ACE: a fast, skillful learned global atmospheric model for climate prediction

Oliver Watt-Meyer, Lead Research Scientist, Climate Modeling

WGNE39-WGSIP 25 Annual Meeting

November 7, 2024

Ai2

How can ML help weather/climate models?

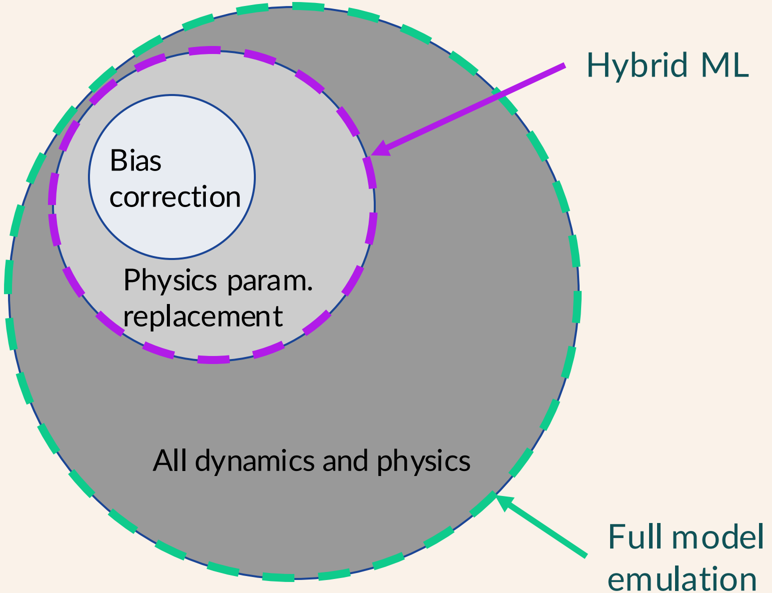
Make global weather and climate models:

- more accurate
- faster
- more affordable/accessible

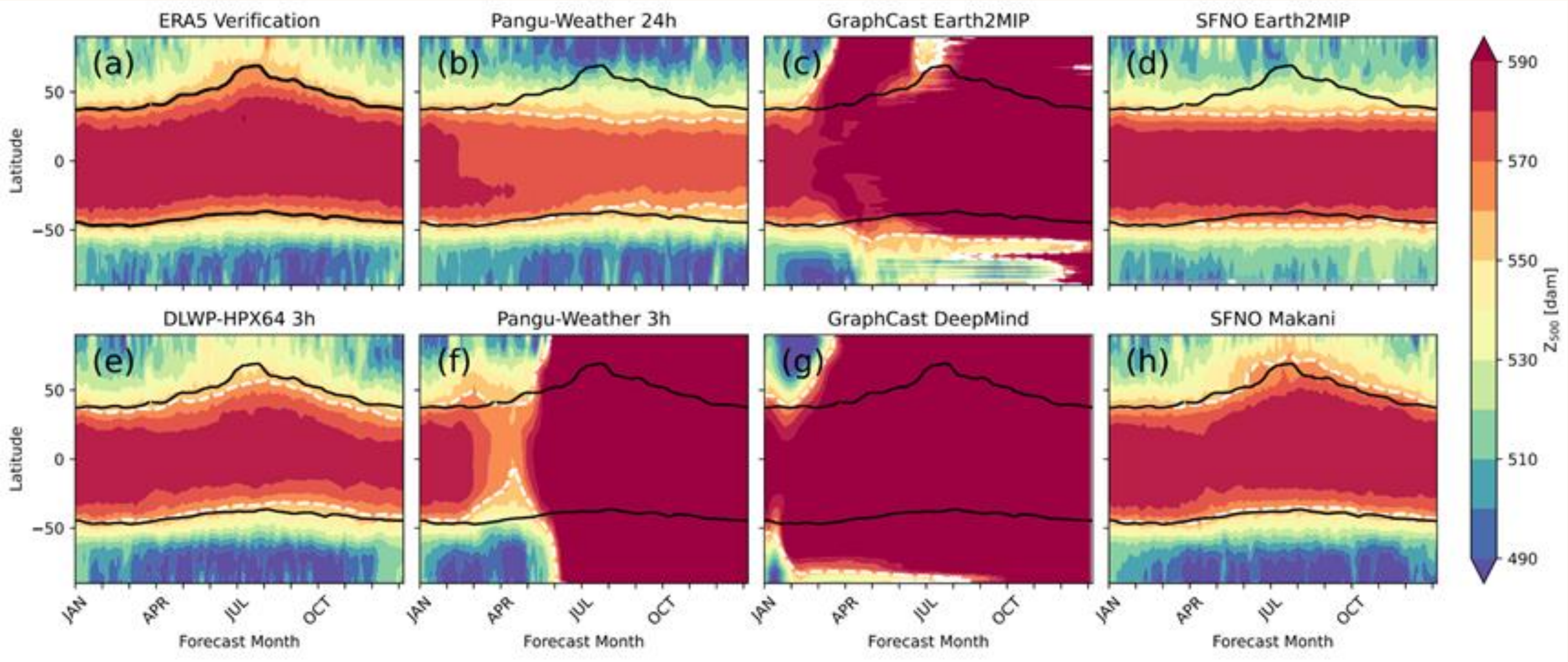
Three general strategies (atmospheric focus):

- Hybrid: ML replaces or corrects parts of the atmospheric model
- ML for post-processing bias correction and/or downscaling
- **Full model emulation: machine learning of entire global atmospheric evolution**

What is the ML responsible for?



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



Karlbauer et al. 2024; https://arxiv.org/abs/2311.06253

Our approach: ACE (Ai2 Climate Emulator)

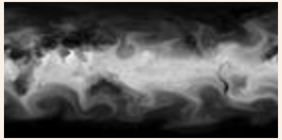
- Train on climate model output, in addition to ERA5
  - Want to use data from diverse range of climates (global warming etc.)
- Start simple… initially we used climatological SST and fixed external forcing
  - For now, use relatively coarse 1° resolution
  - I will also show results with varying external forcings (historical and increased CO_2)
- Provide boundary conditions (insolation, SSTs) as inputs to ML model
- **Focus on long-term stability**

Training Setup

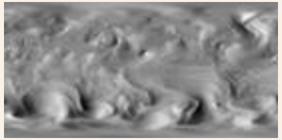
Loss function: mean-squared error of 6-hour forecast computed over all the output variables (for ACE2, accumulated over 2 forward steps)

Input at time *t*

Humidity

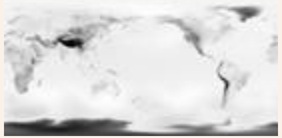


Wind



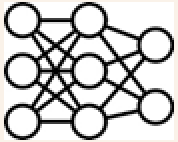
Temperature

Surface pressure



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

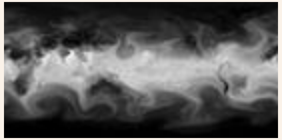
SFNO*



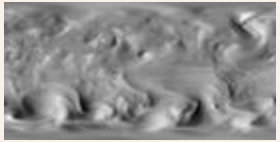
$$f(\mathbf{u};\theta)$$

Prediction of *t* + 6 hours

Humidity

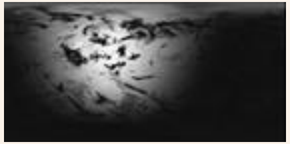


Wind



Temperature

Down SW sfc rad flux



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

Showing subset of inputs/outputs

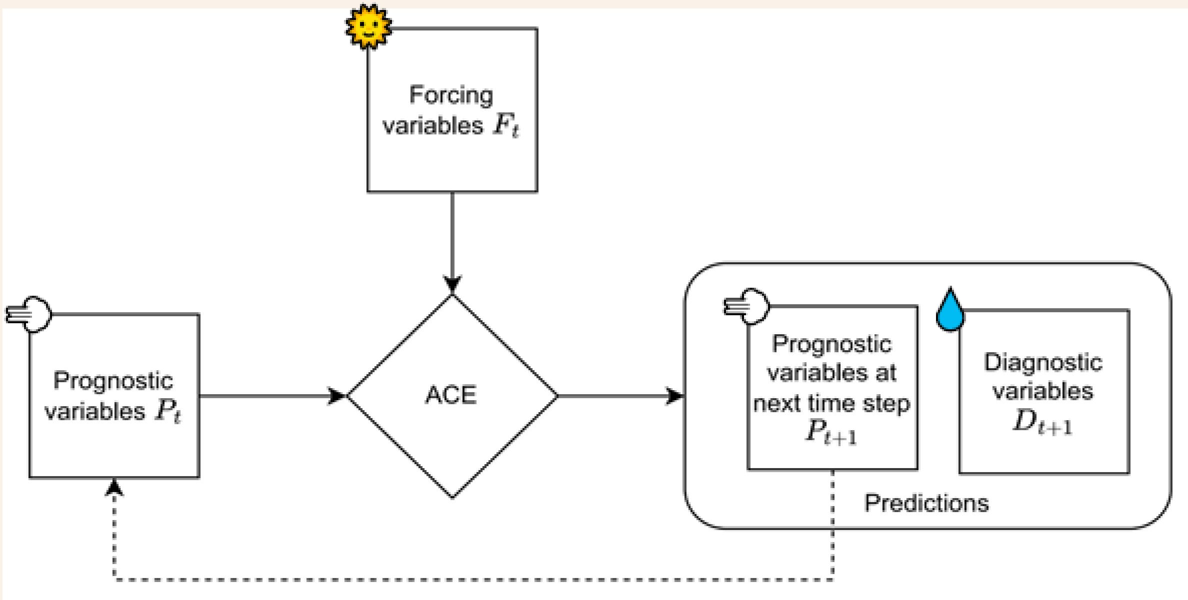
***SFNO**: Spherical Fourier Neural Operator from Bonev et al., 2023 arxiv.org/abs/2306.03838

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

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



Training data

Generate training data by saving output from FV3GFS or SHiELD model runs at C96 (approx 1°) resolution.

We coarsen to **8 terrain-following (hybrid sigma-pressure) finite volume** layers on a 1° Gaussian grid for compatibility with SFNO, better speed, and reduced storage

- Mass and moisture are conserved in vertical coarsening, **allowing calculation of exact budgets.**

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

1. ACE2: Historical SSTs (AMIP)
   - ERA5 or an SHiELD AMIP Ensemble

2. ACE2: Slab ocean with CO_2 forcing (ongoing work)
   - Present-day CO_2, $2xCO_2$ and $4xCO_2$, with a simple slab ocean

ACE (v1) Results

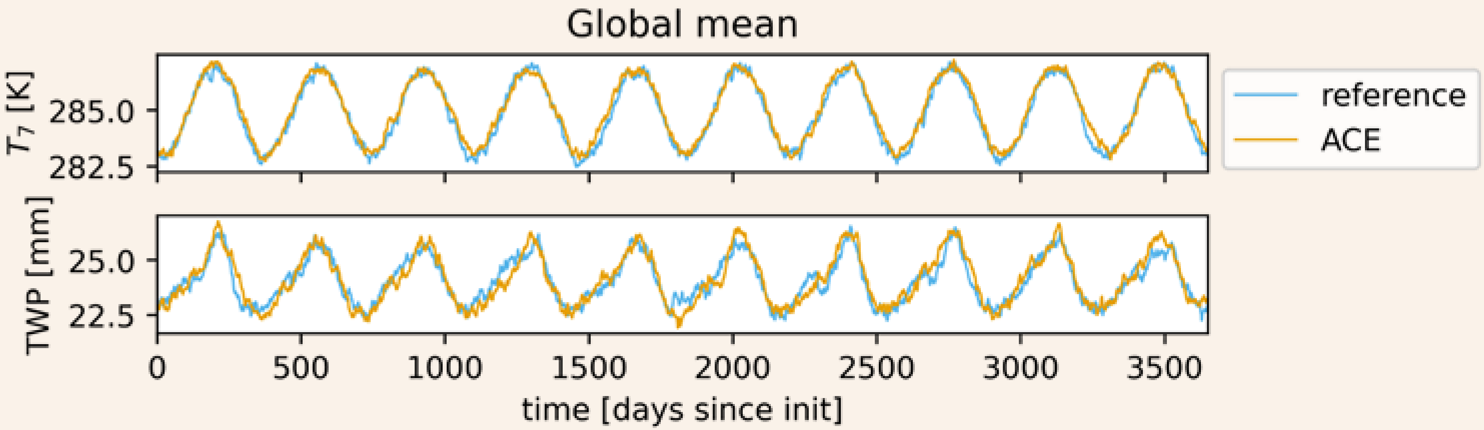
Presented at NeurIPS Tackling Climate Change with AI workshop 2023.
Paper on arxiv: Watt-Meyer et al. (2023)

Ai2

ACE is stable, with an accurate seasonal cycle!

Temperature
(boundary
layer average)

Total Water Path
(vertically integrated
moisture)



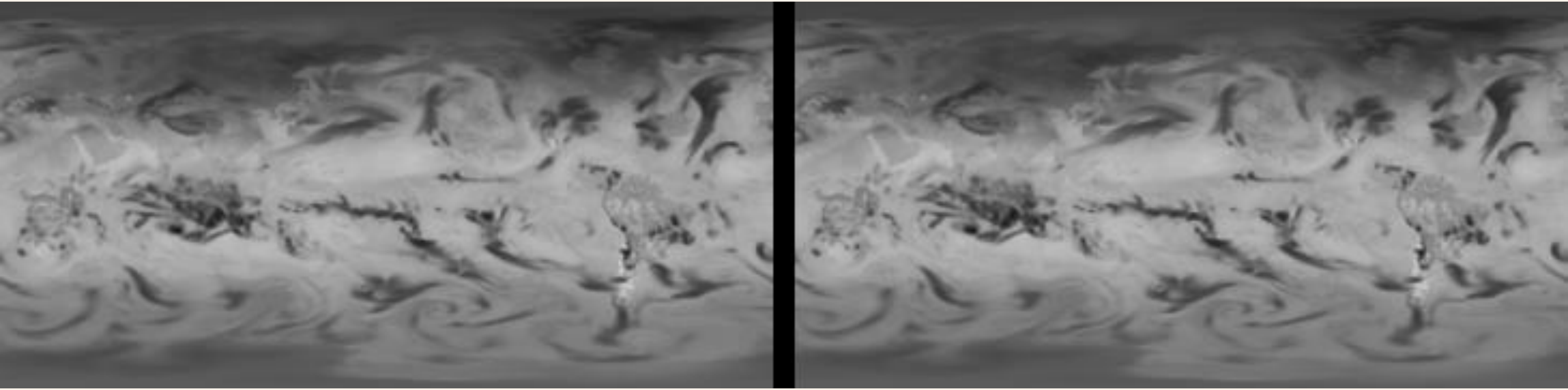
As far as we can tell, indefinitely stable (will show 1000 year simulations later)

Watt-Meyer et al. (2023)

Realistic weather variability

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

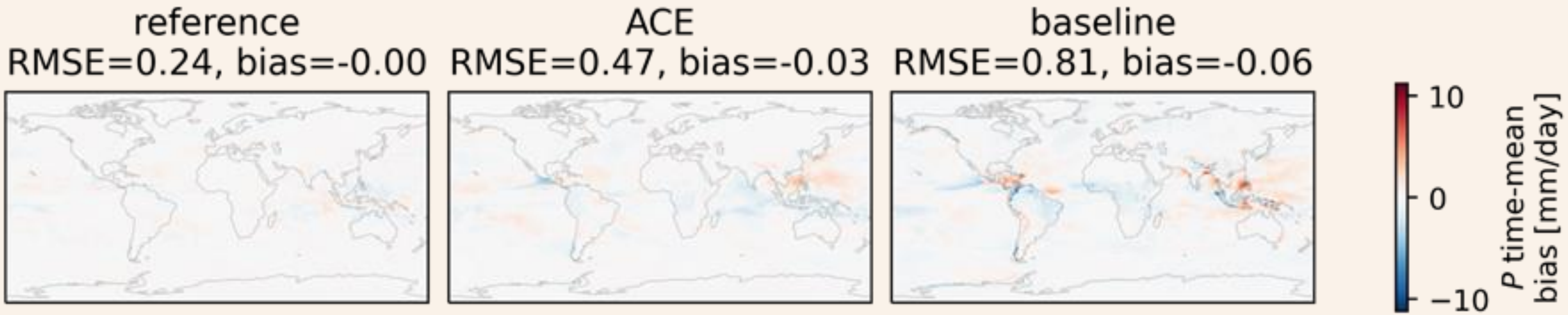
ACE (prediction) FV3GFS (target)



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

Excellent climate accuracy

10 year time-averaged precipitation rate bias



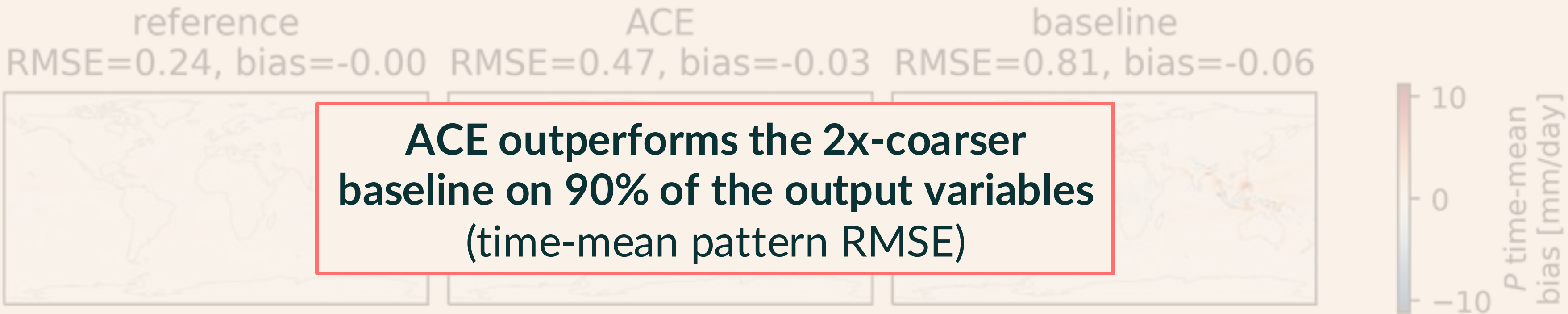
Natural variability of the system (noise floor)

2x coarser physics-based model

Watt-Meyer et al. (2023)

Excellent climate accuracy

10 year time-averaged precipitation rate bias



ACE outperforms the 2x-coarser baseline on 90% of the output variables
(time-mean pattern RMSE)

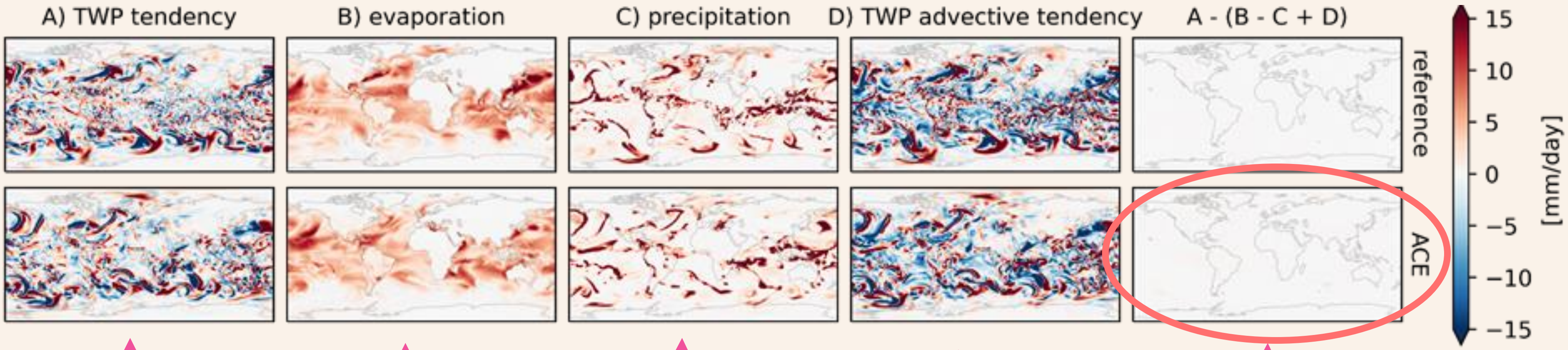
Natural variability of the system (noise floor)

2x coarser physics-based model

Watt-Meyer et al. (2023)

Physical consistency

$$\frac{\partial \mathrm{TWP}}{\partial t} = E - P + \left.\frac{\partial \mathrm{TWP}}{\partial t}\right|_{\mathrm{adv}}$$



Change in Total Water Path

Surface evaporation rate

Surface precipitation rate

Horizontal moisture transport

Balance of all these terms
Should be zero!

$$\mathrm{TWP} = \frac{1}{g}\sum_k q_k \Delta p_k$$

Plot is showing a snapshot 1 year into simulation

Watt-Meyer et al. (2023)

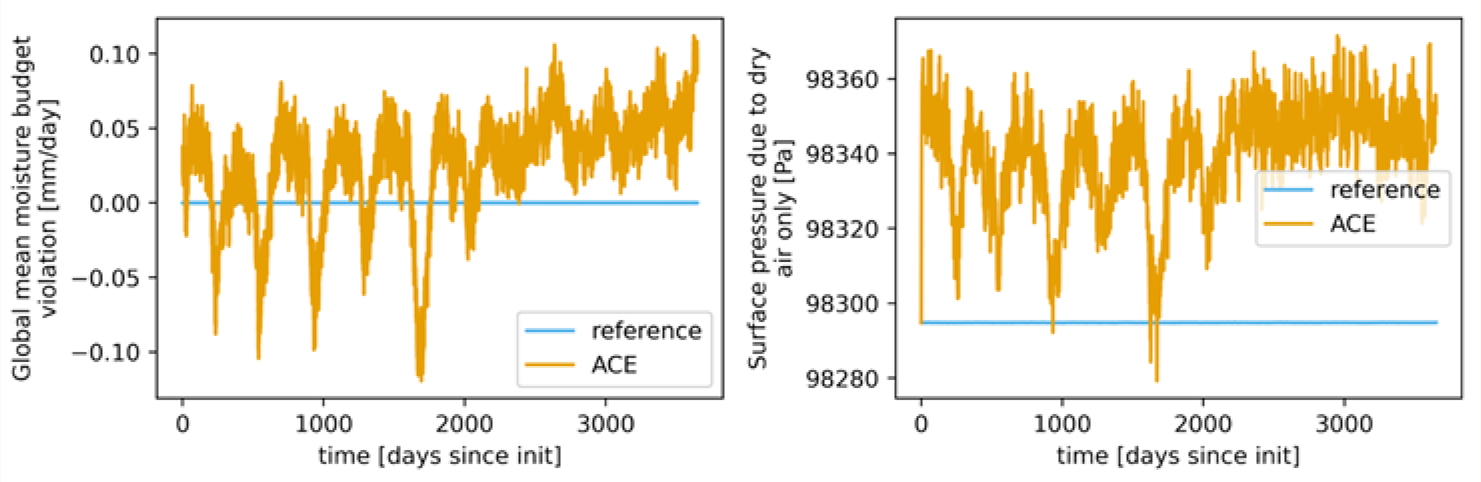
Ai2

Physical consistency — global mean

- Despite the column-budget of moisture being very nearly closed, when we compute the global mean there are non-trivial budget violations

Violation of global mean moisture budget

Global mean surface pressure **due to dry air only**

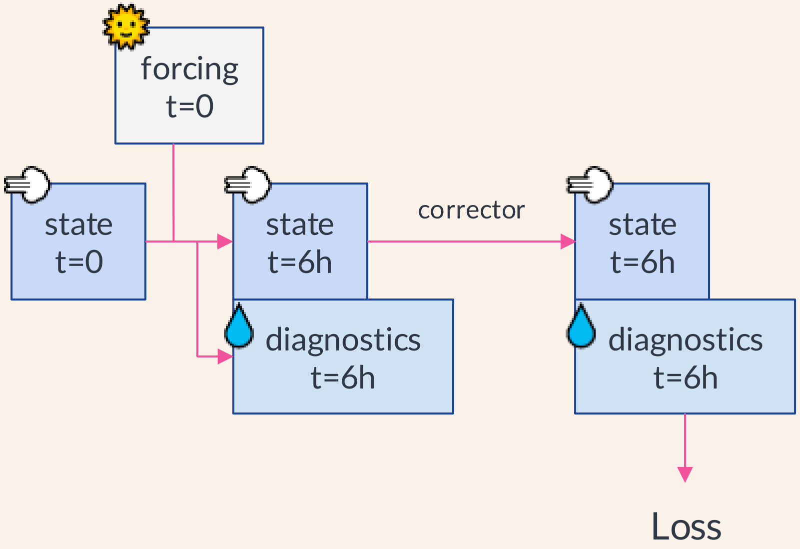


Watt-Meyer et al. (2023)

Enforcing hard budget constraints

Post-step correctors applied to output, during training and inference:

- Mass budget closed by adding a time-dependent spatially-constant correction to surface pressure (very small: ~5 Pa)



Constraint: $\overline{p_s^{dry}(t + \Delta t)} = \overline{p_s^{dry}(t)}$

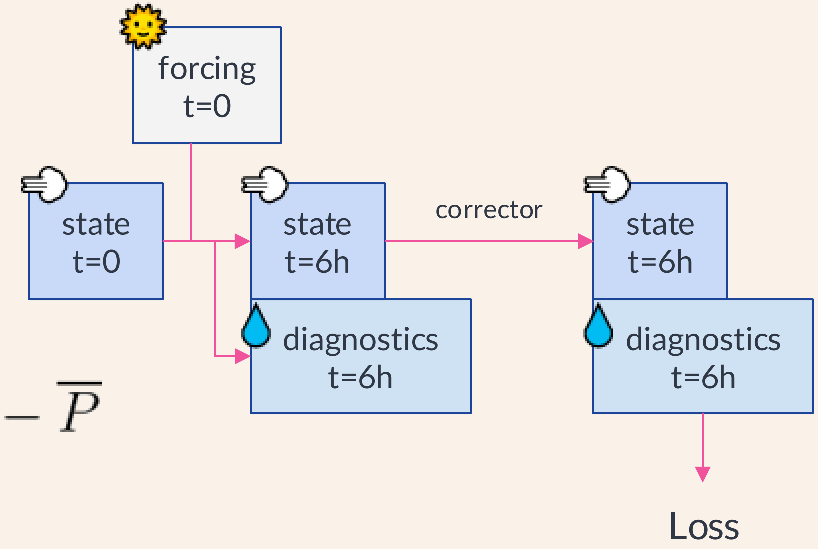
$$p_s^{dry} = p_s - \sum_k q_k \Delta p_k$$

Overline: global mean

Enforcing hard budget constraints

Post-step correctors applied to output, during training and inference:

- Global water budget closed by multiplying precipitation by a constant

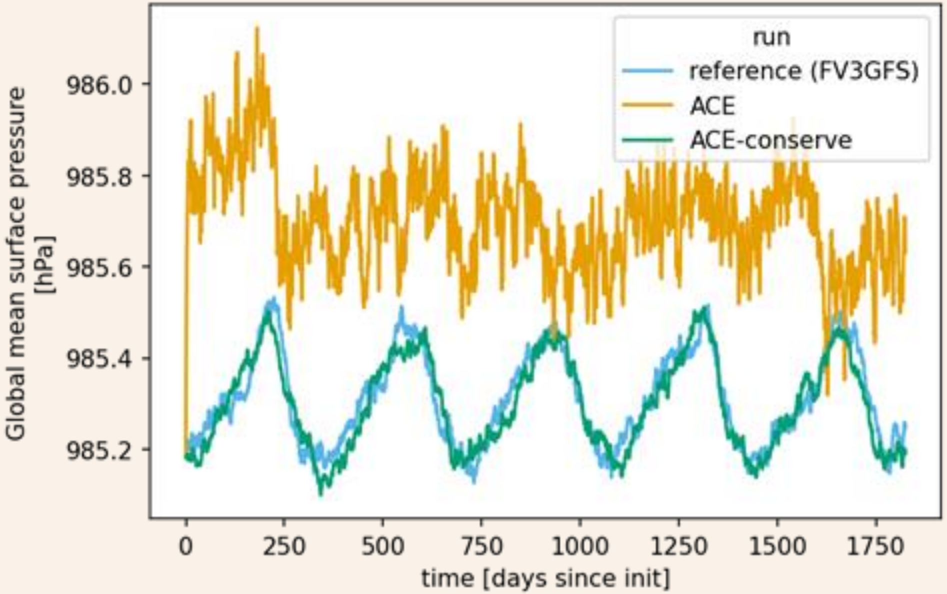


Constraint: $$\frac{\overline{\mathrm{TWP}(t+\Delta t)} - \overline{\mathrm{TWP}(t)}}{\Delta t} = \overline{E} - \overline{P}$$

Overline: global mean

Physical consistency

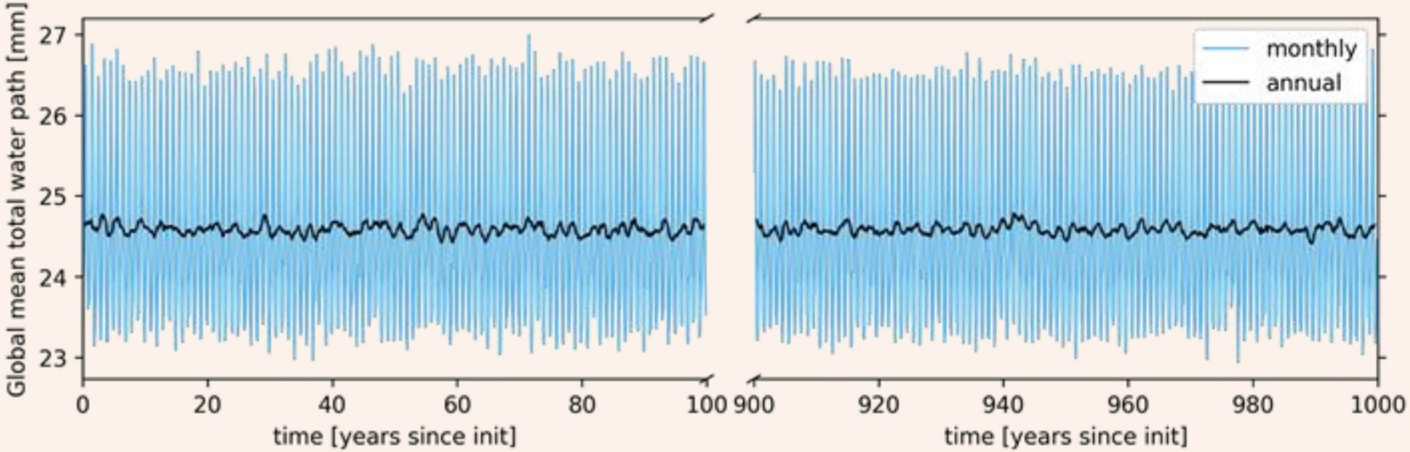
- These constraints (in particular dry air mass constraint!) reduce surface pressure drift



Ai2

Why is ACE stable?

- Architecture made a difference
  - SFNO better than the original 2D FFT based FourCastNet
- “Residual” normalization
  - Account for the different timescales of different variables
- Choice of variables, especially forcing inputs
  - Not surprisingly, insolation is key for seasonal cycle
  - Somewhat more surprisingly, constant topography input also helped



Ai2

ACE2 AMIP-based Emulators

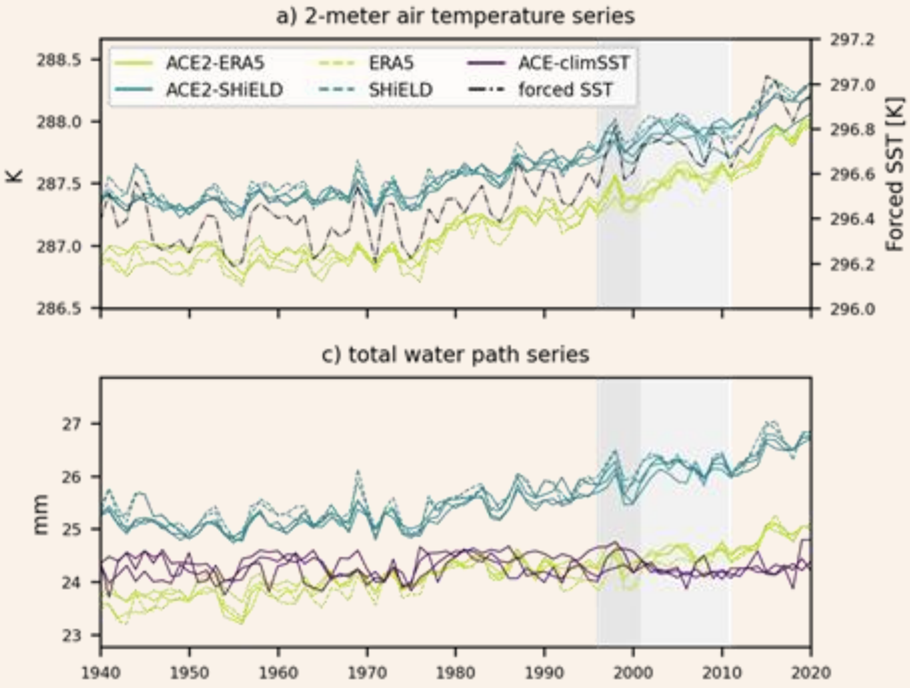
Submitting soon… preprint will be on arxiv in next 2 weeks

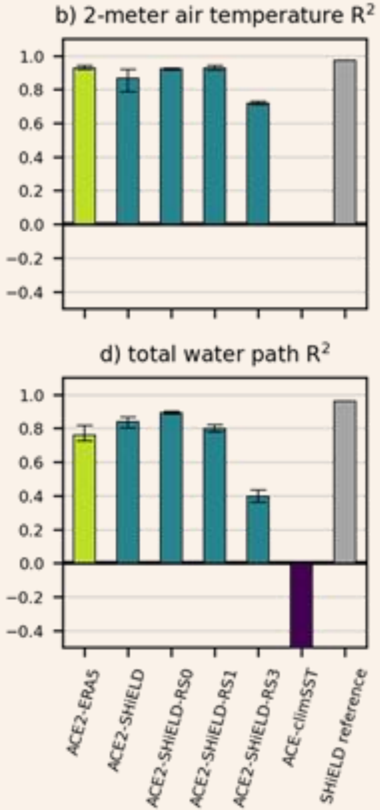
Moving forward: realistic historical SSTs

Now switching two new training dataset:

1. ERA5 over 1940-2020
2. C96 SHiELD AMIP simulation (two realizations) also 1940-2020

Training setup is similar to what was previously described, but now add CO_2 as input variable and also include some new output variables

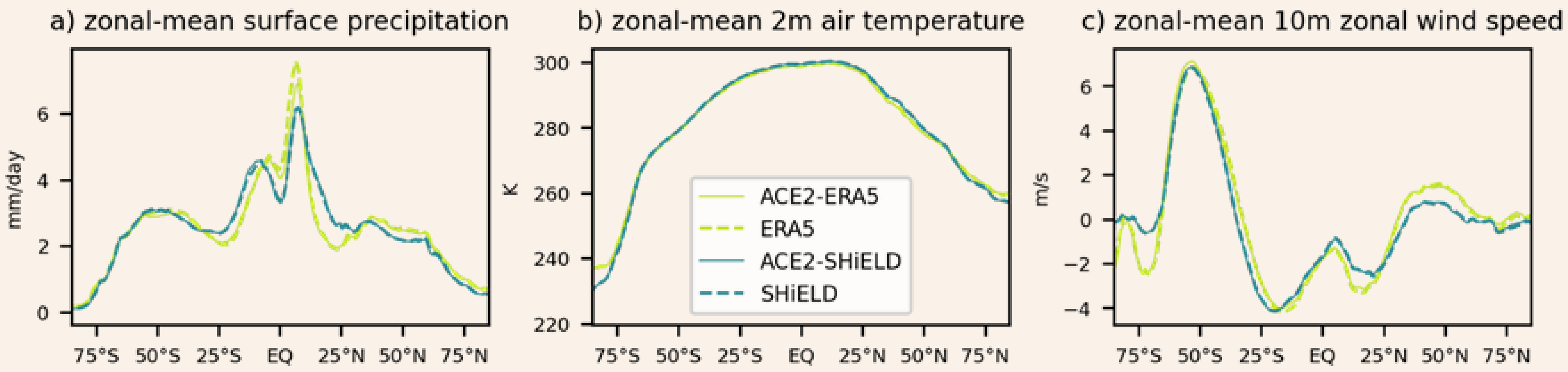




ACE-climSST is the model discussed in previous sections (trained on climSST dataset). Fails to get trend when forced with historical SST.

Ai2

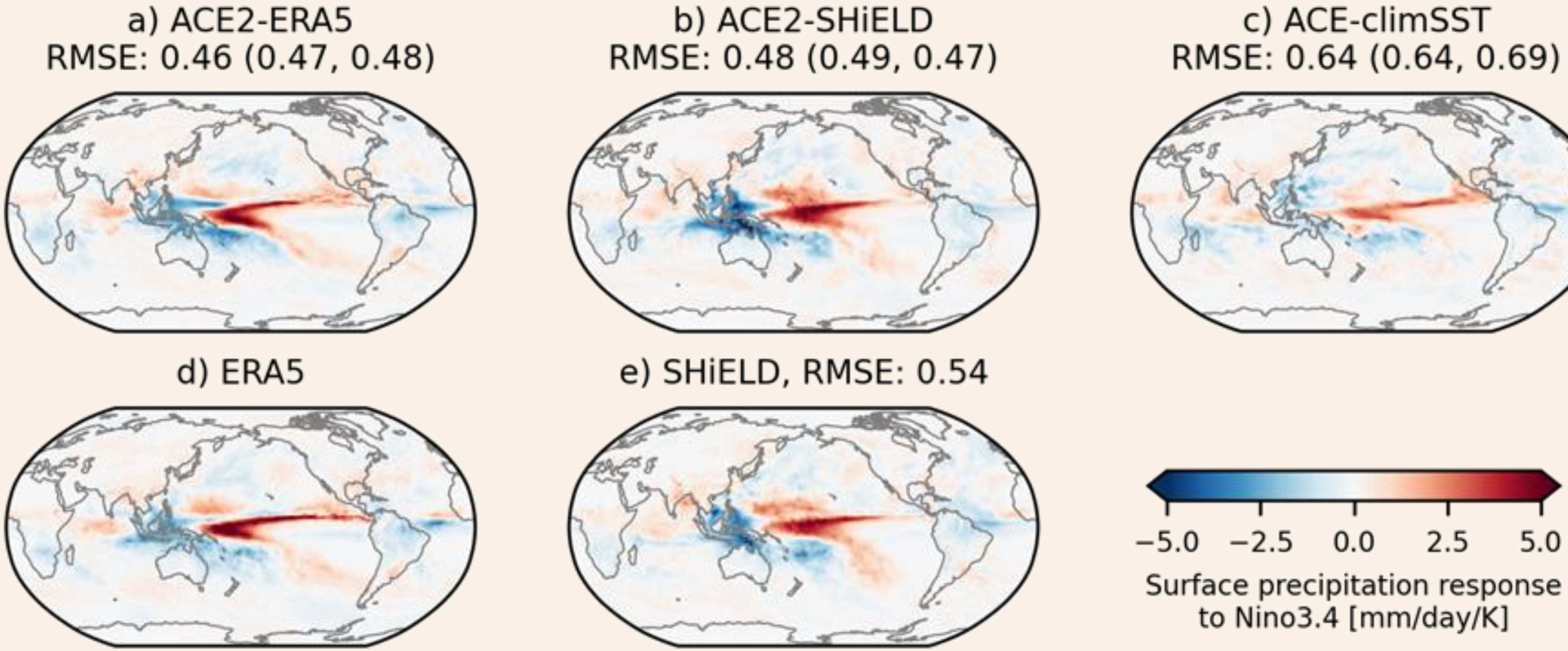
ACE2 has low time-mean AMIP biases on ERA5 and SHiELD



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

Based on 2001-2010 period

Response to El Niño SST variability very accurate



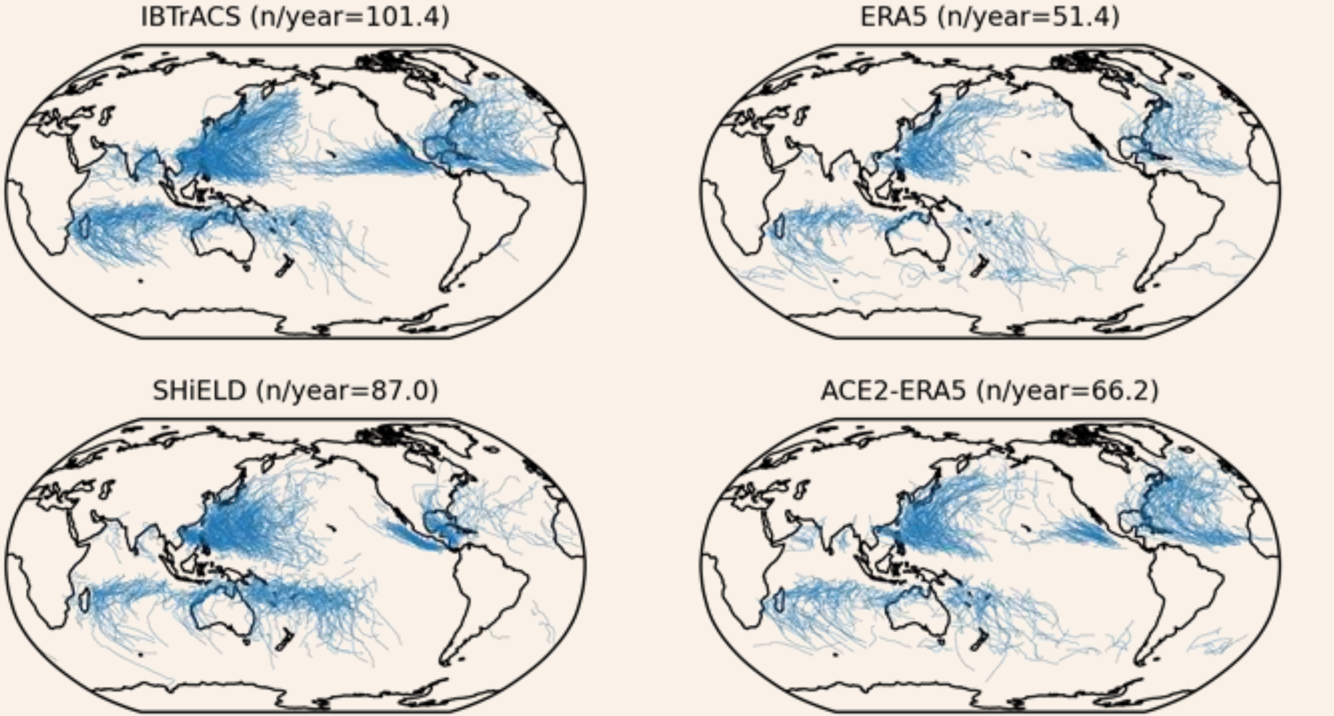
Based on 2001-2010 period

Climate skill - 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.

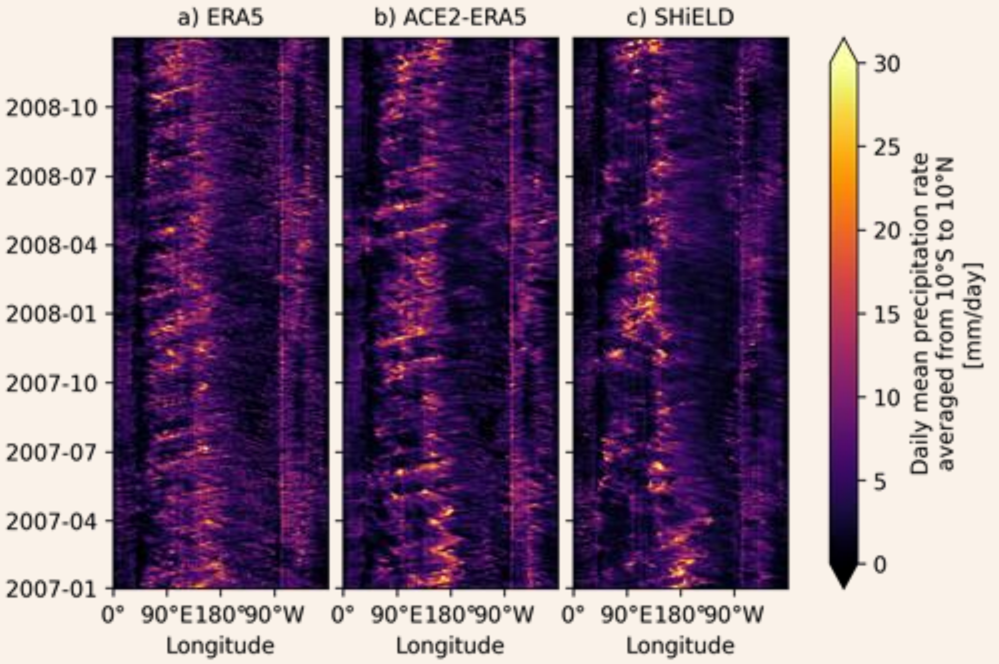
But differences between ERA5, SHiELD and ACE2-ERA5, as well as basin-to-basin differences, are robust to changes in tracking.

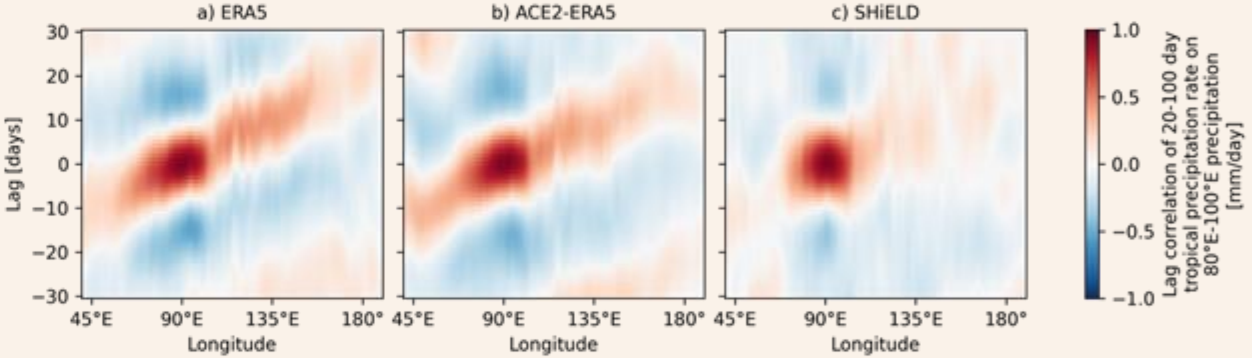


Based on 2001-2010 period

Ai2

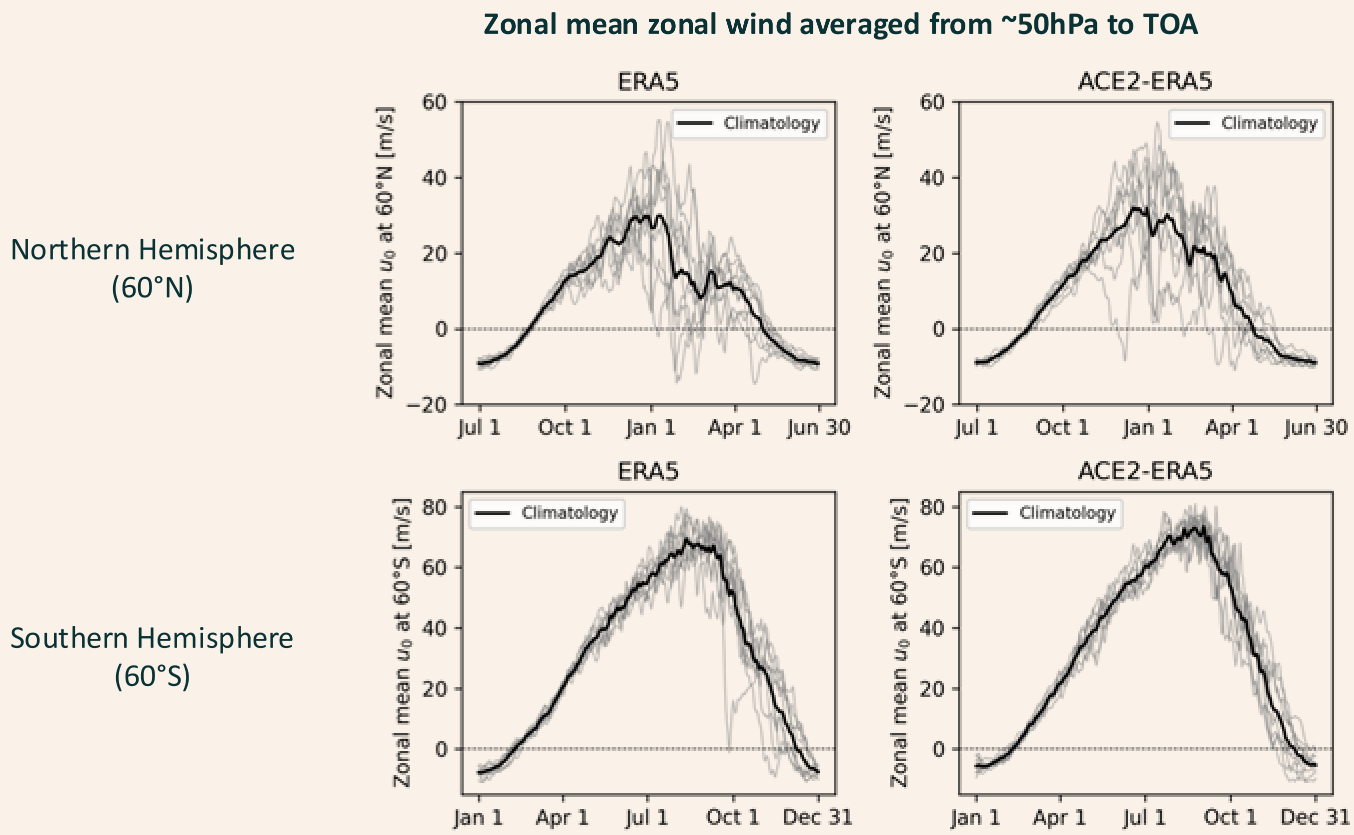
Tropical precip variability





Ai2

Climate skill - ERA5 polar stratosphere



Based on 2001-2010 period

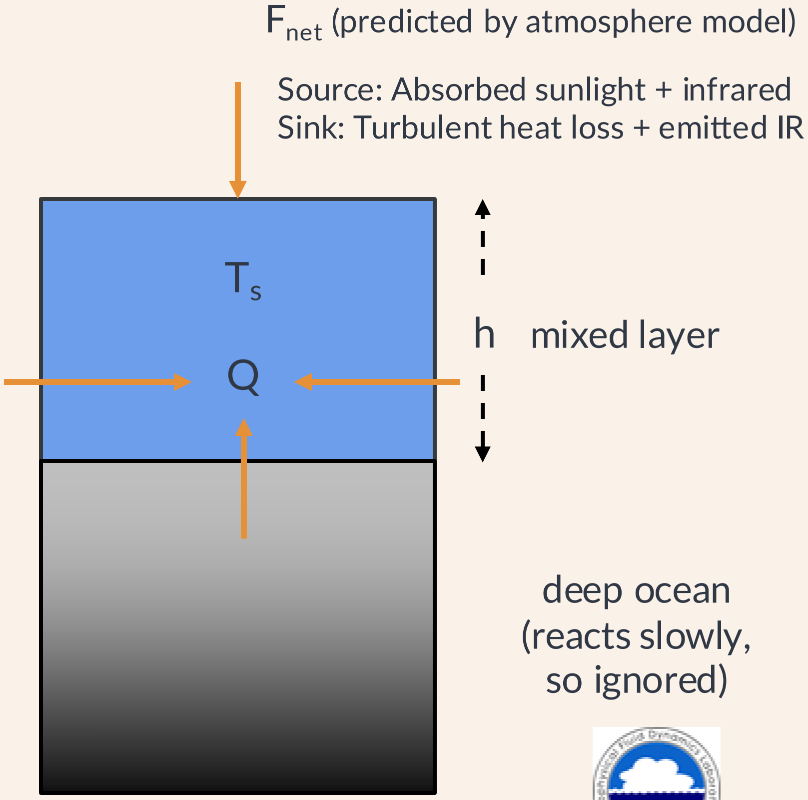
ACE2-SOM (Slab ocean model)

Preprint coming soon.

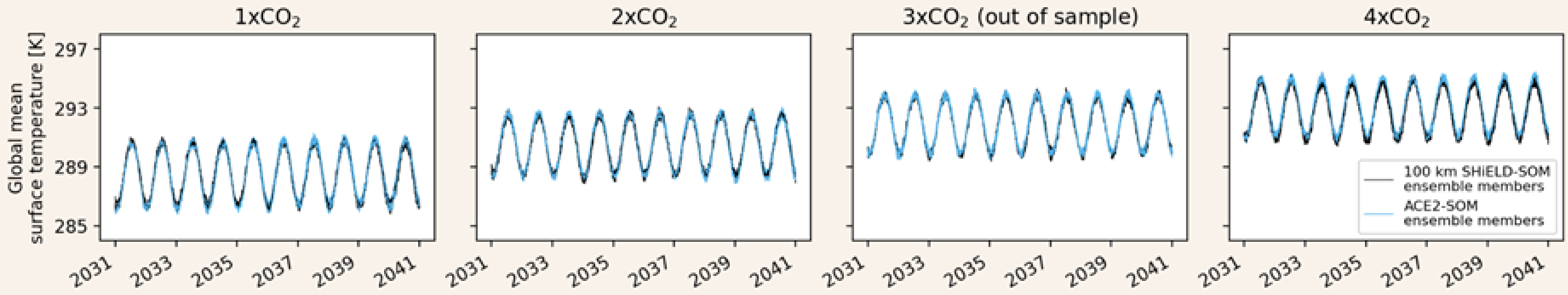
Ai2

Coupled modeling: interactive slab ocean

- Generate training/validation data from three 50-year simulations:
    - 1xCO_2 (present-day CO_2 concentration)
    - 2xCO_2 (two times present-day CO_2 concentration)
    - 4xCO_2 (four times present-day CO_2 concentration)
- Train ACE-SOM with CO_2 concentration added as an input variable.

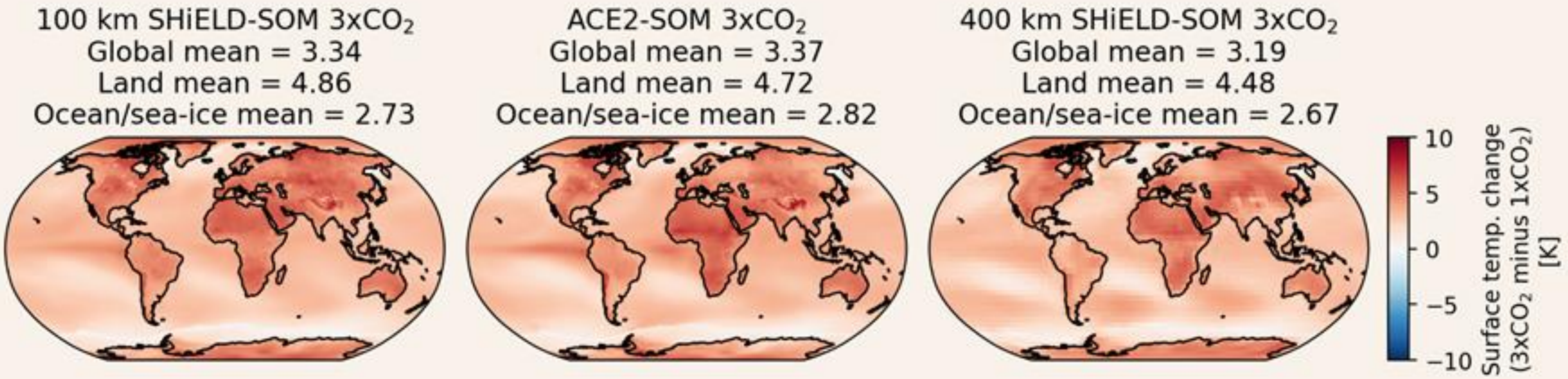


ACE2-SOM is stable in both in-sample and out-of-sample equilibrium climates



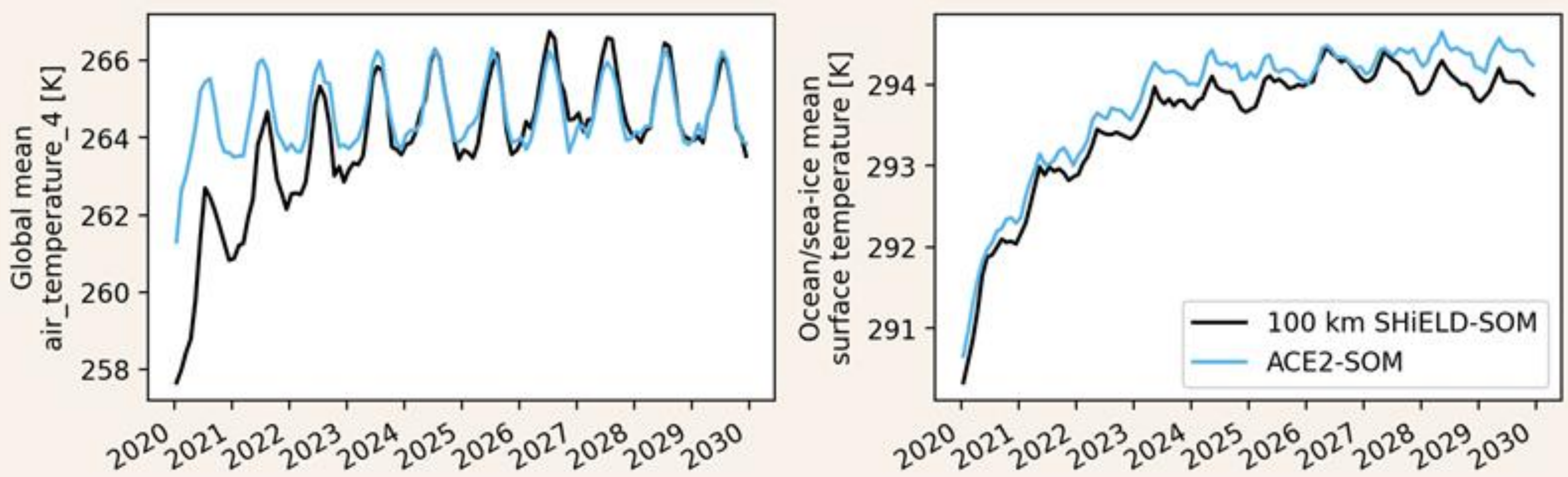
- Results are from five-member initial condition ensembles of 10-year 100 km SHiELD-SOM (our target model) and ACE2-SOM simulations in each climate (20 simulations total per model).
- ACE2-SOM did not see data from the 3xCO_2 climate during training.

ACE-SOM is stable and accurate in multiple climates



The global warming pattern matches that of the physics-based model, capturing robust features like amplified warming over land.

Out of sample test: Abrupt $4xCO_2$ increase



- ML-controlled fields (all but ocean surface temperature) rapidly shift to the $4xCO_2$ regime—clearly unrealistic.
- Ocean surface temperature, aided some by its prescribed thermal inertia, warms more slowly, approximately in accordance with the reference.
- ACE2-SOM still finds the right steady state, helped by its equilibrium $4xCO_2$ training.

Ai2

Computational cost

- Training time for each ACE2 model was about **4.5 days on eight 80GB NVIDIA H100s (240 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

Summary

- We have developed a long-term stable purely ML-based atmospheric model that is suitable for climate prediction
- Faithfully reproduces reference model's climate
  - Tested in climSST, AMIP, and slab-ocean configurations with positive results
  - Captures aspects of forced response to GHG and ocean-atm variability (ENSO)
- Much cheaper/easier to run than reference model itself!

Code, data and model checkpoints are all available!
See github.com/ai2cm/ace

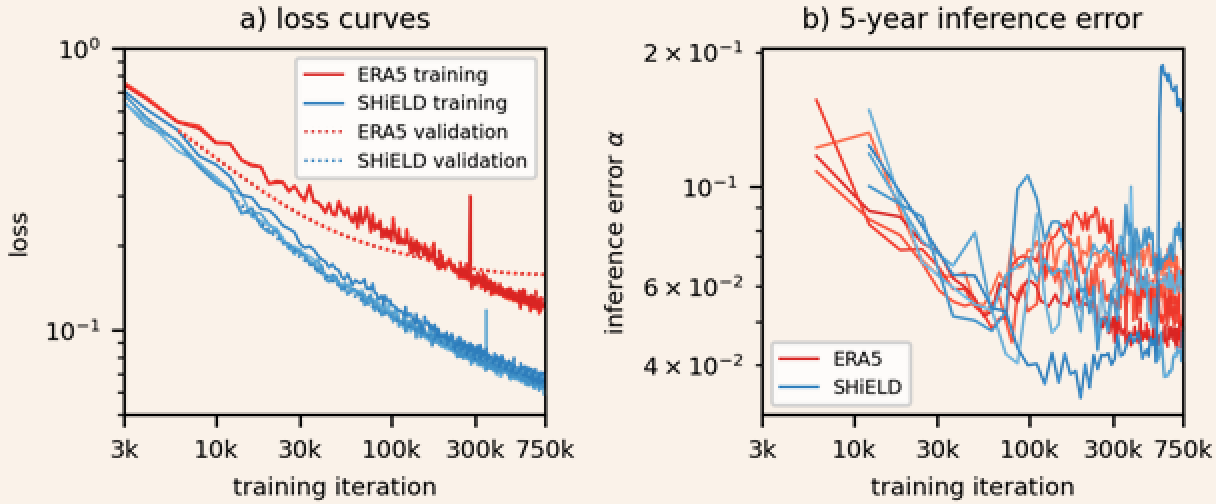
Papers: Watt-Meyer et al. 2023; Duncan et al. 2024

Ai2

Extra Slides

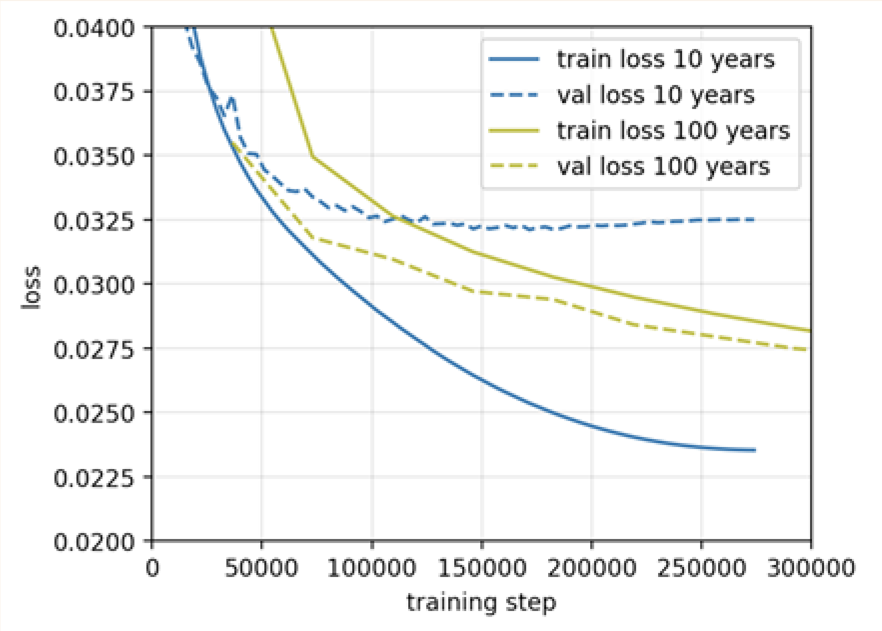
Our training strategy is not perfect

- Climate performance is not entirely robust across random seed or epoch
- A key lesson: track the metric you care about! Throughout training.
- E.g. we do long (up to 5 year) forecasts after every epoch of training
    - Inference is so fast, this typically only adds ~10% to total training time



Ai2

How much data? (more is always better…)



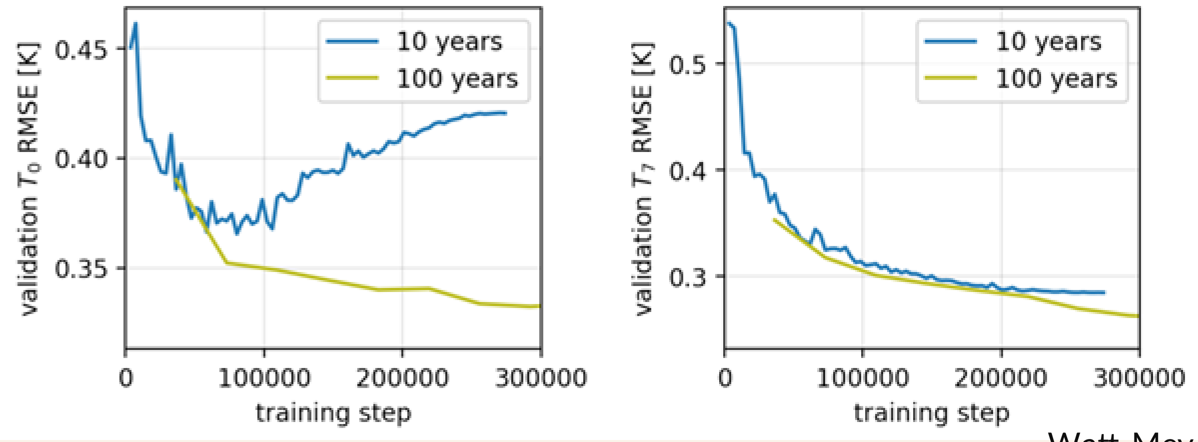
Ai2

Watt-Meyer et al. (2023)

How much data? (more is always better…)

- Depends a bit on the variable
  - Stratosphere varies more slowly (~10-30 day radiative relaxation timescale) than the troposphere (~5 day weather timescale)
  - A fixed # of years of data will have fewer independent samples of stratosphere

Stratospheric temperature 6-hour RMSE **Near-surface temperature 6-hour RMSE**



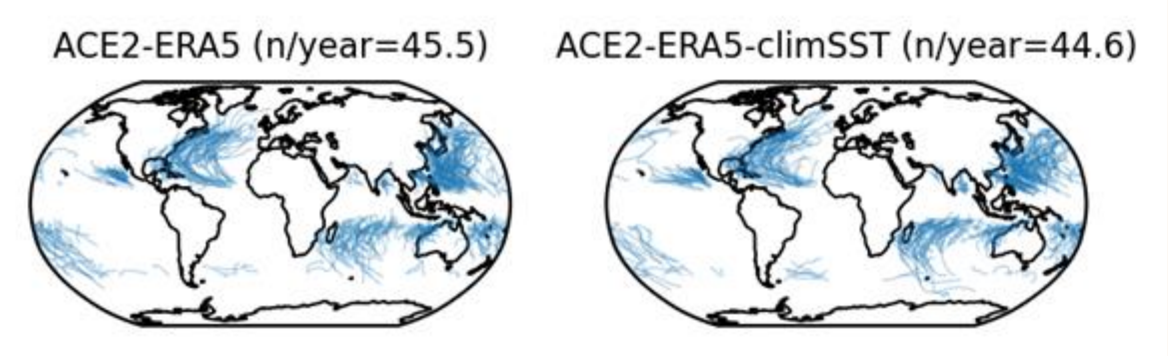
Label indicates length of training dataset

Watt-Meyer et al. (2023)

Climate skill - ERA5 tropical cyclone distribution

This would only go in appendix.

Shows that using the true historical SST dataset versus a climatological SST dataset doesn't make a difference for TC statistics.



Based on 2001-2010 period

notebook