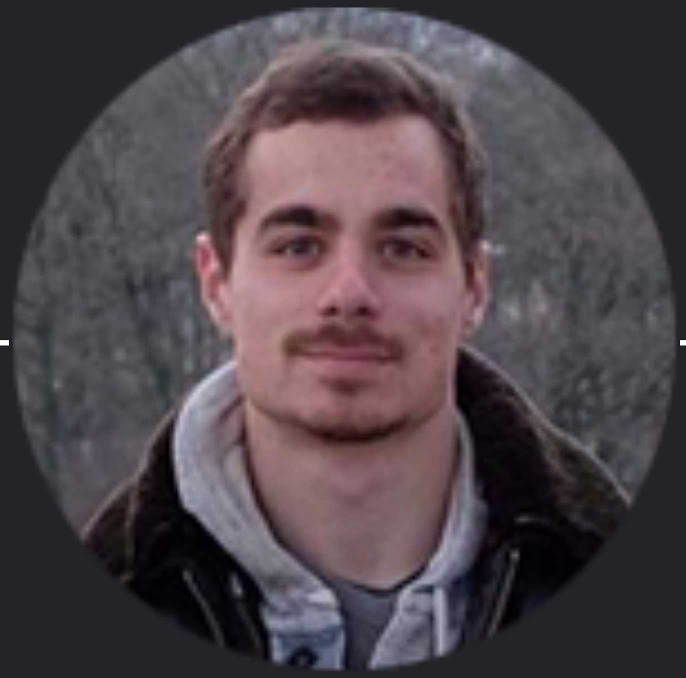
TOWARDS ROBUST NEURAL NETWORK PARAMETERIZATIONS OF CONVECTION

Advances in stability, credibility & software

Mike Pritchard
Associate Professor
University of California, Irvine

Pierre Gentine

Associate Prof,
Columbia



Tom Beucler

UCI Project
Scientist

Griffin Mooers

UCI PhD student

Jordan Ott

UCI PhD student

Pierre Baldi

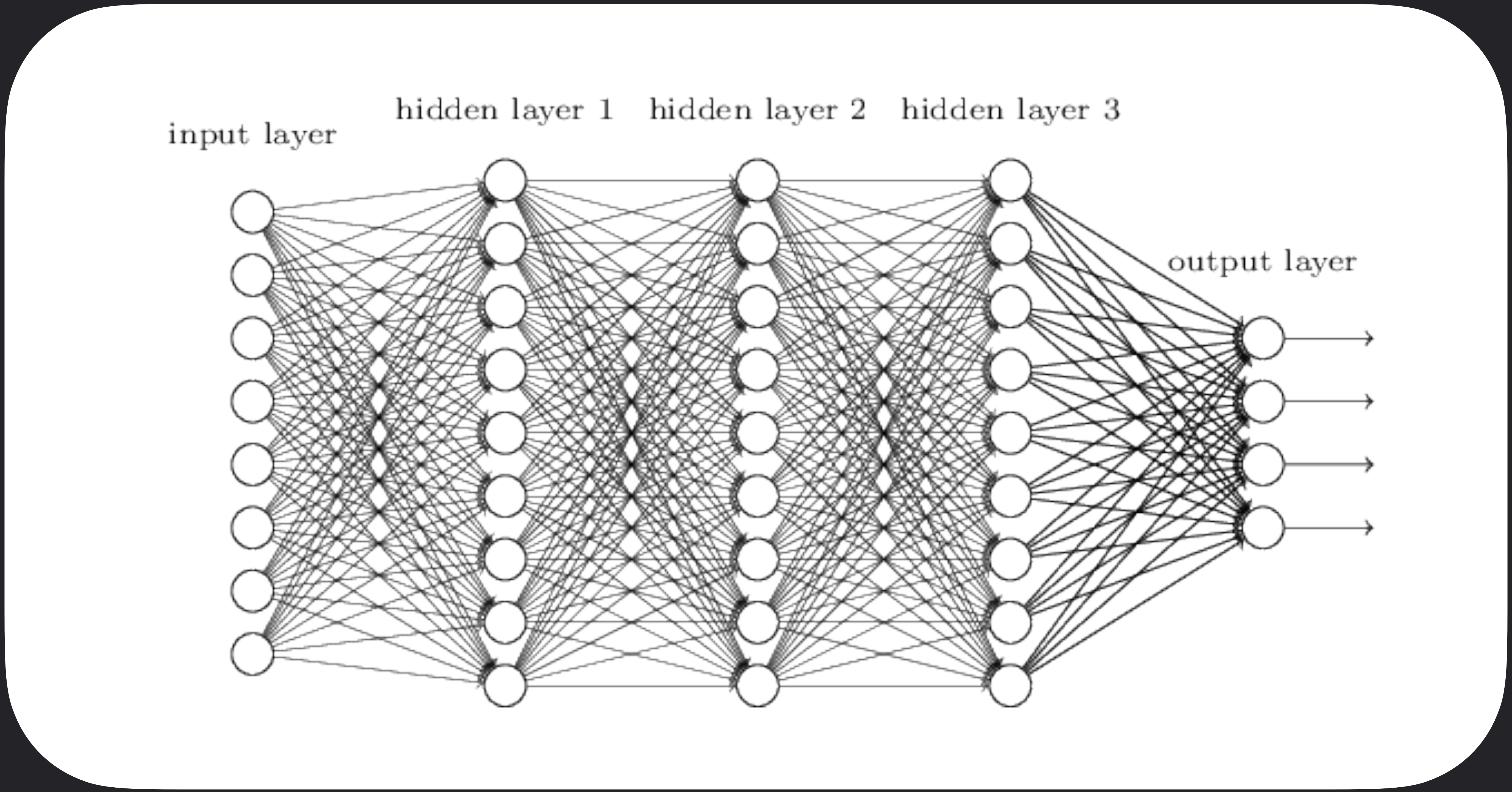
CS Prof, UCI

Stephan Rasp

Former visiting
PhD student

Context.

DNNs: Powerful emulators of high-dimensional nonlinear functions disrupting industry and science.



Schematic of a simple deep feed-forward Neural Network (DNN)

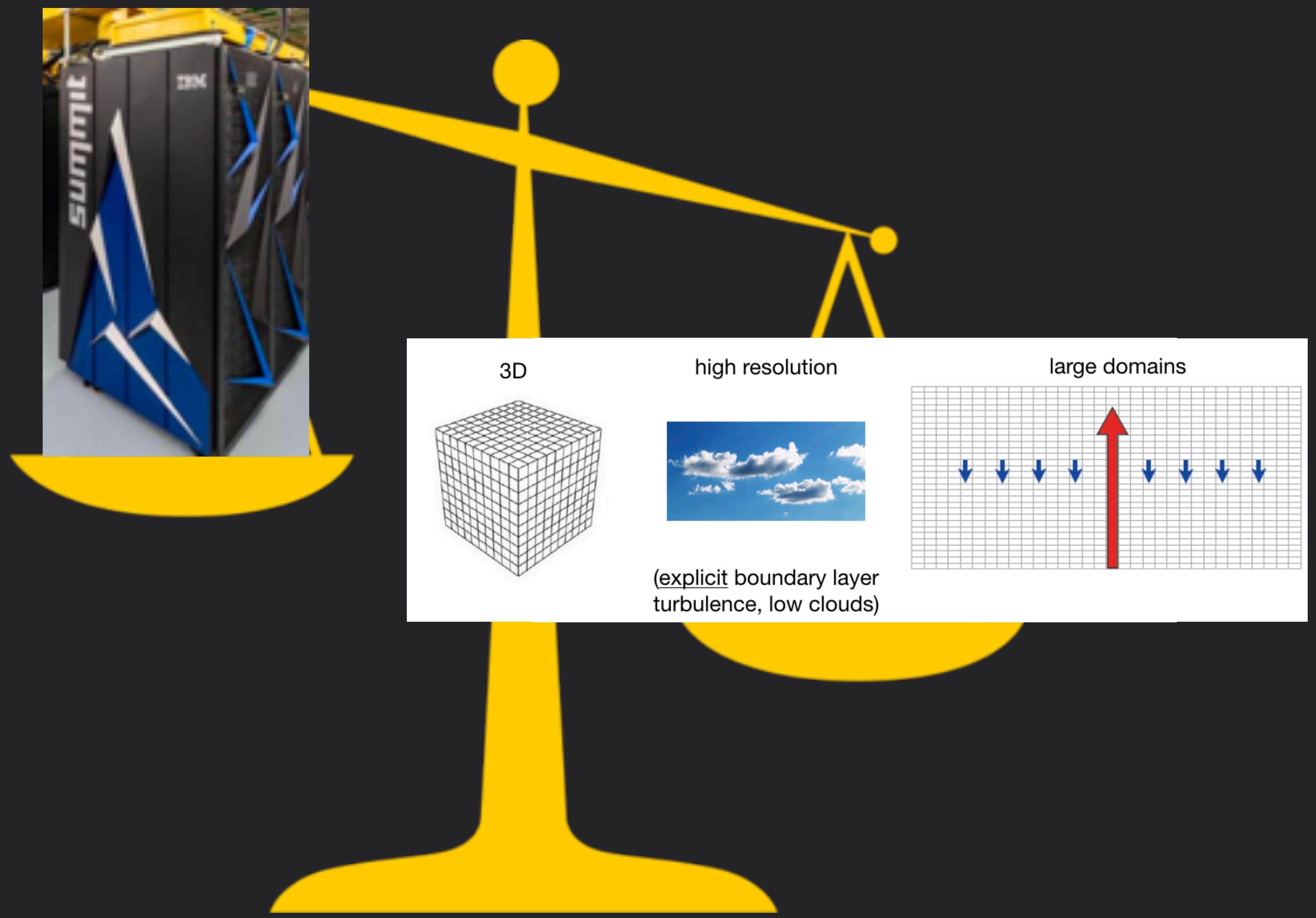
Deep Learning **emulation** **might allow high** definition 3D turbulence ahead of schedule!

If the job is hard, e.g. simulating the whole atmosphere for decades...

...satisfying 3D turbulence calculations can seem too much even for powerful computers.

Deep Learning **emulation** **might allow high** definition 3D turbulence ahead of schedule!

If the job changes to making short simulations just for training machine learning emulators...



...we can do much more justice to turbulence physics.

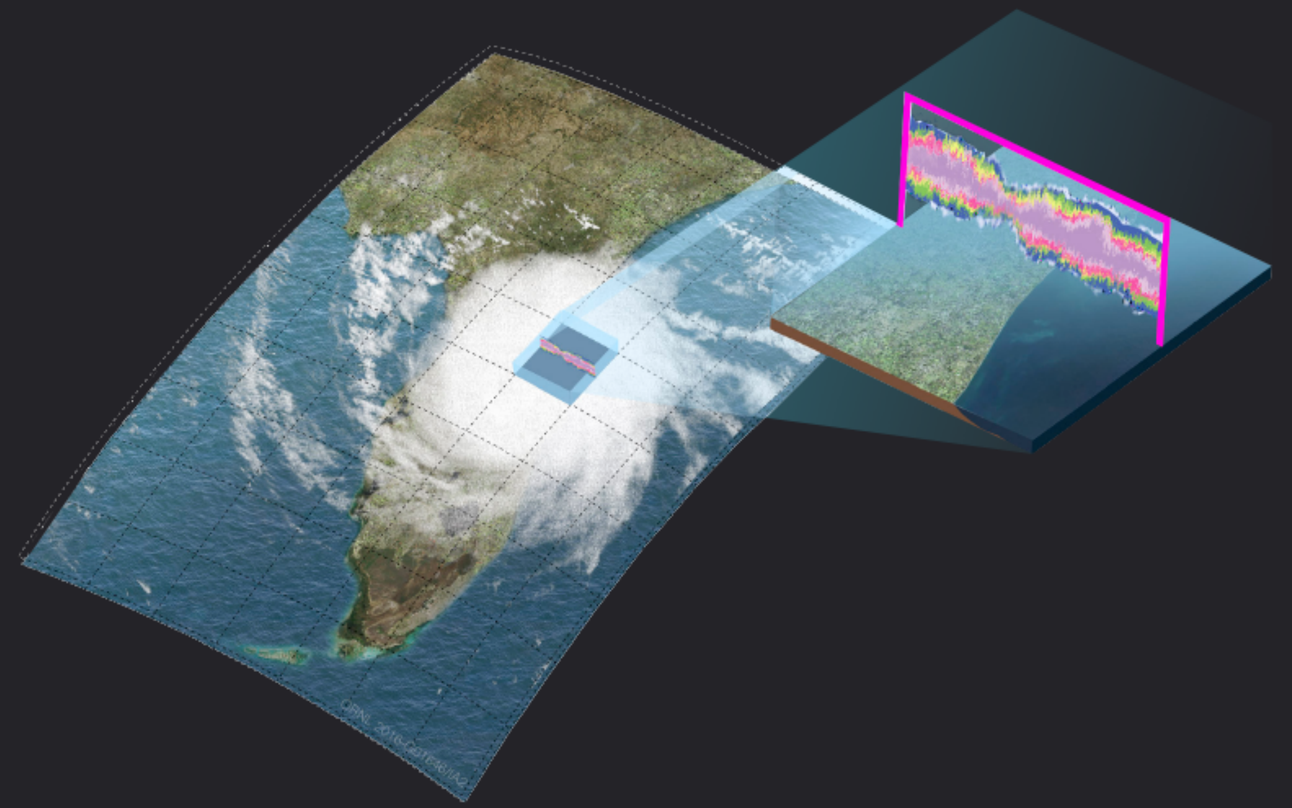
Deep Learning **emulation** can buy performance portability <u>for free</u> and thus access to unbelievable new systems.

"Summit" at Oak Ridge in Tennessee — <u>200</u> petaflops
~ 4,500 NVIDIA Volta V100 GPU nodes
(~ 27,648 research quality GPUs)

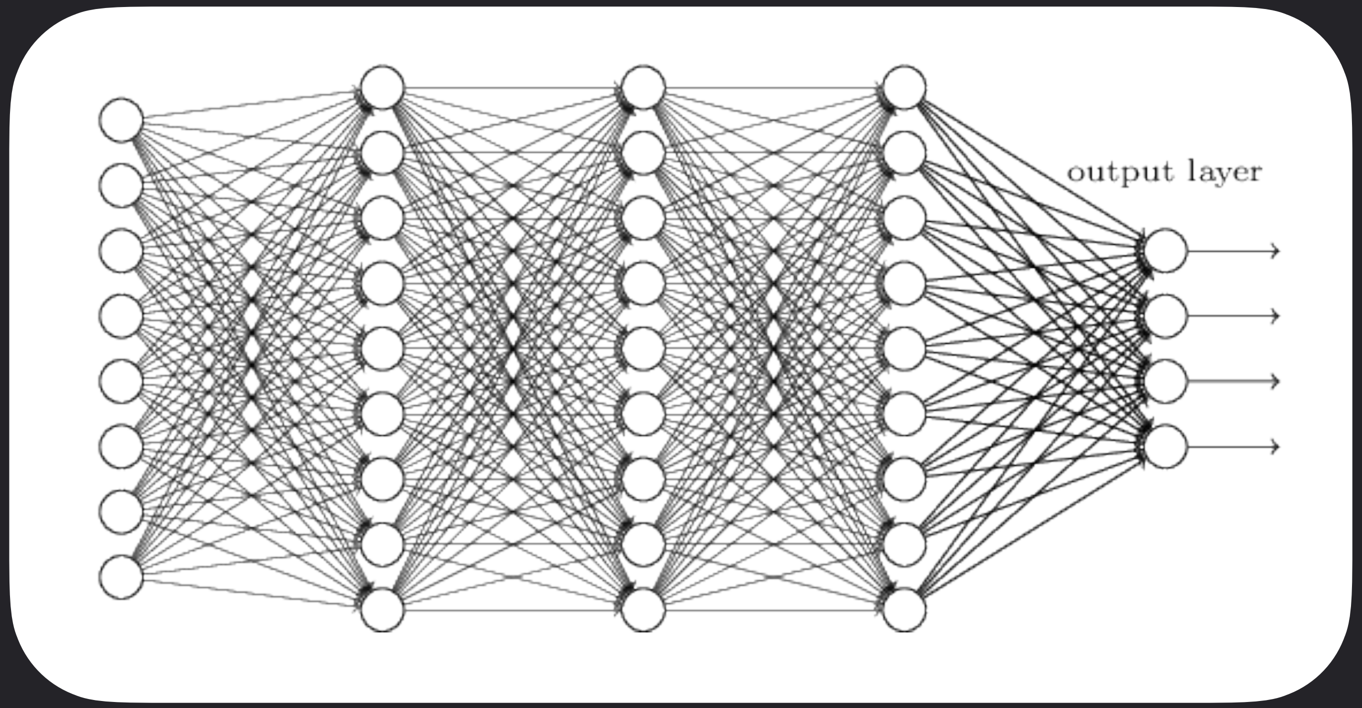
Is deep learning viable for emulating SuperParameterization?

2017: Global aquaplanet SP testbed

Can 140,000,000 outputs from 1 year of ~ 10,000 cloud-resolving models...



Be fit by a deep neural network?



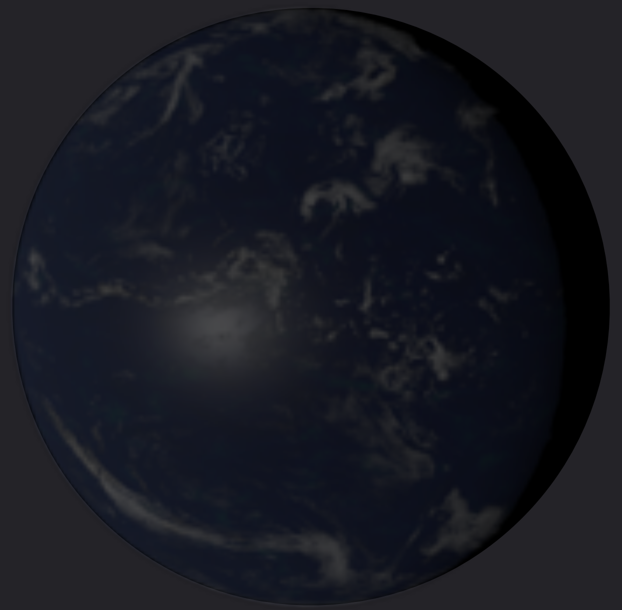
Gentine, Pritchard, Rasp et al., *GRL*, 2018.

Is deep learning viable for emulating SuperParameterization?

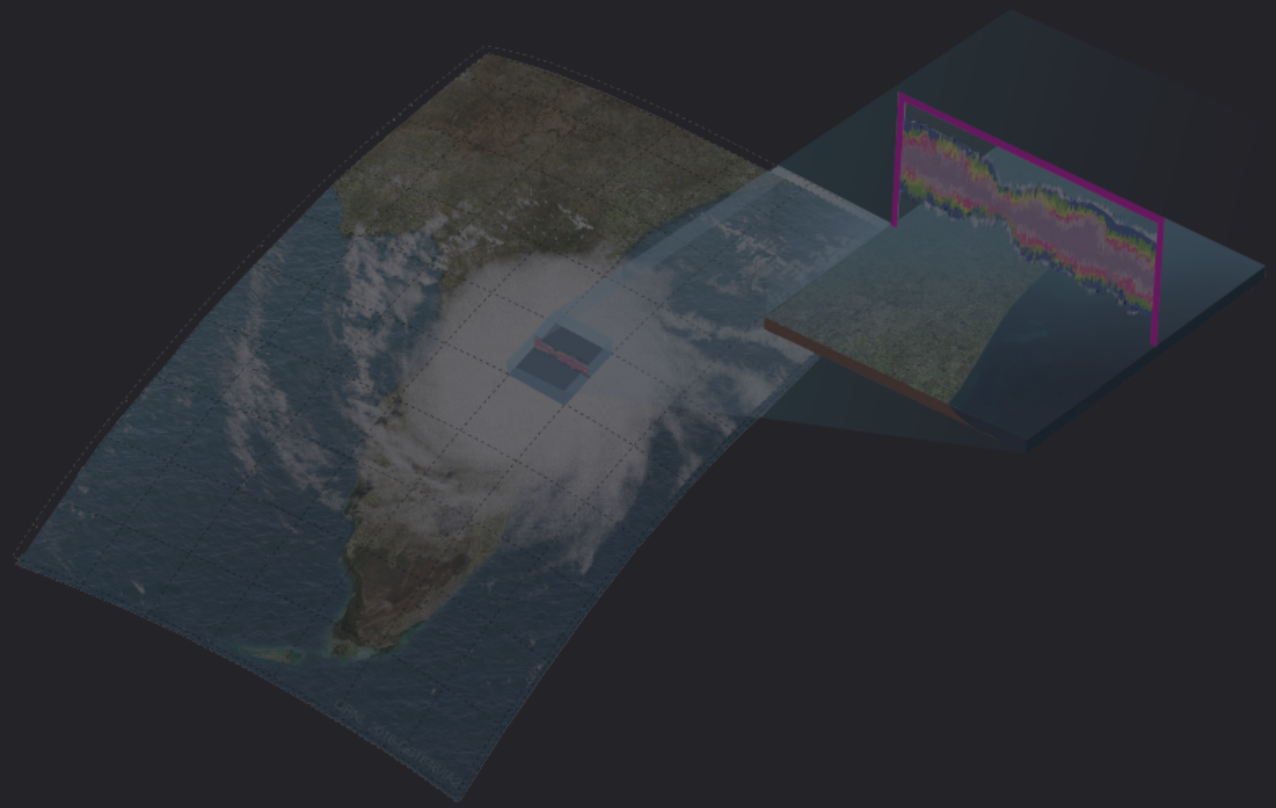
Quite possibly!

The “Cloud Brain”

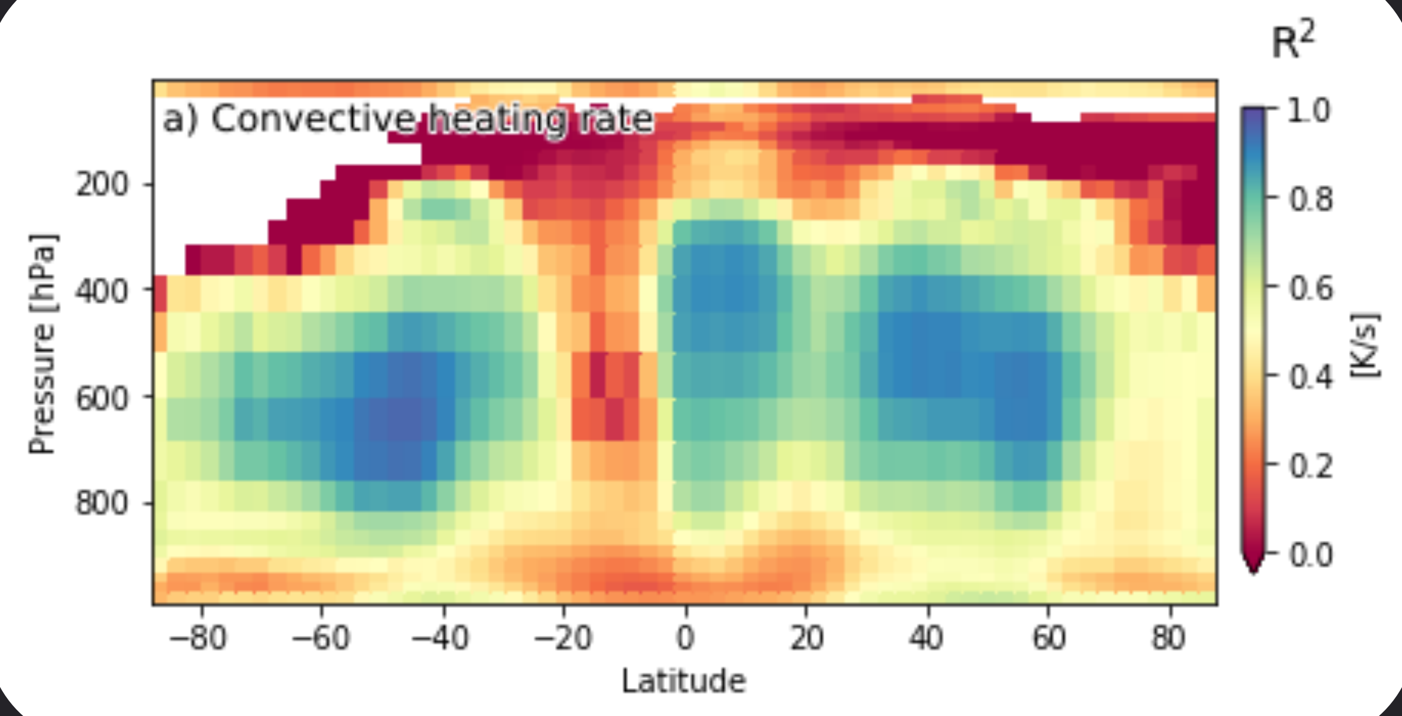
Global aquaplanet testbed



Can 140,000,000 outputs from 1 year of ~ 10,000 cloud-resolving models...

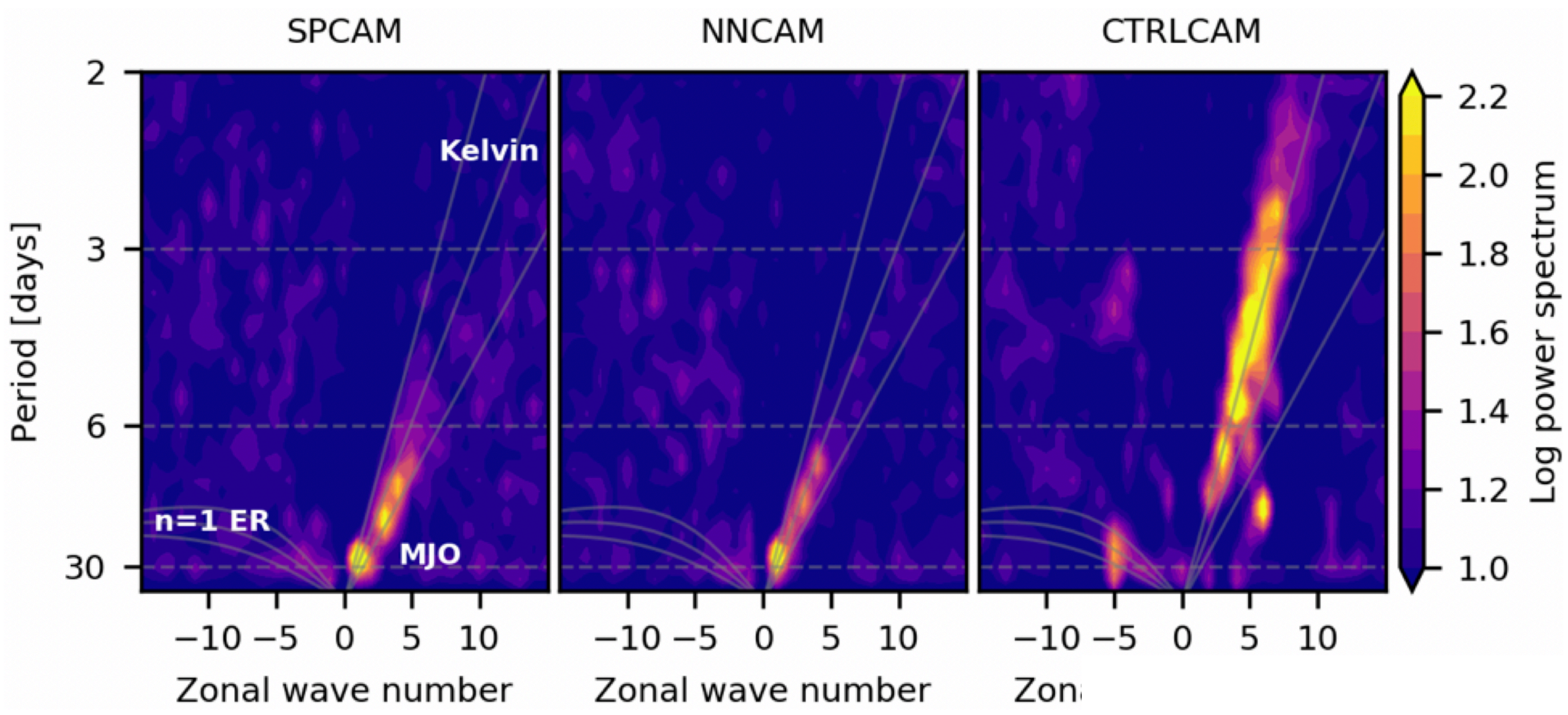


Be fit by a deep neural network?



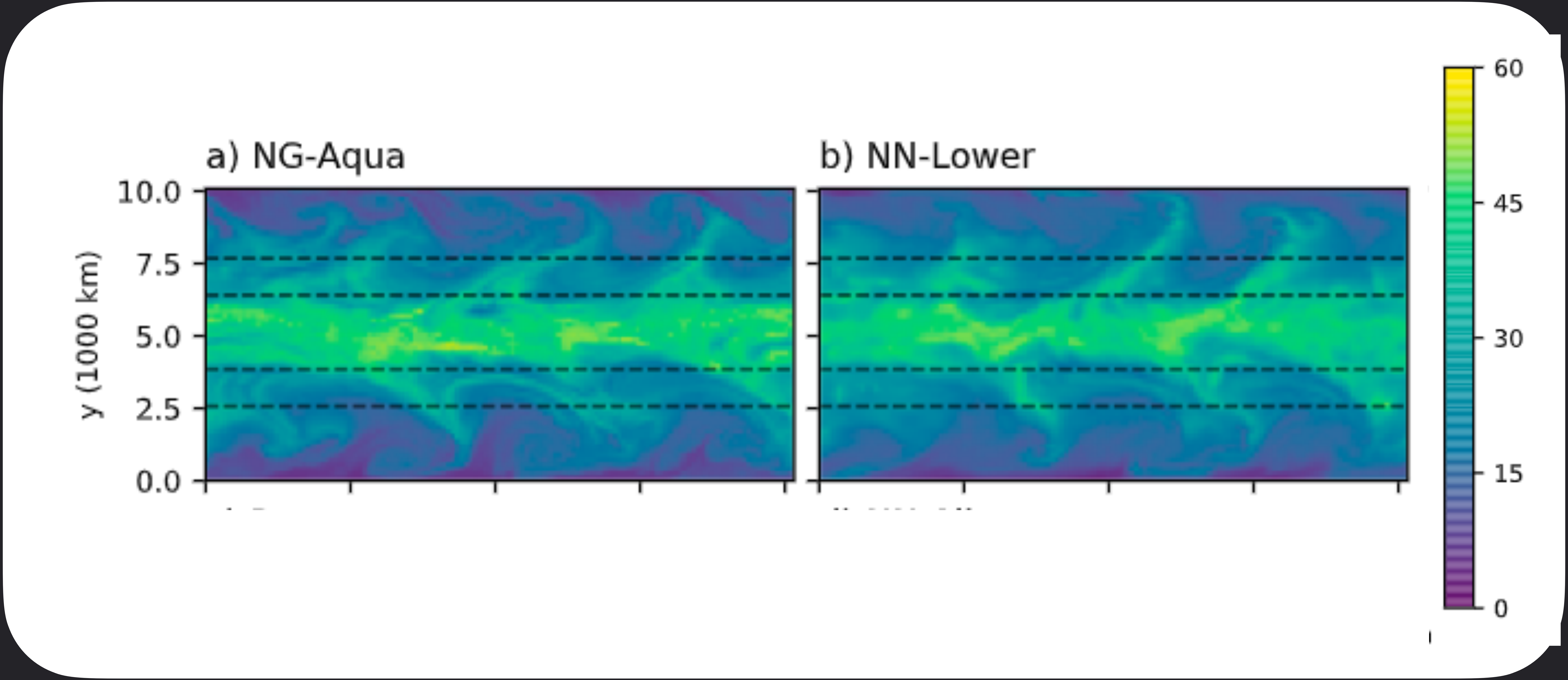
Yes, e.g. $R^2 > 0.7$ for mid-tropospheric heating by convection and radiation.

Promise: Physically credible behavior in multi-year prognostic simulations with NN-emulated convection.



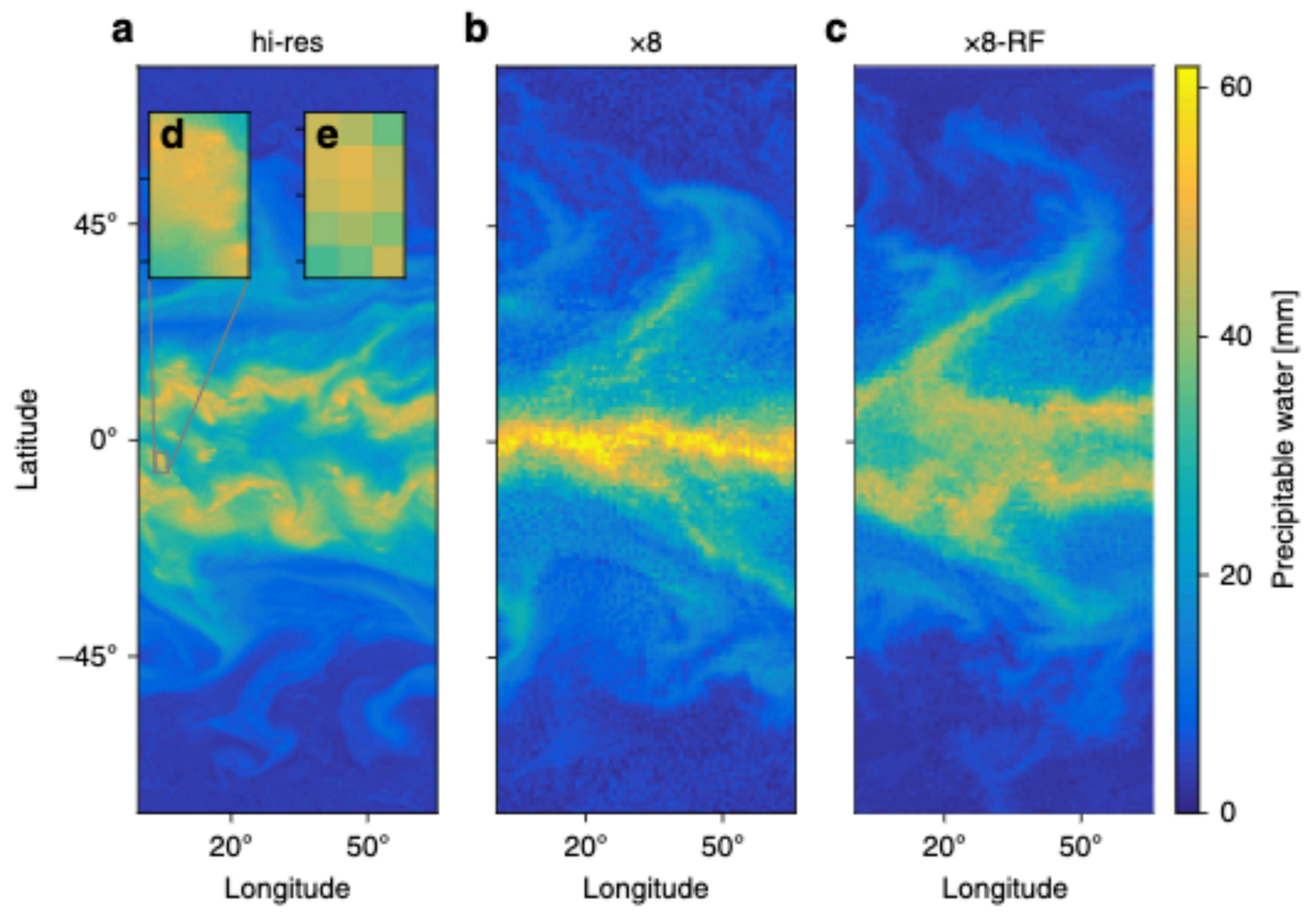
Rasp, Pritchard and Gentine, *PNAS*, 2018.

Promise: NN trained on coarse-grained global aquaplanet behaving well 5 days into a prognostic forecast.



Brenowitz and Bretherton, *JAMES*, 2019.

Promise: Random-Forest trained carefully via coarse-graining GCRM aquaplanet slice at equilibrium (prognostic)



Yuval & O'Gorman, *Nature Communications*, 2020.

Problem 1: NNs attractive but don't obey constraints.

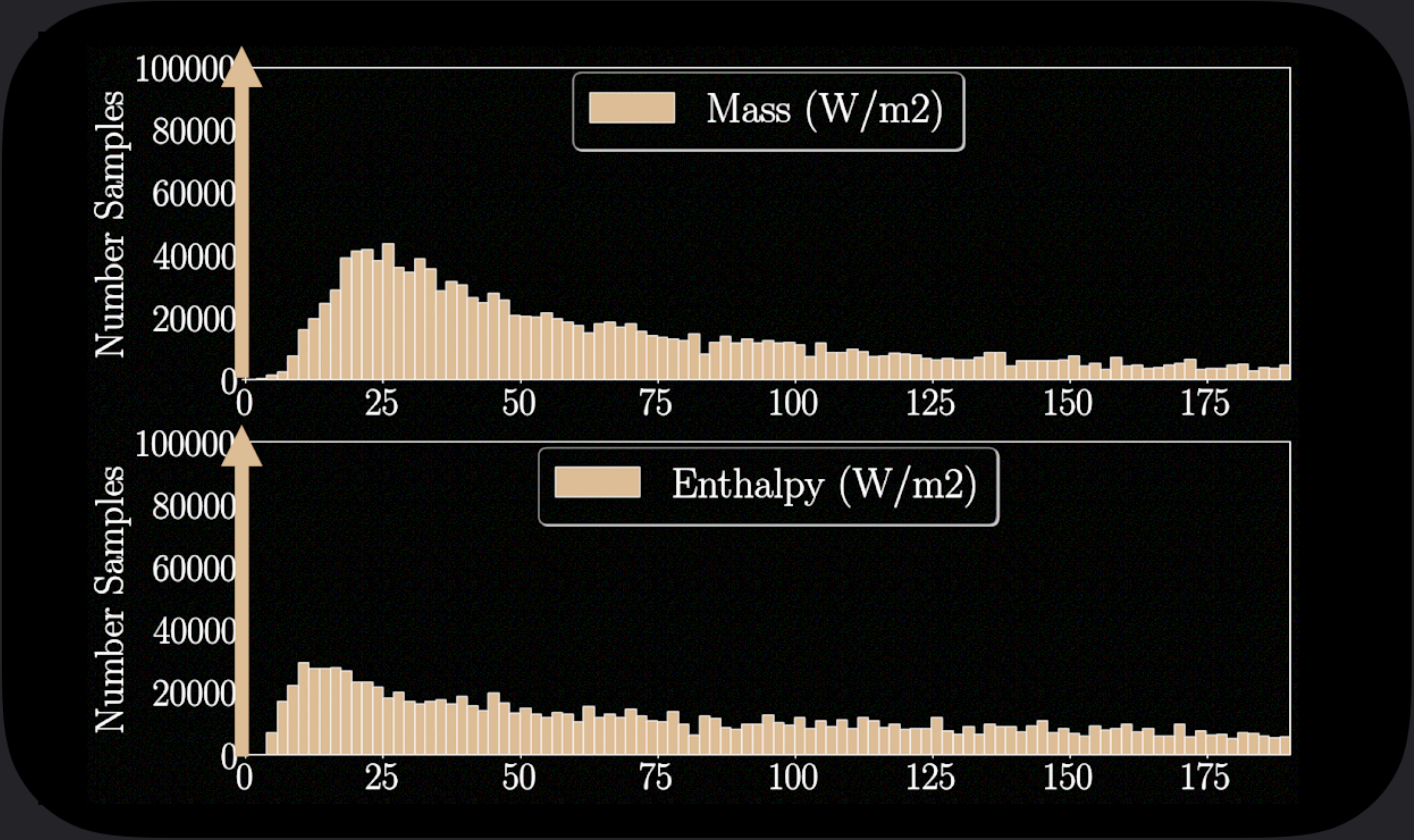


Figure courtesy of Tom Beucler, UCI.

Problem 2: Does the idea work beyond aquaplanets?

Problem 3: Instabilities abound and stable runs are rare.

Example of the neural network blowing up in prognostic mode.

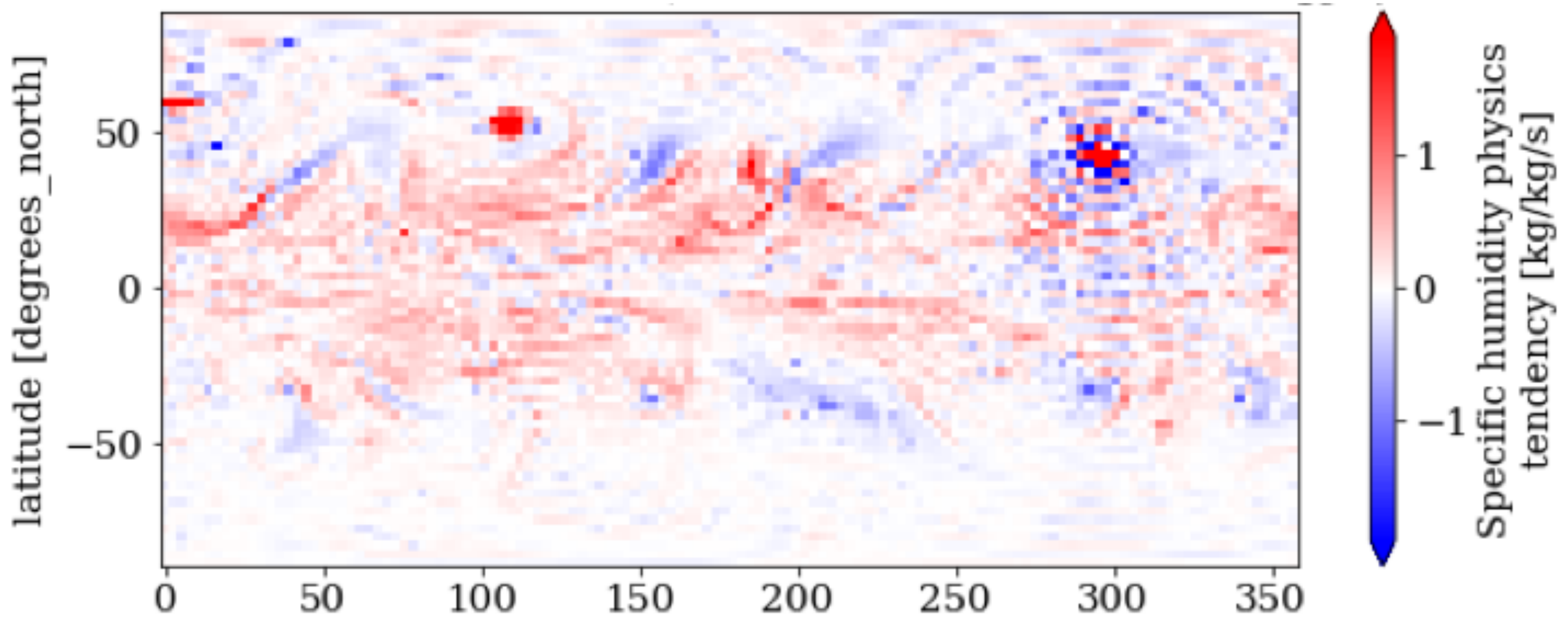


Figure courtesy of Tom Beucler, UCI

ROAD MAP

→ I. ADDING PHYSICAL CONSTRAINTS

Adapting the Rasp et al. NN to conserve column mass, enthalpy and radiation to precision.

II. FINDING QUALITY FITS IN MORE REALISTIC DATA

"SHERPA": A formal hyperparameter tuning package uncovers skill in a real-geography setting.

III. REPRODUCING ONLINE STABILITY.

"FKB" software that is helping probe the link between offline validation skill and online performance.

ROAD MAP

I. ADDING PHYSICAL CONSTRAINTS

→ *Adapting the Rasp et al. NN to conserve column mass, enthalpy and radiation to precision.*

II. FINDING QUALITY FITS IN MORE REALISTIC DATA

"SHERPA": A formal hyperparameter tuning package uncovers skill in a real-geography setting.

III. REPRODUCING ONLINE STABILITY.

"FKB" software that is helping probe the link between offline validation skill and online performance.

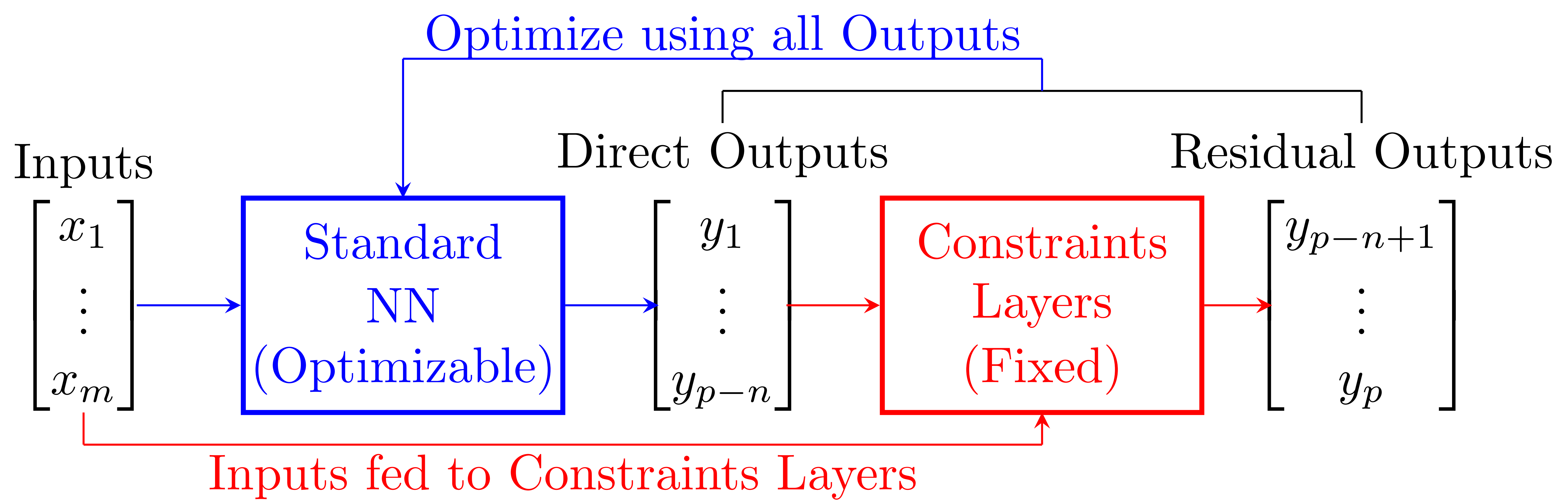
How to physically constrain neural network parameterizations?

Option #1:

Through the loss function:

$$\text{Loss} = \alpha \left| \boldsymbol{C} \begin{bmatrix} x \\ y \end{bmatrix} \right|_2 + (1 - \alpha)\,(\text{Mean} - \text{squared error}) \quad , \quad \alpha \in [0, 1]$$

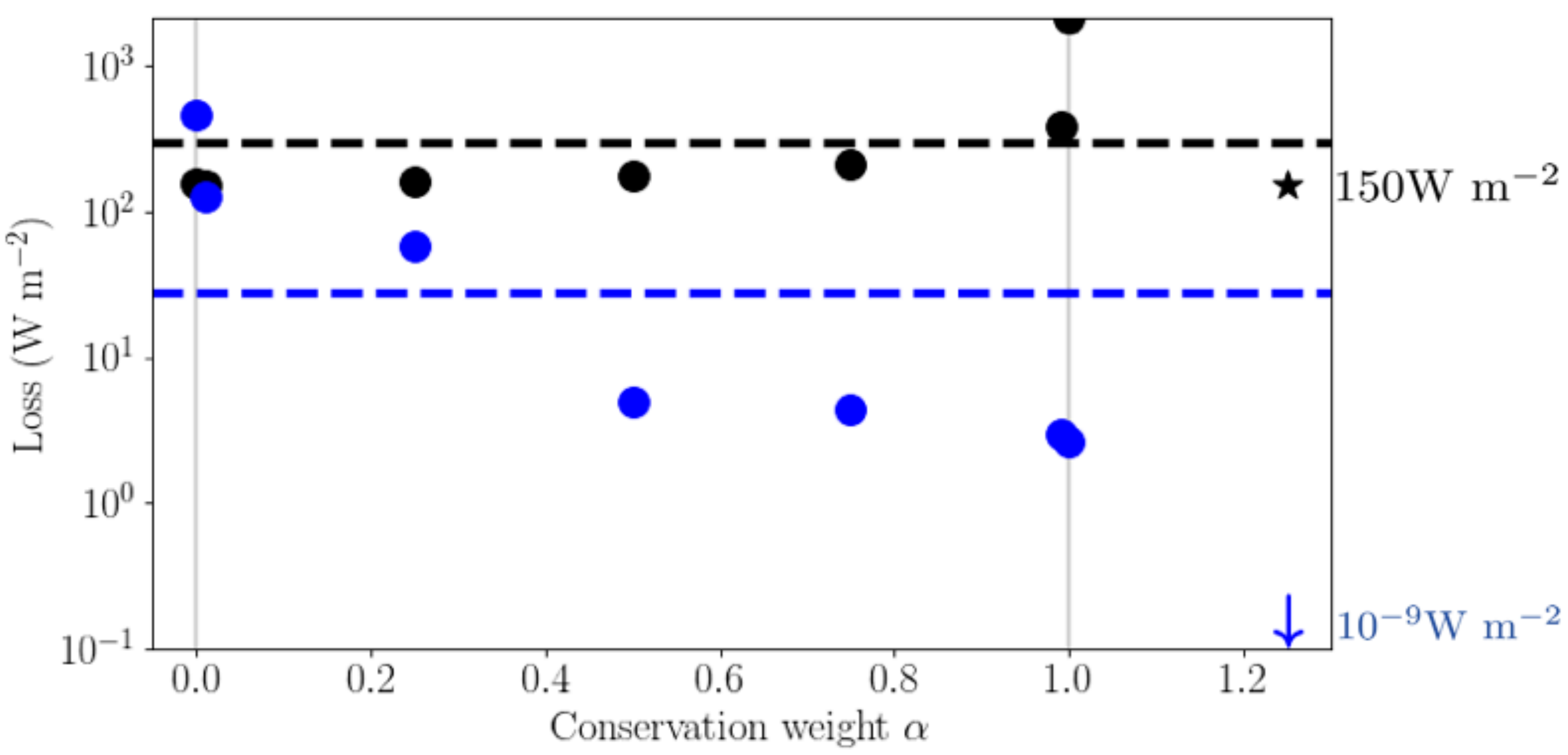
Option #2:
Hard constraints in the architecture:



Tom Beucler's idea: Enforce *n* constraints ***within*** the neural net architecture.

Beucler, Pritchard et al., 2020 arXiv:1909.00912

Architecture constrained DNNs perform well.



Loss: Trade-off between **physical constraints** and **performance**

Architecture: **Constraints enforced & competitive performance**

Beucler, Pritchard et al., 2020 arXiv:1909.00912

ROAD MAP

I. ADDING PHYSICAL CONSTRAINTS

→ *Adapting the Rasp et al. NN to conserve column mass, enthalpy and radiation to precision.*

II. FINDING QUALITY FITS IN MORE REALISTIC DATA

“SHERPA”: A formal hyperparameter tuning package uncovers skill in a real-geography setting.

III. REPRODUCING ONLINE STABILITY.

“FKB” software that is helping probe the link between offline validation skill and online performance.

ROAD MAP

I. ADDING PHYSICAL CONSTRAINTS

Adapting the Rasp et al. NN to conserve column mass, enthalpy and radiation to precision.

→ II. FINDING QUALITY FITS IN MORE REALISTIC DATA

“Sherpa”: A formal hyperparameter tuning package uncovers skill in a real-geography setting.

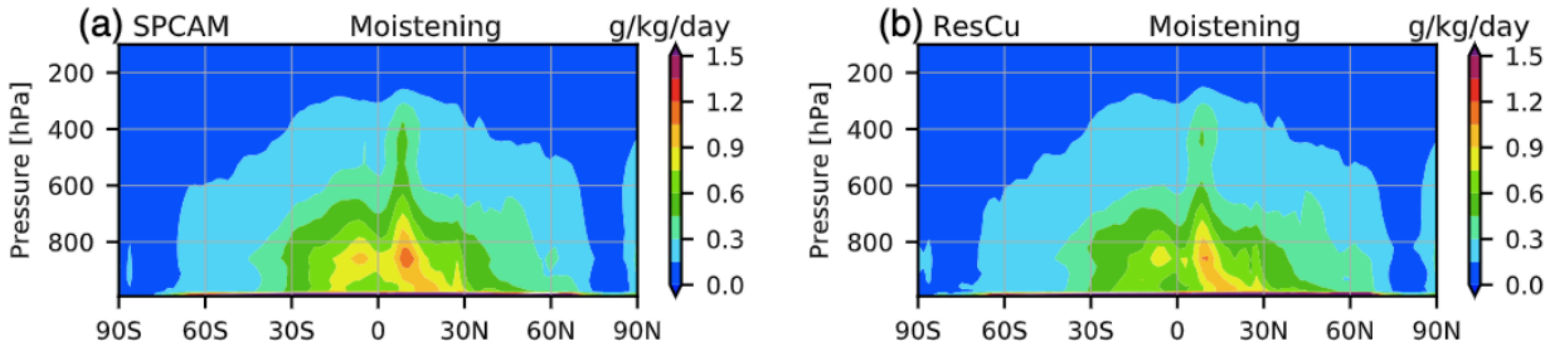
III. REPRODUCING ONLINE STABILITY.

“FKB” software that is helping probe the link between offline validation skill and online performance.

Problem 2: Does the idea work beyond aquaplanets?

Han et al. 2020: First real-geography offline fits of a SP-GCM

Success requires “resnet” NN architectures + vertical convolution + memory back in time.



Is NN parameterizability harder than it has appeared?

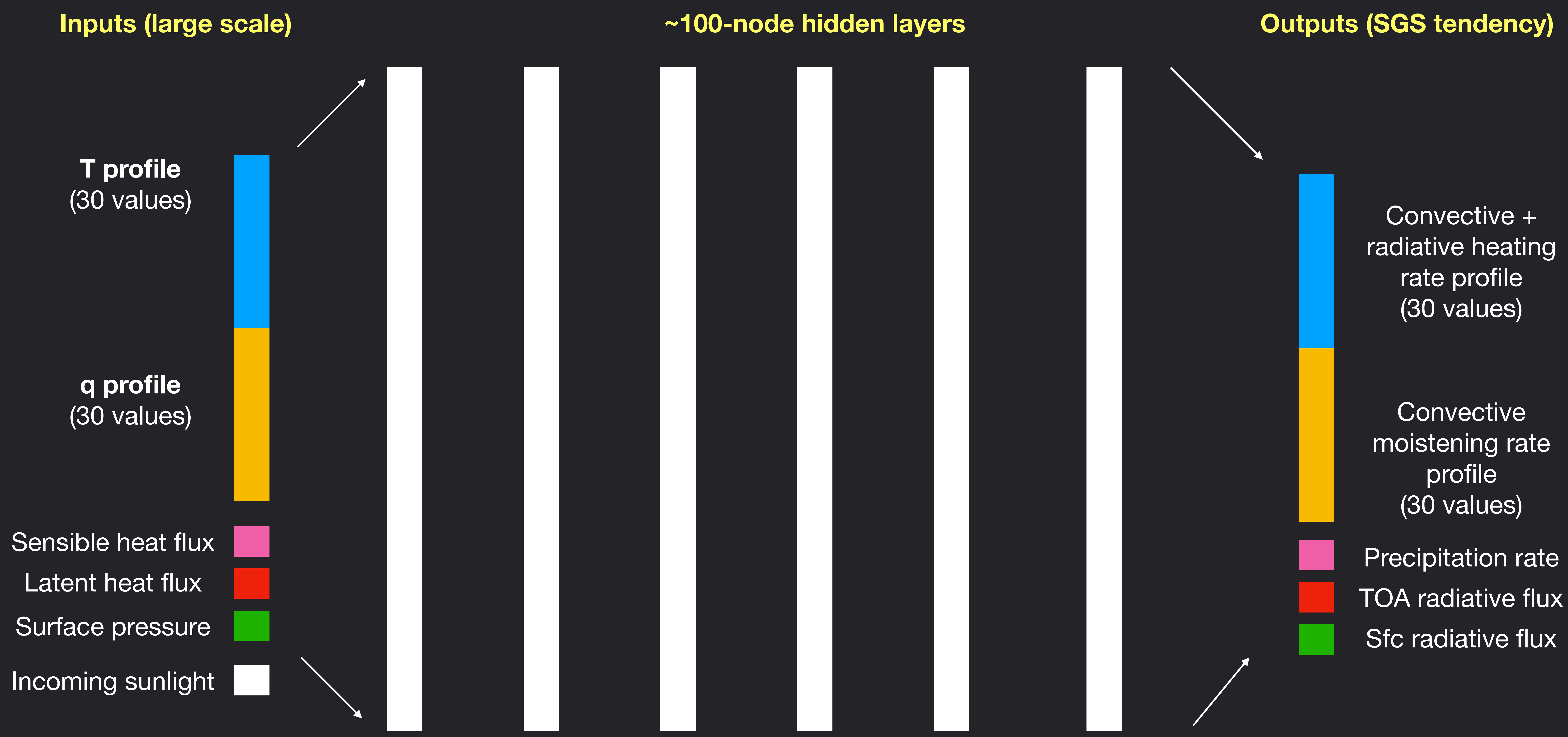
Han et al., *JAMES*, 2020.

Revisiting our own (simpler) DNN fits after relaxing aquaplanet idealizations

Model version: SPCAM3.0	→	SPCAM5
Dynamical core: Spectral + semi-Lagrangian	→	Finite-volume, 2-deg
Physics columns: ~8k	→	~14k
No geography or land	→	Real geography & land
No seasonality	→	Full seasonality
Weak oceanic diurnal cycles	→	Realistic diurnal convection cycles
Zonal symmetry	→	Walker cells, asymmetric storm tracks, etc.

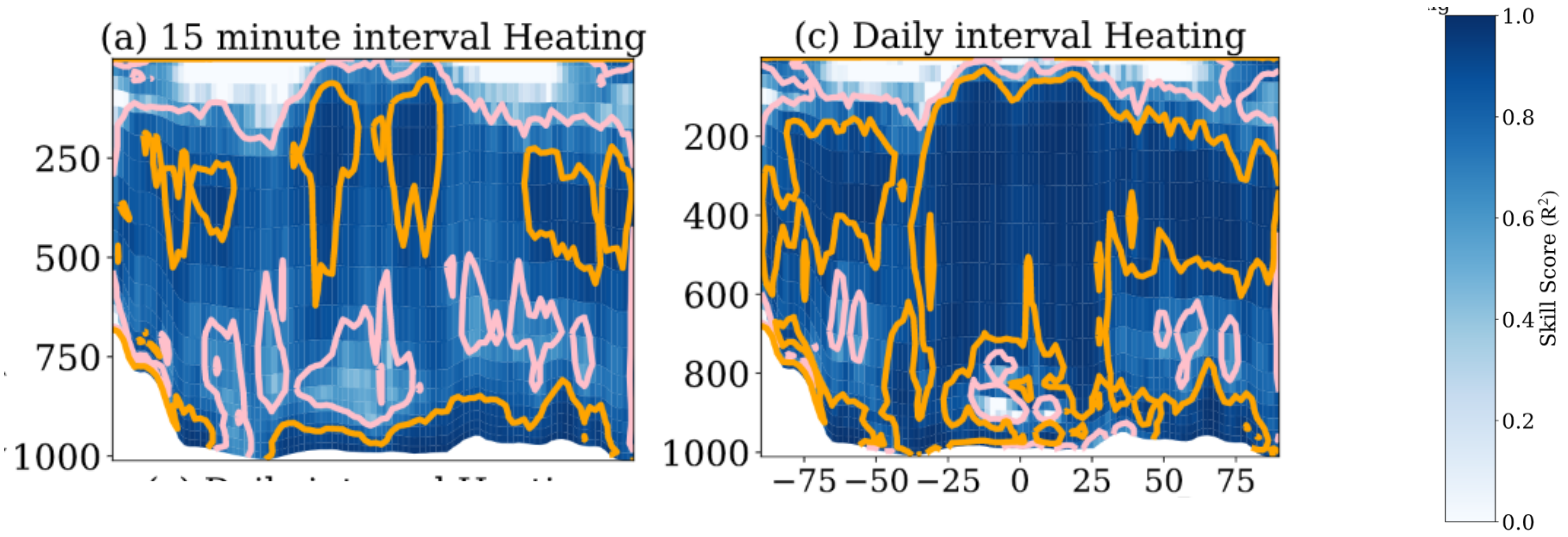
Mooers, Pritchard et al., in review (arXiv:2010.12996)

Schematic of the crude NN we will use.



Lessons learned in the real-geography limit.

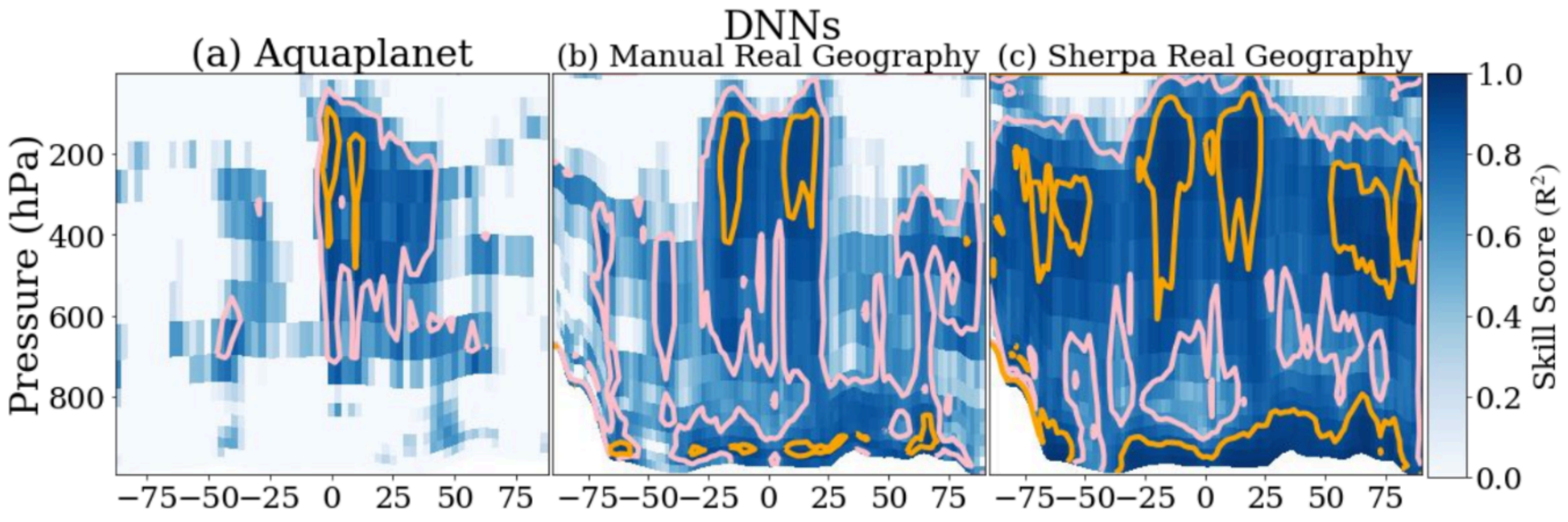
Competitive skill is possible even with crude DNNs



> 90 % temporal variance of zonal mean heating can be fit most places.

Mooers, Pritchard et al., in review (arXiv:2010.12996)

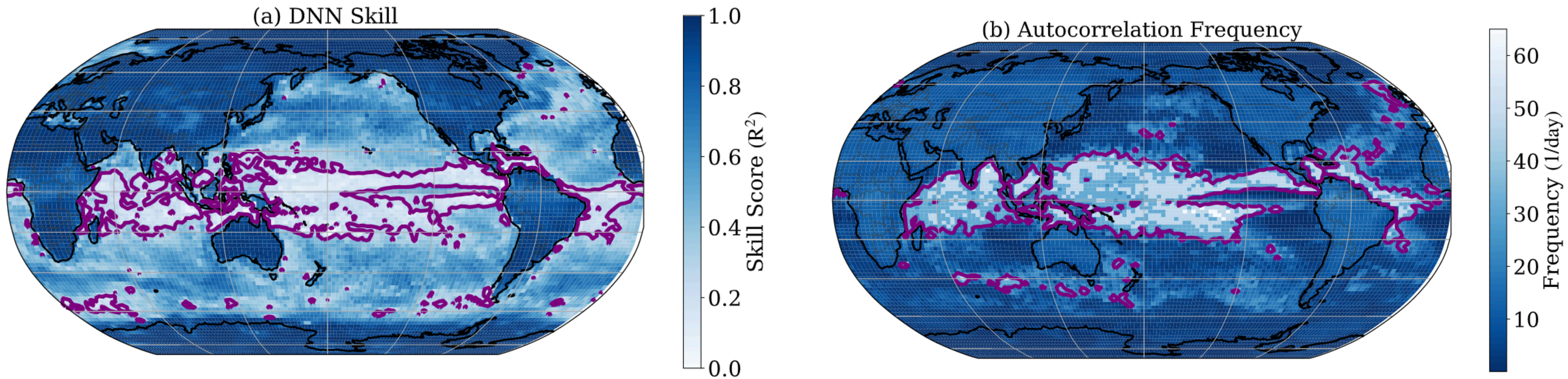
In-depth, automated hyperparameter tuning matters.



Sherpa: Easy to use semi-automated hyperparameter tuning software

Mooers, Pritchard et al., in review (arXiv:2010.12996)

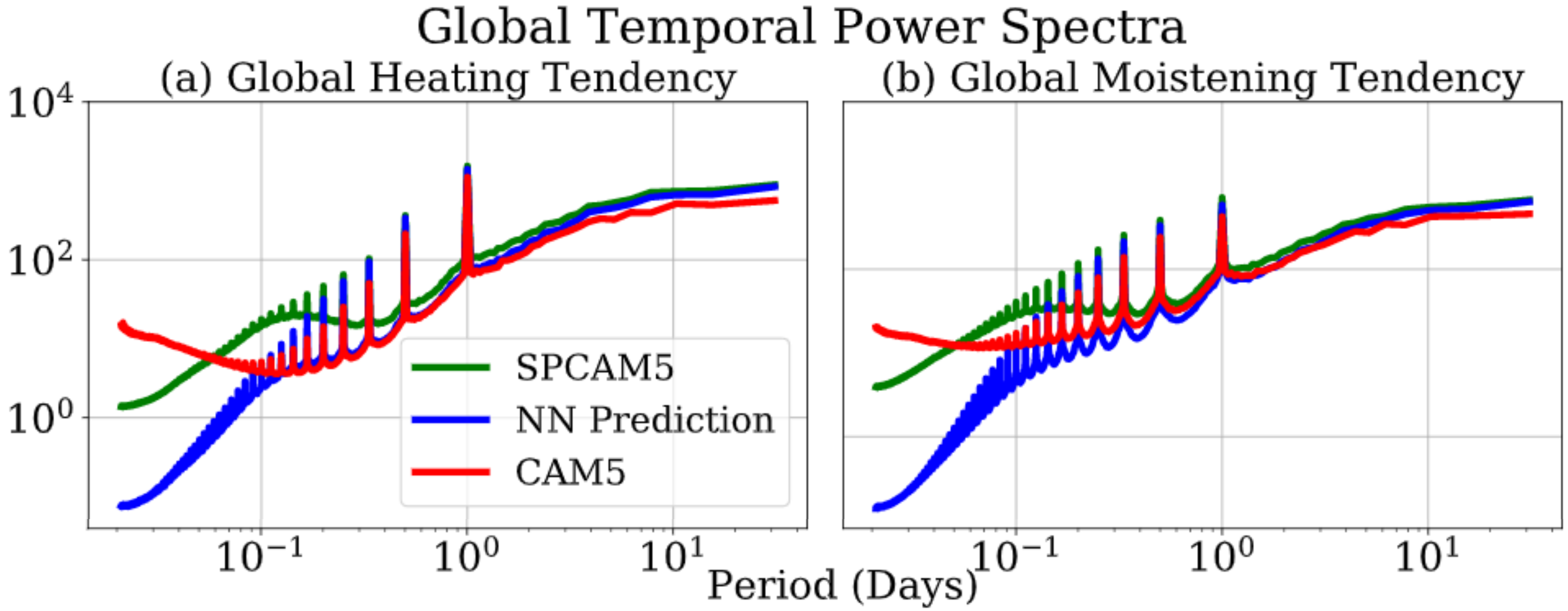
Least skill where signals decorrelate rapidly (e.g. tropical marine boundary layer)



Heating tendency skill at lowest model level

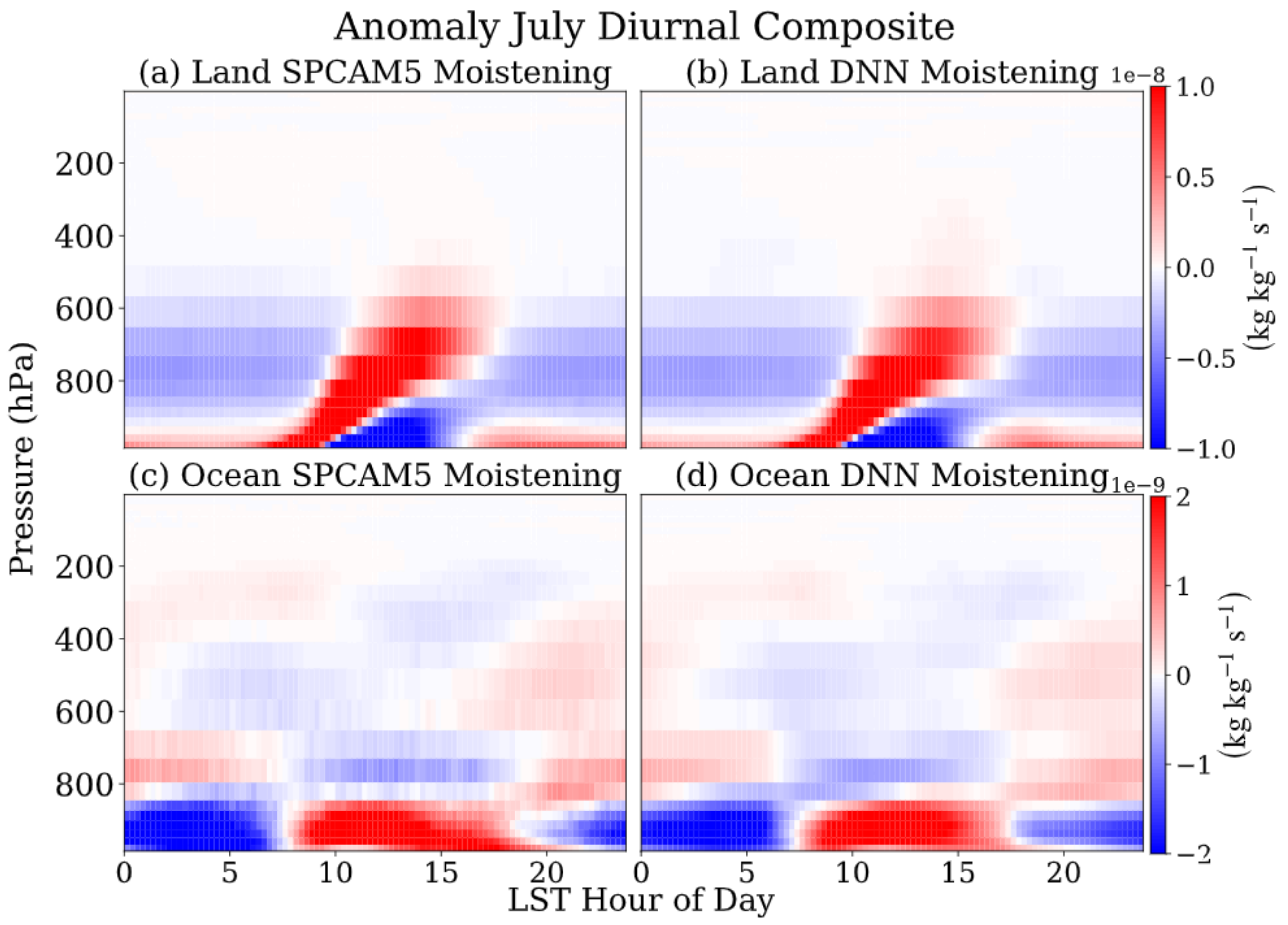
Mooers, Pritchard et al., in review (arXiv:2010.12996)

Synoptic and diurnal harmonics equally emulatable.



Mooers, Pritchard et al., in review (arXiv:2010.12996)

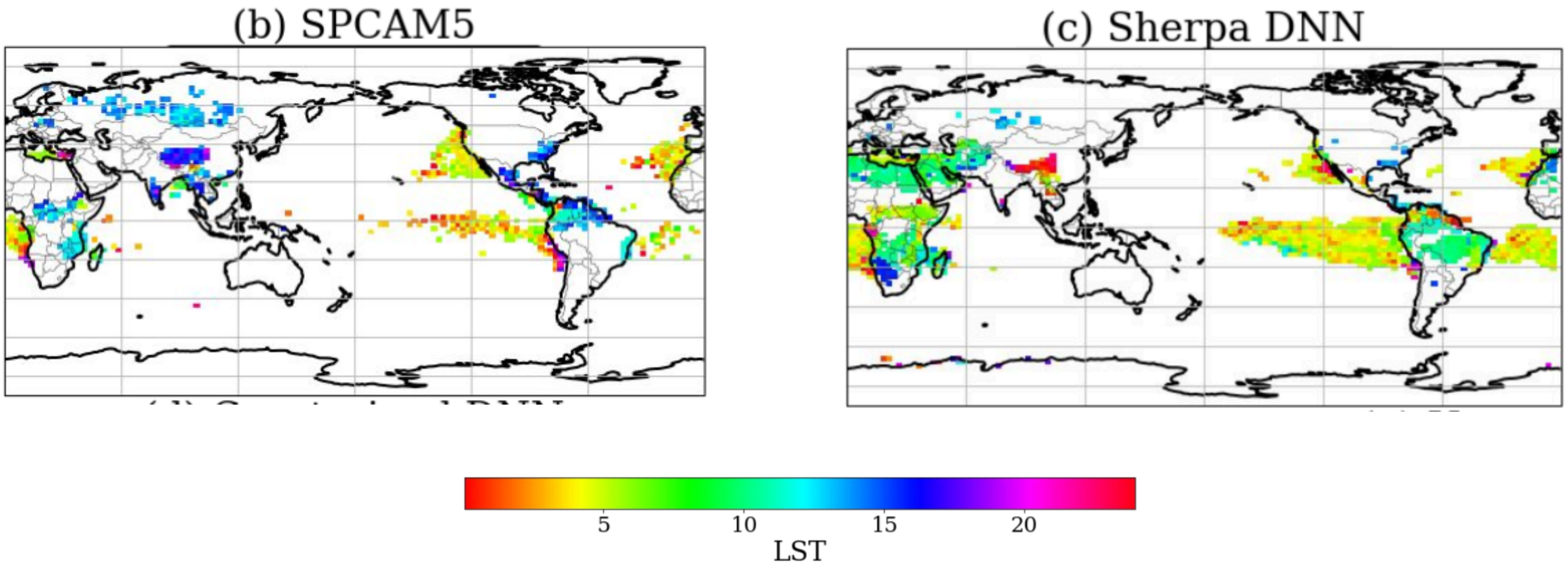
For heating & moistening, excellent diurnal composites.



Mooers, Pritchard et al., in review (arXiv:2010.12996)

For diurnal precipitation, at first a curious conundrum.

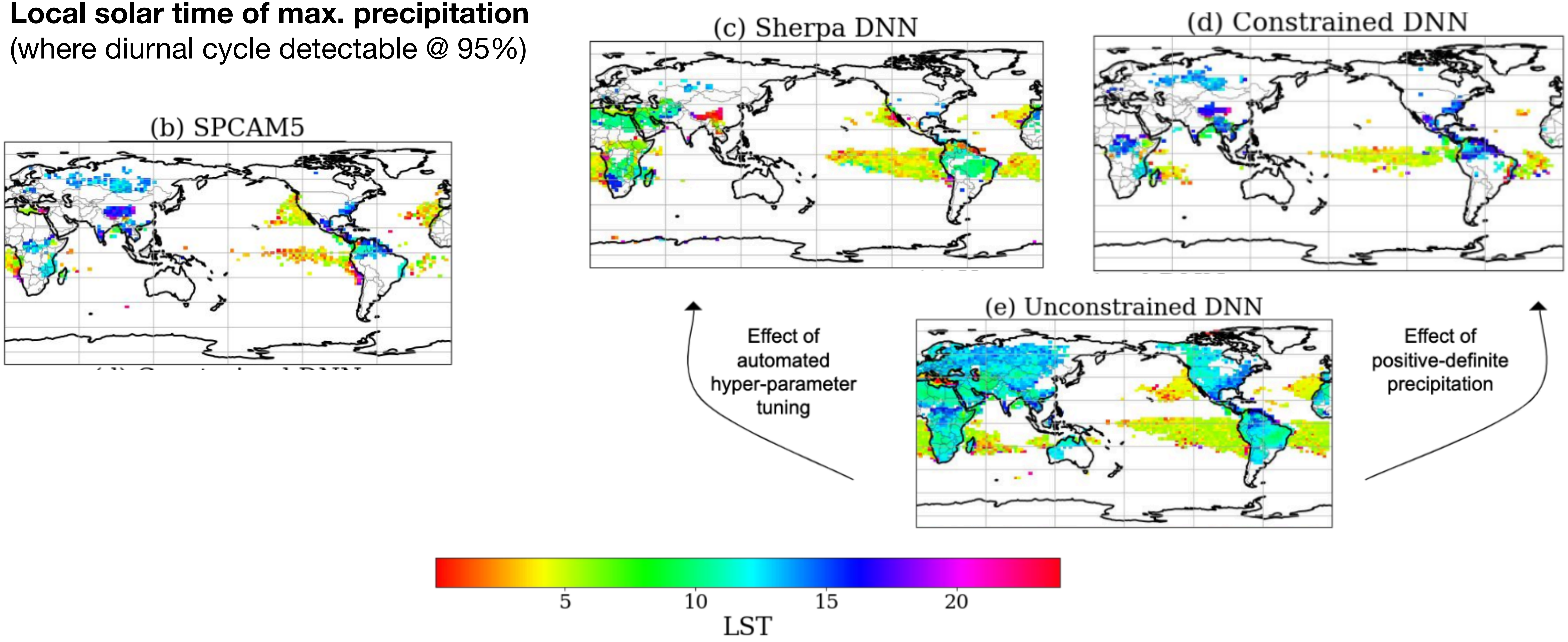
Diurnal rainfall:
Local solar time of max. precipitation
(where diurnal cycle detectable @ 95%)



Mooers, Pritchard et al., in review (arXiv:2010.12996)

Physical constraints vs intensive hyperparameter tuning can have complementary effects.

Diurnal rainfall:
Local solar time of max. precipitation
(where diurnal cycle detectable @ 95%)



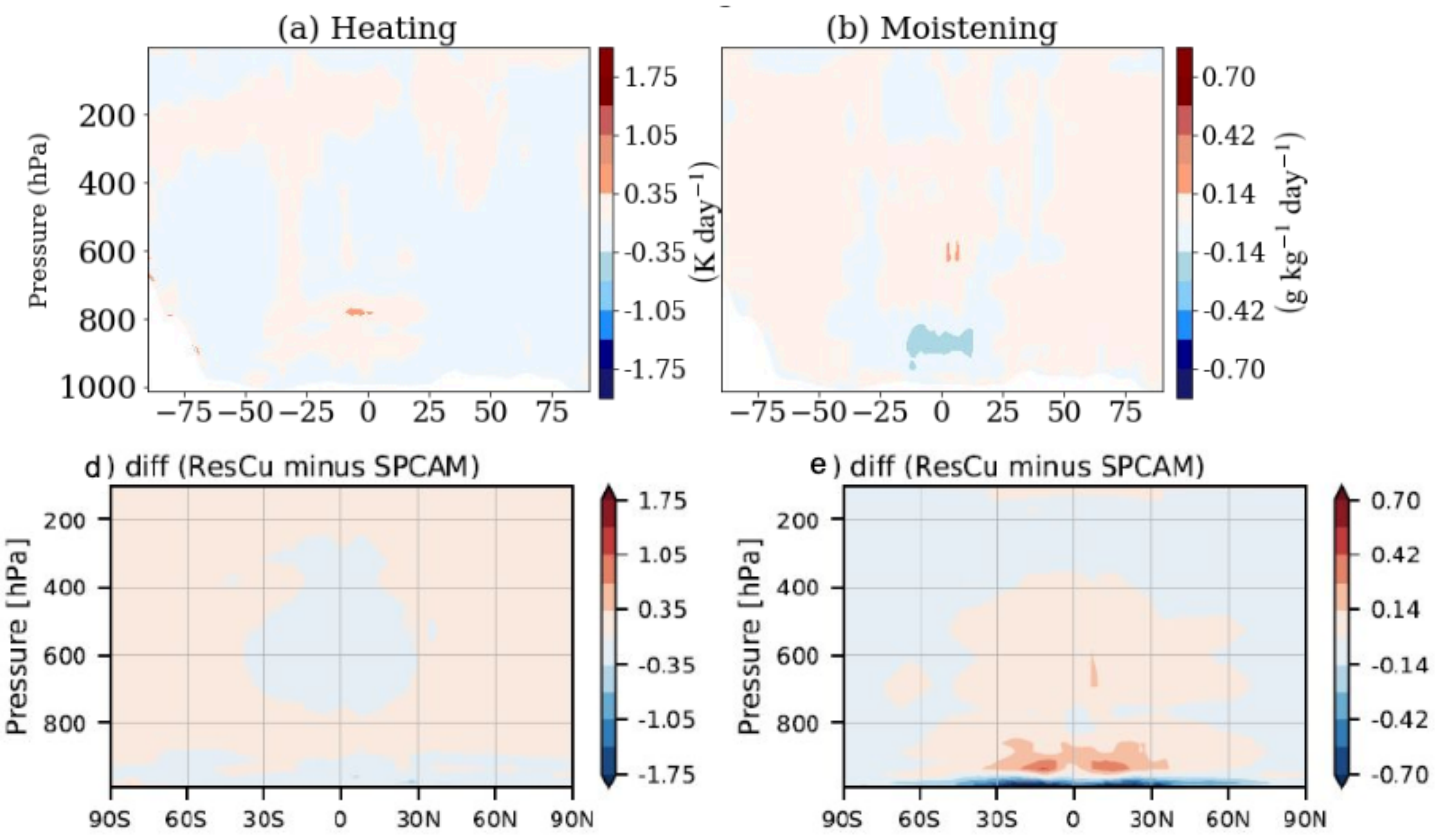
Mooers, Pritchard et al., in review (arXiv:2010.12996)

Bottom line: SP convection still seems parametrizable with NNs locally in time & without learning vertical basis functions.

Biases in the zonal mean heating & moistening rate (offline)

These results
(Crude DNN using current state only)

Han et al. 2020
(Sophisticated resnet including past state & vertical convolution)



Mooers, Pritchard et al., in review (arXiv:2010.12996)

ROAD MAP

I. ADDING PHYSICAL CONSTRAINTS

Adapting the Rasp et al. NN to conserve column mass, enthalpy and radiation to precision.

→ II. FINDING QUALITY FITS IN MORE REALISTIC DATA

“Sherpa”: A formal hyperparameter tuning package uncovers skill in a real-geography setting.

III. REPRODUCING ONLINE STABILITY.

“FKB” software that is helping probe the link between offline validation skill and online performance.

ROAD MAP

I. ADDING PHYSICAL CONSTRAINTS

Adapting the Rasp et al. NN to conserve column mass, enthalpy and radiation to precision.

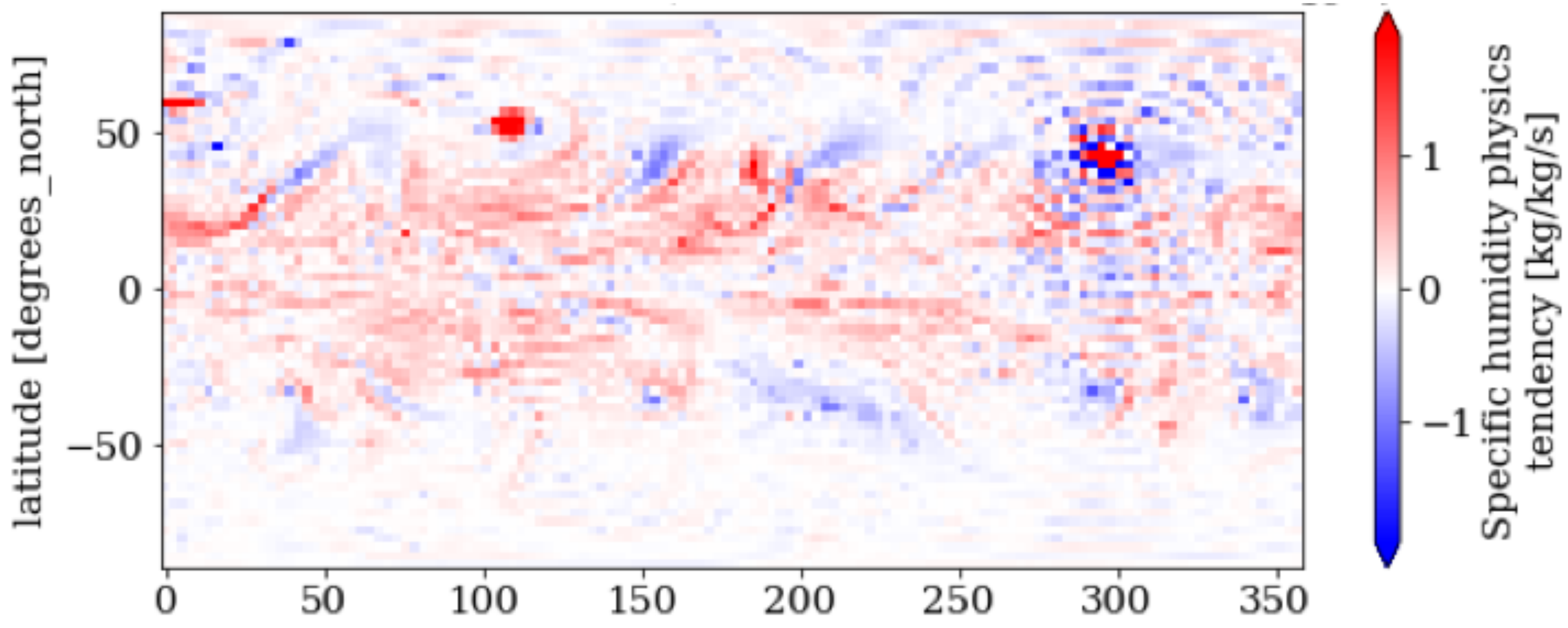
II. FINDING QUALITY FITS IN MORE REALISTIC DATA

“Sherpa”: A formal hyperparameter tuning package uncovers skill in a real-geography setting.

→ III. REPRODUCING ONLINE STABILITY.

“FKB” software that is helping probe the link between offline validation skill and online performance.

Instability crisis: minor enhancements of Rasp et al. 2018 aquaplanet training data have not succeeded prognostically.



More in: Brenowitz, Beucler, Pritchard & Bretherton, *JAS*, 2020.

Is machine learning emulation of subgrid cloud physics viable for operational climate simulation?

Can past successes be <u>reproduced</u> with enhanced training data?

Are instabilities controllable? Does offline NN skill predict online *coupled* performance?

?

Barriers

Tuning NNs is painful, 100's of tests.

Translating NNs to fortran kernels is hard.

Is machine learning emulation of subgrid cloud physics viable for operational climate simulation?

Can past successes be <u>reproduced</u> with enhanced training data?

Are instabilities controllable?
Does offline NN skill predict
online *coupled* performance?

?

Barriers

Tuning NNs is painful,
100’s of tests.

Translating NNs
to fortran kernels is hard.

New software…

“SHERPA”
Reliably autotuning NN parameters

“Fortran-Keras Bridge”
Simplifies embedding
NNs in climate models

Test dozens of candidate NNs in prognostic mode.

The Fortran-Keras Bridge (FKB)

https://github.com/scientific-computing/FKB

Takes the pain out of testing / training simple NNs within E3SM

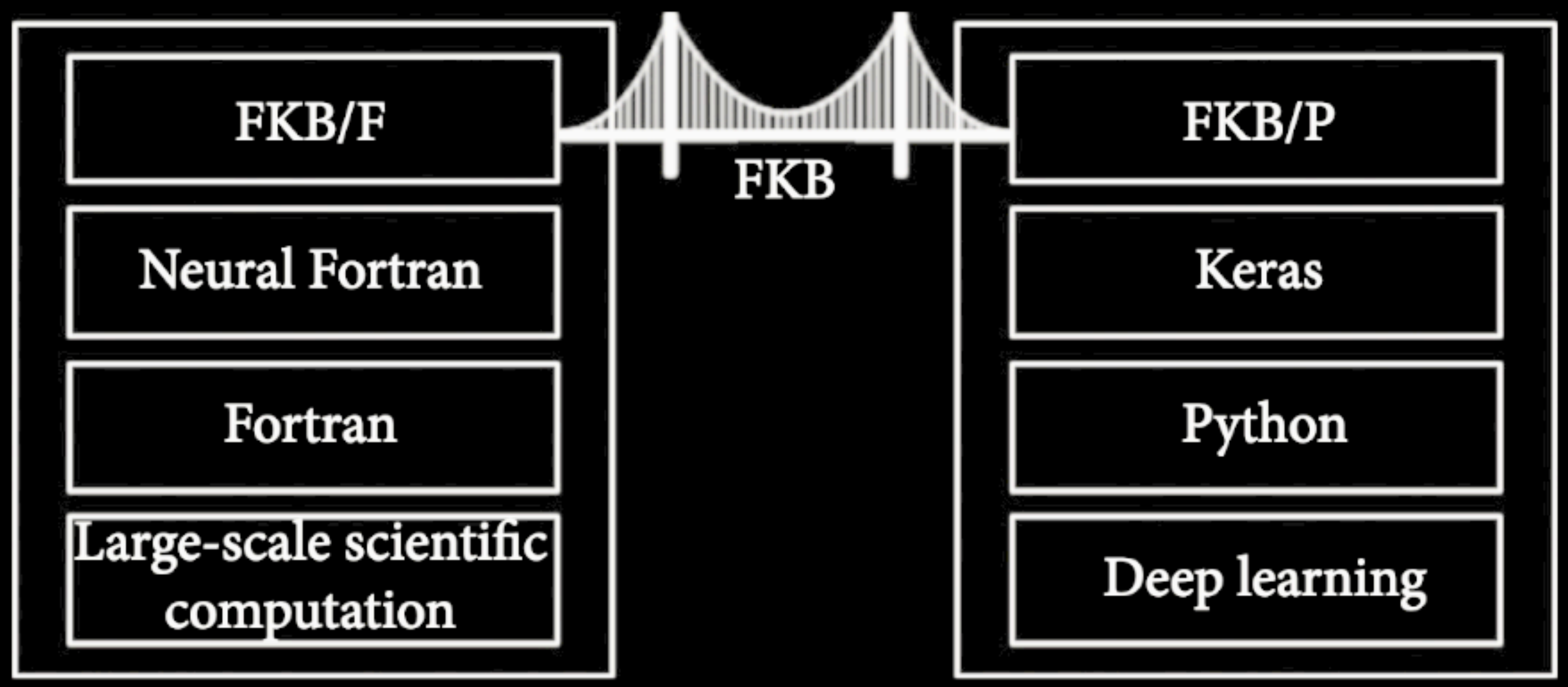
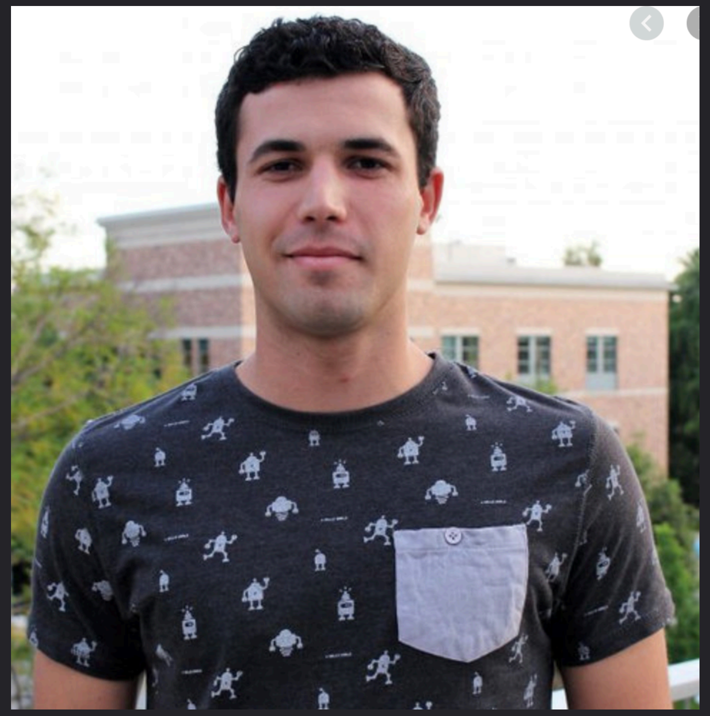


FIGURE 2: Positioning of FKB within Fortran and Python ecosystems.

Milan Curcic | Pierre Baldi | Jordan Ott

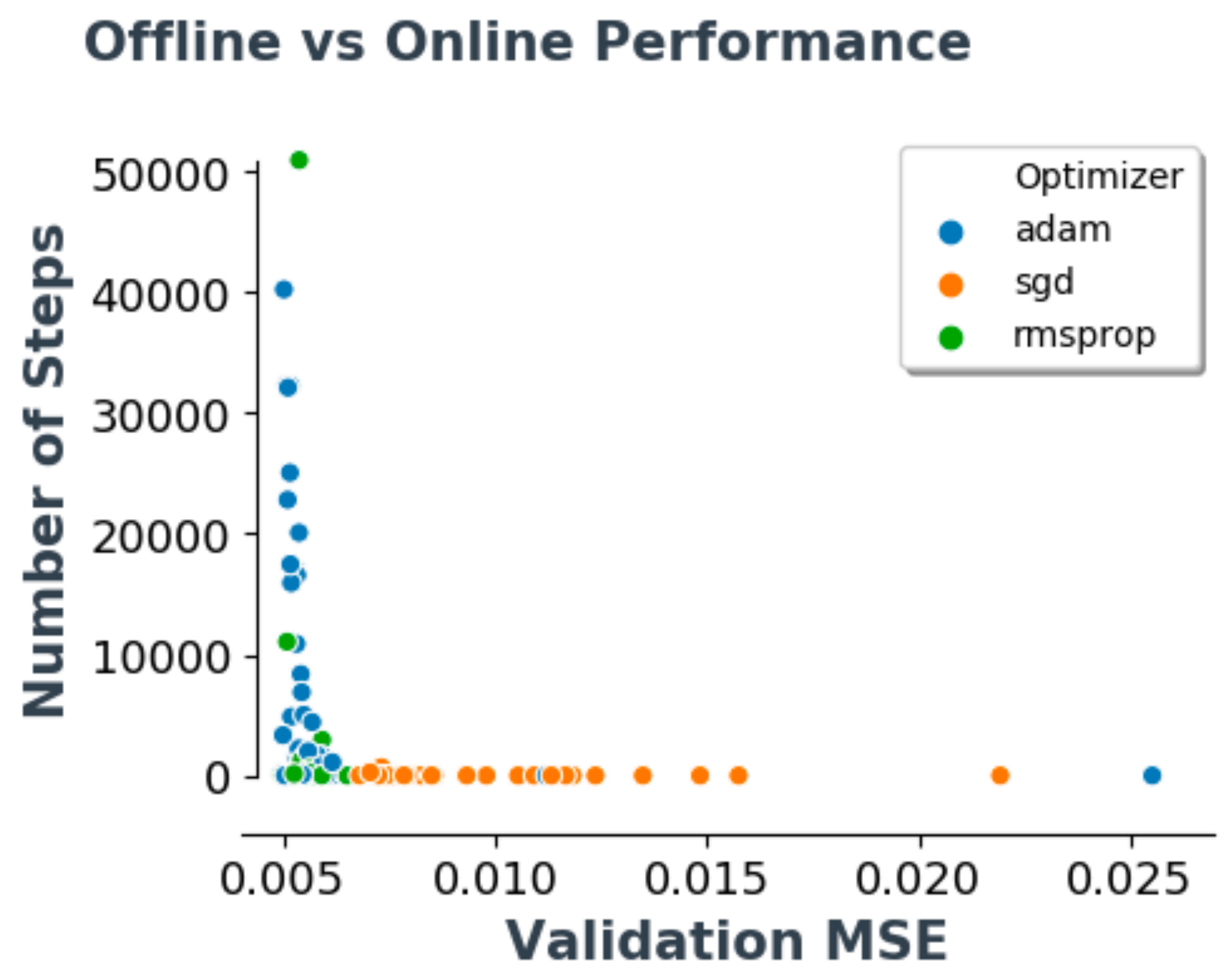


U. Miami | (UCI Computer Science)

Ott, Pritchard et al., *Scientific Programming*, 2020.

First results from a large ensemble of NN-coupled climate model tests (aquaplanet) sampling diverse architectures

Offline fit skill has predictive value & optimizer choice matters. (32-col SP-aquaplanet)



Ott, Pritchard et al., *Scientific Programming*, 2020.

Widespread hyperparameter tuning worth it but must be paired with widespread prognostic testing.

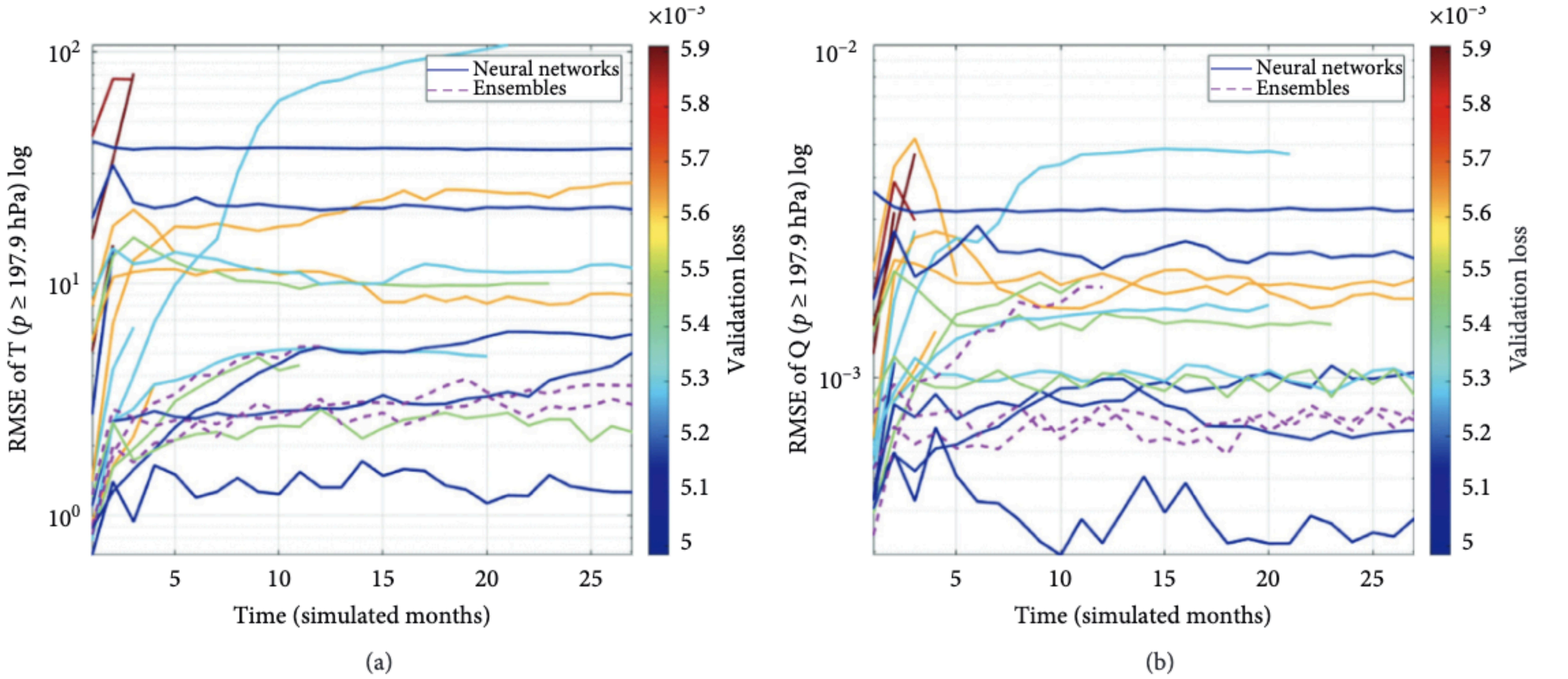


FIGURE 4: The time-evolution of the tropospheric (a) temperature and (b) humidity biases, colorized by the offline validation error.

Ott, Pritchard et al., *Scientific Programming*, 2020.

Is machine learning emulation of subgrid cloud physics viable for operational climate simulation?

Can past successes be reproduced with enhanced training data?

Yes. Formal hyperparameter tuning via Sherpa has yielded offline fits in real-geography mode that look competitive.

Will Sherpa+FKB prove reliable for success in prognostic real-geography mode?
(stay tuned!)

Are instabilities controllable? Does offline NN skill predict online *coupled* performance?

Perhaps. A stubborn instability when SP-AQUA is retrained on 32-col CRM data can be avoided with formal tuning.

While many candidate NNs drift to unphysical coupled attractors, optimal fits are performant

First reproduction of Rasp et al. success.

THANKS

It is an exciting time for numerical climate dynamics!