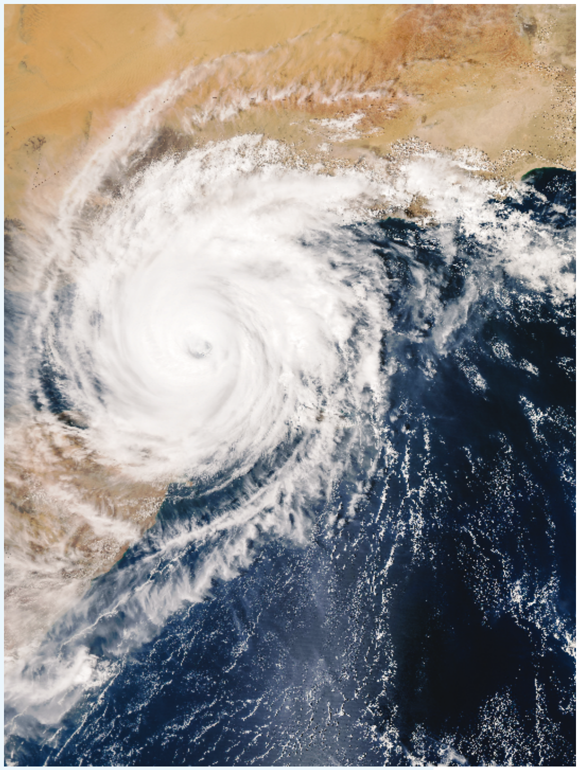


VULCAN CLIMATE MODELING

Improving weather forecast skill and rainfall climatology of FV3GFS using machine learning

Chris Bretherton, Senior Director of Climate Modeling, Vulcan Inc., ChrisBr@vulcan.com or breth@uw.edu

VCM ML group: Noah Brenowitz, Spencer Clark, Brian Henn, Anna Kwa, Jeremy McGibbon, Andre Perkins, Oli Watt-Meyer

GFDL collaborator: Lucas Harris

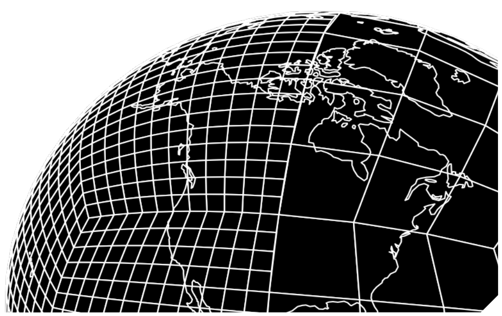
February 2021

VCM pioneers software to improve weather and climate models

VCM = Vulcan Climate Modeling, a philanthropic, open-source project of Vulcan Inc. in Seattle
https://www.vulcan.com/Our-Work/Climate/Advancing-Climate-Science.aspx

Two interlocking groups spun up in mid-2019, partnering with NOAA's Geophysical Fluid Dynamics Lab, using next-gen version of US FV3GFS global weather forecast model:

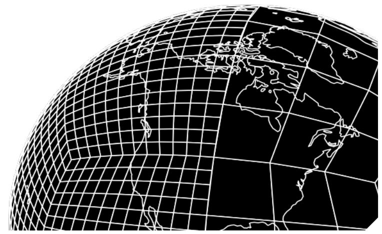
- **"Faster"** (led by Oli Fuhrer): Use a domain-specific language (DSL) to rewrite the model to run faster on modern supercomputers (CPU or GPU), enabling multiyear climate simulations with 1-3 km grids
- **"Better"** (led by Chris Bretherton): Train machine learning (ML) on these simulations and other reference data sets to increase accuracy of rainfall predictions by affordable 25-200 km-grid GCMs. Unique ML niche: Improve full-complexity global atmospheric model (decades of FORTRAN)



Vulcan Climate Modeling (VCM): Who we are

ML group

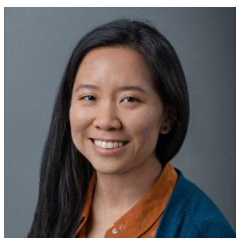
Chris Bretherton



Oli Fuhrer

DSL group

Noah Brenowitz



Anna Kwa

Oli Watt-Meyer

Jeremy McGibbon

Johann Dahm

Eddie Davis

Rhea George

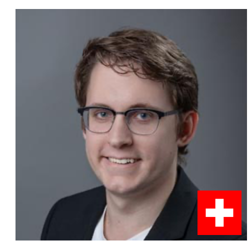
Spencer Clark

Andre Perkins

Brian Henn

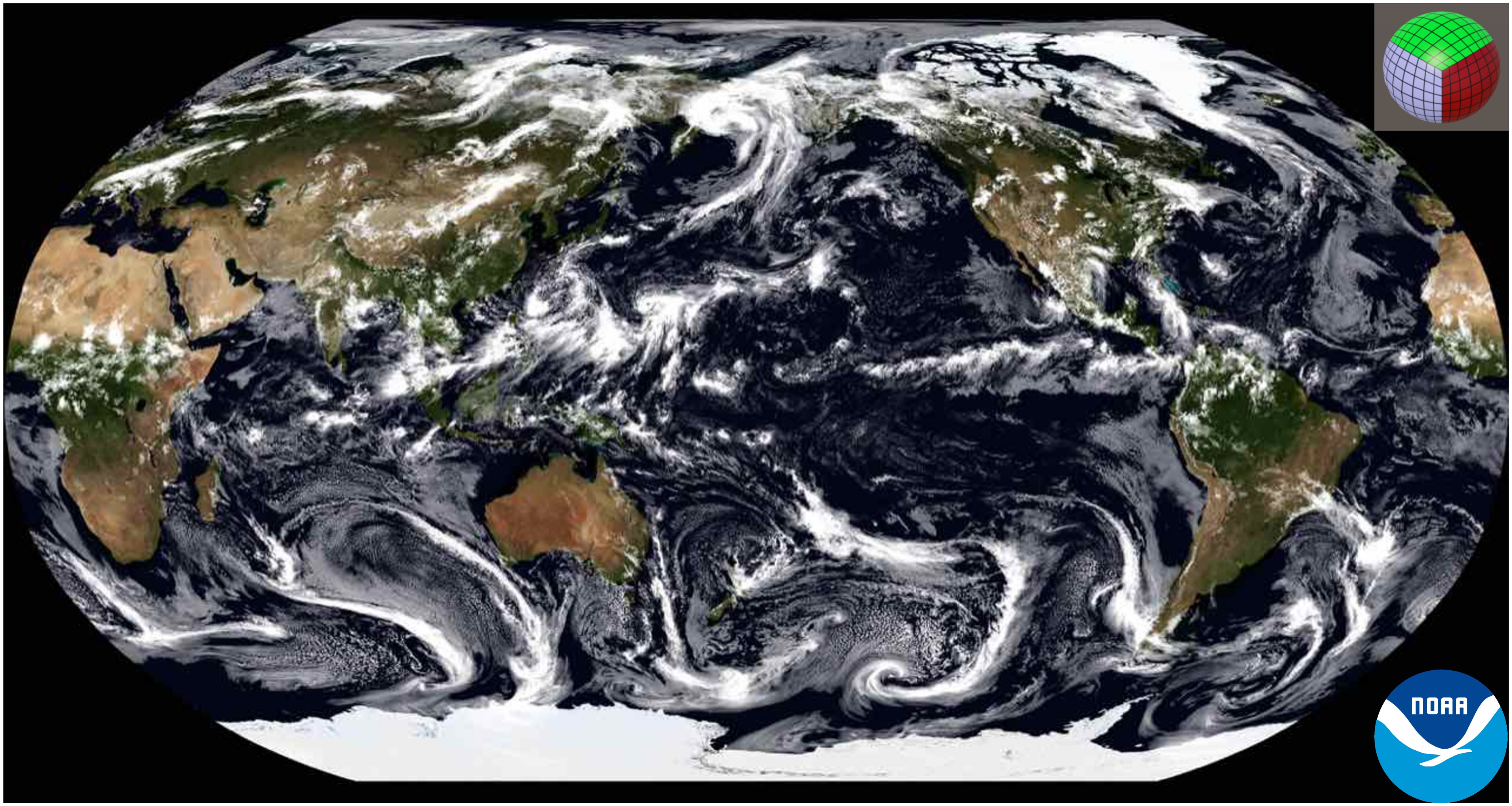
NASA Collaborator: Chris Kung

Oliver Elbert



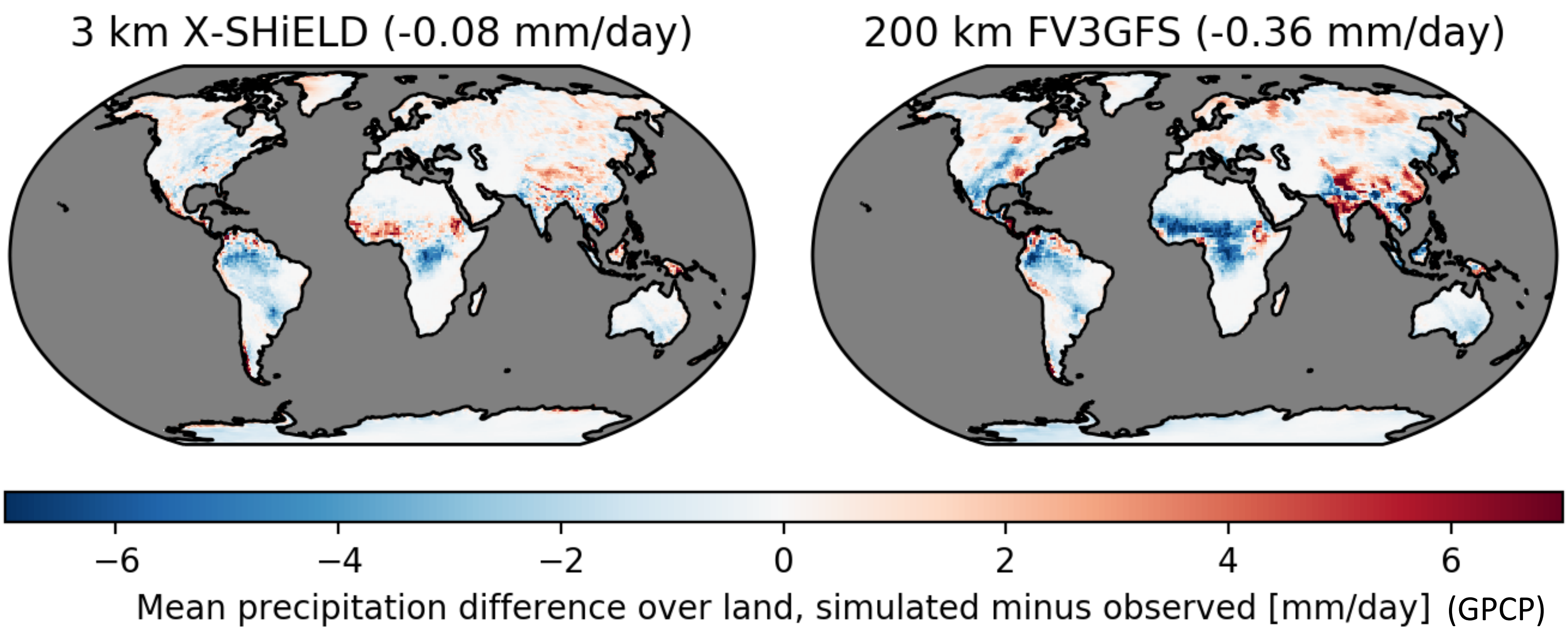
Tobias Wicky

NOAA's global model: FV3GFS (coarse)/X-SHiELD (3 km reference)



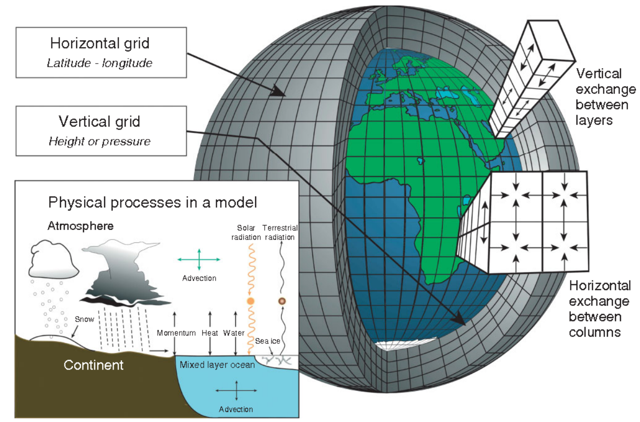
SHiELD 3 km 40-day DYAMOND run, S.-J. Lin and Xi Chen, GFDL

3 km grid enables better mean rainfall simulation than 200 km grid



- 3 km 40-day mean rainfall bias is much smaller over sub-Saharan Africa and Himalayas
- Can we use ML trained on the 3 km simulation to improve coarser FV3GFS simulations?

GCM atmospheric dynamics and parameterized physics



$$s = T + \frac{g}{c_p} z \qquad q = \frac{\text{Mass water vapor}}{\text{Mass dry air}}$$

$$\frac{\partial \overline{s}}{\partial t} + \overline{\mathbf{v}} \cdot \overline{\nabla} \overline{s} = Q_1$$

Apparent heating (K/day)

SW+ LW radiation, latent heating, turbulence etc.

$$\frac{\partial \overline{q}}{\partial t} + \overline{\mathbf{v}} \cdot \overline{\nabla} \overline{q} = Q_2$$

Apparent moistening (g/kg/day)

$$\frac{\partial \overline{\mathbf{u}}}{\partial t} + \overline{\mathbf{v}} \cdot \overline{\nabla} \overline{\mathbf{u}} + \mathbf{f} \times \overline{\mathbf{u}} - \frac{1}{\rho} \nabla \overline{p} = Q_{u,v}$$

Apparent momentum source (m/s/day)

Advection (numerical solver)

ML can be used for all or part of the apparent sources

v is 3D wind, **u** = (*u*, *v*) is horizontal wind, also need z-mom, mass conservation, state eqns.

Column physics approximation

Physical parameterizations depend on the local atmospheric column conditions

- Typical parameterization inputs
  - Profiles of humidity, temperature, winds
  - sunlight, surface properties…
- Typical parameterization outputs:
  - Tendencies of humidity, temperature, winds
  - clouds, rain, snow….

This simplification is most appropriate when grid boxes are much wider than they are high.

We also use it for our machine learning (ML)

Single Atmospheric Column

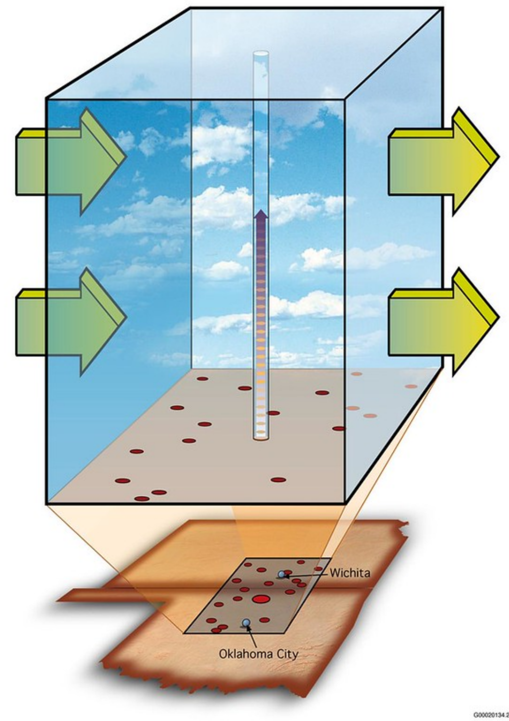


Image courtesy of the U.S. Department of Energy Atmospheric Radiation Measurement (ARM) user facility.

ML Reference data sets: Observations and fine-grid simulations

Goal: Train a column-wise ML correction (or replacement) to supplement physical parameterizations of a coarse-grid climate model so it evolves like a more accurate 'reference' data set

Two possible reference data set types:

- Observational gridded analyses of the atmospheric state (good for weather)

  A weather forecast model (GFS) synthesizes and interpolates countless diverse observations

  Limited to recent weather variability in the satellite era; best if assimilating model = coarse model

- Fine-grid global model (good for climate and weather)

  Coarse-grain the fine-grid data to coarse grid using horizontal block averaging

  Can leverage hyper-realistic but computationally expensive global reference simulations

  Can do short reference simulations over a range of climates for ML training and testing

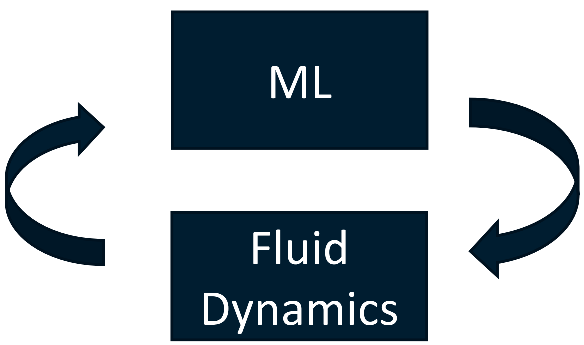
  Prior work suggests this can work in idealized settings such as a flat water-covered earth

  Only as good as the fine-grid simulations, which will still have biases

ML Evaluation: Online ≠ Offline

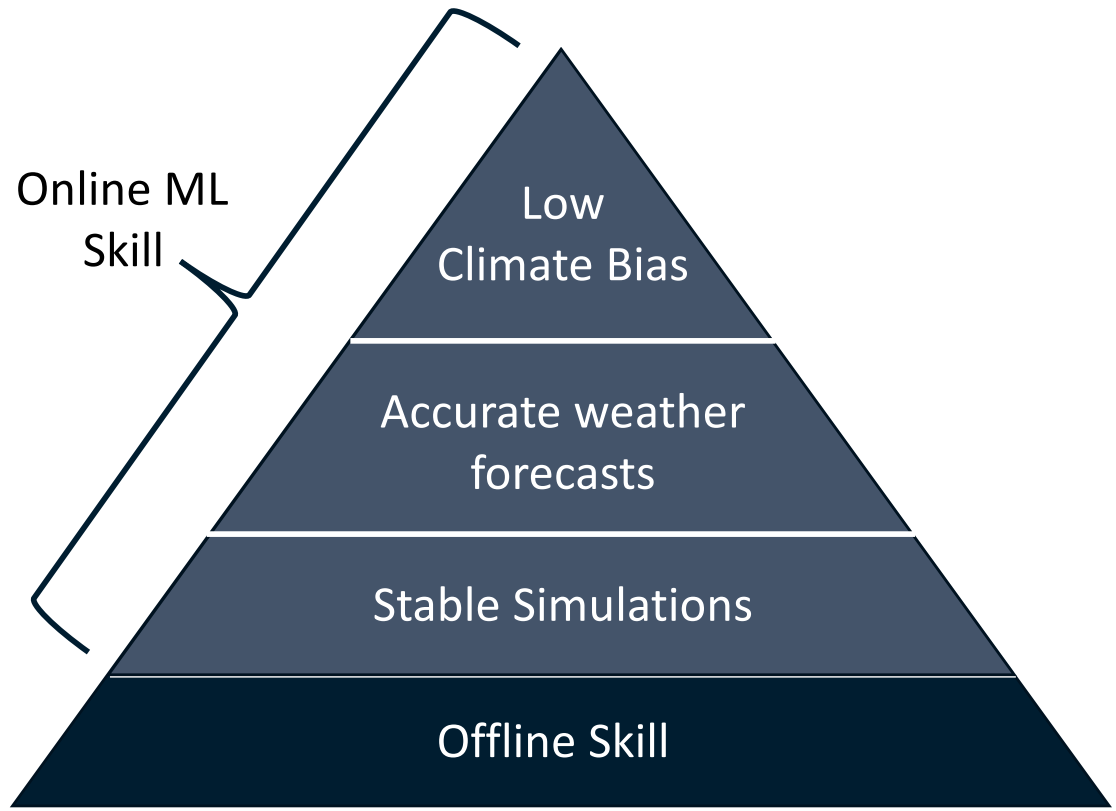
Application of ‘Hybrid’ ML, i. e. interactively coupled to other model components

Coupled to Fluid Dynamics



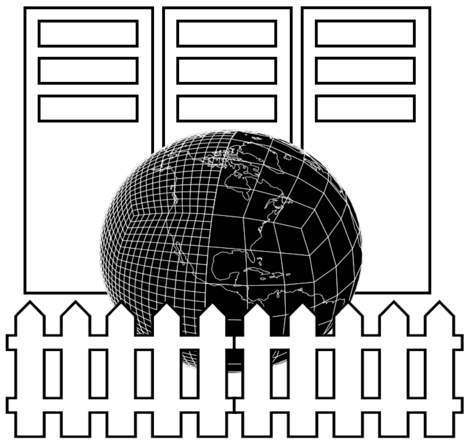
Training ≠ Testing
(offline) (online)

ML has a challenging ‘hierarchy of needs’



VCM Workflow/Cloud computing overview

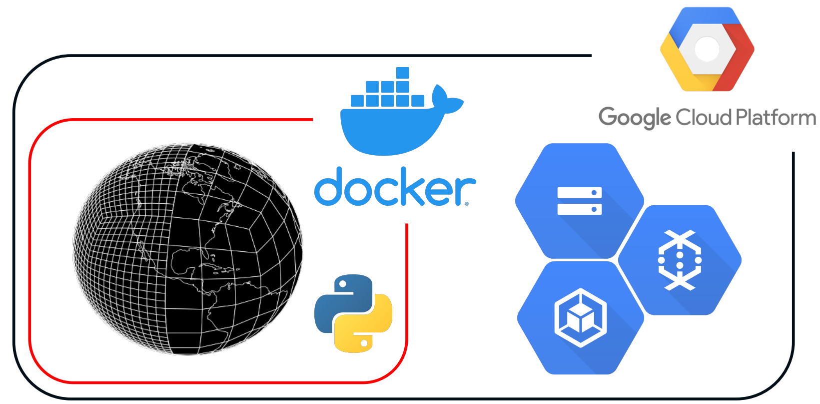
GAEA



Constraints

- GAEA is inaccessible to Vulcan due to NOAA's HPC firewall
- FV3GFS is a UFS community model; X-SHiELD is GFDL-customized
- Fortran is not ML language of choice
- Vulcan has no compute infrastructure

VCM-ML Needs: Reproducibility and Scalability



Solutions

- VCM ported public release of FV3GFS to reproducible/portable docker images
- We wrapped it in Python for flexibility [*McGibbon et al. 2021, ACP, submitted*]
- In-line SHiELD coarsening on GAEA piped into ML workflow on Google Cloud Platform

Tendency-difference method

- $a(t, x, y, s)$: advected scalar (e.g. humidity). $a^r(t, x, y, s)$ is from the reference model
- Uncorrected coarse model ($*$ = no ML):

$$\frac{\partial a^*}{\partial t} = -\mathbf{v}^* \cdot \nabla a^* + Q_a^p,$$

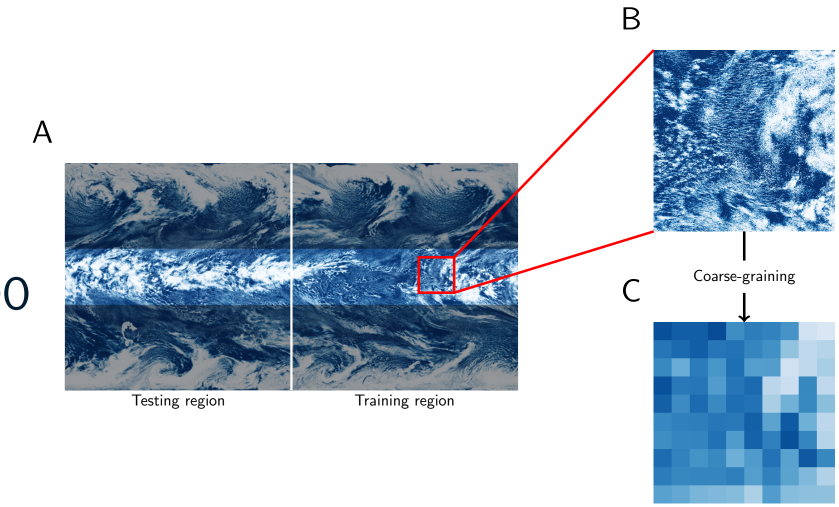
can include no physics (Q_a^p = 0), all physics on, or a subset of parameterizations.

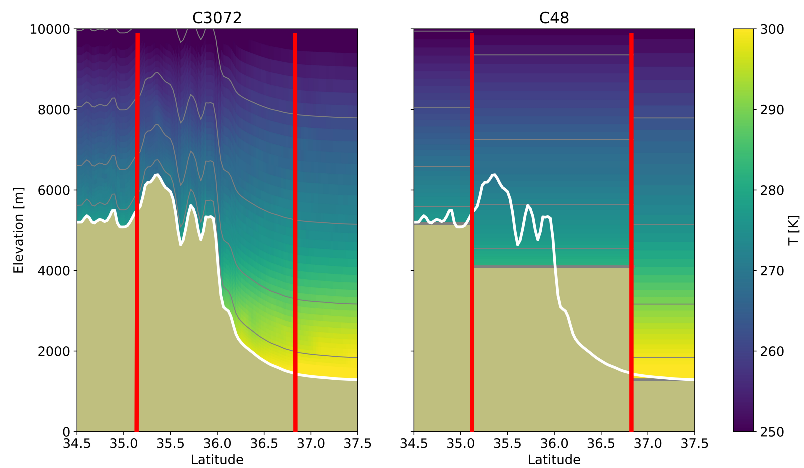
- Define the **tendency difference** $\Delta Q_a = \frac{\partial a^r}{\partial t} - \frac{\partial a^*}{\partial t}$ of identically initialized reference – coarse models
- Machine-learn it as a function of atmospheric column state: ΔQ_a^{ML}
- Then ML-corrected model evolution should approximately match the reference model:

$$\rightarrow \qquad \left(\frac{da}{dt}\right)^{ML} = \frac{\partial a^*}{\partial t} + \Delta Q_a^{ML} \cong \frac{\partial a^r}{\partial t}$$

VCM historical progression

- Starting point: Aquaplanet results of Brenowitz and Bretherton (2018, 2019) using TD method
- Initial ML goal: Similar approach for real-geography 200 km FV3GFS, trained with coarsened 3 km X-SHiELD
  - Use random forests rather than neural nets for stability
  - Learn the combined physics, or just a correction to the physics
  - Coarsening over orography was a challenge
  - Large initialization transients in w over orography contaminate ML
- Method abandoned in mid-2020 in favor of a corrective ML approach trained by nudging the target coarse model to a reference model.

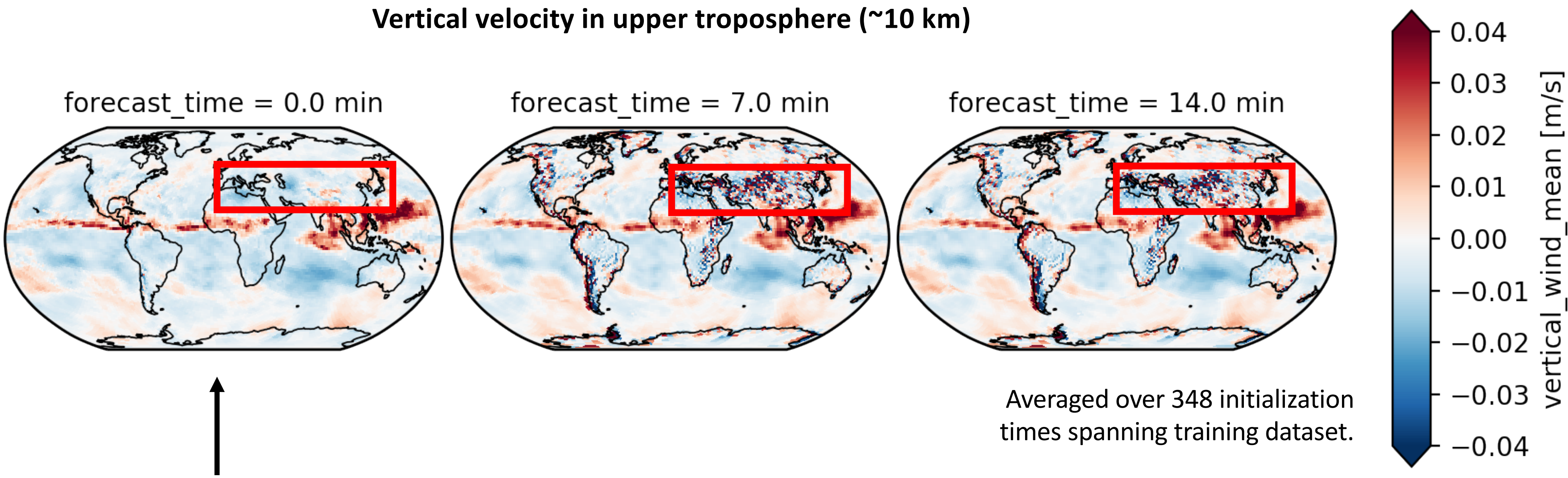




Initialization shock

- When initialized with coarsened fine-grid data, the coarse model winds take several hours to settle into a realistic 'slow manifold' evolution around mountains
- This adds 'noise' to the fine-coarse tendency difference, degrading ML performance

Vertical velocity in upper troposphere (~10 km)



Averaged over 348 initialization times spanning training dataset.

Fine resolution model coarsened to 200 km resolution

Nudging tendency method for ML training

Goal: coarse model in which scalars $a^c(t, x, y, \eta)$ track reference $a^r(t, x, y, \eta)$:

$$\frac{\partial a^c}{\partial t} = -\mathbf{v}^c \cdot \nabla a^c + Q_a^p + \Delta Q_a^{ML} \cong \frac{\partial a^r}{\partial t}$$

Method: Nudge no-ML coarse model to the reference:

$$\frac{\partial a^n}{\partial t} = \underset{\text{Coarse advection}}{-\mathbf{v}^n \cdot \nabla a^n} + \underset{\text{Physics params}}{Q_a^p} + \underbrace{\frac{a^r - a^n}{\tau}}_{\Delta Q_a^n}$$

- Machine-learn the nudging tendencies ΔQ_a^n of temperature, humidity and wind profiles as function of coarse model column state

- Nudging time scale $\boldsymbol{\tau}$: A few hours for a 200 km coarse model
  - slow enough to damp relaxation 'whip-lash' in a^n
  - fast enough to keep a^n near the reference state a^r

Nudge-to-obs approach

[Watt-Meyer et al. 2021, GRL, submitted]

- Use NCEP observational analysis as reference (accurate for satellite era 1980-present).
- Analysis is available 6-hourly; use nudging timescale $\boldsymbol{\tau}$ = 6 h.
- Train on a full year of data (2015), including nudging tendencies of u, v as well as T, q.
- Limiter to prevent ML correction from creating negative humidity in forecasts.
- Performance using random forest encouraging compared to baseline 200 km FV3GFS:
  - ✓ ML-corrected forecast model improves weather skill and reduces rainfall bias
  - ✓ It runs skillfully over an independent 2016 validation period with different SSTs
  - ✓ Other ‘climate’ biases (e. g. surface T) are mostly comparable to baseline simulation
  - ✓ We are starting to achieve comparable prognostic performance with neural net-based NN; this has required more effort in regularization and hyperparameter tuning.

Machine learning: model training

From available training output (e. g. nudging tendencies every 5 hrs for 365 d), subset to:

Training set = 2.2M samples (160 times x 13824 grid columns)

Test set = 1.2M samples (90 times x 13824 grid columns)

Features

air temp vertical profile

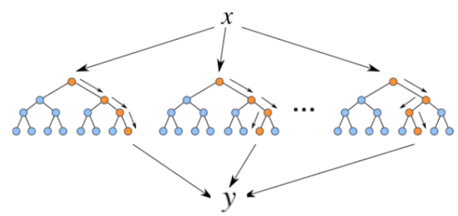
specific humidity vertical profile

additional potential features, ex.

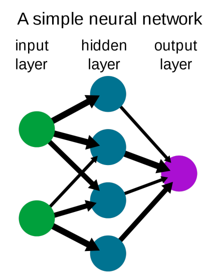
insolation
land/sea mask
surface elevation

Profiles of u, v

Random Forest



Neural Network



Predictions

dQ1 vertical profile

dQ2 vertical profile

Profiles of dQu,v

ΔQ_a^{ML}

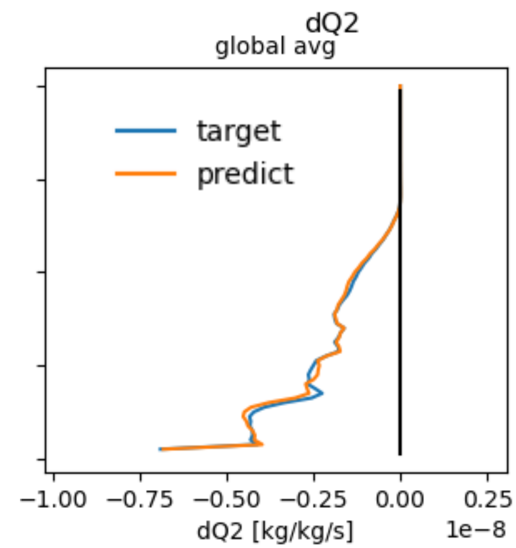
Results shown use RF, which involves less TLC;
we can now achieve comparable forecast skill with NN

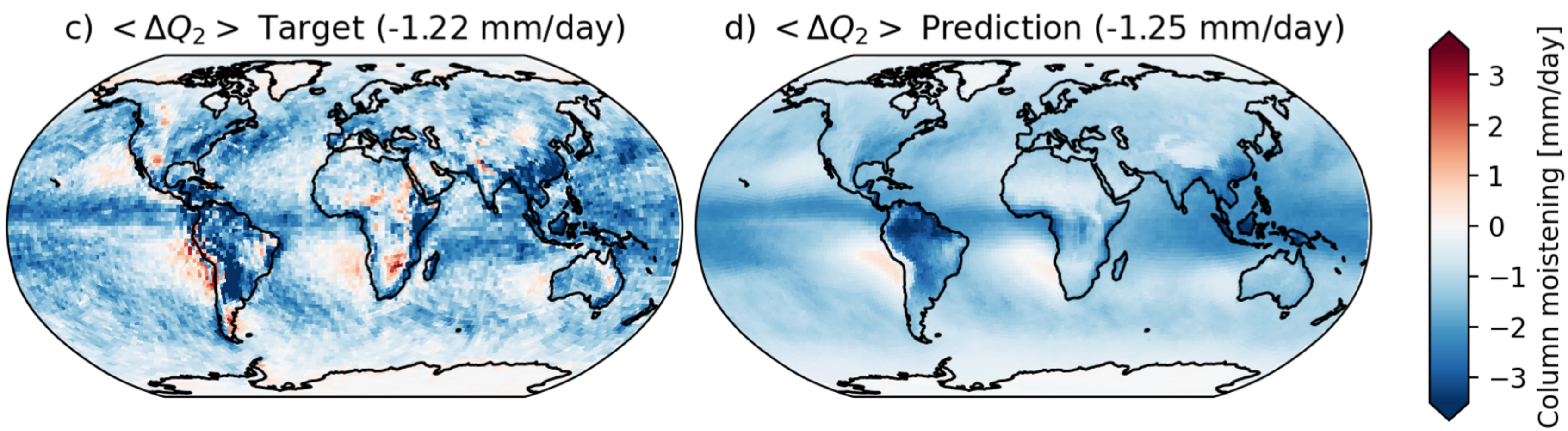
Offline comparison of random forest skill for nudge-to-obs

'Target' corrective q nudging tendency profile:

- Well-learned by random forest
- Large (1 mm/d) global-mean column drying; baseline model rains too little in real climate
- Only explains 5-20% of dQ2 variance

Use 2016 for offline testing:
Global means from a sample of times



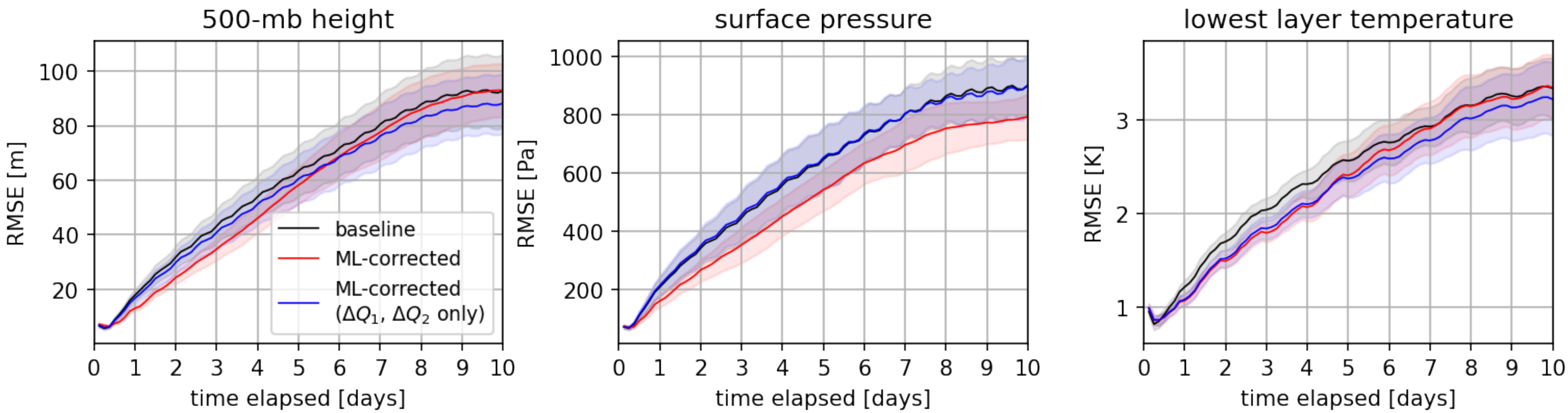


Animation of physics and nudging tendencies in a 10-day forecast

- Physical parameterization tendencies (pQ) reflect radiation, PBL, and convection/clouds
- ML learns the mean spatial dQ patterns well and variability about those patterns less well
- Corrective ML tendencies are smooth and slowly varying functions of column state
- They prevent drift of the baseline model to its biased internal climatology

Improved weather forecast skill

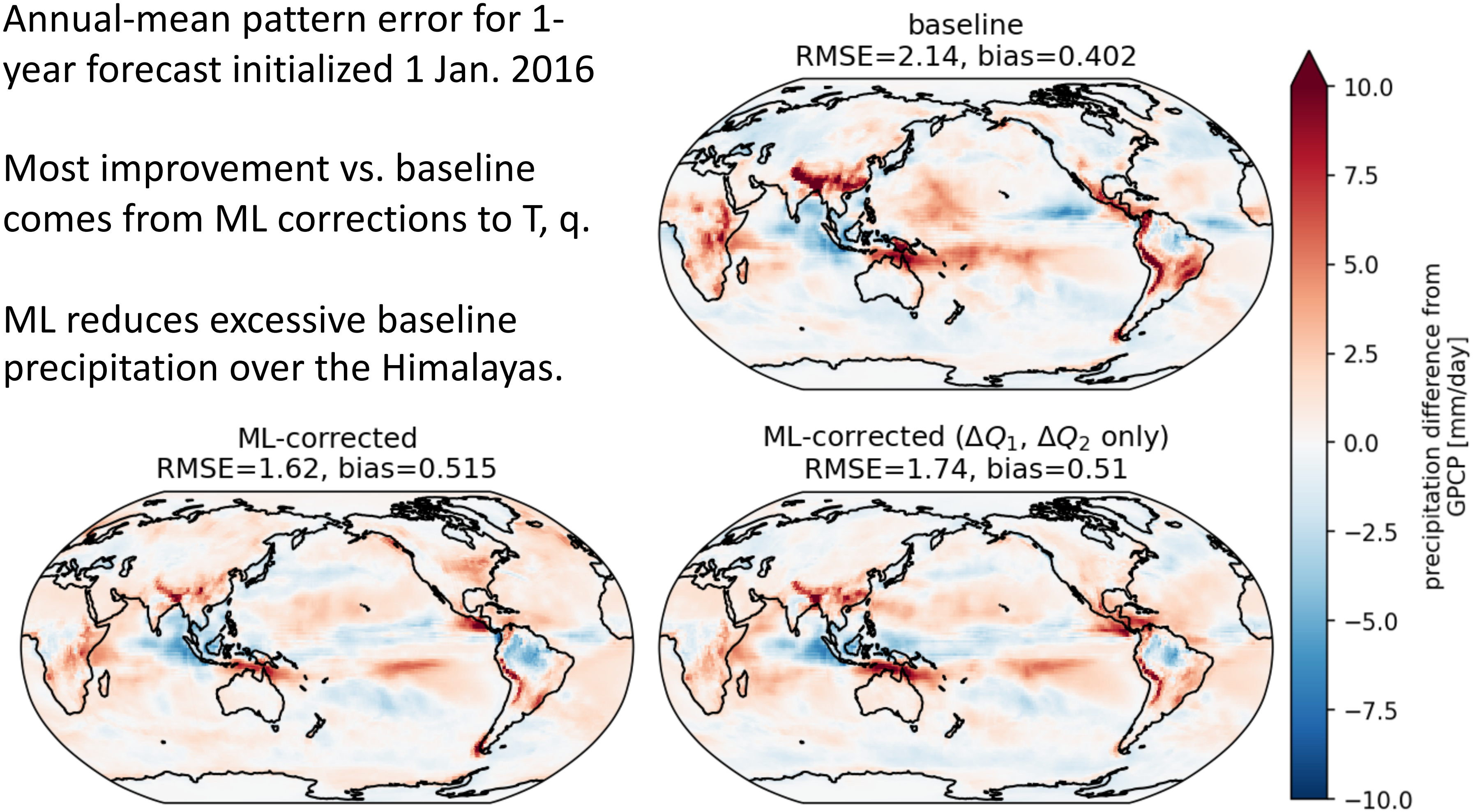
RMS error growth of 12 forecasts, initialized on the first day of each month of 2016



- ML of T, q, u, v nudging trounces baseline forecast of Z500, surf. pressure, Tsurf.
- ML of T, q nudging ($\Delta Q_1, \Delta Q_2$ only) improves forecast of Tsurf and (slightly) Z500.

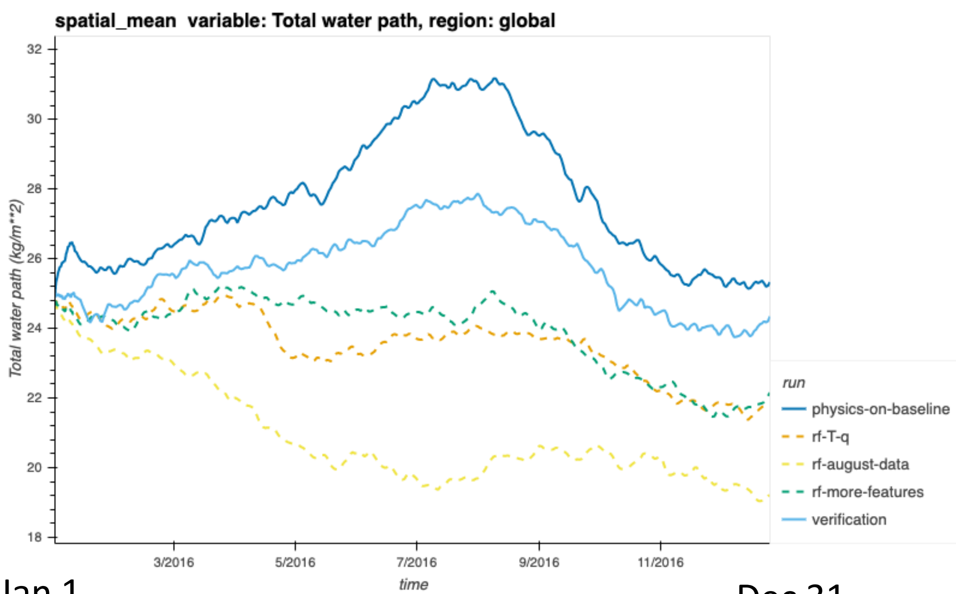
Nudge-to-obs reduces annual mean precipitation errors by 20%

- Annual-mean pattern error for 1-year forecast initialized 1 Jan. 2016
- Most improvement vs. baseline comes from ML corrections to T, q.
- ML reduces excessive baseline precipitation over the Himalayas.



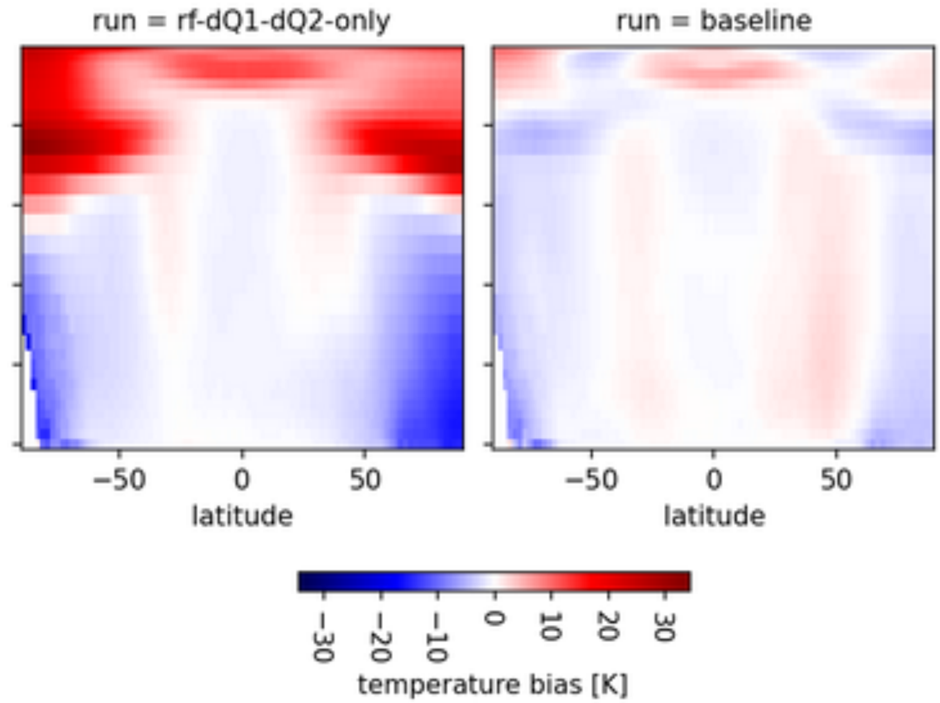
Some other nudge-to-obs climate biases are small, some large

Precipitable water (good)

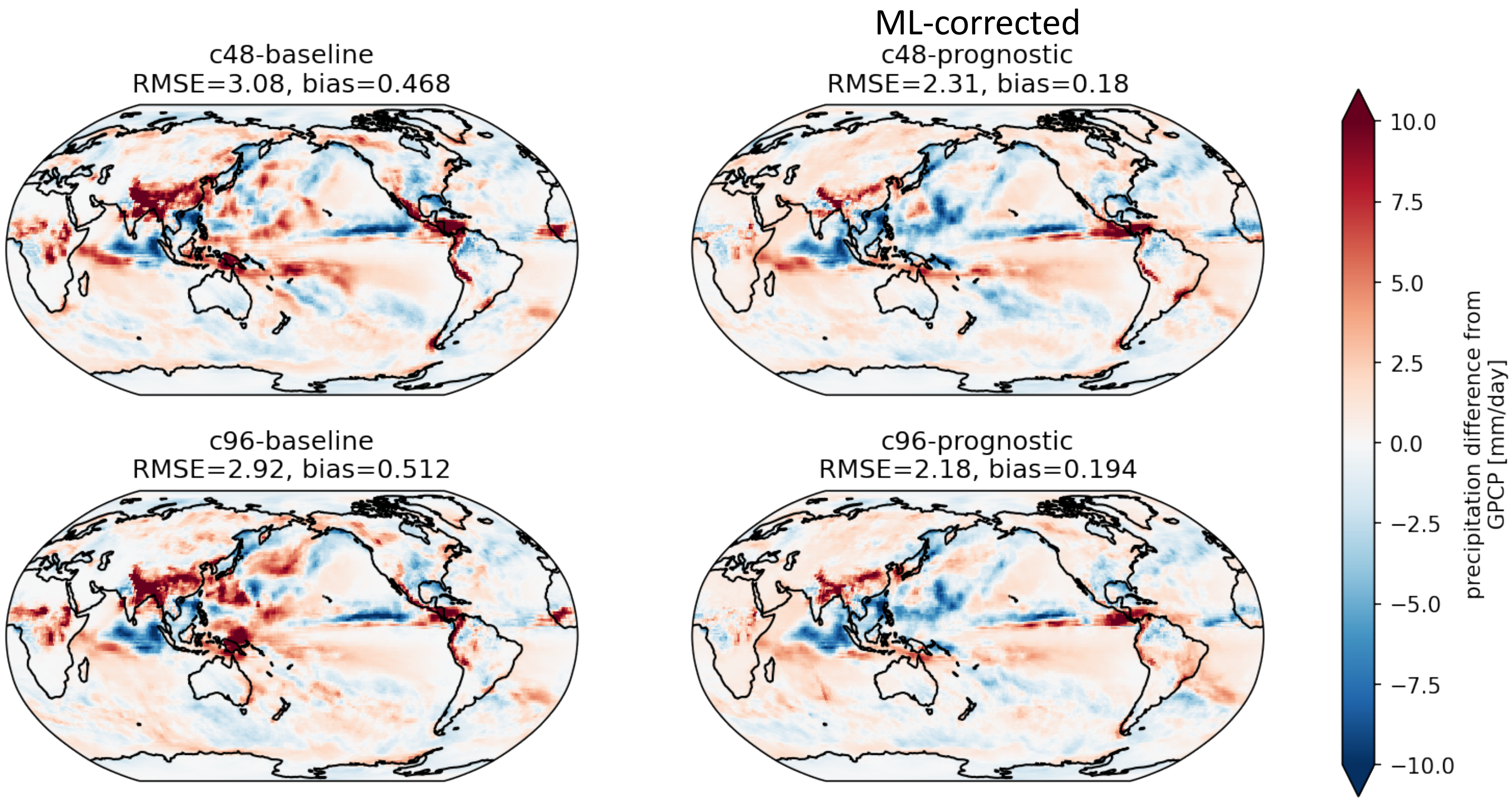


Jan 1 Dec 31

Polar temperature (bad)



Also works at 100 km (C96) coarse resolution



- 25% reduction in RMSE of 1x1° time-mean precipitation bias over 1 Aug-9 Sept 2016.
- Currently working on C384 testing

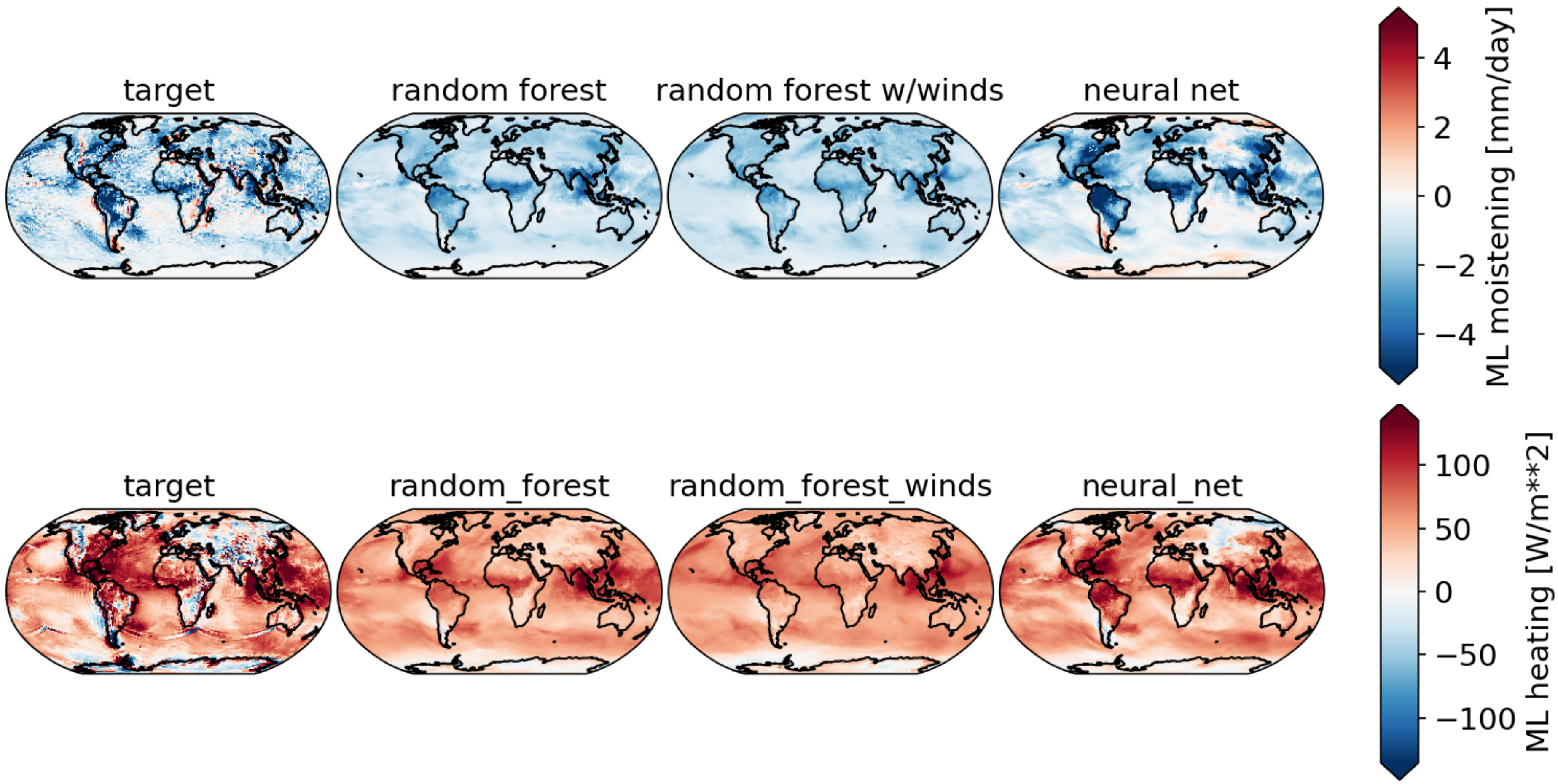
Nudge-to-fine corrective ML

- Like nudge-to-obs, but reference is pressure-level coarsened output from a fine-grid simulation (3 km X- SHiELD or 25 km SHiELD/FV3GFS) at 15 min – 3 hr frequency:
  - 3 km: Thunderstorm and mountain-resolving, no deep cumulus or GWD scheme
  - 25 km: Fine-grid conventional AGCM, can be economically run in multiple climates for years.

- Nudging time scale **τ** = 3 hrs chosen for C48 coarse model
  - Slow enough to keep a^n near its 'slow manifold'
  - Fast enough to keep a^n near the coarsened fine model $\bar{a}$
  - Machine learn the nudging tendency ΔQ_a^N as function of coarse model column state
  - Longer nudging timescales up to 24 hours reduce bias but blur the diurnal cycle of rainfall
- Results with 3 km fine model have ML corrections to heating, moistening, not winds

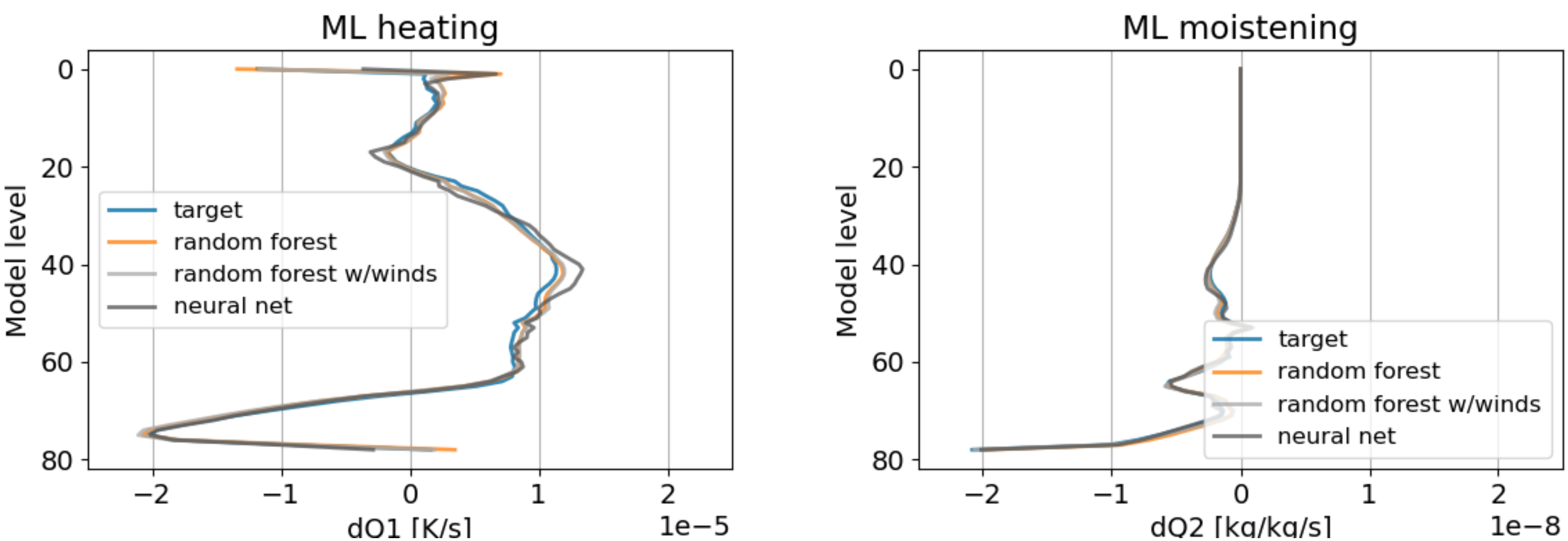
Offline comparison of RF and NN skill for 200 km/3 km nudge-to-fine

Vertical Day 6-40 column-integrated time means



Offline comparison of NN and RF skill for 200km/3km nudge-to-fine

Global time-mean vertical profiles

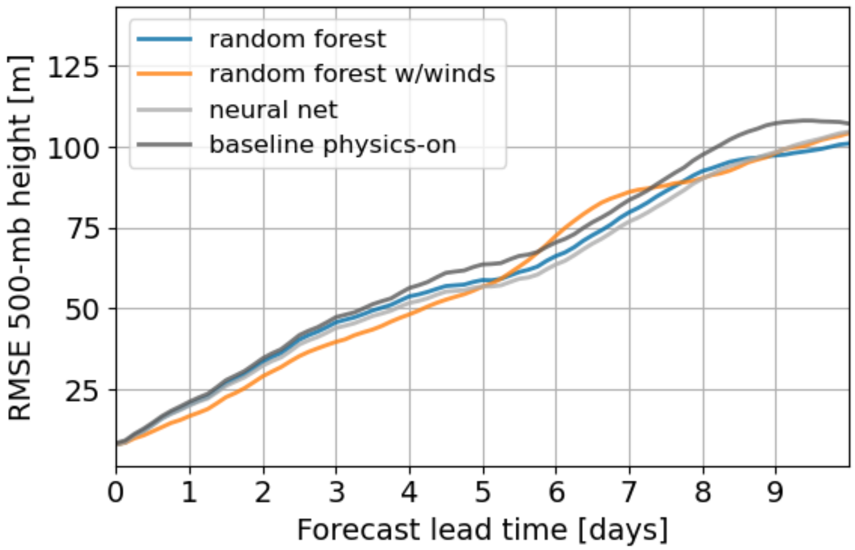


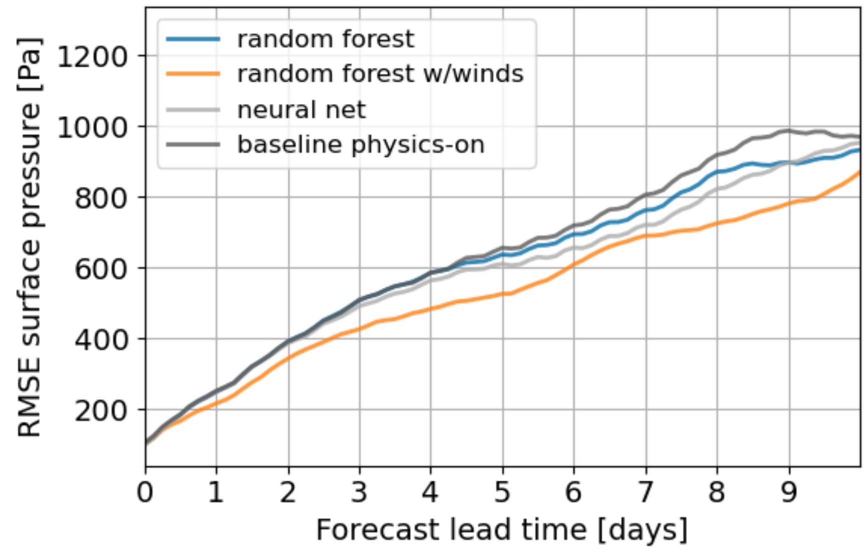
Even in global-time-mean, the ML corrective tendency profiles are significant (net drying and mid-tropospheric heating) and are well-learned by both RF and NN

200 km/3 km nudge-to-fine forecast results

- Verification, like training is limited to the 40 d of 3 km fine-res GSRM simulation
- Both forms of ML increase coarse-model forecast skill in:
  - 10 day weather forecasts
  - Day 5-40 mean rainfall distribution and daily cycle over land
- Both forms of ML still lead to undesirable mean-state biases, e. g. too little ice cloud but too much rain over most land areas, in part associated with land surface feedbacks. Improved ML training protocol will address these biases.

Improved weather forecast skill

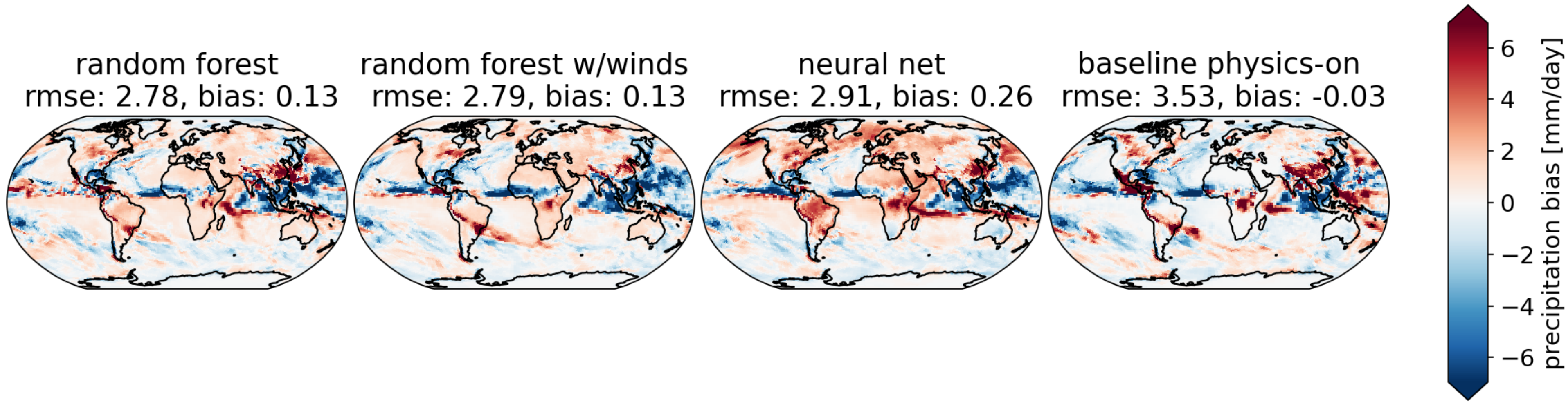




200 km simulations initialized with coarsened state of 3 km model on August 5, 2016

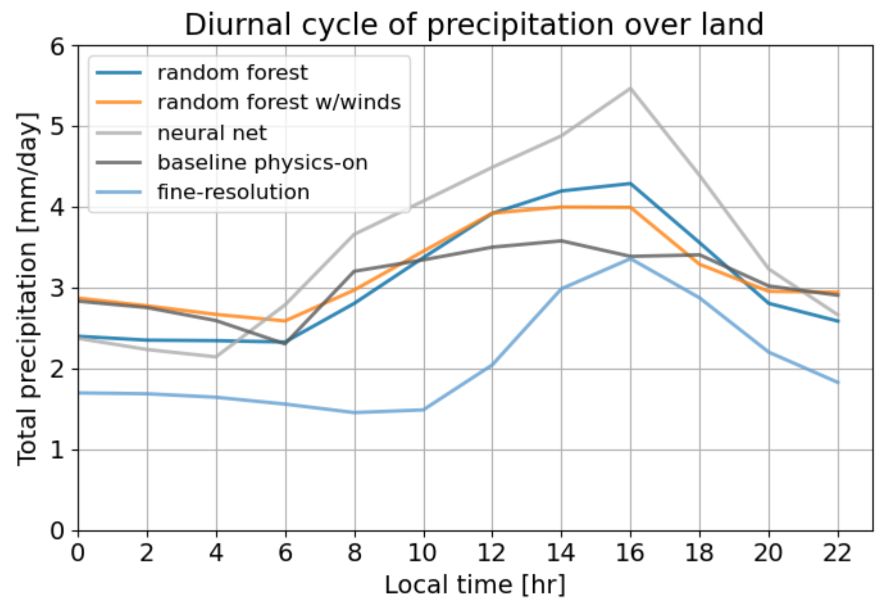
Improvements in precipitation bias

- RMSE of 35-day mean precipitation pattern improved by 20% (RF)/15 % (NN)
- However, errors increase in dry subtropics, and both have global wet biases



Improvement to diurnal cycle over land

- Afternoon rainfall maximum over warm land is strong in the fine simulation, but poorly represented by the baseline coarse model
- Ideally, the ML should be able to correct coarse model to the fine-res diurnal cycle
- Since weather variability in rainfall dominates its diurnal cycle, there is no guarantee this will work
- Encouragingly, the nudge-to-fine approach makes an afternoon rainfall peak, though with a high bias in time-mean rainfall



Progress toward climate change prediction?

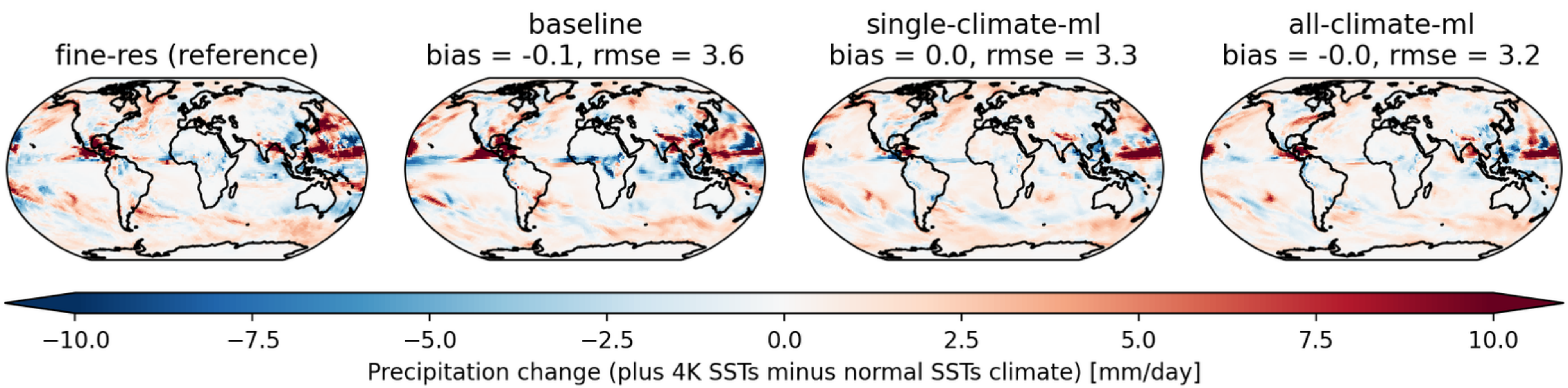
- 40-day 25 km fine simulations with control, warm (+4 K), and cold (-4 K) SSTs (GAEA)
- Goal: Train one nudge-to-25 km fine ML scheme across a range of climates to achieve consistent coarse-res (200 km) climate and climate change response
- T,q,u,v nudging tendencies learned and used correctively in coarse simulations

Results:

- Forecast skill: RF trained on samples of data around the annual cycle of all three runs beats 200 km baseline Z500, SLP skill vs. fine-res in all 3 climates
- Climate skill: RF-corrected run has a moist bias in rainfall over land vs. fine-res reference, same as with 3 km fine res reference.
- We have extended this protocol to 1 year training and testing in the 3 climates, but some forecasts experience rapid ML-induced climate drift after 6-9 months.

200 km/25 km nudge-to-fine 4K – ctrl rainfall trends

- Some apparent improvement over baseline in capturing fine-res trends

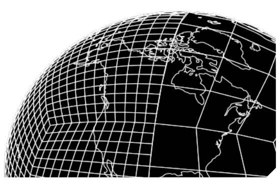


Emulation

- We have also worked on emulation of FV3GFS physical parameterizations using ML (mostly NNs)
  - TKE-EDMF PBL scheme
  - Deep convective trigger
- We have not found it straightforward to develop emulations that work skillfully and without bias in a forecast setting, but we continue to explore this, inspired by more successful results of A. Belochitski and V. Krasnopolsky at EMC

Conclusions

- Vulcan Climate Modeling, in partnership with GFDL, has developed a sophisticated cloud-based workflow for applying Python-based machine learning to FV3GFS, a full-complexity operational global atmospheric model.
- We have developed corrective machine learning schemes that improve the skill of global weather forecasts and their time-mean rainfall distribution vs. reference datasets from global observational analyses or fine-res simulations.
- The ML correction helps an affordable coarse-grid model capture precipitation trends in an expensive fine-grid model due to a climate change.
- Current short-term goals:
  - reduce TOA and surface flux biases that impact coupling to land and ocean
  - systematically reduce other climate biases (e. g. 200 hPa T)
  - demonstrate our ML methods at higher coarse-model grid resolution
  - progress further with neural net learning.



Submitted papers in review:

Nudge-to-obs method: Watt-Meyer et al. 2021, *GRL*,
Preprint doi:10.1002/essoar.10505959.1
Python wrapper for FV3GFS: McGibbon et al. 2021, *GMD*