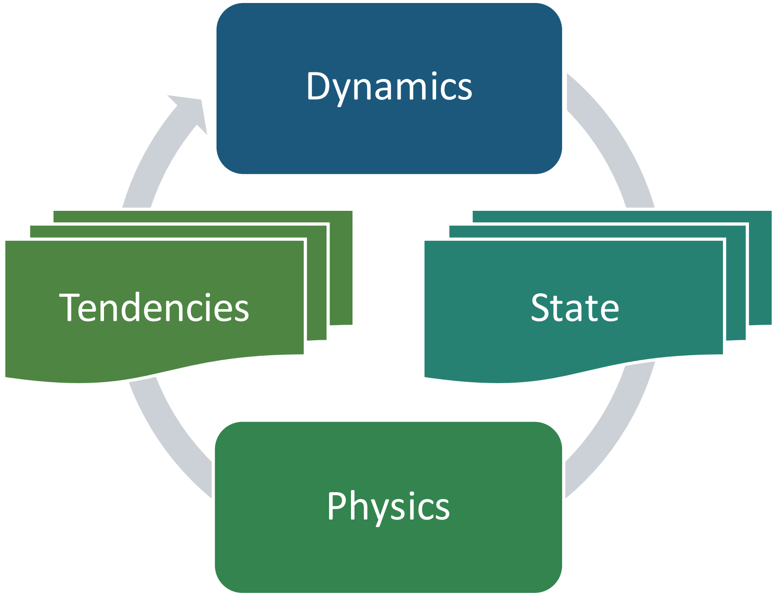
The optimization dichotomy

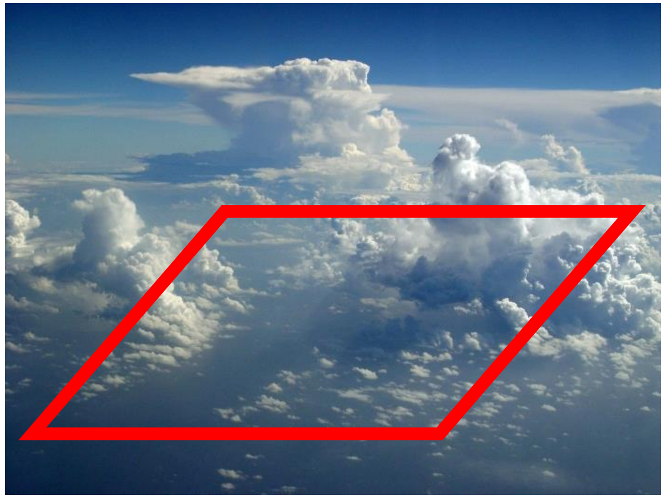
Why is it so hard to improve climate/weather models with machine learning?

Stephan Rasp

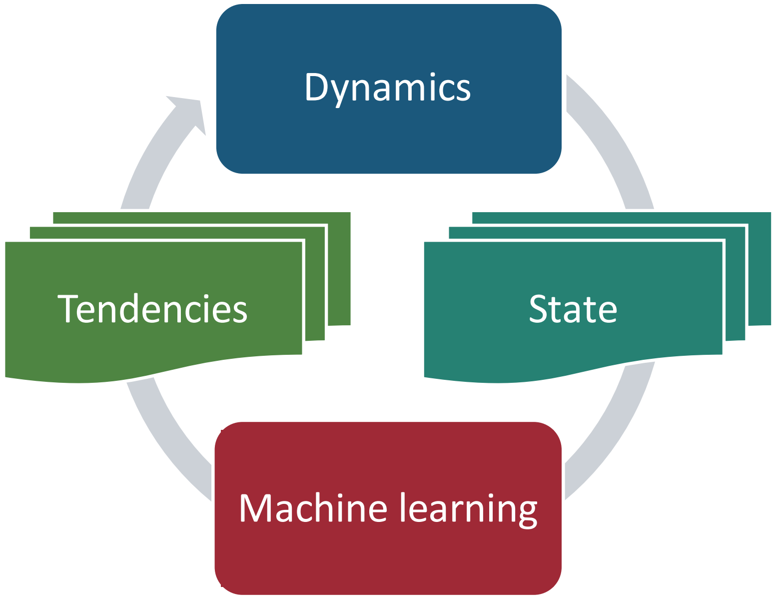
The premise

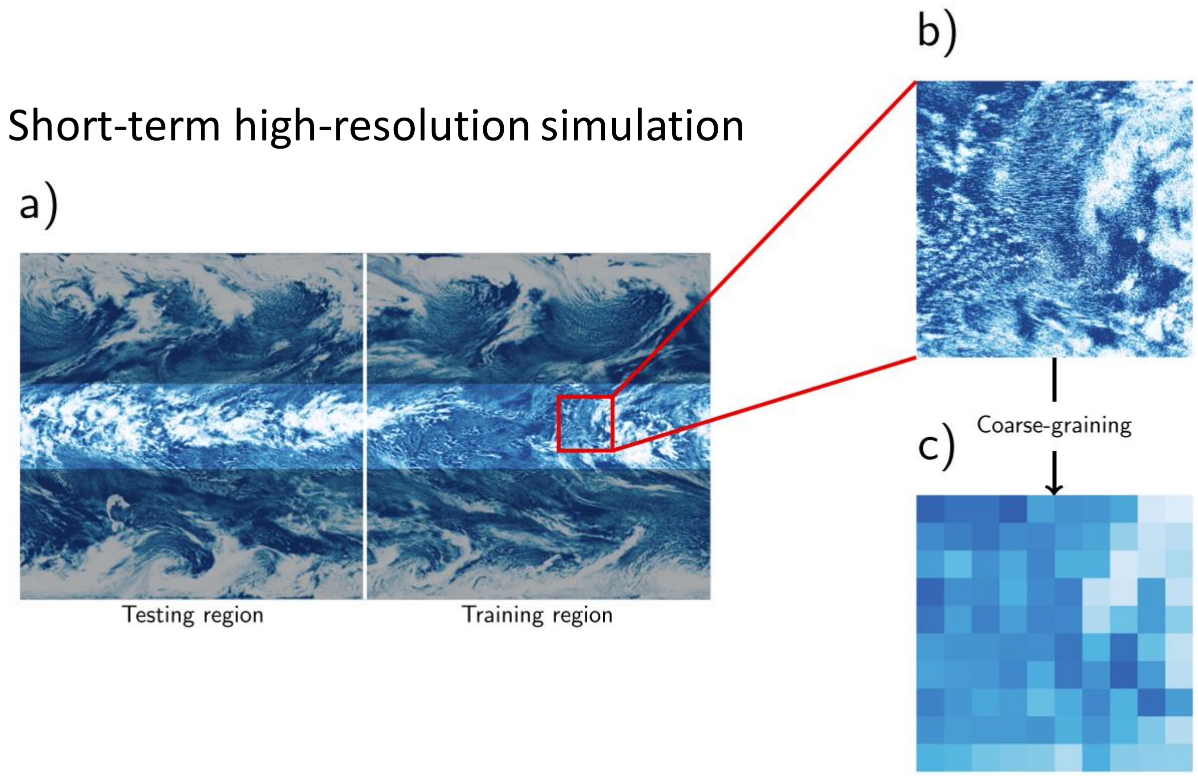
- Climate models continue to have large uncertainties
- Mainly caused by the parameterization of clouds
- Cloud-resolving simulations (km-scale) appear to have reduced uncertainties
- But will remain computationally too expensive for decades
- Could machine learning provide a short-cut?





ML to the rescue?





Challenge 1: Creating the training dataset

Super-parameterization

- *Rasp et al. 2018*
- Emulating an embedded 2D CRM
- Exact problem definition
- Not the real deal

Approximate coarse-graining

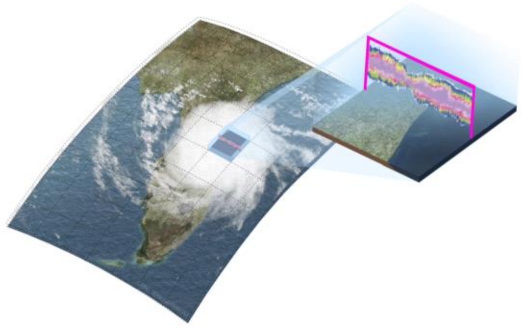
- *Brenowitz and Bretherton 2018/19*
- Compute residual tendencies by subtracting coarse-grained advection
- Model-agnostic
- Sensitive, no conservation properties

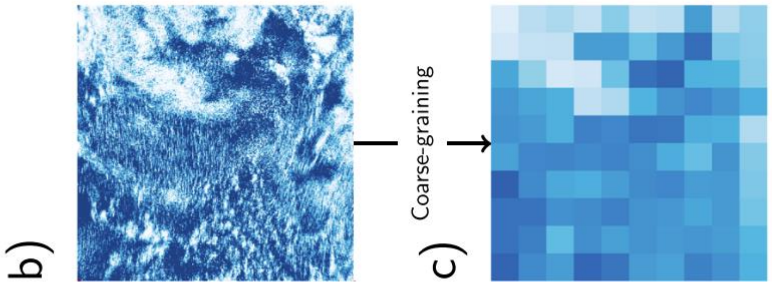
Exact coarse-graining

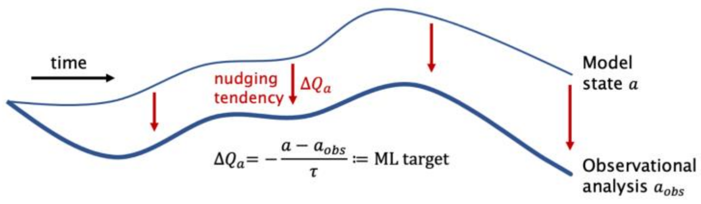
- *Yuval and O'Gorman 2020*
- Compute difference between LR and HR model versions after one time step
- Follows physical constraints
- Yet to test for long time steps, different models

Nudging tendencies

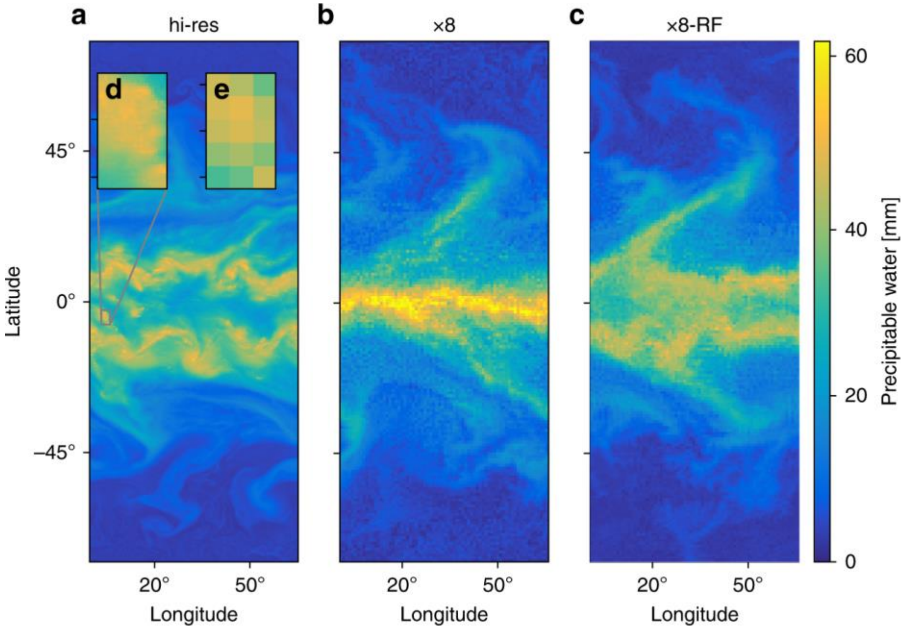
- *Watt-Meyer et al. 2021*
- Nudge model towards reanalysis. Learn nudging tendencies.
- Learns "observations", only correction
- No conservation properties

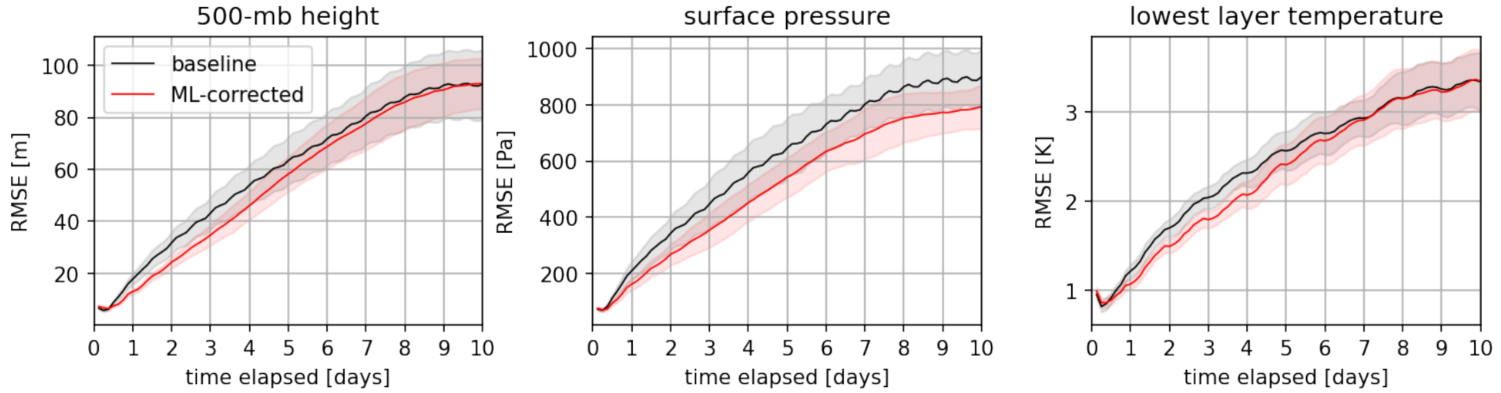






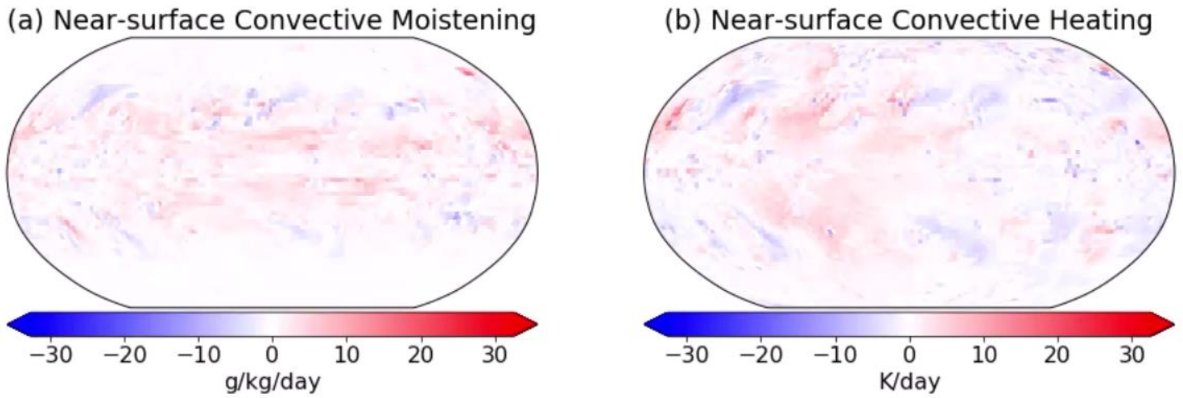
Results in a nutshell: It works…

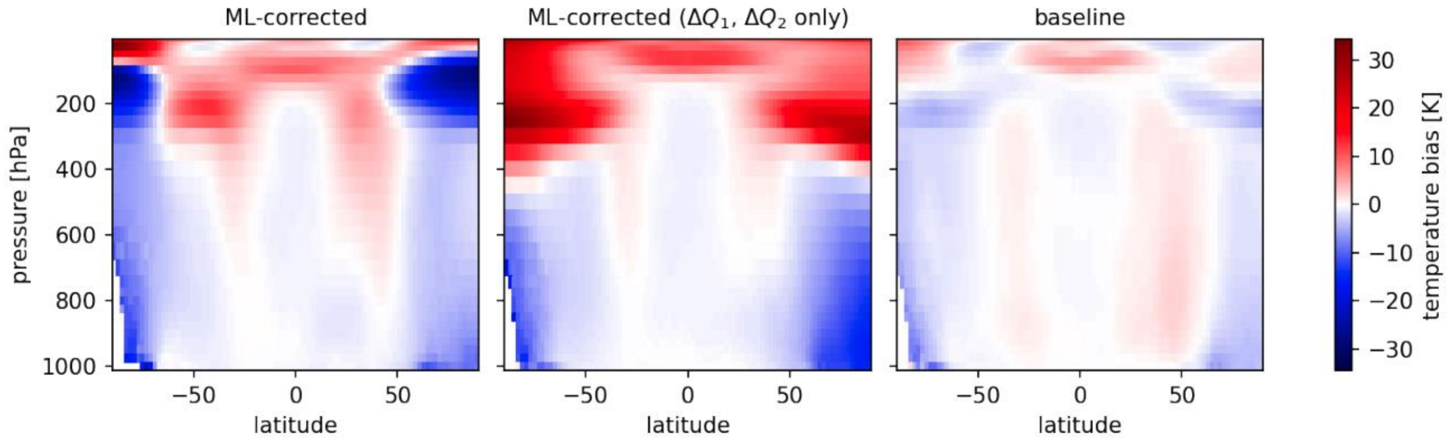




…but with problems.

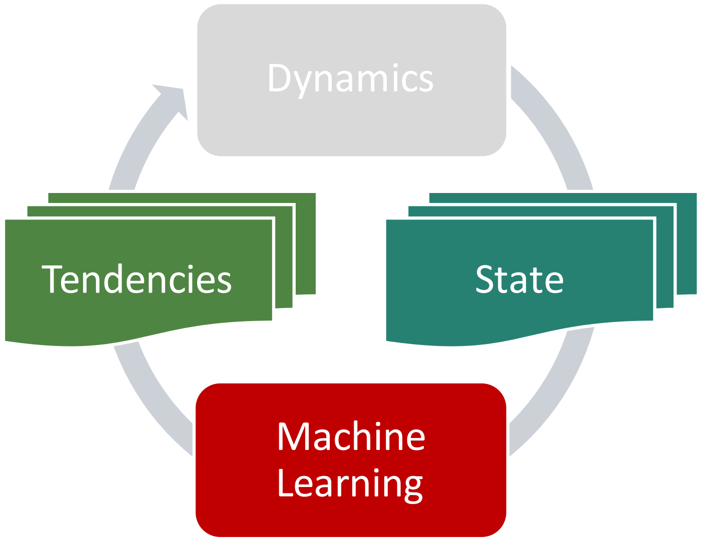
Time to Crash: 1.2day





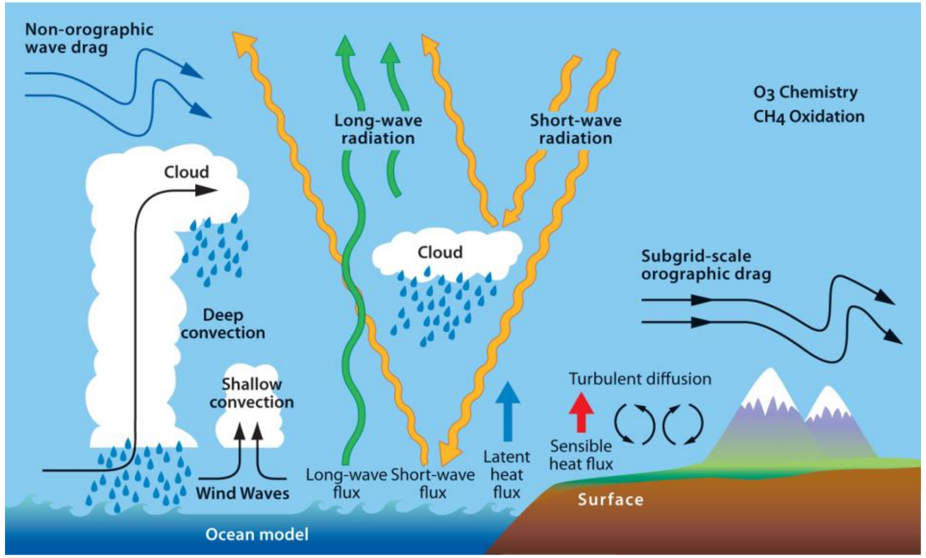
Challenge 2: Optimization dichotomy

Goal:
Better weather/climate forecasts



- We optimize for time step sub-grid tendencies
- We want good weather forecasts or climate statistics
- In-between the two are thousands of time steps with ML-physics interaction
- Traditional parameterizations can be tuned with “physical” parameters
- Can’t do that with ML models

The ML parameterization continuum



Pure emulation

- Replacing existing, slow parameterization with fast ML emulator
- Radiation (ECMWF-Nvidia, etc.)
- Microphysics (Gettelman et al. 2020)
- Gravity waves (Chantry et al. 2021)

High-resolution target

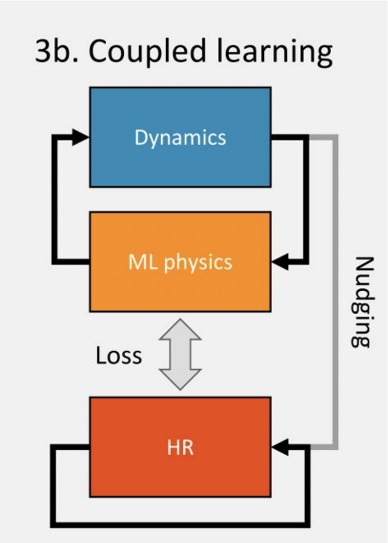
- Coarse-graining is difficult
- Resulting simulations struggle with biases and crashes
- High resolution better but still biased

Observations

- Nudging a potential method
- But still not optimizing for the actual target

Paths forward

Coupled Learning



Rasp, 2020. GMD

Differentiable Physics

∇_θ 

$\downarrow$

$$\nabla_\theta \mathcal{L}$$

cf. 4DVAR (Alan Geer 2021)

Gradient-free approaches (EnKF)

$$\theta_1, \theta_2, \ldots, \theta_n$$

$\downarrow$

$$\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_n$$

$\downarrow$

$$cov(\theta', \mathcal{L}')$$

CliMA
CLIMATE MODELING ALLIANCE

Big picture: Hybrid ML-Physics modeling

Neural networks are very capable function approximators

- (Theoretically) large potential for speeding up/improving complex simulations
- "There exists a neural network that creates better climate/weather simulations."

Neural networks are only good at exactly what they are trained for

- Optimization dichotomy: Currently not the case
- Offline skill $\neq$ Online skill

Challenge: Train a NN in an environment that is close to reality

- Training within the coupled ML-Physics model

Not Secure | mldata.pangeo.io

PANGEO ML DATASETS   Site

WEATHER AND CLIMATE DATASETS FOR AI RESEARCH

PREPROCESSED DATASETS

This is a list of weather and climate datasets preprocessed for AI research. This can include benchmarks, competitions or ML papers with published data. The list is in alphabetical order.

AI FOR EARTH SYSTEM SCIENCE SUMMER SCHOOL HACKATHON

- *Code and Data*: https://github.com/NCAR/ai4ess-hackathon-2020
- *Source*: NCAR, Lawrence Berkeley Lab, and NOAA
- *Description*: 5 challenge problems related to prediction and emulation. GOES challenge problem focuses on predicting lightning from GOES-16 satellite imagery. GECKO-A challenge problem focuses on emulating the GECKO-A chemistry model from a large set of model time series. Microphysics challenge problem focuses on emulating the TAU bin microphysics scheme. HOLODEC challenge problem focuses on estimating rain drop distribution properties from synthetic holographic diffraction patterns. ENSO challenge problem focuses on predicting ENSO from gridded model output.

AMS SOLAR ENERGY PREDICTION CONTEST

- *Code and data*: https://www.kaggle.com/c/ams-2014-solar-energy-prediction-contest
- *Source data*: GEFS forecasts and Mesonet solar observations
- *Description*: Predict total daily solar irradiance from GEFS and Oklahoma Mesonet Data