First steps toward a machine-learning based moist physics parameterization by coarse-graining

Noah Brenowitz*, Chris Bretherton*, Spencer Clark*+, Lucas Harris+, Brian Henn*, Anna Kwa*, **Jeremy McGibbon***, Andre Perkins*, Oli Watt-Meyer*

* Vulcan Climate Modeling ML Group
+ GFDL

jeremym@vulcan.com

October 1, 2020

Introducing the Problem

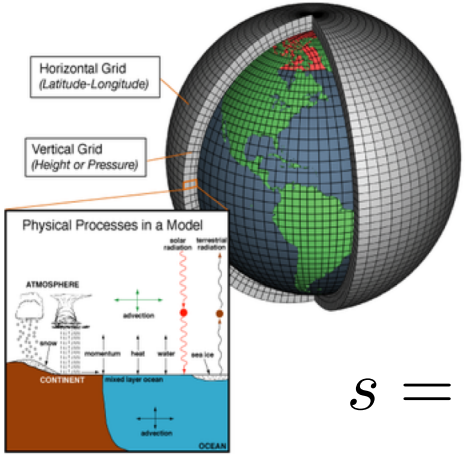
Goal: Improving a climate model to improve rainfall predictions using machine learning (ML)

A global storm-resolving model (GSRM) with a finer grid of 1-3 km may (with work) do great things, but it is very expensive!

Goal:

Use a realistic GSRM to train a skillful machine-learning based parameterization of subgrid clouds and precipitation for a coarser-grid global climate model.

Coarse-resolution 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, etc

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

Apparent moistening (g/kg/day)

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

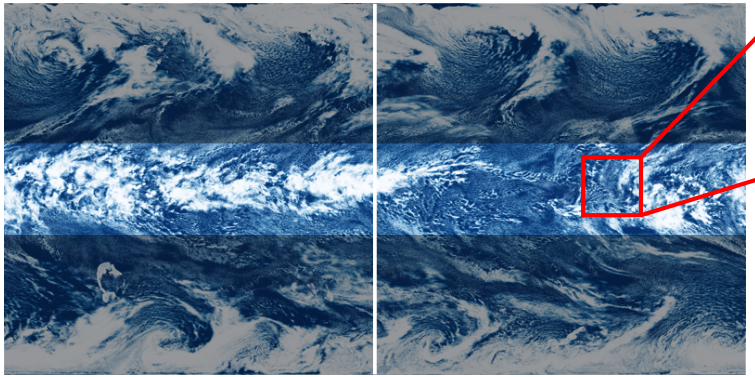
Apparent momentum source
(for now rely on coarse model parameterizations of PBL, GWD, etc.)

Past work: Training ML using a coarse-grained 4 km tropical channel simulation

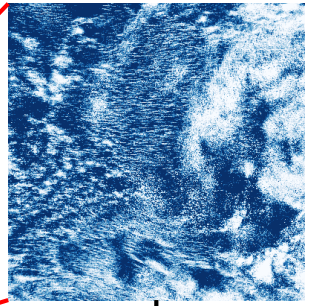
Brenowitz and Bretherton 2018, 2019; Rasp et al. 2018; O'Gorman and Yuval 2020

- Use 80-day 4 km aquaplanet run as 'truth' to machine-learn moist physics parameterization for the low-res model.
- Goal: forecast with low-res dycore + ML param should match hi-res run.

A

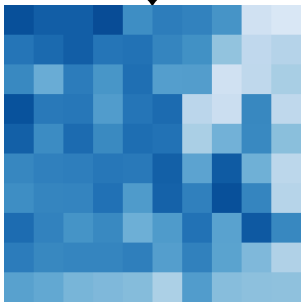


B



Coarse-graining

C

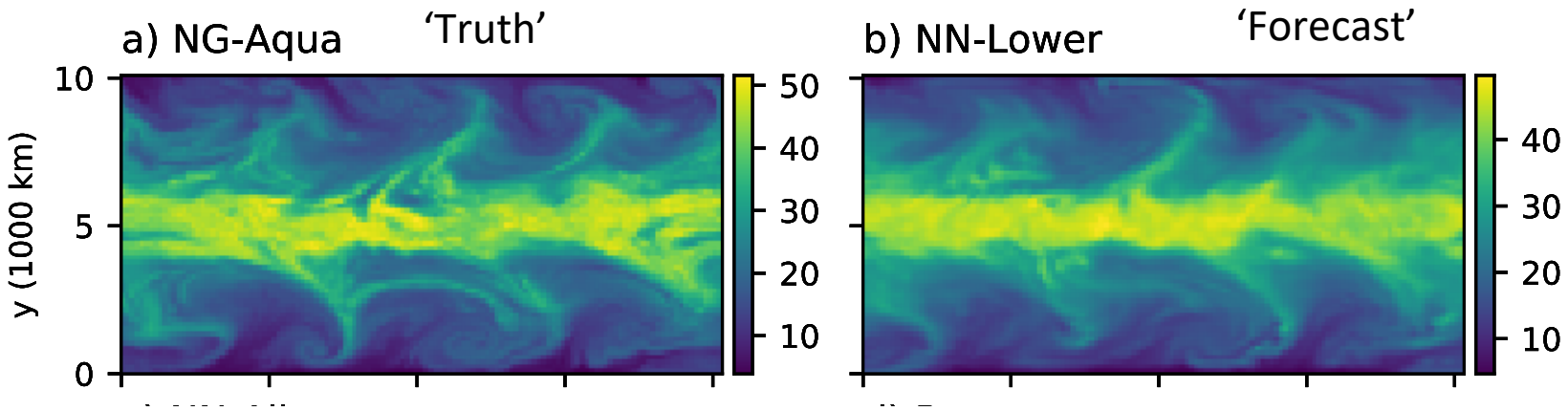


10^6 training boxes from 80-day simulation

- 160 km coarse (low-res) grid
- Calculate $Q_{1,2}(\mathbf{r}, t)$ (coarse-grid 'moist physics' tendencies including radiation) as residuals of dynamical equations.
- Unified moist physics, turbulence and radiation parameterization: Learn $Q_{1,2}$ as functions of local column conditions using a neural net.

Couple the ANN to the flow solver on 160 km grid

If inputs and error metric are carefully designed to prevent rapid model blow-up, hi-res model is skillfully forecast by low-res model with NN parameterization



Precipitable water after 5 days

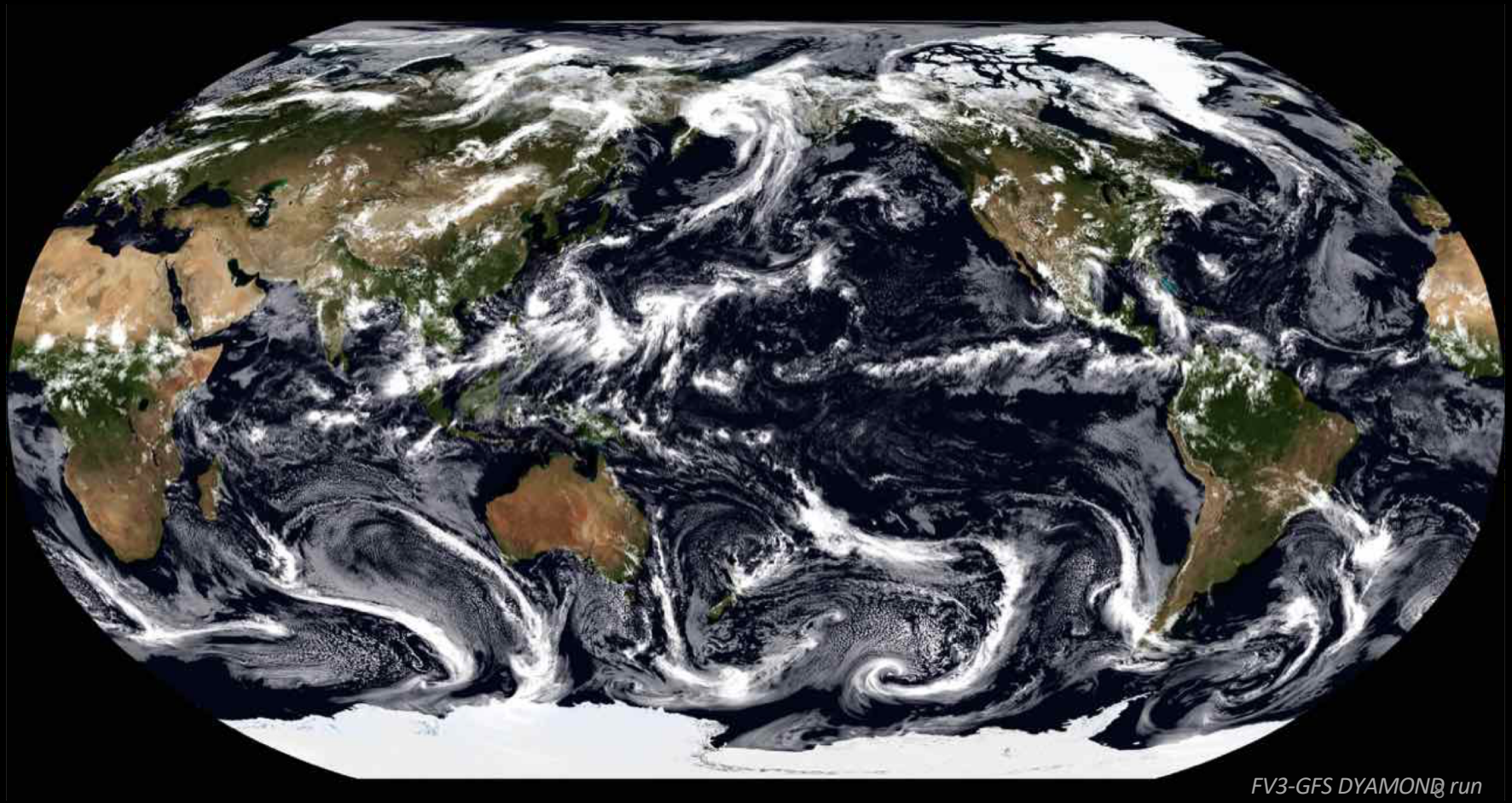
Brenowitz and Bretherton (2019)

…but the 'climate' slowly drifts after 10 days toward a weaker ITCZ

See Rasp et al. (2018, GRL) and O'Gorman and Yuval (2020, arXiv) for other aquaplanet successes with similar methods applied to related models.

Target model: FV3GFS/Shield

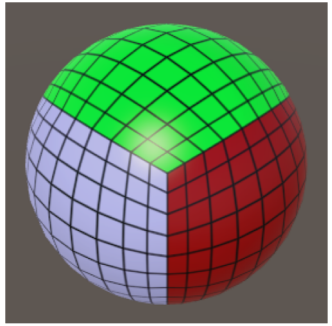
Can we apply same ML approach to GFDL's 3 km FV3-GFS global atmospheric model?



FV3-GFS DYAMOND run
S.-J. Lin and Xi Chen, GFDL

FV3GFS and SHiELD[1] global weather/climate models

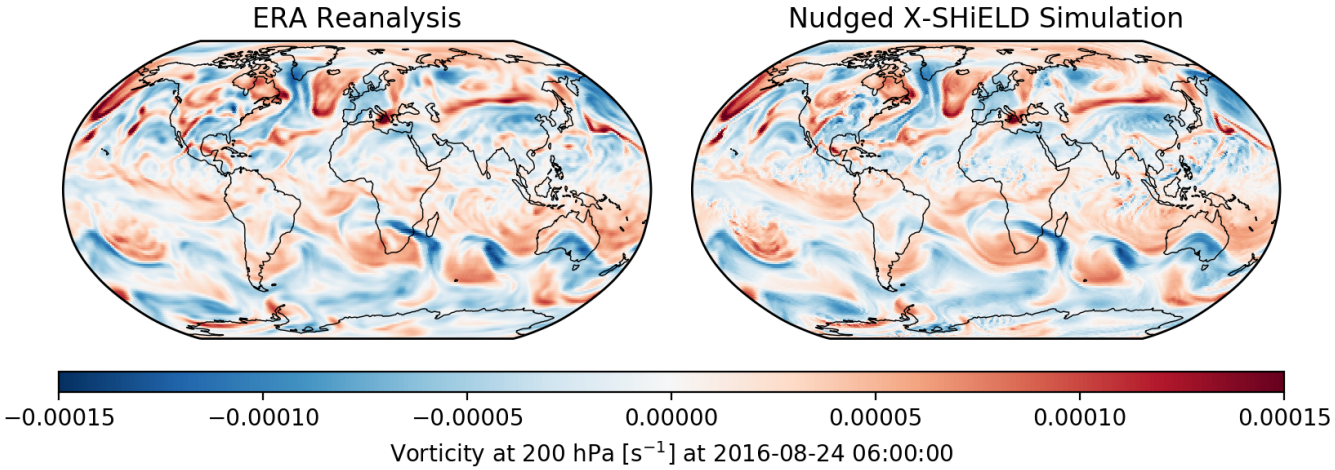
- FV3GFS: Open-source global atmosphere model used by NOAA for operational weather forecasts
- FV3 dycore – Customized D-grid finite volume method on cubed sphere.
- Nonhydrostatic by default, 80 vertical levels used here.
- Specified time-varying sea-surface temperature used here
- **Horizontal grid resolutions:**
  - 3 km (C3072) No deep cumulus parameterization or gravity-wave drag
  - 13 km Used for NCEP's current operational global weather forecasts
  - 25 km Finest grid currently practical for climate simulations of many decades
  - 200 km (C48) Typical coarse climate model grid – good for prototyping or millennial runs.
- **Physical parameterizations:**
  - Land surface and surface fluxes (NOAH)
  - Radiation (RRTMG)
  - Gravity-wave drag
  - Boundary-layer (including shallow clouds) and shallow Cu (Han-Bretherton, Han-Pan)
  - Cloud microphysics and subgrid variability (GFDL one-moment)
  - Deep cumulus convection (SAS)



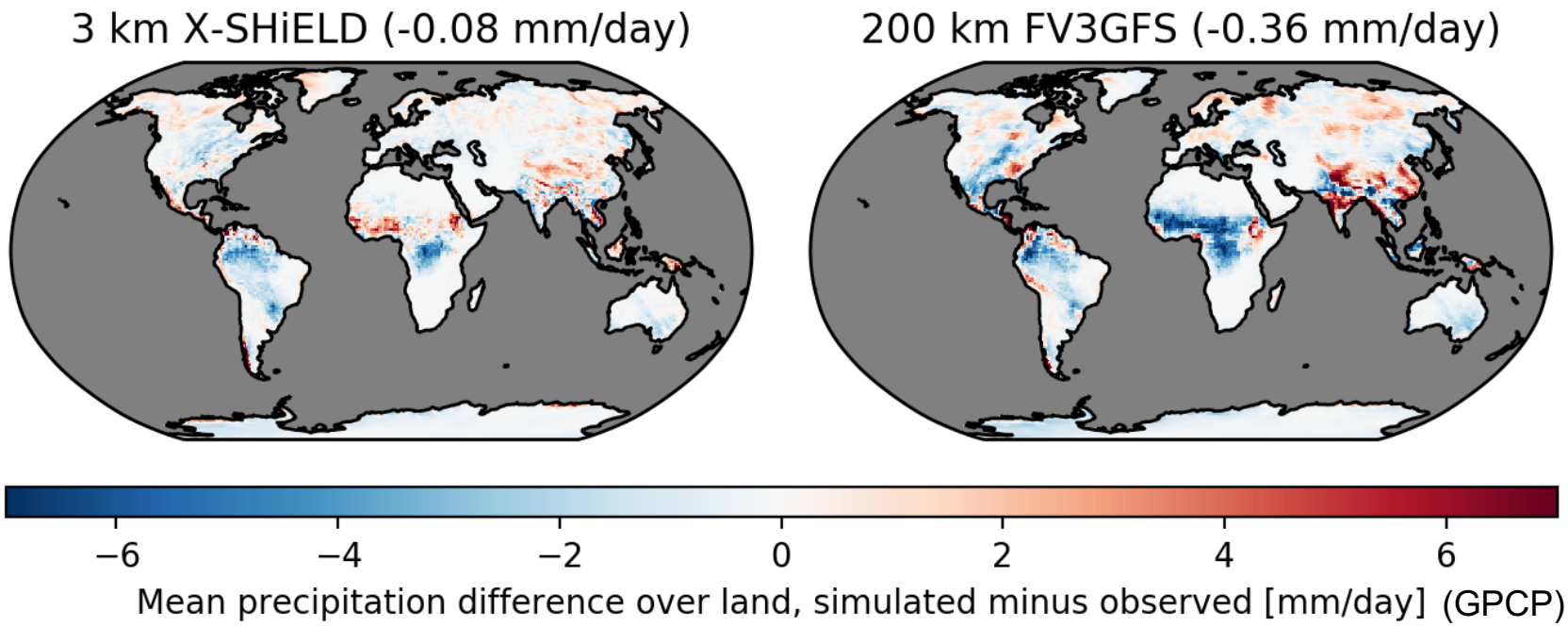
[1] GFDL's SHiELD is FV3GFS with modest changes to cloud physics and advection and is not open-source.

Training dataset: nudged 3 km SHiELD (modified FV3-GFS)

- Training dataset: 40 d 'nudged DYAMOND' simulation on GAEA (1 Aug to 9 Sep 2016):
  - Observed SSTs
  - Light nudging (τ = 1 day) of 3 km T/u/v/p_s to ERA5 reanalysis keeps meteorology 'data-aware'. Nudging tendencies are considered to be part of the learned physics
  - Store atmospheric and land-surface restart fields coarse-grained to 25 km every 15 min



40 d mean precipitation bias over land: 3 km SHiELD vs. 200 km FV3GFS



3 km rainfall bias much smaller over sub-Saharan Africa and Himalayas
Diurnal cycle of precipitation over land is also greatly improved in SHiELD

Our Approach(es)

Coarse model physics

We run ML on top of four configurations of the coarse-resolution model:

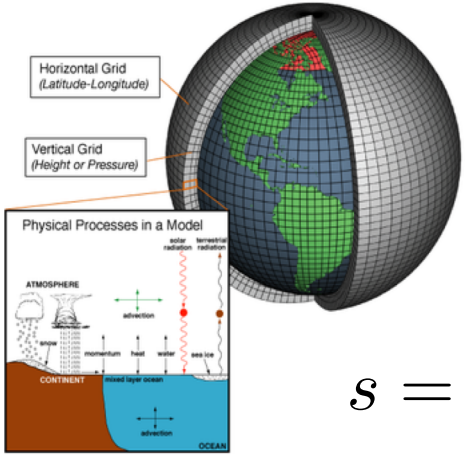
1. **physics-on**
   - All physical parameterizations on
     (land surface, boundary layer, convection, radiation, microphysics, gravity wave drag)
2. **deep-off**
   - Turn off deep convection scheme
3. **clouds-off**
   - Deep and shallow convection schemes off
   - No microphysics
   - Use clear-sky radiation only
4. **physics-off**
   - Run only dynamical core

Coarse model physics

We run ML on top of four configurations of the coarse-resolution model:

1. **physics-on**
   - All physical parameterizations on
     (land surface, boundary layer, convection, radiation, microphysics, gravity wave drag)
2. **deep-off**
   - Turn off deep convection scheme
3. **clouds-off**
   - Deep and shallow convection schemes off
   - No microphysics
   - Use clear-sky radiation only
4. **physics-off**
   - Run only dynamical core

Coarse-resolution 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, 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
(for now rely on coarse model parameterizations of PBL, GWD, etc.)

Coarse-graining and tendency-difference method

- $a_f(t, x, y, \sigma)$: field (e.g. humidity) at fine resolution. $a_c(t, x, y, \sigma)$ is coarse-res
- Coarse-graining operator: $\bar{\cdot}$ (horizontal averaging)
- Coarse model (200 km FV3GFS) should match fine model (3 km SHiELD) starting at $a_c = \overline{a_f}$:

$$\frac{\partial a_c}{\partial t} \approx \frac{\partial \overline{a_f}}{\partial t}$$

- Uncorrected coarse model:

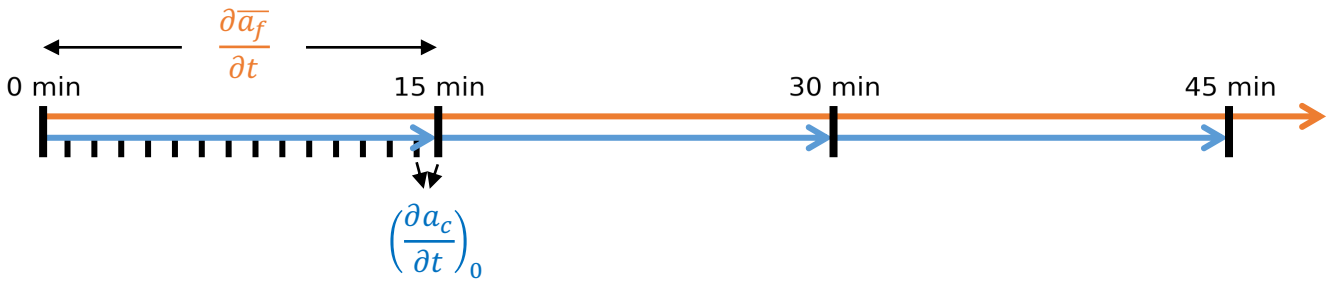
$$\left(\frac{\partial a_c}{\partial t}\right)_0 = A_c + Q_a^p, \qquad A_c = -\mathbf{u}_c \cdot \nabla a_c$$

- Coarse model can include no physics ($Q_a^p = 0$) or a subset of parameterized physical processes (e.g. turbulence, radiation, clouds, Cu parameterization).
- Machine-learn a state-dependent corrective source ΔQ_a for the coarse model:

$$\Delta Q_a = \frac{\partial \overline{a_f}}{\partial t} - \left(\frac{\partial a_c}{\partial t}\right)_0 \qquad \rightarrow \qquad \left(\frac{da}{dt}\right)_c = \left(\frac{da_c}{dt}\right)_0 + \Delta Q_a^{ML} \cong \frac{\partial \overline{a_f}}{\partial t}$$

Original “One-Step” tendency difference method

Coarsened state of fine-resolution model saved every 15 minutes.
Fine-res tendencies computed from these snapshots.



Coarse-resolution model initialized from each coarsened high-resolution snapshot and run forward for 15 minutes, with a 1-minute timestep.

Low-res tendencies computed from final minute.

Apparent source:

$$\Delta Q_a = \frac{\partial \overline{a_f}}{\partial t} - \left(\frac{\partial a_c}{\partial t}\right)_0$$

Original “One-Step” tendency difference method

Coarsened state of fi solution m l saved every 15 minutes.
Fine-res ten mpu hese snapshots.

$\frac{\partial \overline{a_f}}{\partial t}$

0 min

45 min

Coarse-resolutio
coarsened hi
forward for 15 min

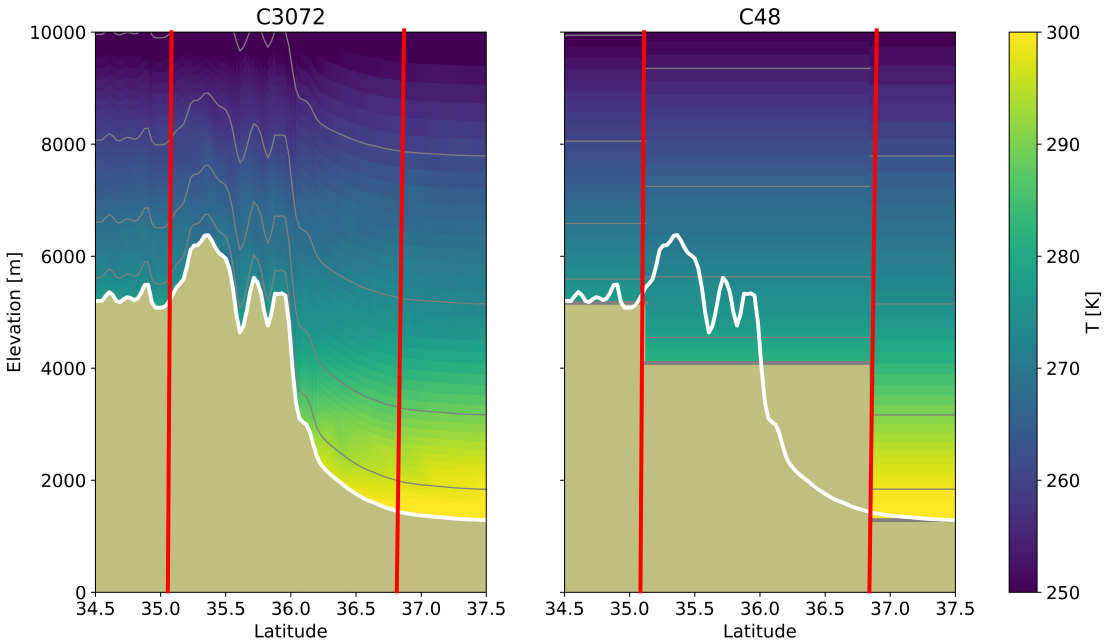
Low-res tendencies comp fina

Apparent source:

$$\Delta Q_a = \frac{\partial \overline{a_f}}{\partial t} - \left(\frac{\partial a_c}{\partial t}\right)_0$$

Conceptual issues over topography

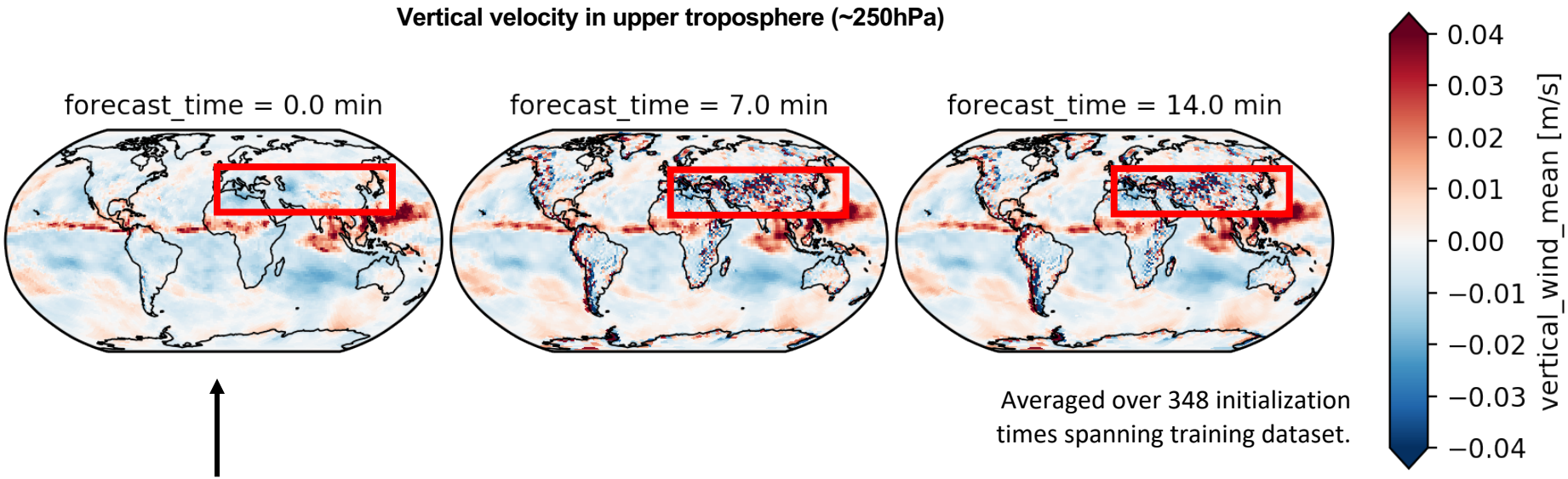
- Consider 3 km -> 200 km coarse-graining over the Himalayas



- We coarse-grain to obtain vertical **profiles** and **apparent sources** of T, q, etc.
- 5 km relief within a coarse cell
- Most fields are much more constant along a pressure surface than along a terrain-following model surface

→ Coarse-grain on pressure levels, not model levels

Vertical velocity noise over topography



Averaged over 348 initialization times spanning training dataset.

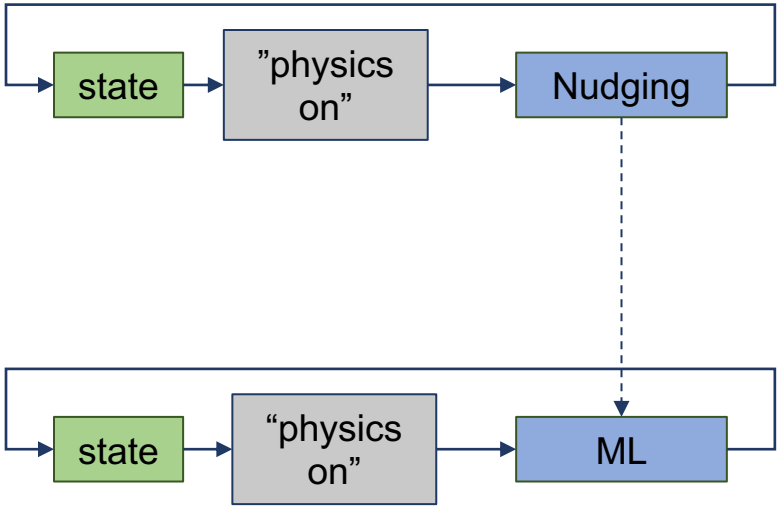
Fine resolution model coarsened to 200km resolution

Nudging method for coarse-graining

- Strongly nudge coarse model to high-resolution model

$$\frac{\partial a_n}{\partial t} = \frac{\partial a_c}{\partial t} + \Delta Q_a^N \cong \frac{\partial \overline{a_f}}{\partial t}$$

$$\Delta Q_a^N = \frac{\overline{a_f} - a_c}{\tau}$$, (τ=3h in results here)



“nudge to fine-res”:

$$\Delta Q_a^{ML} = \Delta Q_a^N$$

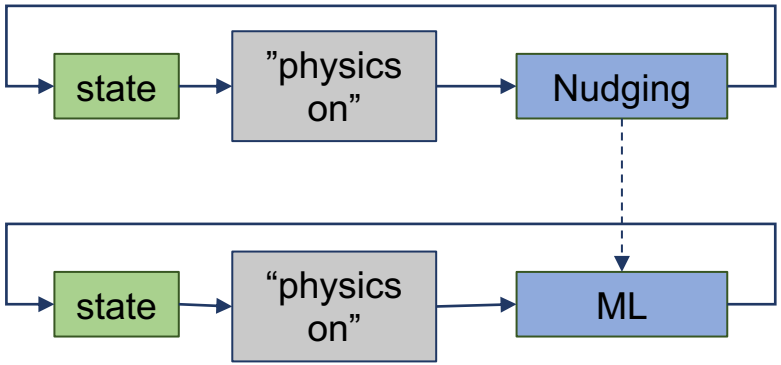
“Nudge to obs” method

- Not a coarse-graining method, no fine-res model involved
  - Enables much longer simulations
- Nudge coarse model to observations (analysis), predict the nudging tendency

$$\frac{\partial a_n}{\partial t} = \frac{\partial a_c}{\partial t} + \Delta Q_a^N \cong \frac{\partial \overline{a_f}}{\partial t}$$

$$\Delta Q_a^N = \frac{\overline{a_f} - a_c}{\tau}$$, (τ=6h in results here)

$$\Delta Q_a^{ML} = \Delta Q_a^N$$ “nudge to obs”

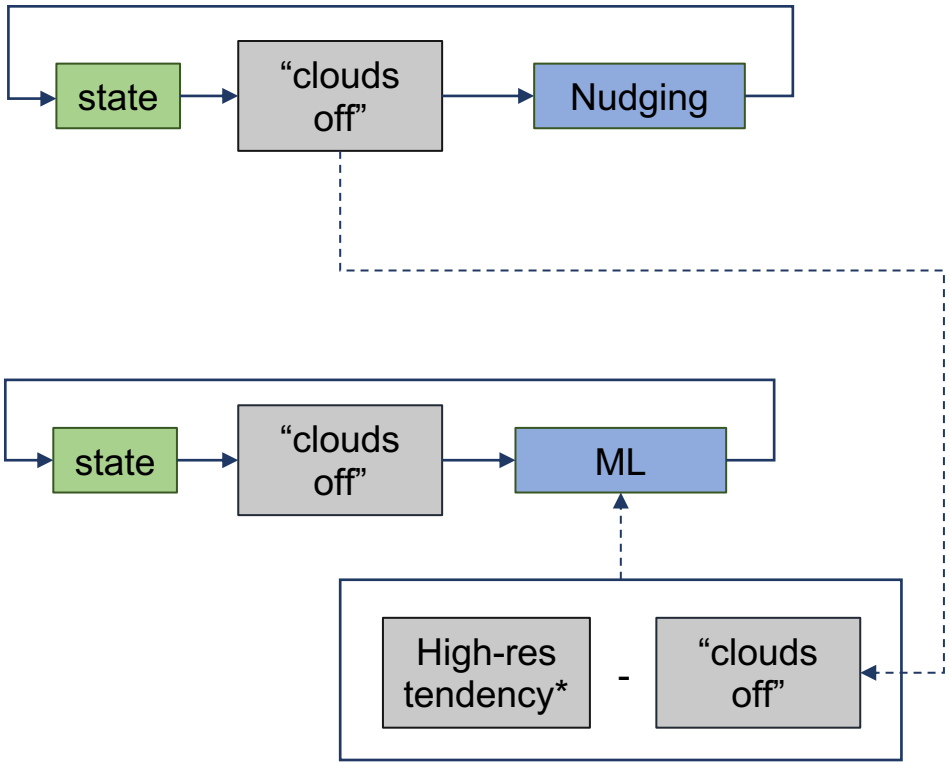


“hybrid” method

- Strongly nudge coarse model to high-resolution model

$$\frac{\partial a_n}{\partial t} = \frac{\partial a_c}{\partial t} + \Delta Q_a^N \cong \frac{\partial \overline{a_f}}{\partial t}$$

$$\Delta Q_a^N = \frac{\overline{a_f} - a_c}{\tau}, \ (\tau\text{=3h in results here})$$



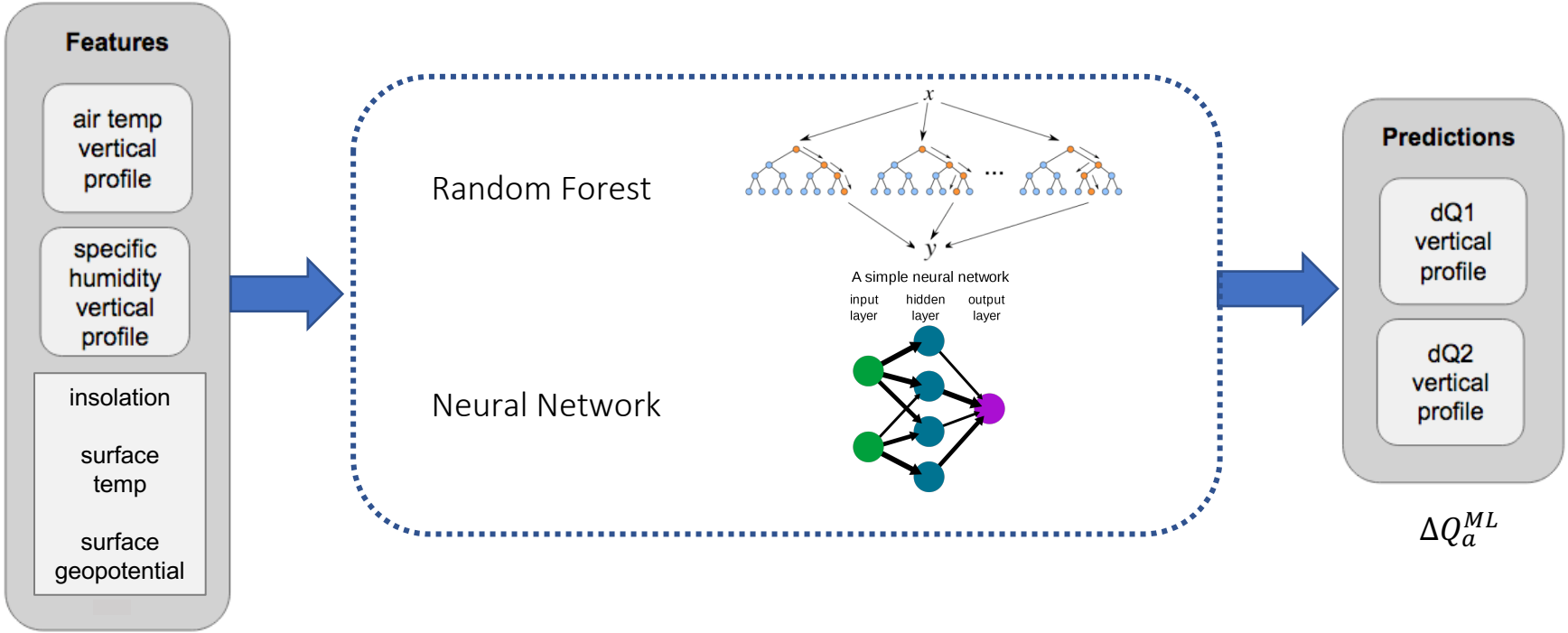
“hybrid fine-res” or “hybrid”:

$$\Delta Q_a^{ML} = \frac{\partial \overline{a_f}}{\partial t} - \frac{\partial a_c}{\partial t}$$

* Computed eddy flux + physics

Machine learning: model training

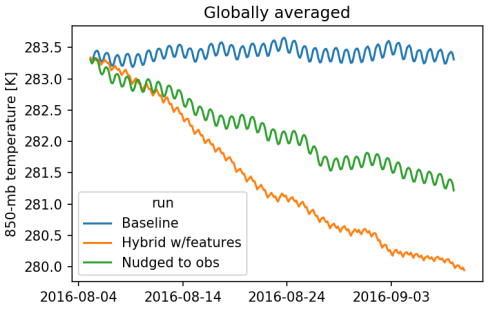
Training set = 1.7M samples (130 initializations x 13824 grid points)
Test set = 660K samples (48 initializations x 13824 grid points)
Train/test data separated by split date to minimize correlated data across sets

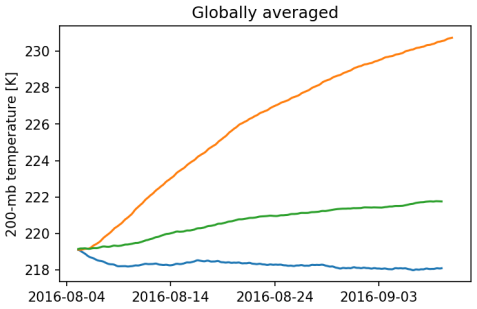


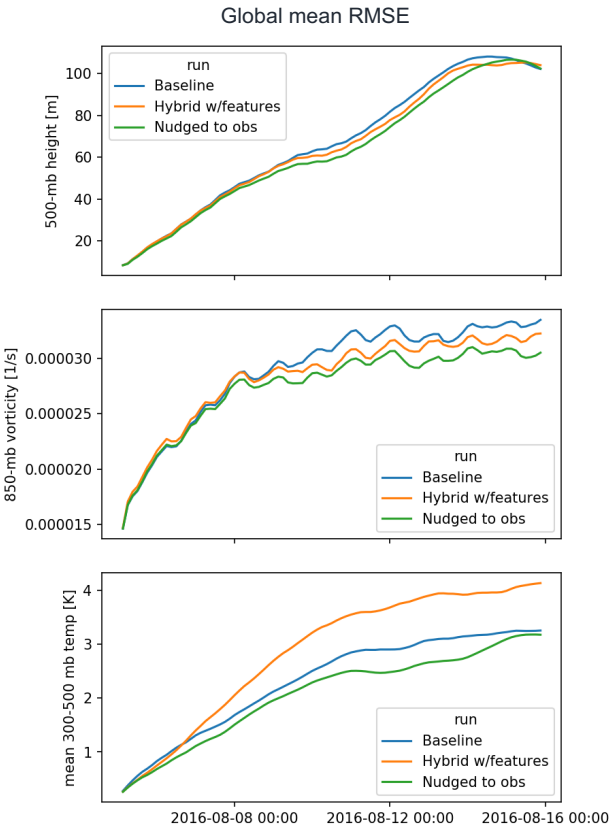
Early Results

Prognostic run improves on baseline, but drifts

- Skill improvements for several variables on weather prediction timescales
- Drift causing issues
  - Solutions being investigated

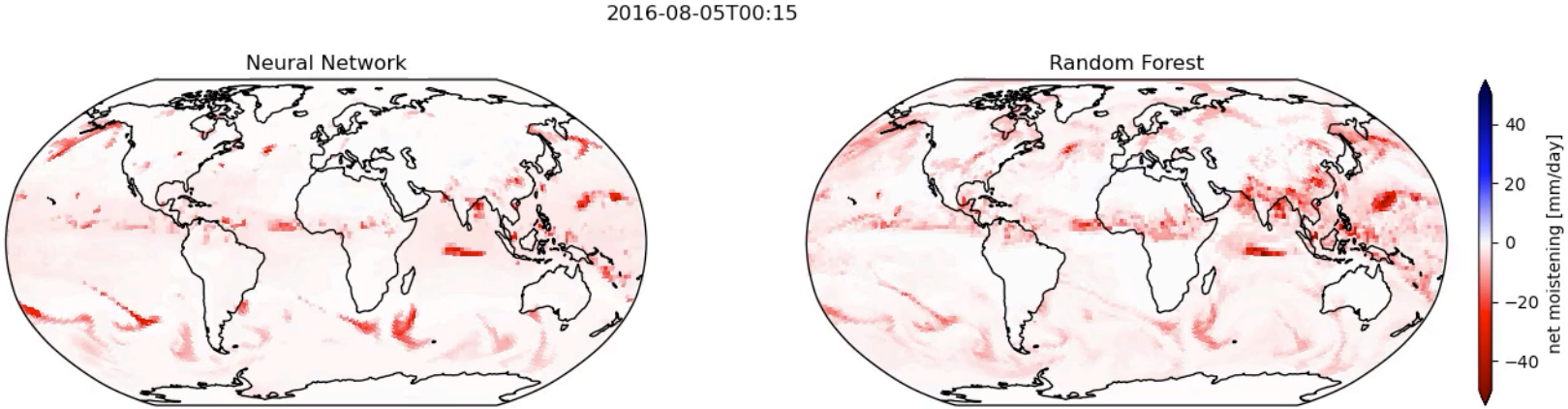






Random forests more stable out of the box

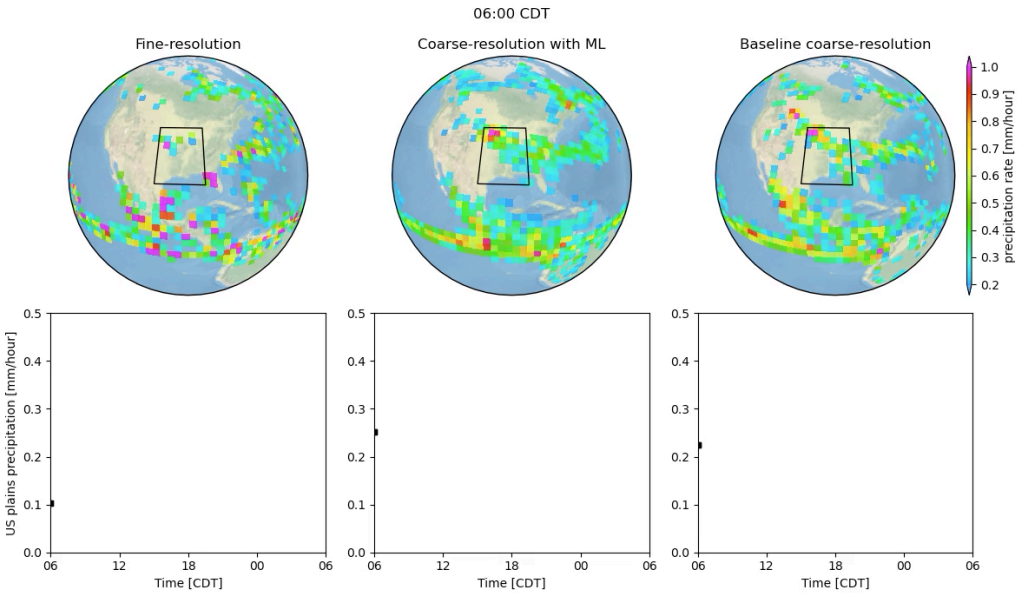
- This model with NN “blows up” after ~6 days
- Random forest outputs are bounded, neural network outputs are not
- Exploring several possibilities to stabilize NNs



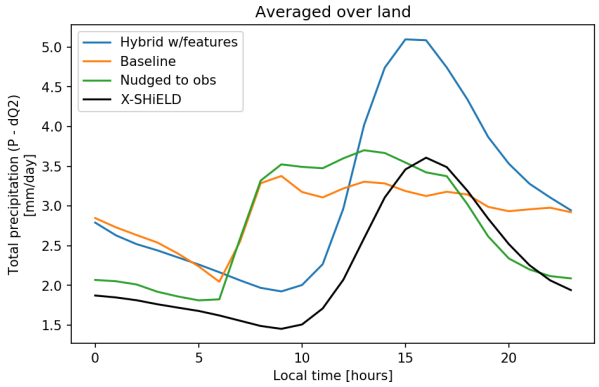
Negative values = precipitation (mm/s)

Some improvement to diurnal cycle over land

- nudged and hybrid approaches improve diurnal cycle of precipitation over land
- Reproduces afternoon maximum of precipitation



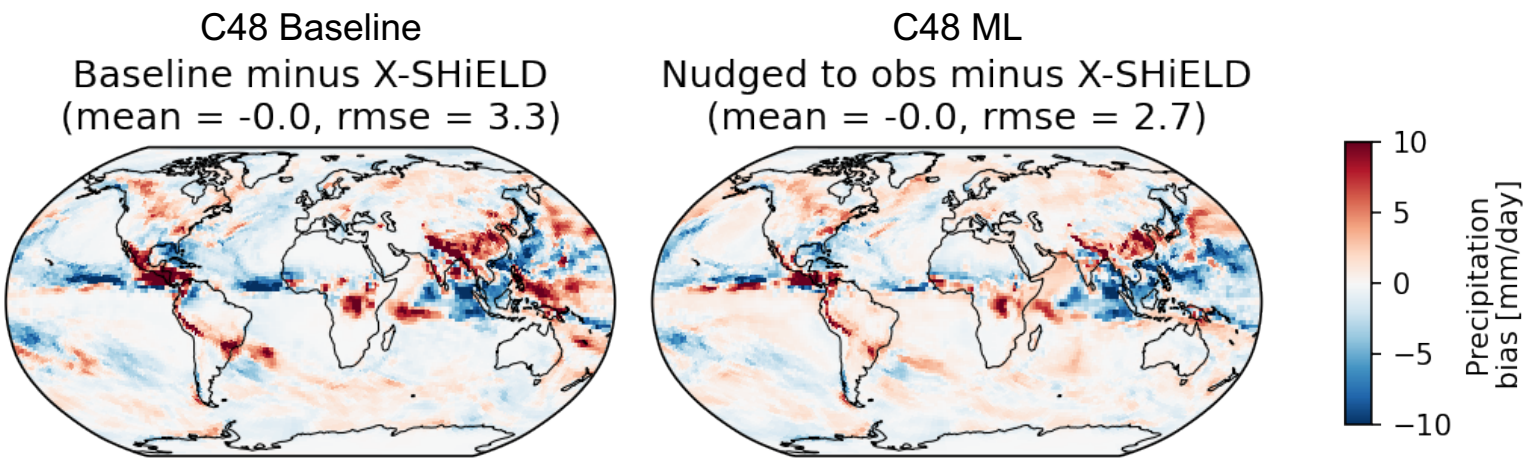
Composite over first 5 days of prognostic run



Average over 10-day August prognostic run, minus 1 day spin-up

Improvements in precipitation bias

- ML trained on “nudge to obs” improves precipitation bias
- Improvements in several areas
- Small large-scale bias increase in other areas, working to address



Average over August 5, 2016 – September 10, 2016

Conclusions and Outlook

- VCM has developed a unique cloud-based workflow for training a ML correction to the coarse-resolution FV3GFS model based on fine-resolution X-SHIELD simulations from GFDL
- We have trained stable ML schemes that can make skillful global rainfall forecasts over land and ocean for 5 days or longer given specified SST
- “nudge to obs”, “nudge to fine”, and “hybrid” approaches all improve short-term skill, with varying degrees of climate drift
- Improvements are made to short-term diurnal cycle and bias of precipitation
- Investigating approaches to minimize drift and further improve model skill
- Investigating ways to further stabilize neural-network-based simulations