AI for Climate on the NRP

Duncan Watson-Parris, Scripps/HDSI (UC San Diego)

7NRP Workshop, 7th May 2026

climate-analytics-lab.github.io – dwatsonparris@ucsd.edu

Climate Analytics Lab

SCRIPPS INSTITUTION OF OCEANOGRAPHY

UC San Diego

HALICIOĞLU DATA SCIENCE INSTITUTE

Climate Analytics Lab

Varan Madan

Giorgia Nicolaou

Rishabh Patni

Andrew Williams

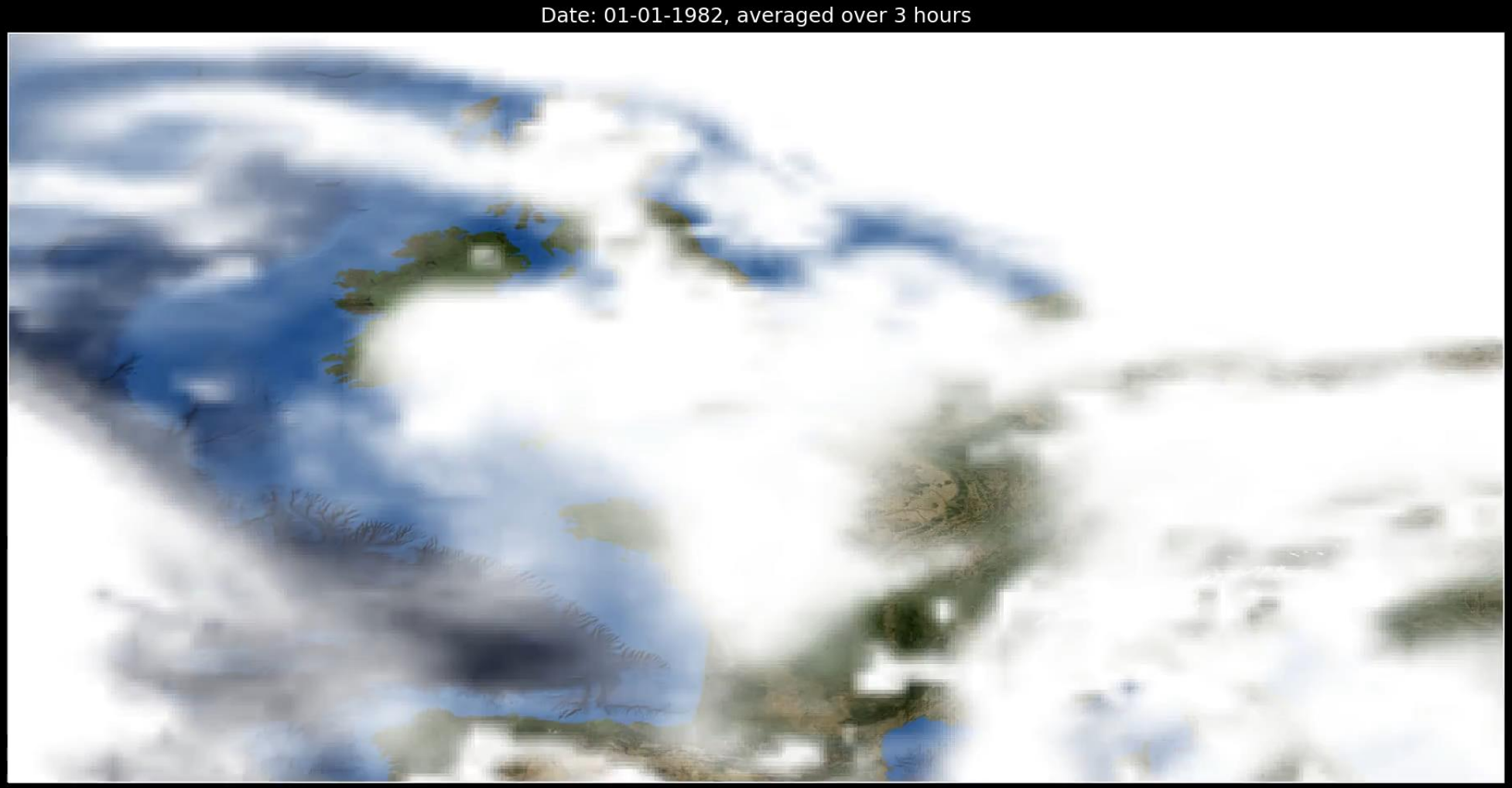
Willa Tobin

Sophia Wynn

Shreejith

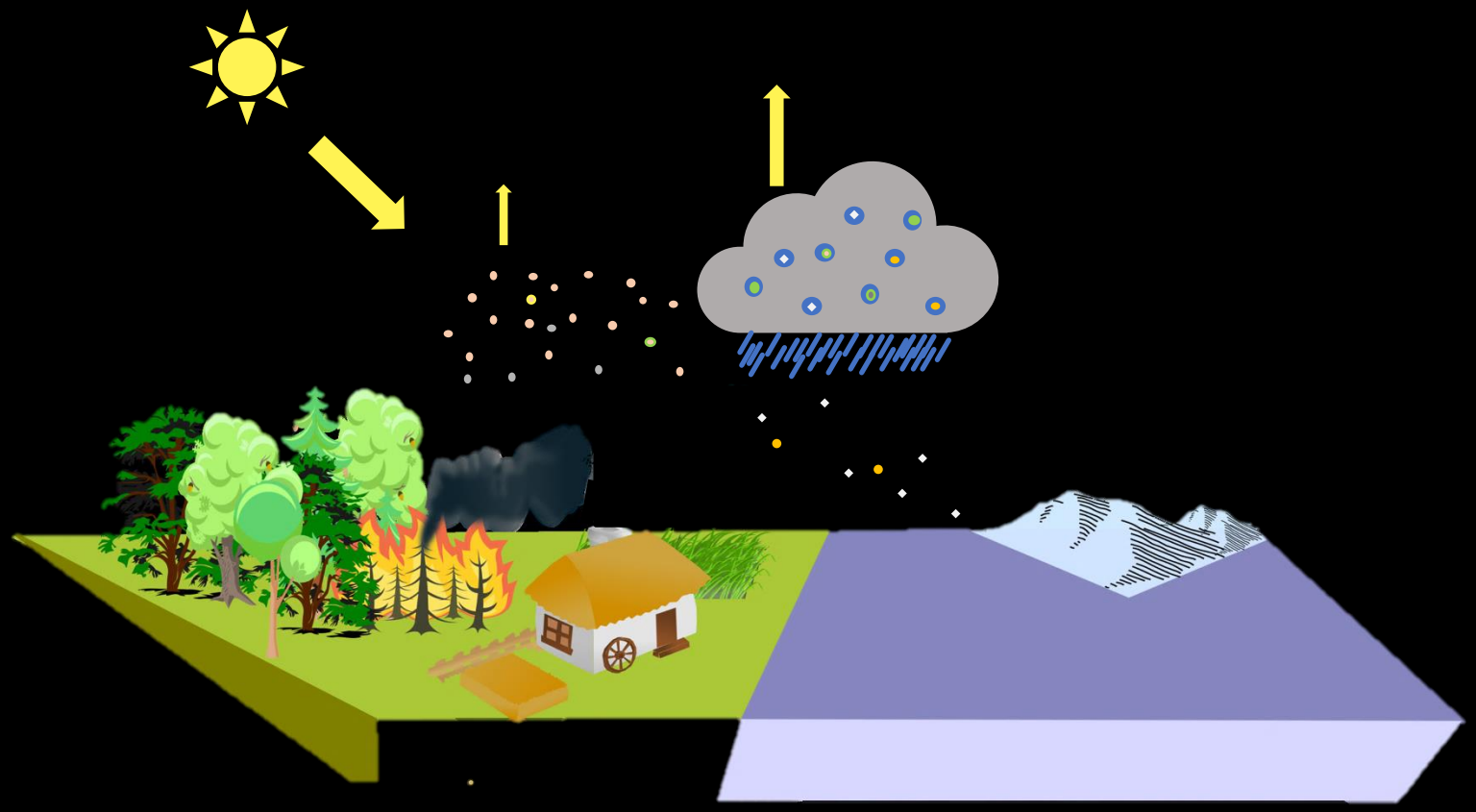
Jackson Wilke

ML for Weather and Climate are worlds apart



c.f. Watson-Parris 2021

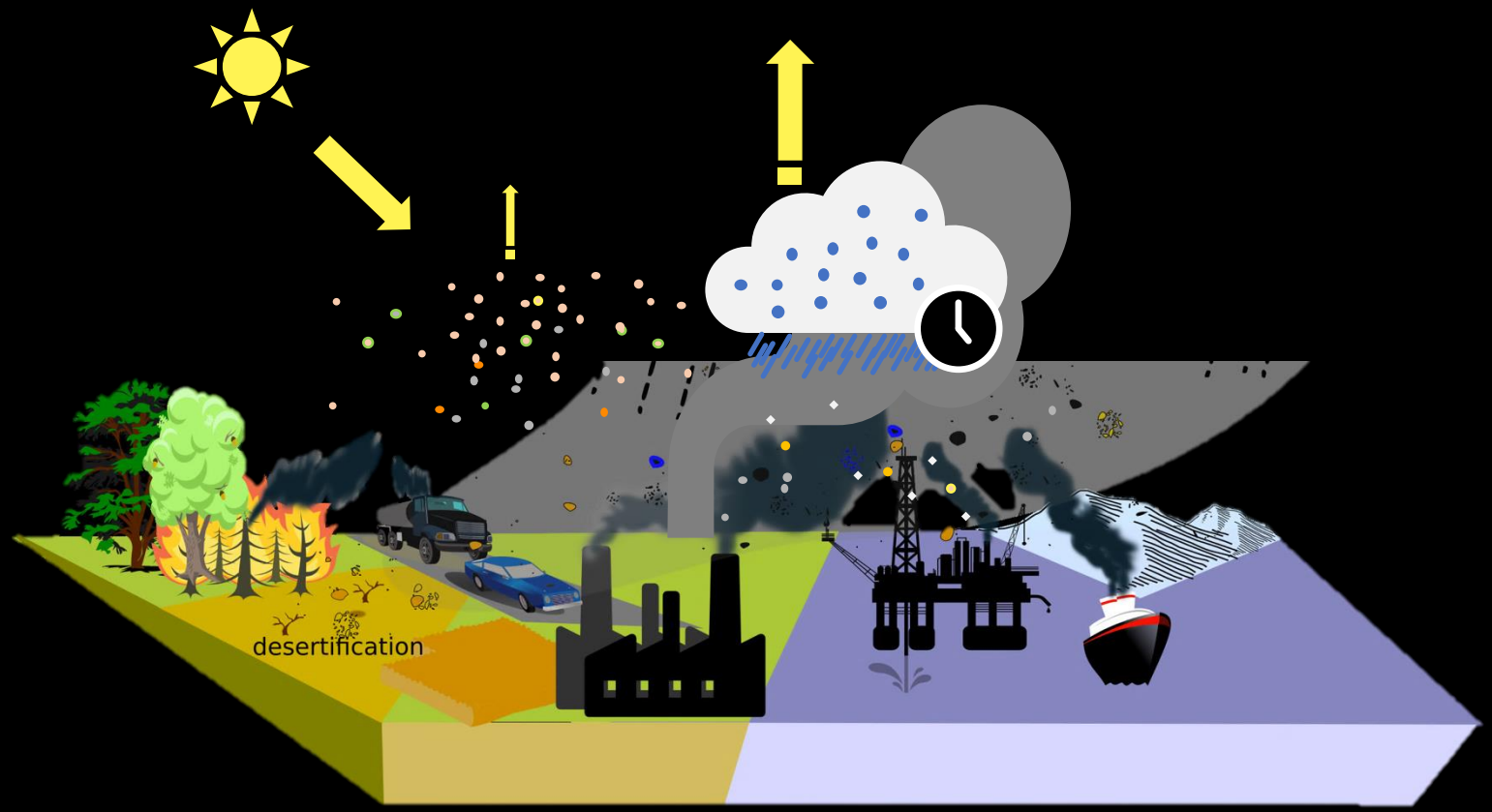
Natural aerosol



Pre-industrial

After Bellouin et al. *Rev. Geophysics* (2019)

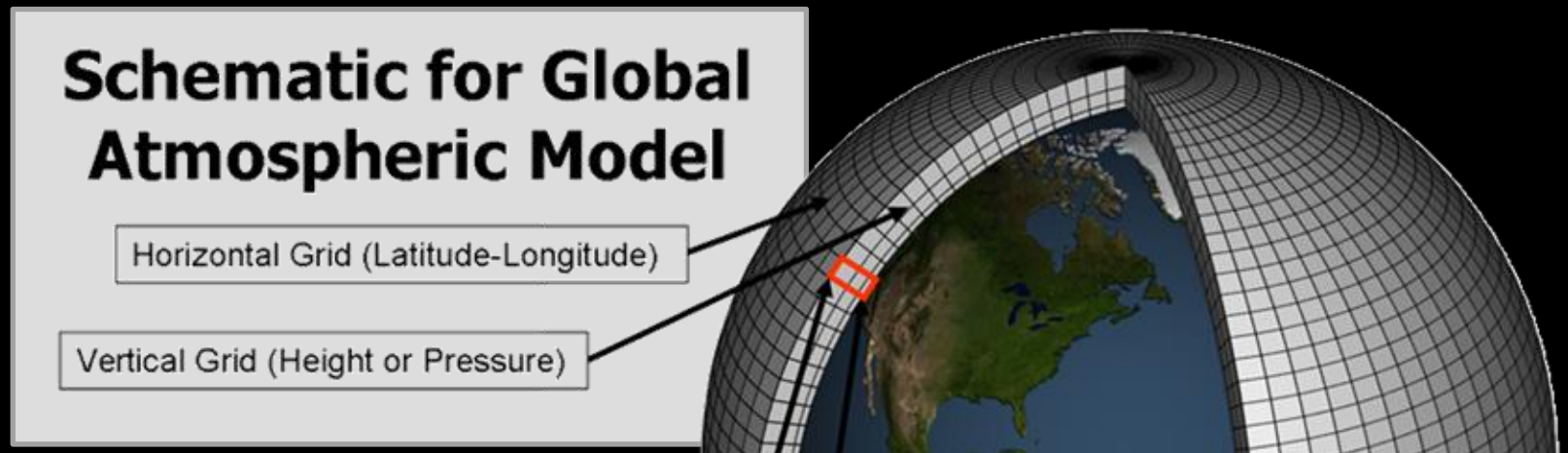
Aerosol Forcing

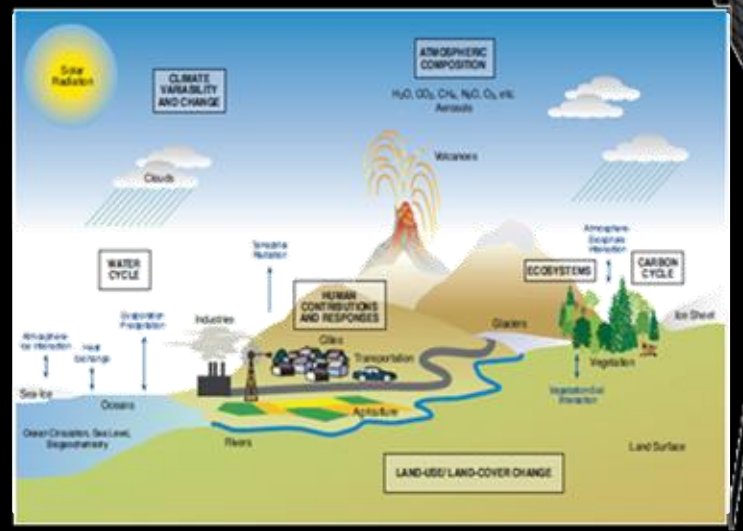


Present day

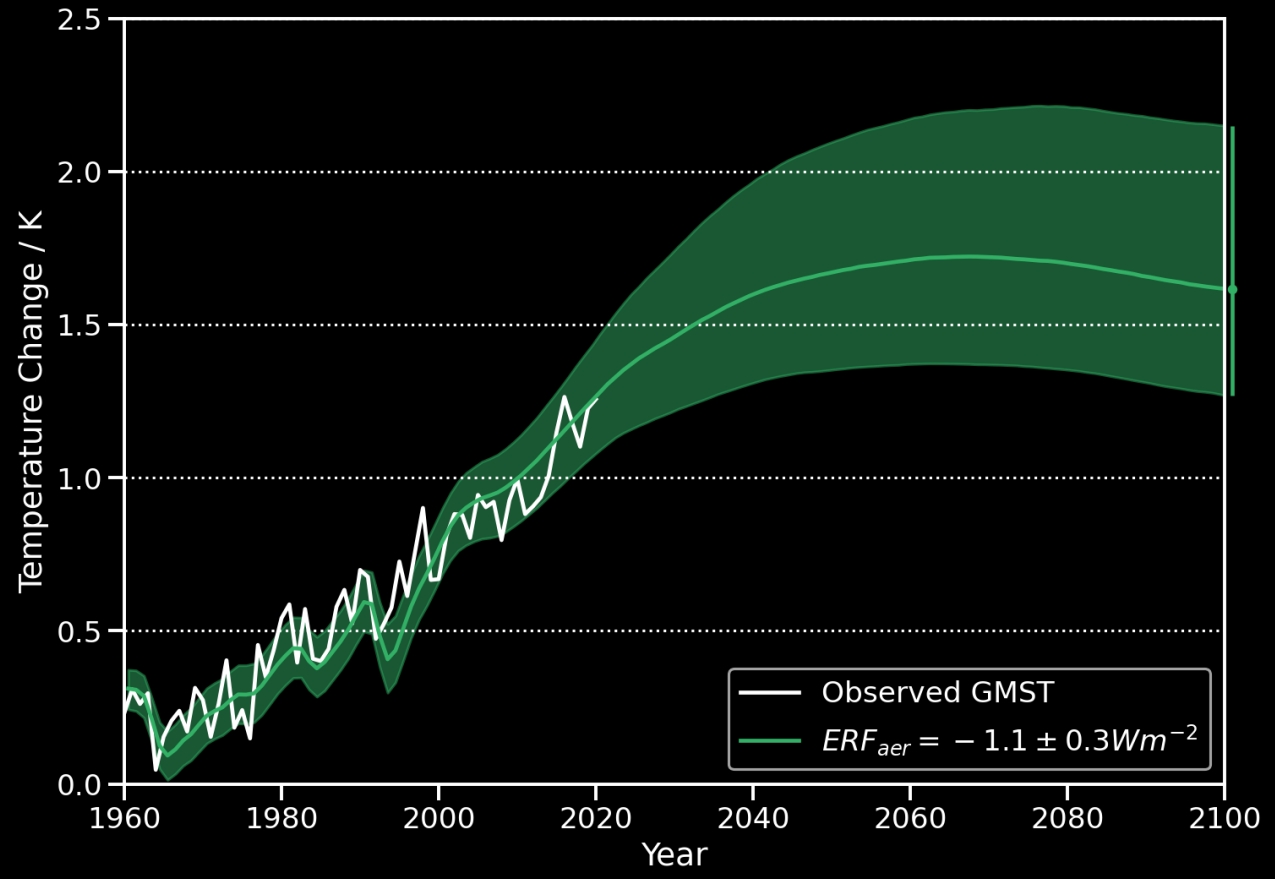
After Bellouin et al. *Rev. Geophysics* (2019)

Climate models



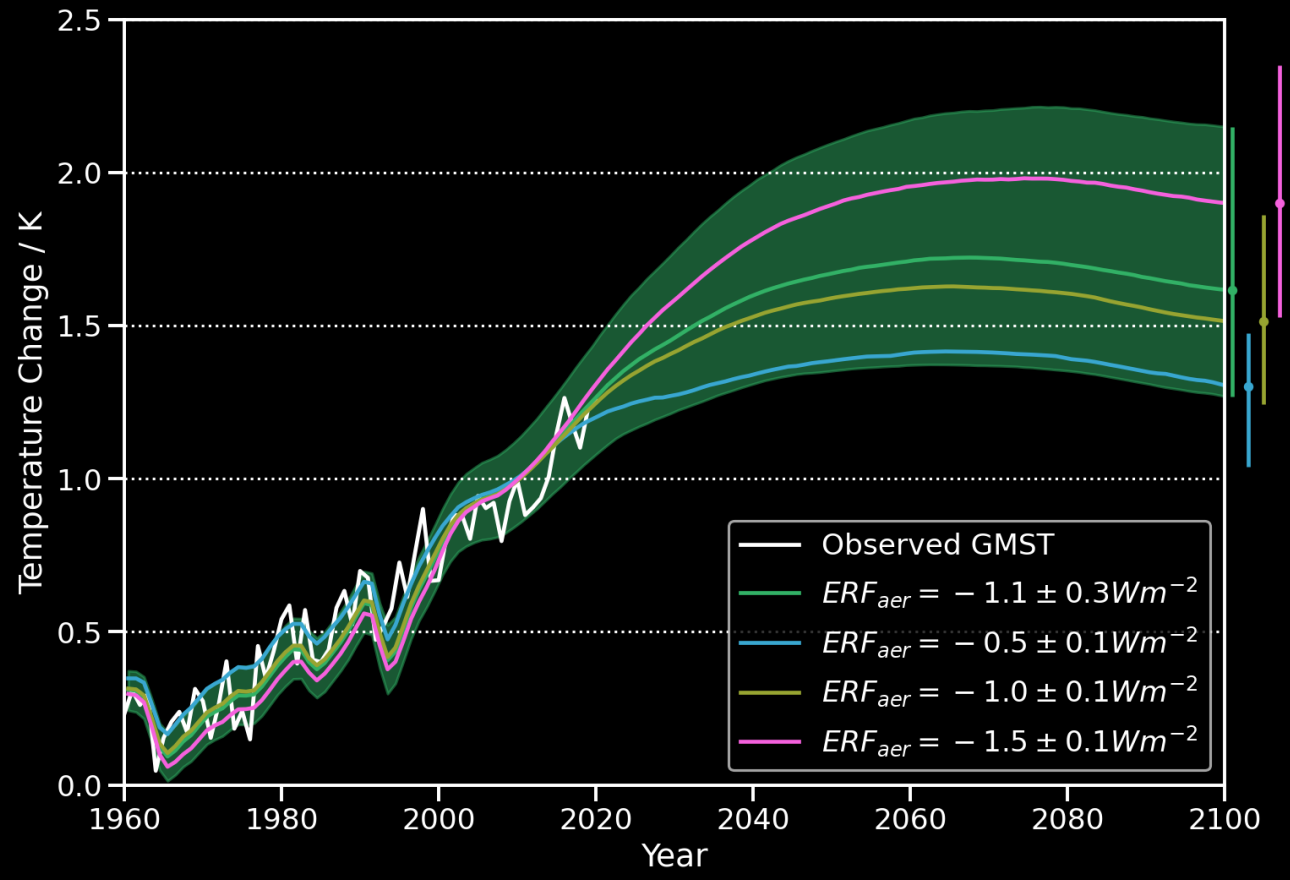


Uncertain Aerosol Forcing - Uncertain Climate Projections



Watson-Parris & Smith *Nature Clim. Change* (2024)

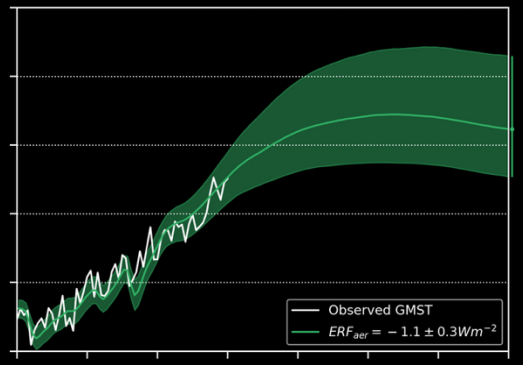
Uncertain Aerosol Forcing - Uncertain Climate Projections



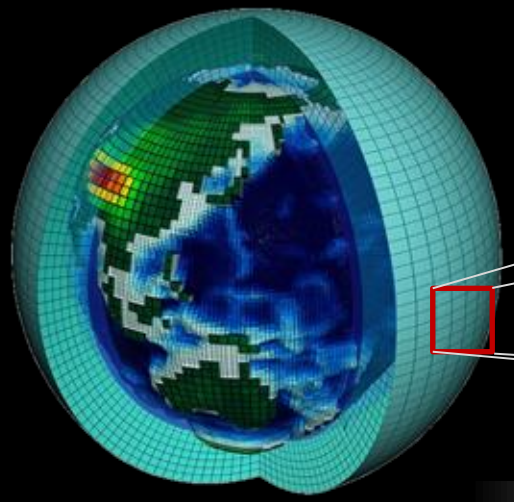
Watson-Parris & Smith *Nature Clim. Change* (2024)

Climate is a multiscale problem

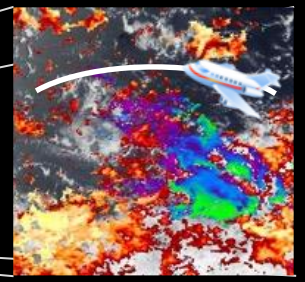
Energy balance model:
~10,000km



Climate model:
Δx ~ 100km



Observations:
Δx < 1km

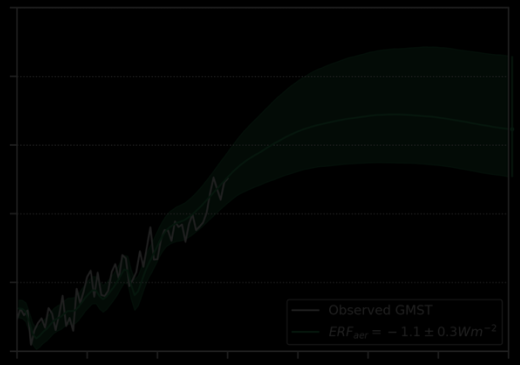


Energetic constraints

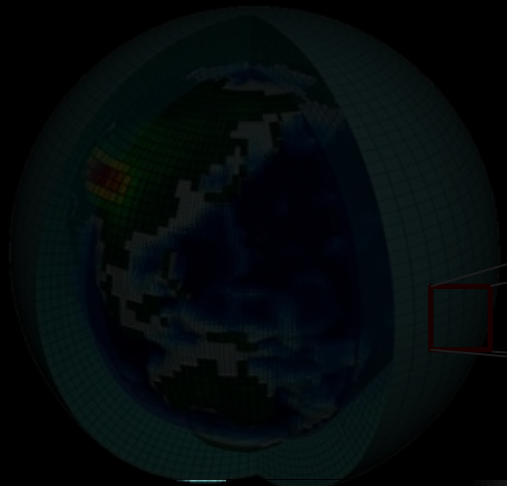
Process knowledge

Climate is a multiscale problem

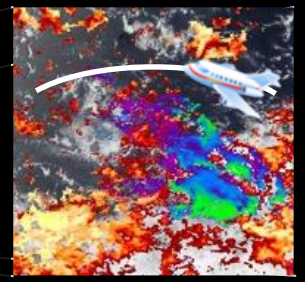
Energy balance model:
~10,000km



Climate model:
Δx ~ 100km



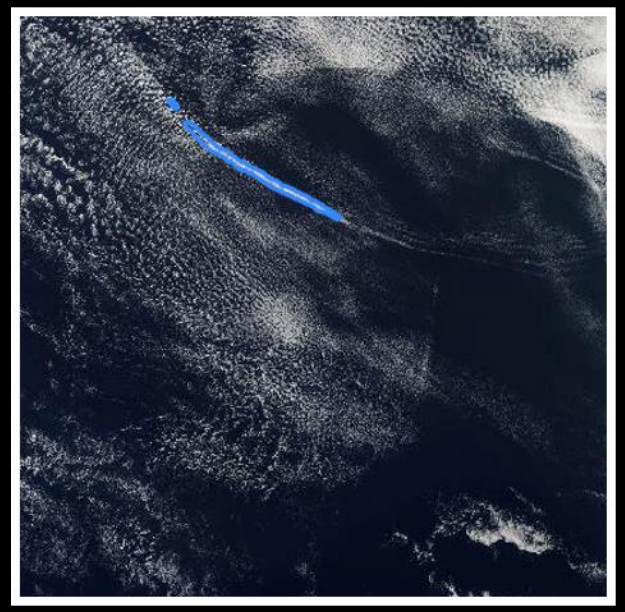
Observations:
Δx < 1km



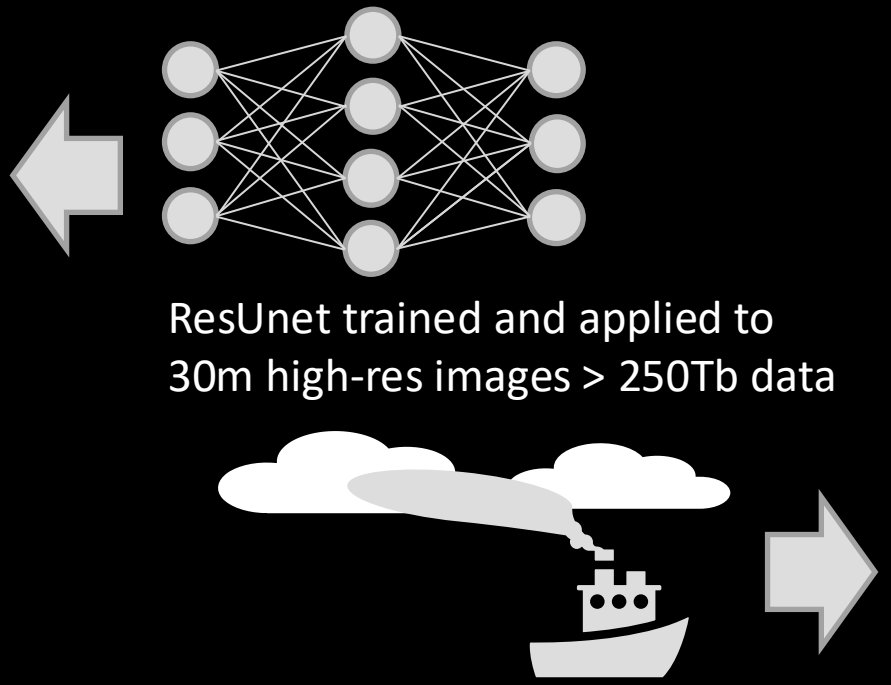
Energetic constraints

Process knowledge

Bridging the scale gap: Observations

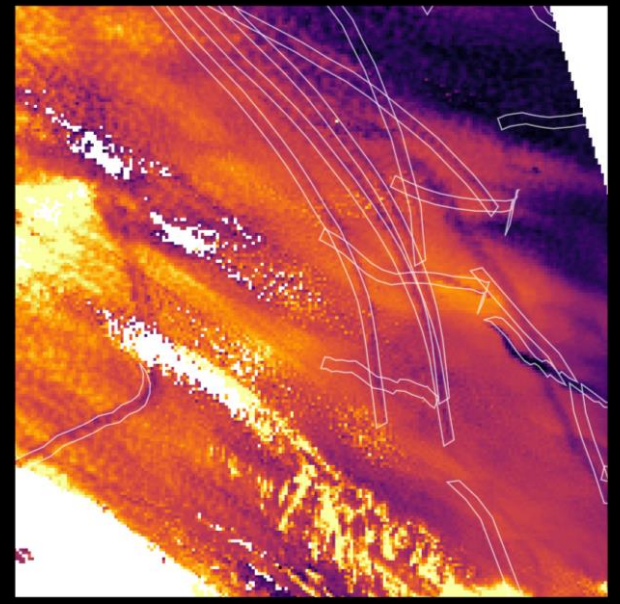


Watson-Parris et al. *PNAS* (2022)



ResUnet trained and applied to
30m high-res images > 250Tb data

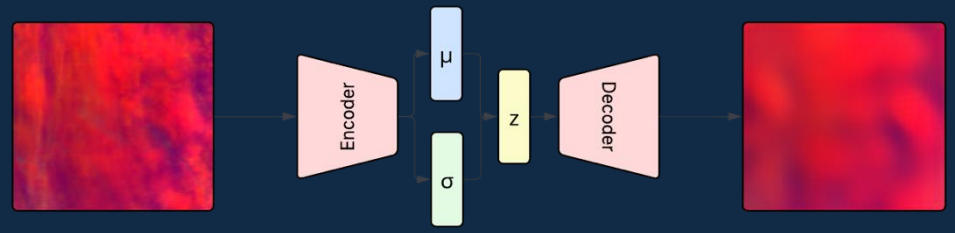
4m Lagrangian trajectories
calculated and collocated



Manshausen et al. *Nature* (2022)

Continuous Representation of Cloud Morphology

Deep learning on 1M+ MODIS tiles reveals cloud types form a continuum, not discrete categories



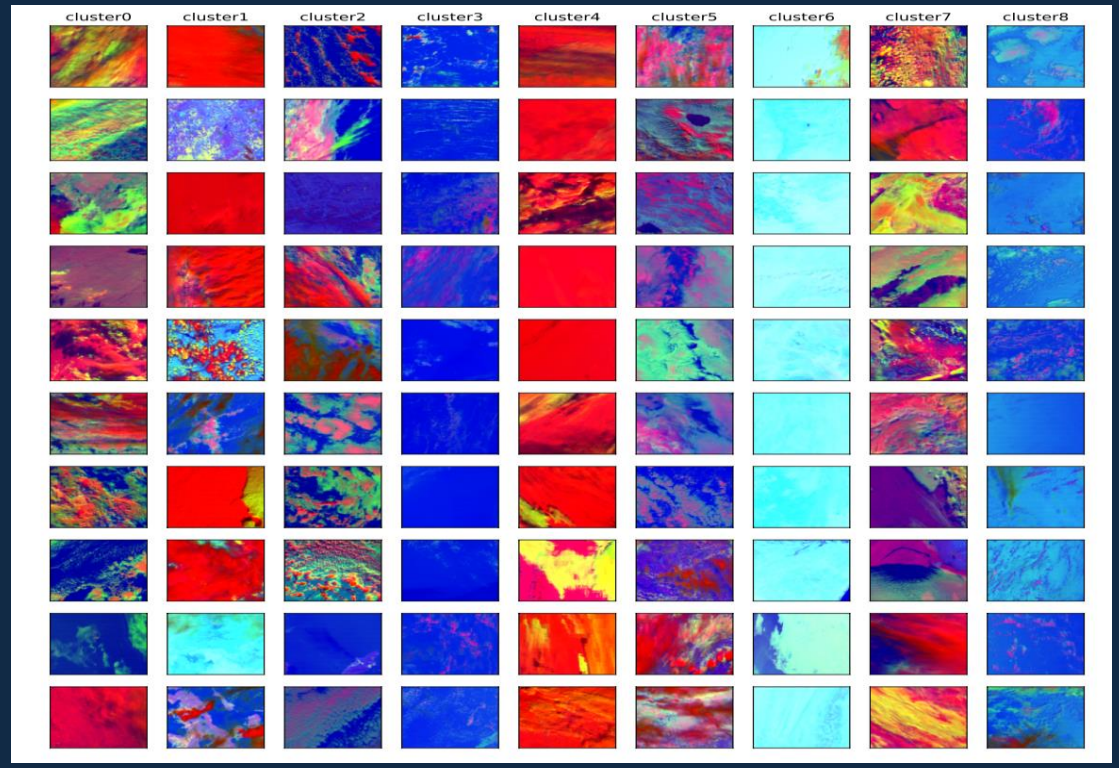
Variational Autoencoder learns continuous 2048-dim latent space from satellite patches

Tile2Vec

Unsupervised triplet learning on MODIS Day Microphysics composites; 50-dim spatial embeddings with FAISS similarity search

VAE

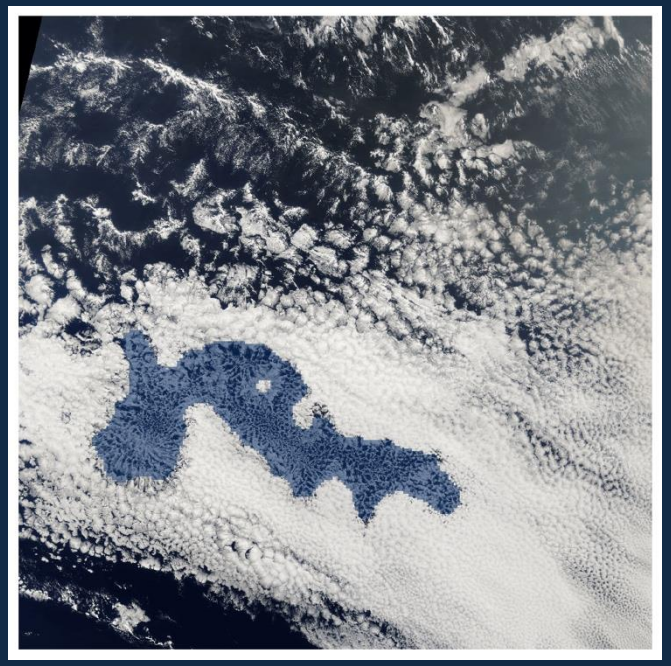
Generative latent space enabling smooth interpolation between cloud types and dimensionality reduction via PCA/t-SNE



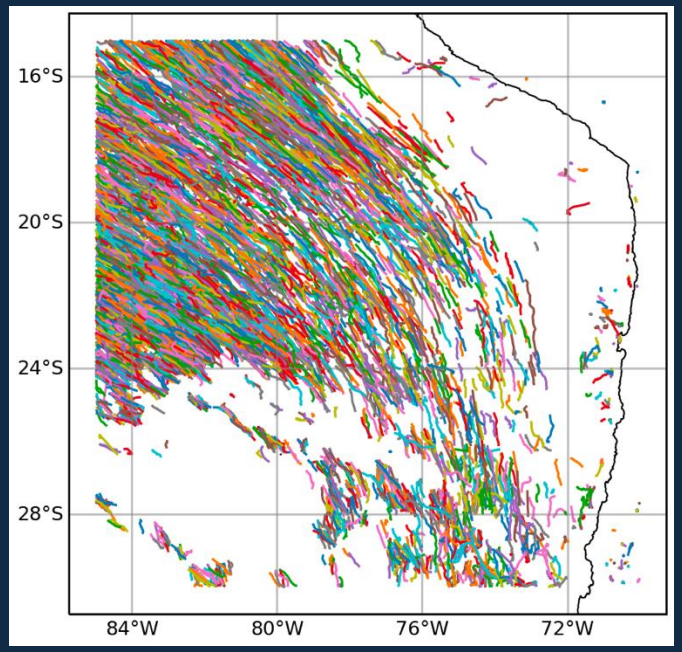
Key result: K-means reveals no stable optimal cluster count — cloud morphology occupies a continuous manifold, enabling richer characterization of radiative effects and transitions.

Tracking Open Cells in Marine Stratocumulus

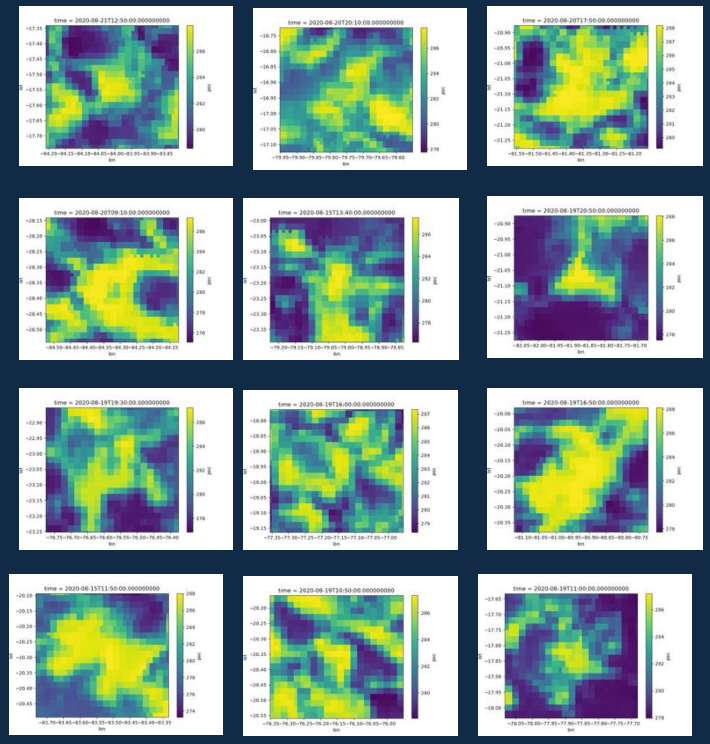
Automated segmentation & tracking of >500k Pockets of Open Cells from GOES-16 reveals new diurnal dynamics



UNet trained to segment Open Cellular clouds



500k+ cells tracked with TOBAC



Time series of tracked, centered cells

Diurnal cycle

First observational evidence of strong diurnal modulation of open-cell extent, peaking at midday

SST dependence

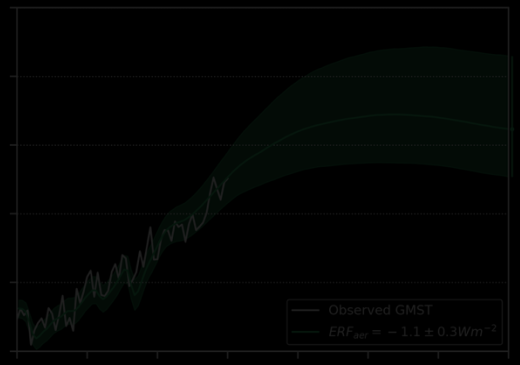
Cell area scales with sea surface temperature via Rayleigh-Bernard convection

Microphysics

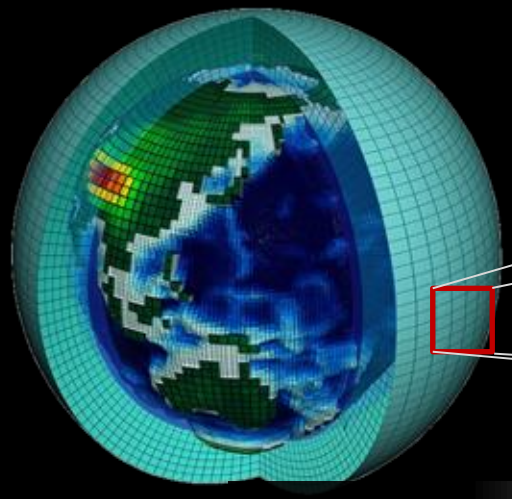
Particle size peaks before cell formation then declines as cells expand

Climate is a multiscale problem

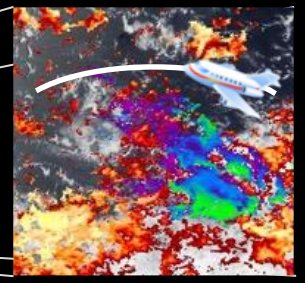
Energy balance model:
~10,000km



Climate model:
Δx ~ 100km



Observations:
Δx < 1km



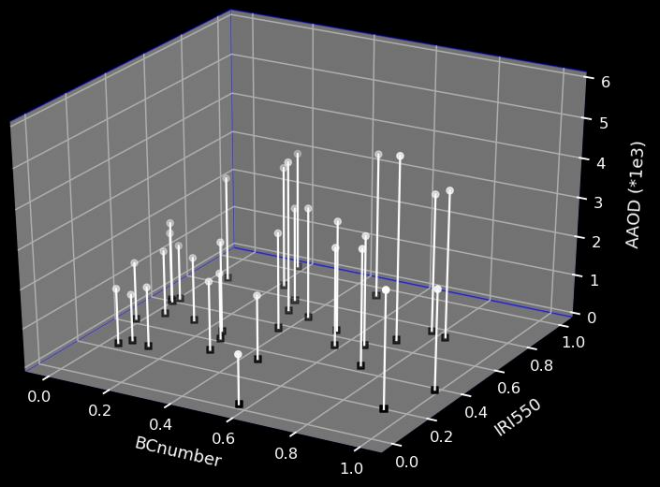
Energetic constraints

Process knowledge

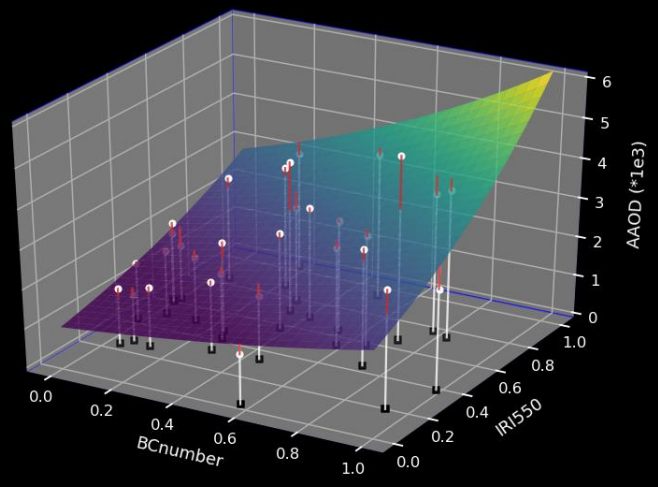
Climate model emulation for calibration

How do we choose the best parameters for our climate model?

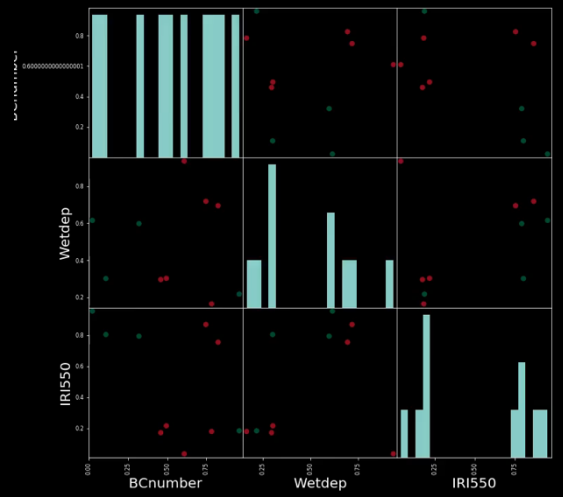
Model data



Emulation



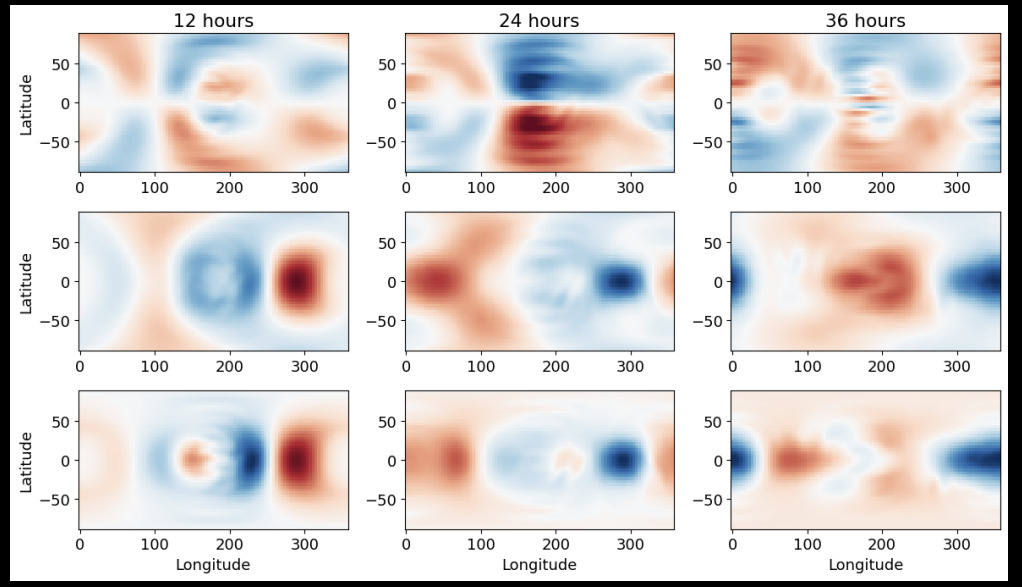
Inference / Calibration



Hybrid model prototyping with JAX-GCM

- **Fully differentiable physical** atmospheric Python model with gradients from JAX for calibration and sensitivity analysis
- **Plug-and-play** parameterizations enable flexible experimentation and online learning
- **Compatible** with CPUs, GPUs, and TPUs, with automatic parallelization for scalability
- **Validated** against Fortran baseline, with example optimization notebooks for tuning and experimentation

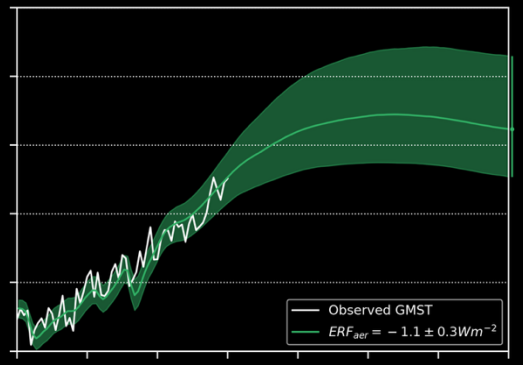
```
from jcm.model import Model

# Initialize a default (aquaplanet) model
model = Model()
predictions = model.run(total_time=30)

model.predictions_to_xarray(predictions)
```



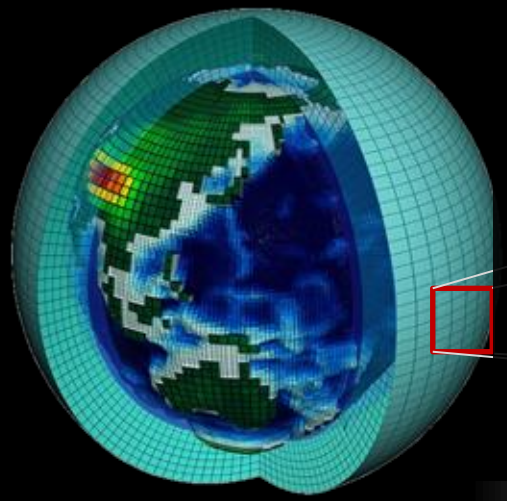
Maddan, Davenport et al. *(in review.)*

Climate is a multiscale problem

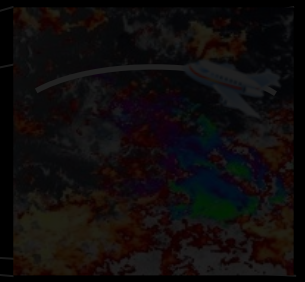
Energy balance model:
~10,000km



Climate model:
Δx ~ 100km



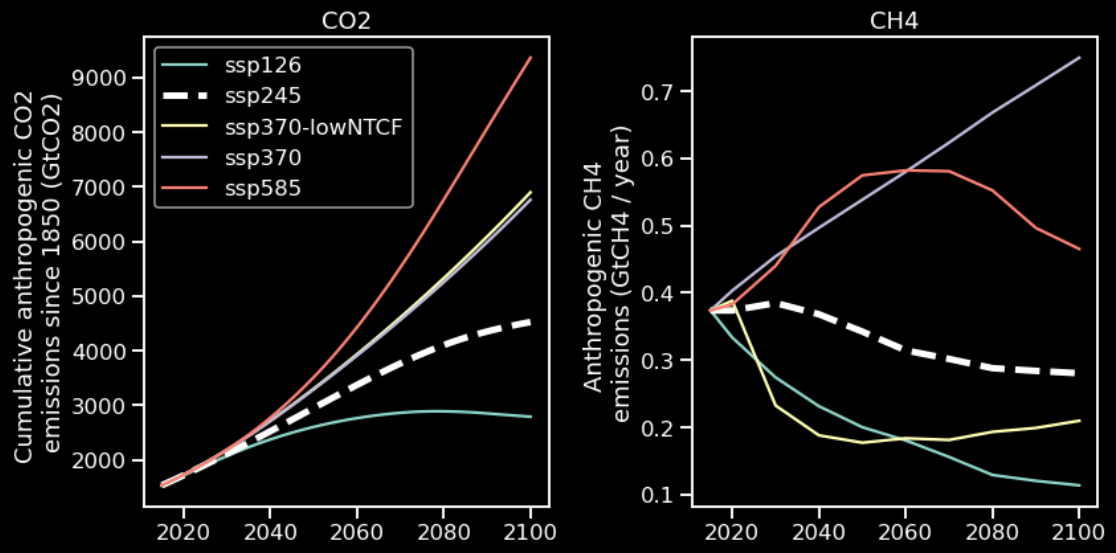
Observations:
Δx < 1km



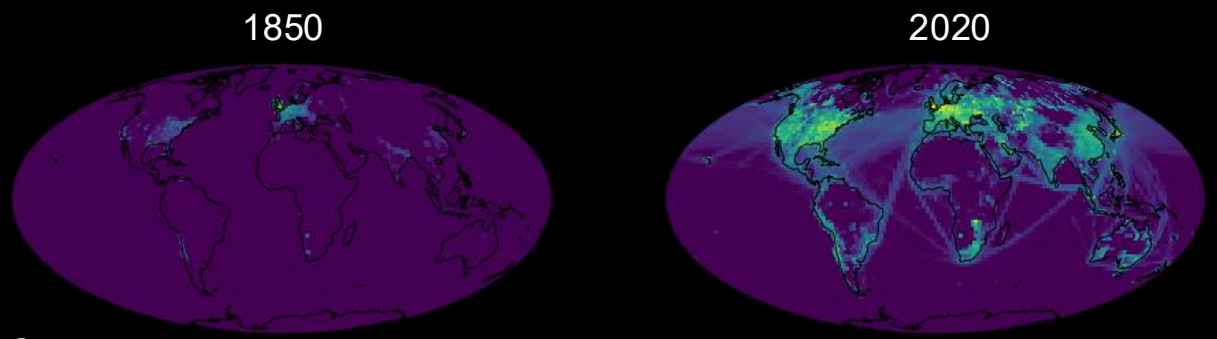
Energetic constraints

Process knowledge

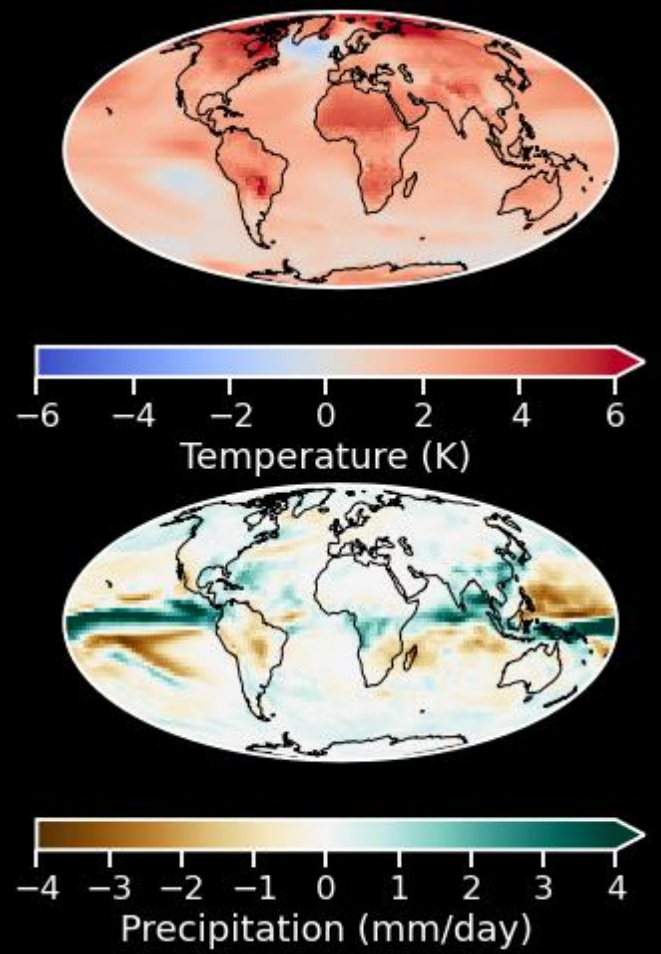
Exploring scenario uncertainty: ClimateBench



Spatially resolved emissions of SO2 and BC

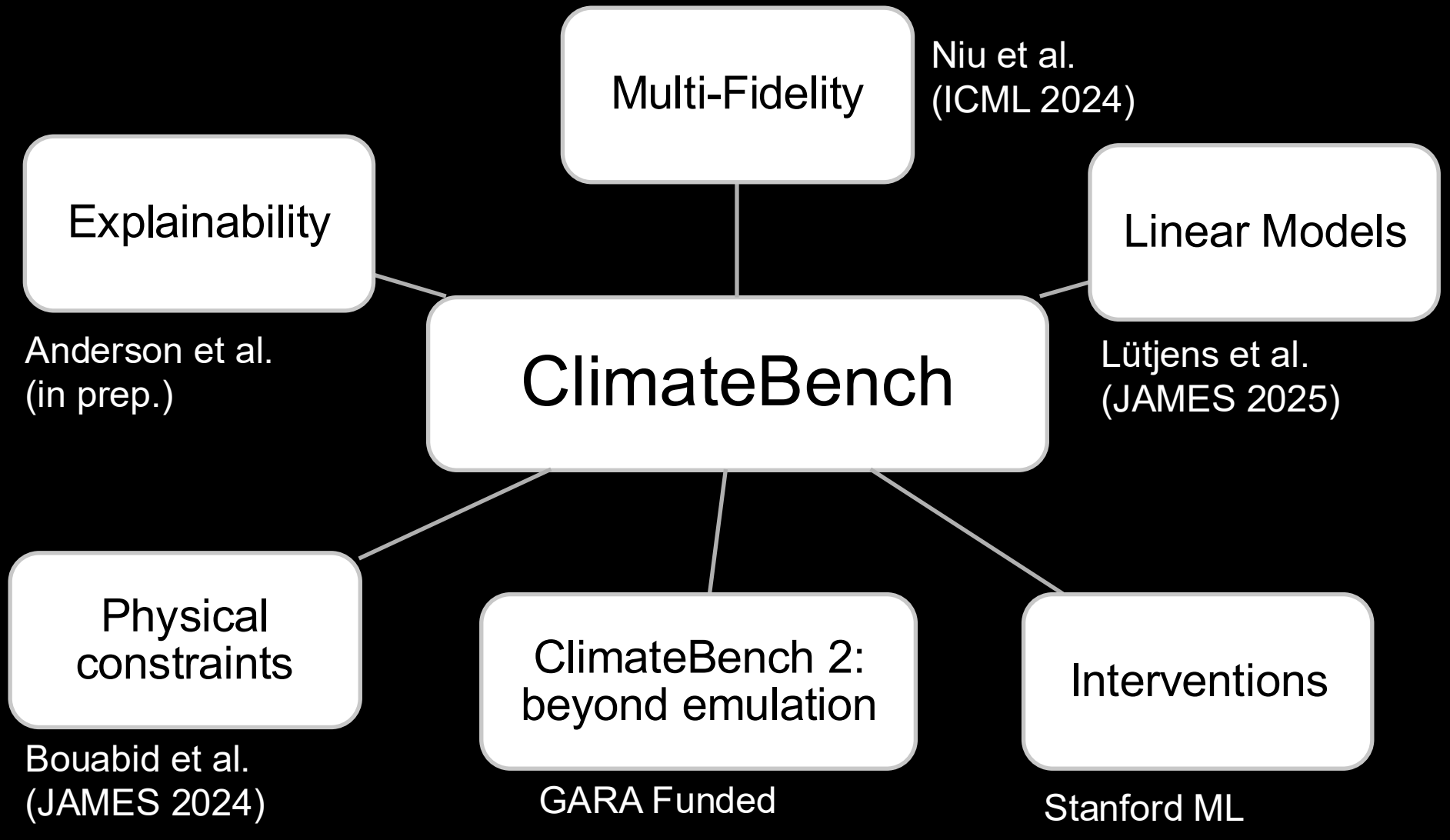


Emulate GCM

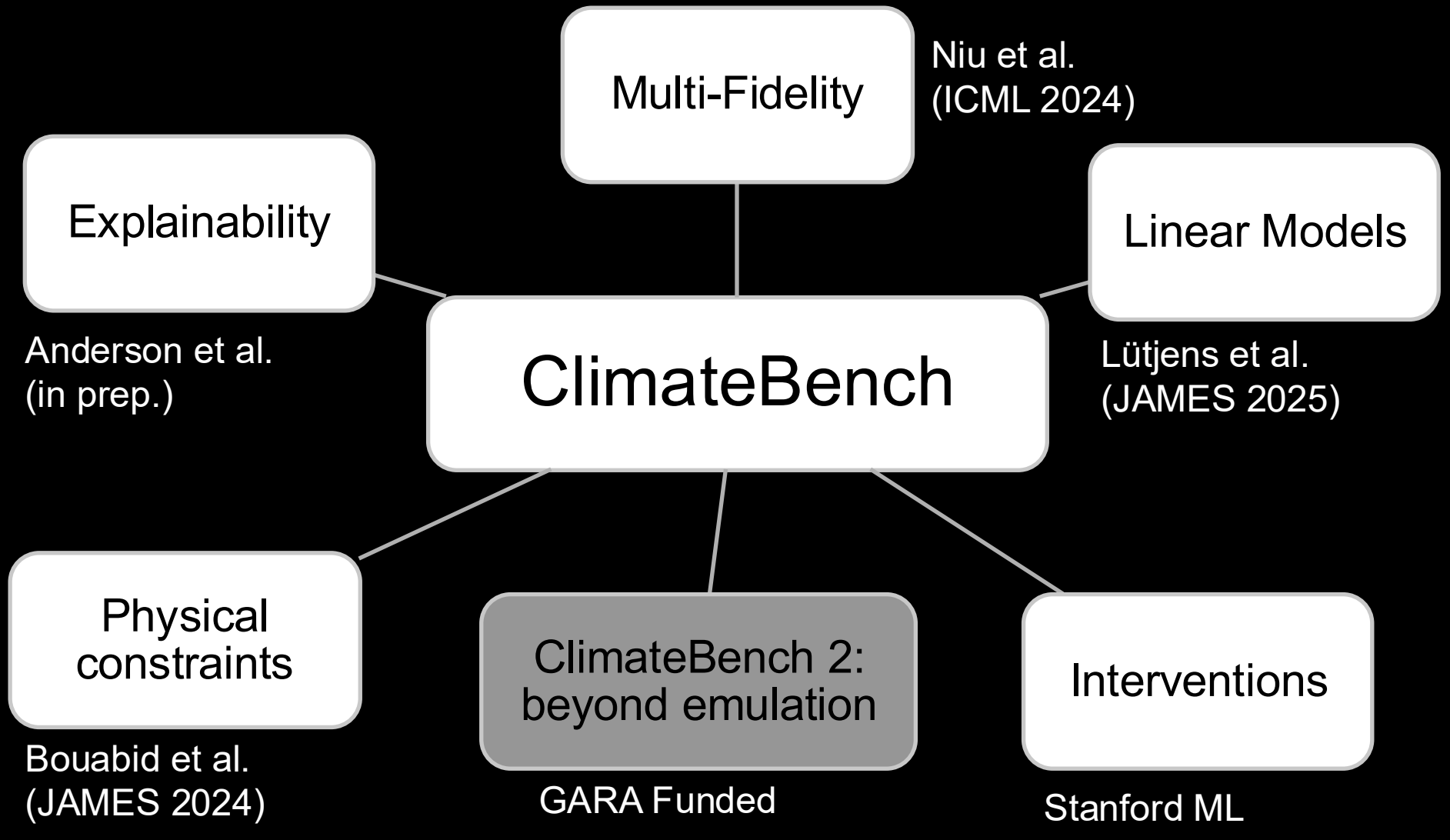


Watson-Parris et al. *JAMES* (2022)

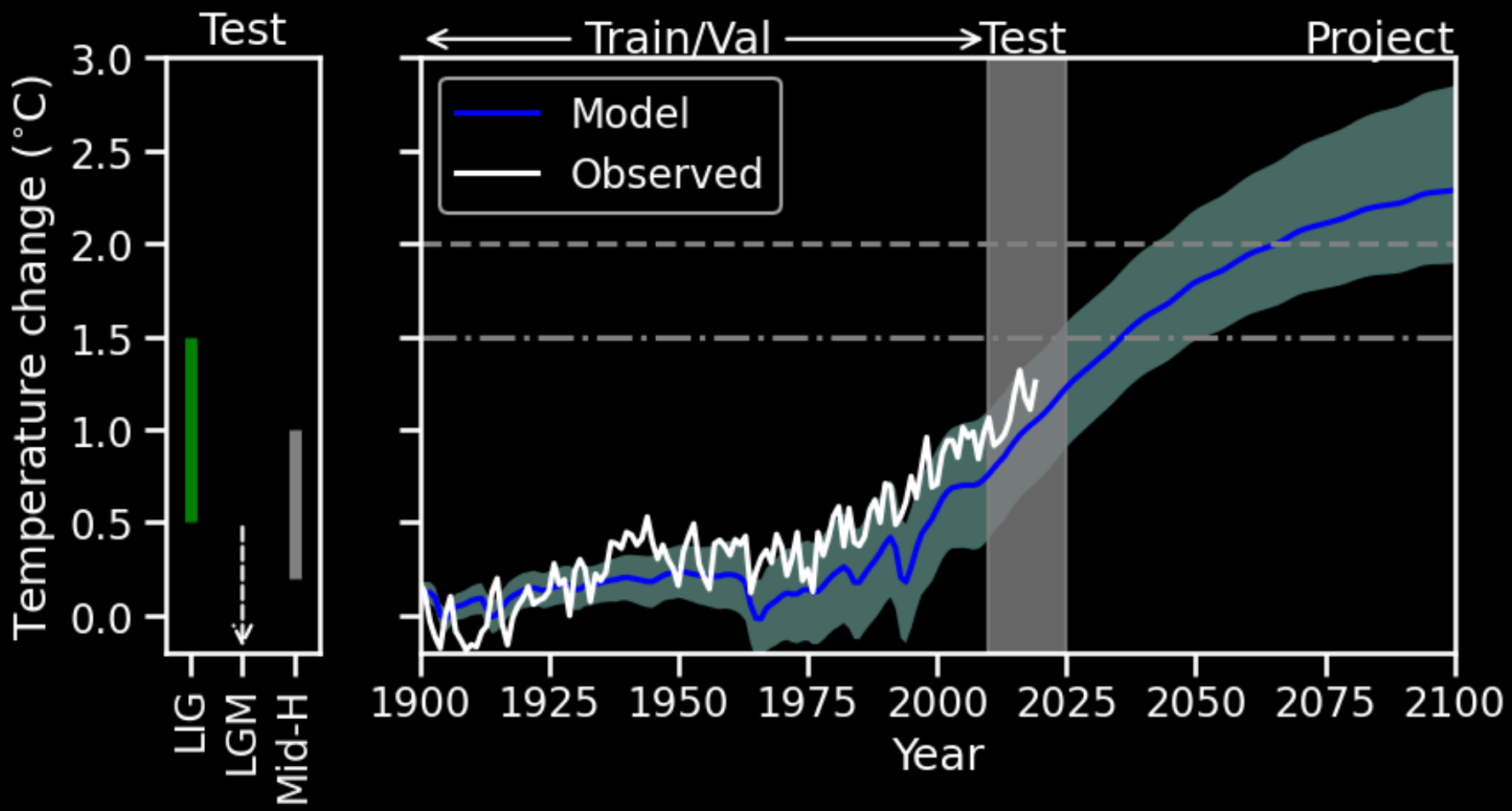
ClimateBench follow-ups



ClimateBench follow-ups



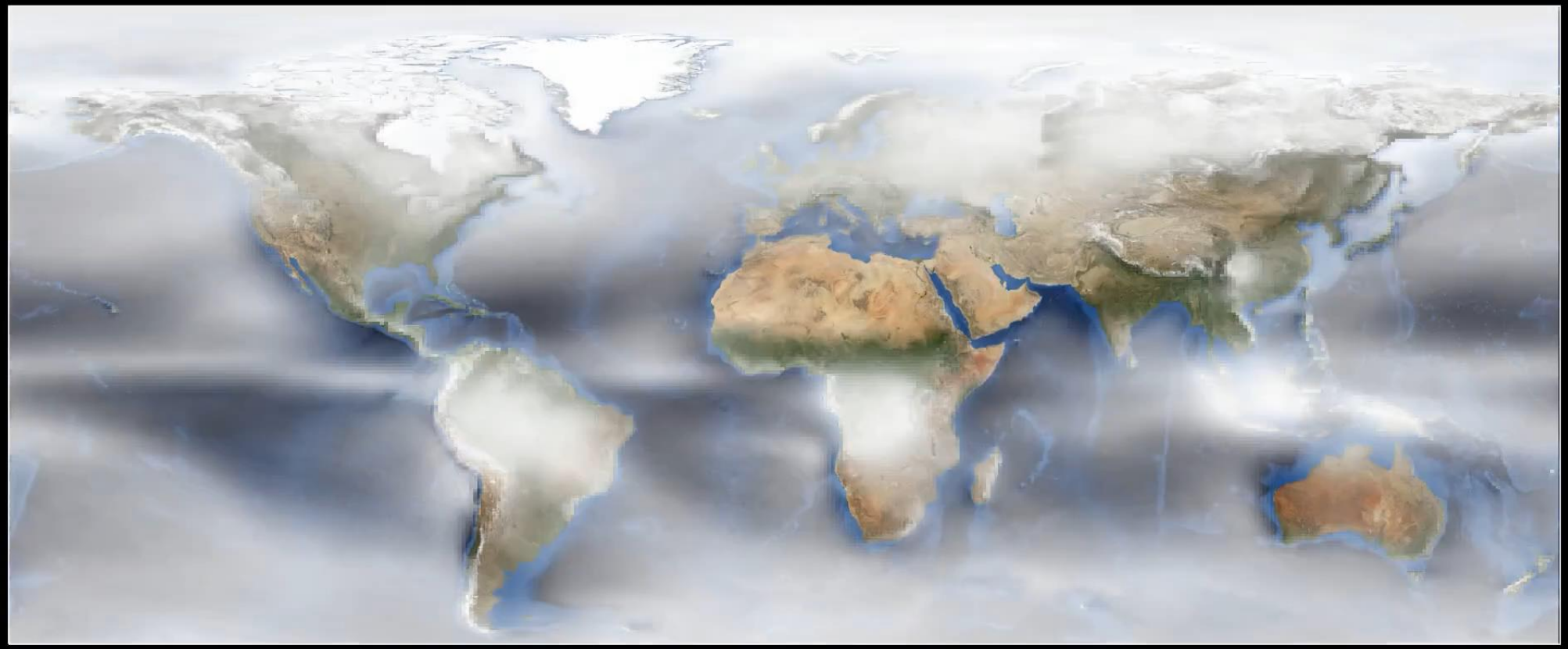
ClimateBench2: Beyond emulation



GARA

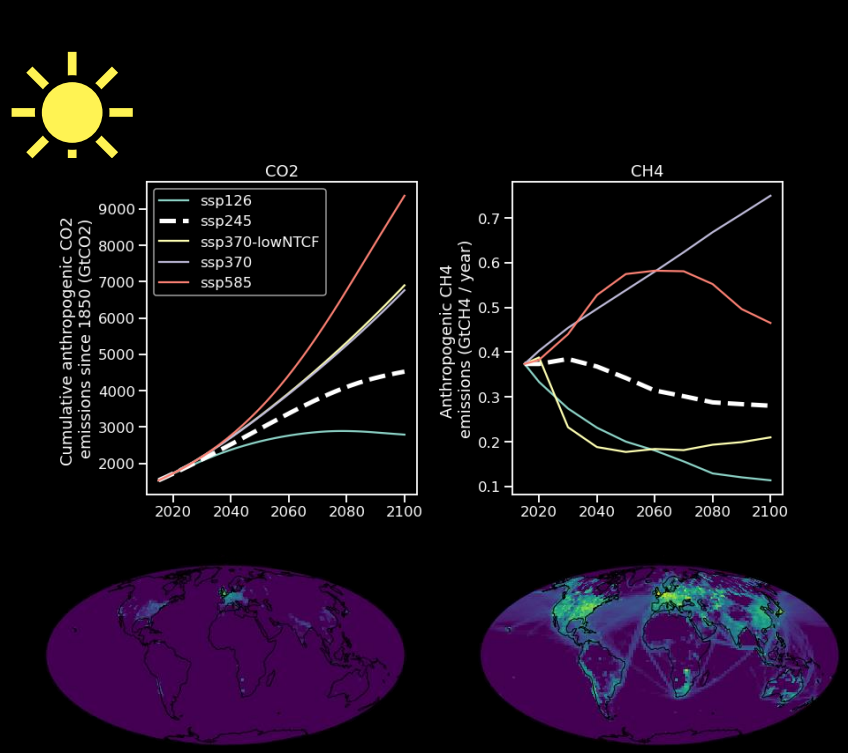
Watson-Parris et al. *(in prep.)*

Generative Weather Models

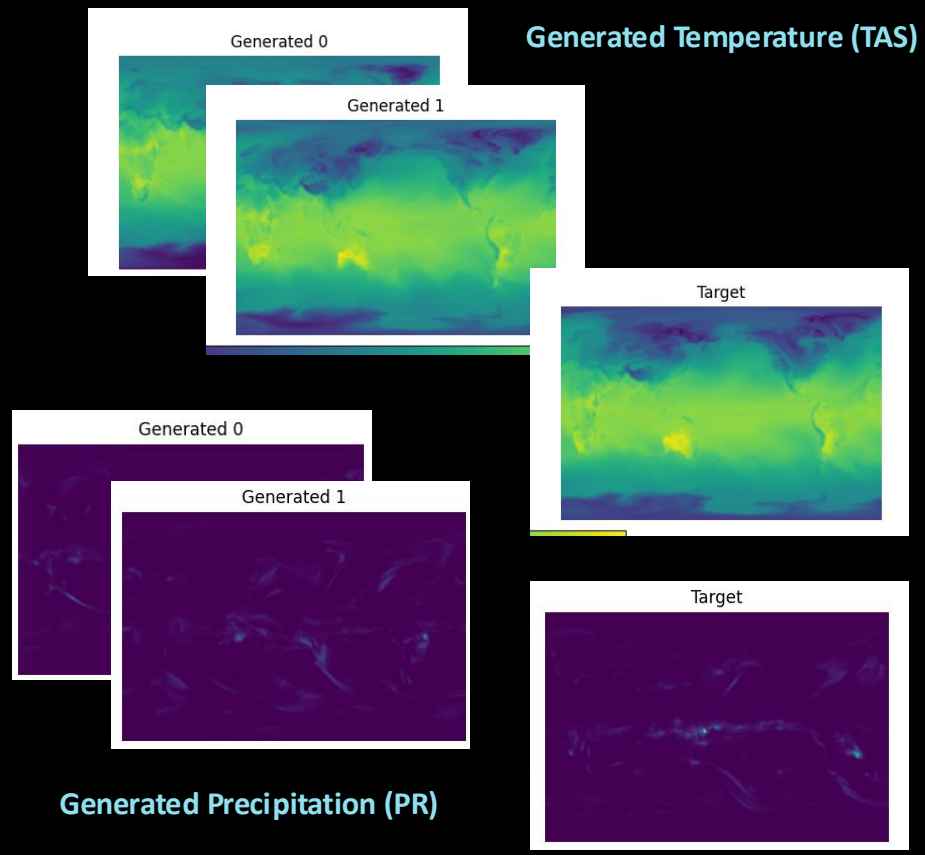


ClimForGe: Daily Climate Emulation with Diffusion Models

Generating 50 years of global daily temperature & precipitation fields in ~100 minutes on a single GPU



Emulate $\mathcal{F}(X, \theta)$



365x	Higher temporal resolution than existing emulators
2x	Spatial resolution vs ClimateBench baselines
8	NRP GPUs for training (V100s); 1 A100 for inference

Climate Analytics Lab

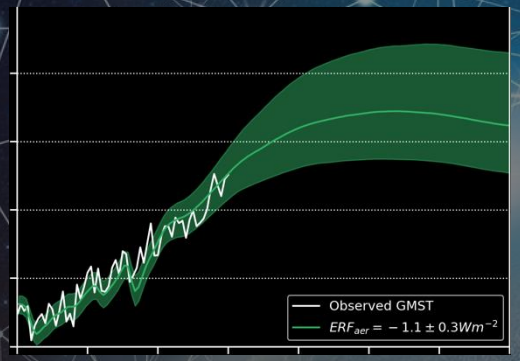
SCRIPPS INSTITUTION OF OCEANOGRAPHY

UC San Diego

HALICIOĞLU DATA SCIENCE INSTITUTE

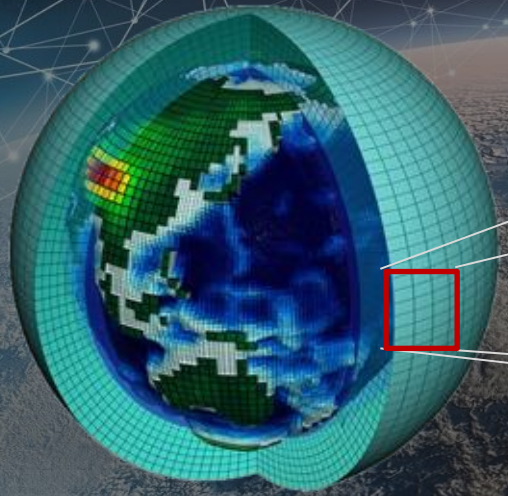
Climate Analytics Lab

Energy balance models



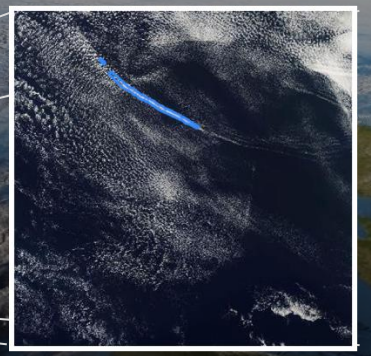
Emulation

Climate models



Distillation / calibration

Observations



NSF

CMIP Coupled Model Intercomparison Project

DARPA

GARA

climate-analytics-lab.github.io – dwatsonparris@ucsd.edu