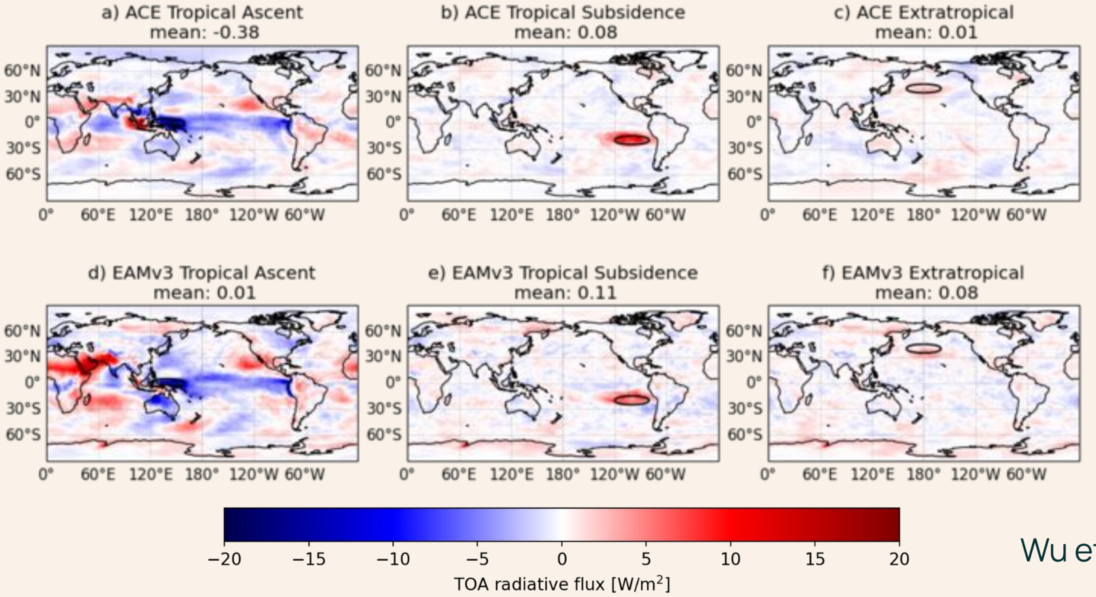
Bloch-Johnson et al. 2024

Assessing importance of SST warming pattern

- Strategy: train ACE on an AMIP simulation with DOE's EAMv3, then do full suite of Green's function protocol experiments and compare **(not training directly on Green's function experiments!)**
- ACE generally captures the expected responses but details can differ in some regions

ACE

EAMv3



See also Van Loon et al. 2025 which applies Green's function methodology to published ACE models including ACE2-ERA5.

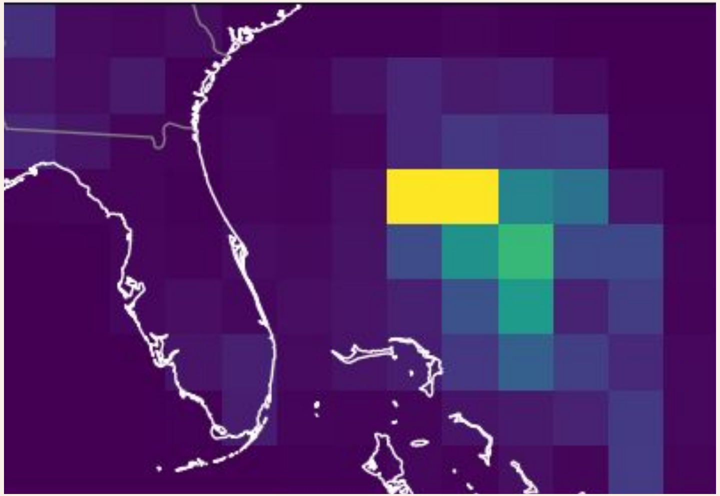
Wu et al. in prep

Local climate data on the fly

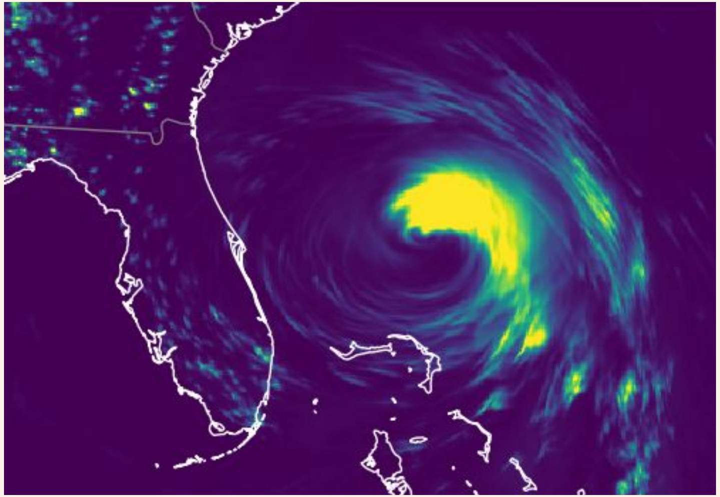
- AI climate emulators will make it far cheaper and easier to:

  1) do simulations across a wide range of future scenarios
  2) de-bias and downscale on the fly for regions/periods of interest

Example 6hr rainfall prediction at 100km



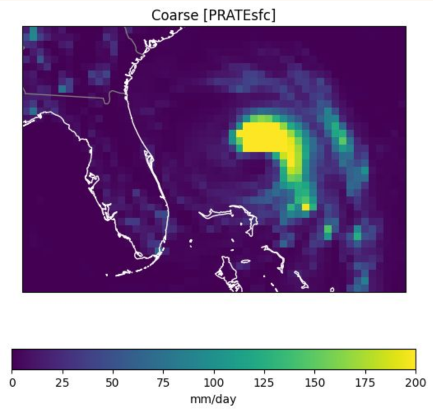
Example 6hr rainfall prediction at 3km

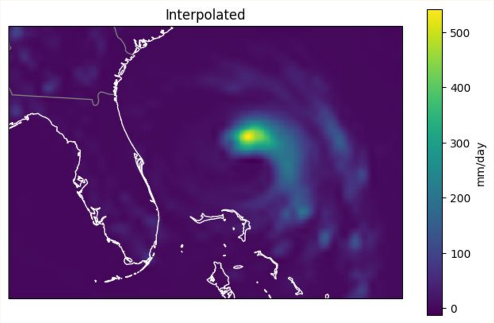


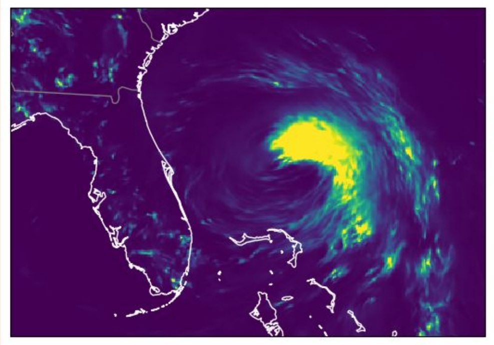
Ai2

Local climate data on the fly

- We are developing a downscaling module that can be plugged into ACE
    - Will be trained on output from a global 3-km simulation with GFDL's X-SHiELD model







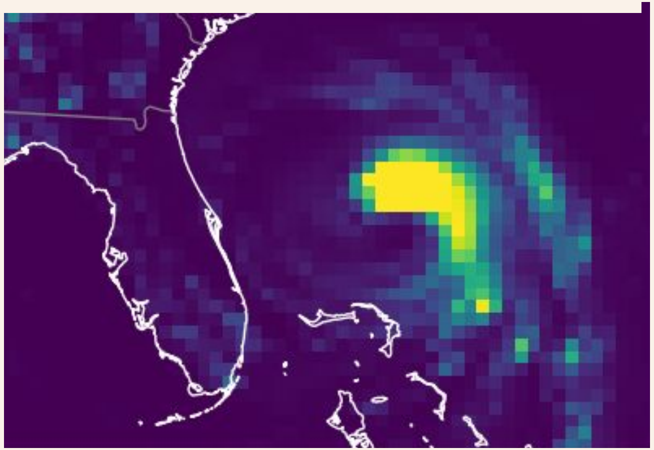
EDM Diffusion
(e.g., Karras et al.)
UNet Architecture

Applying NVIDIA's "Residual Corrective Diffusion Modeling" approach (arXiv)

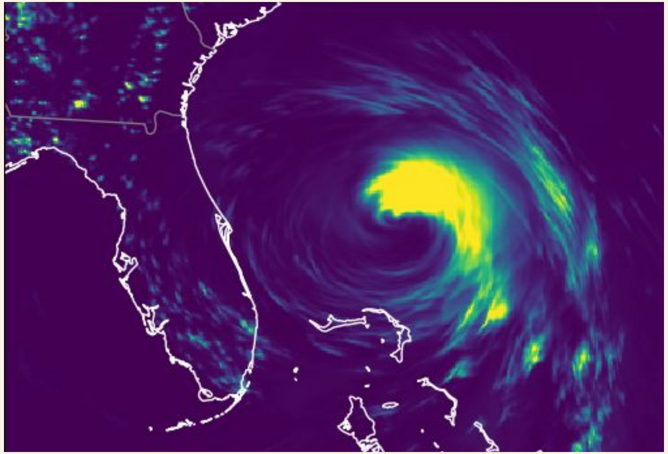
Initial downscaling results

8x downscaling (25→3km) of tropical cyclone in NW Atlantic with diffusion modeling

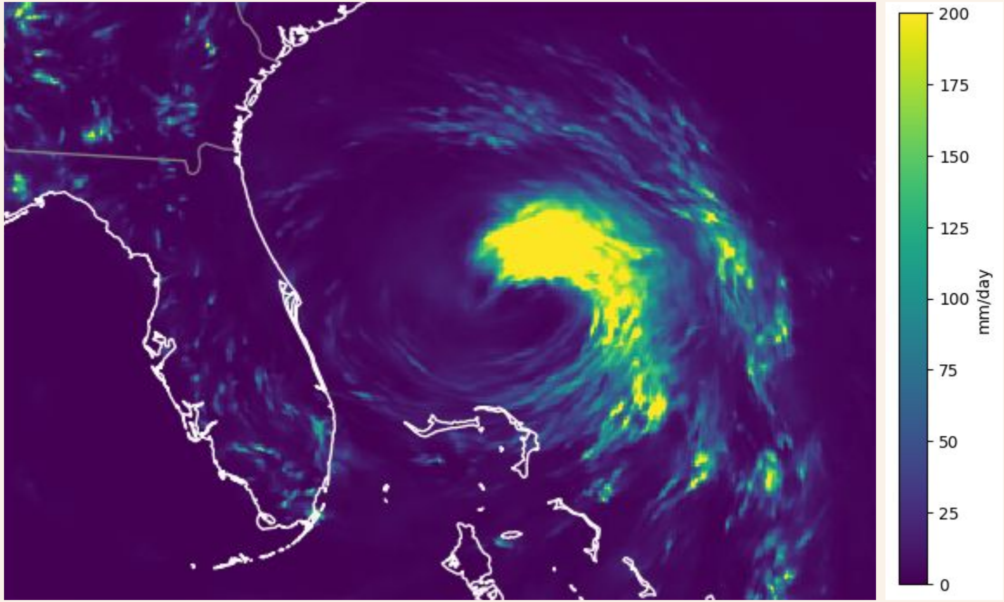
25km Coarse Input



Target 3km X-SHiELD



Generated 3km output



Summary

ACE

- The Ai2 Climate Emulator (ACE) is a stable, accurate ML-based atmospheric model that is nearing suitability for climate prediction
- Faithfully reproduces reference model's climate
  - Skillful in climSST, AMIP, and slab-ocean configurations
  - Captures forced response to greenhouse gas and SST perturbations (ENSO, Green's functions)
- Easy to use and 100-1000x faster to run than the reference model!
- Ongoing work
  - Energy conservation & accurate learning of CO_2 effects on radiative heating
  - Coupling to an emulator of a dynamical ocean model
  - Downscaling module for assessing local impacts

Open-source code, data and model checkpoints: github.com/ai2cm/ace

ACE: Watt-Meyer et al. 2023; Duncan et al. 2024; ACE2: Watt-Meyer et al. 2024; ACE2-SOM: Clark et al. 2024