Pretraining and fine-tuning AI surrogates for GCMs

Abhnil Prasad[1], Kirill Trapeznikov[2], Steven Sherwood[1], David Fuchs[1], Jim Gimlett[2]

[1]Climate Change Research Centre, University of New South Wales, Sydney, Australia

[2]STR, Woburn, MA, USA

Introduction

- Developing a hybrid climate model with AI surrogates emulating physical parameterizations requires a dataset that captures subgrid dynamics.
- Most physics-emulators require large and accurate training data, but existing Cloud-Resolving Simulations (CRMs) do not fully resolve convection, observations from field campaigns do not measure all the relevant variables and Large-Eddy Simulations (LES) use only a limited domain.
- We propose and test a pretraining and fine-tuning strategy for combining the benefits of the different types of training data available to obtain better physics emulator using the concept of transfer learning, where a neural network is trained sequentially on two different but closely related datasets.

Datasets

- Four years of coarse-resolution climate data using the Community Atmosphere Model (CAM4) within CESM as the GCM.
- Three years of finer resolution data using the superparameterized CAM (SPCAM) with 16 columns of SAM replacing CAM4's conventional parameterization.

Neural Network Setup

- 7 fully connected (FC) layer blocks.
- Hidden dimension of each layer, $H = 512$.
- Weighted mse (optimization loss).

Pretraining, benchmarking and fine-tuning strategy

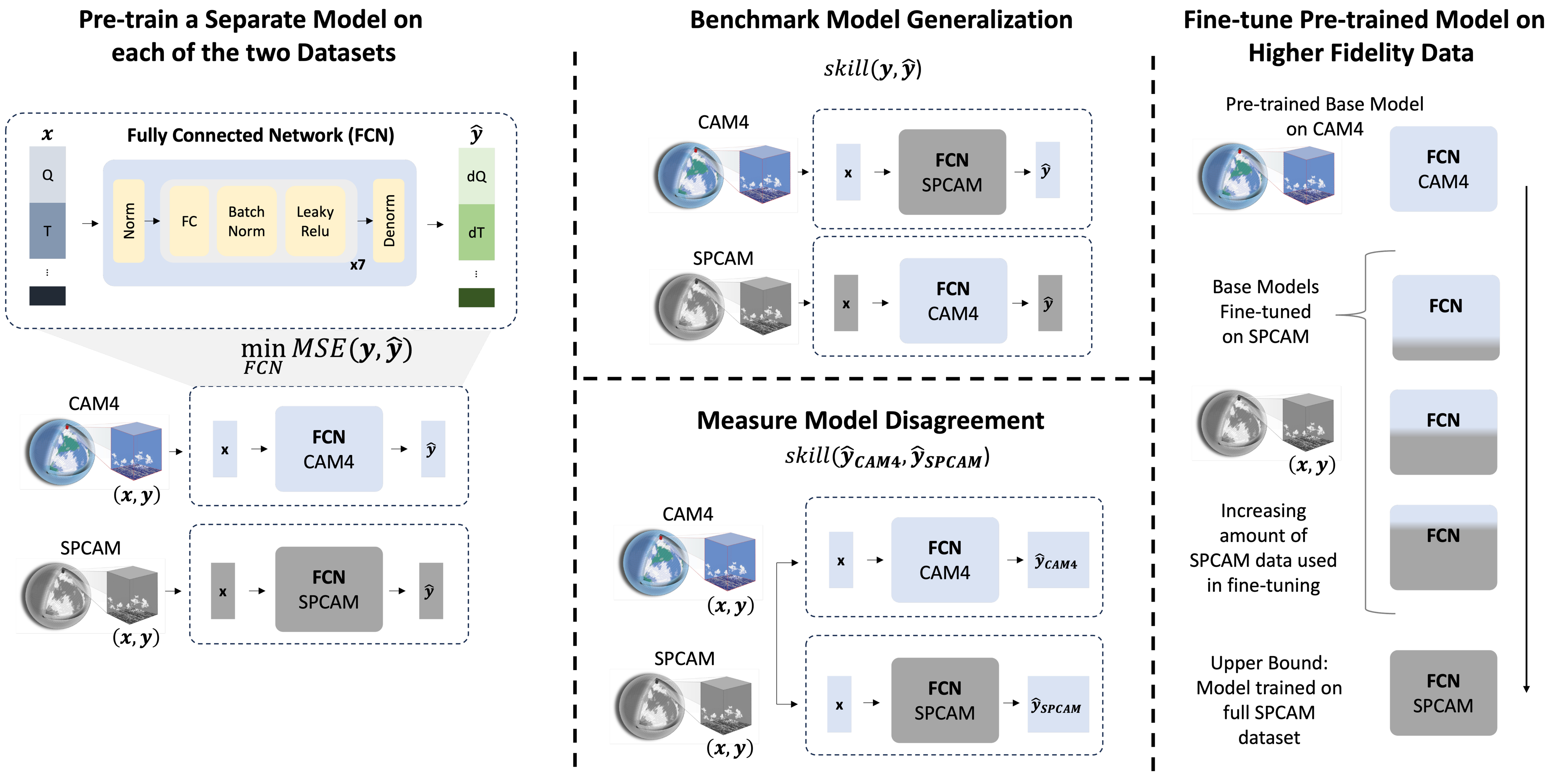

FIG1:A summary of experiments conducted with the surrogate.

Metrics

1. Truncated skill:

We compute the truncated skill analogously (where $\sigma^2(\cdot)$ is the variance averaged over the same dimensions as the mse)

$$skill(\mathbf{y}, \hat{\mathbf{y}}) = \max\left[0, 1 - \frac{mse(\mathbf{y}, \hat{\mathbf{y}})}{\sigma^2(\mathbf{y})}\right] \quad (1)$$

2. Disagreement skill:

We measure model disagreement on a specific dataset. Given samples from a $(\mathbf{x}, \mathbf{y})$, and two models (one trained on a dataset a, f^a and another trained on another dataset b, f^b), we compute disagreement skill by simply evaluating both models on the same inputs and normalizing by the variance of the matching output.

$$skill_D(f^a(\mathbf{x}), f^b(\mathbf{x}), \mathbf{y}) = \max\left[0, 1 - \frac{mse(f^a(\mathbf{x}), f^b(\mathbf{x}))}{\sigma^2(\mathbf{y})}\right] \quad (2)$$

Results

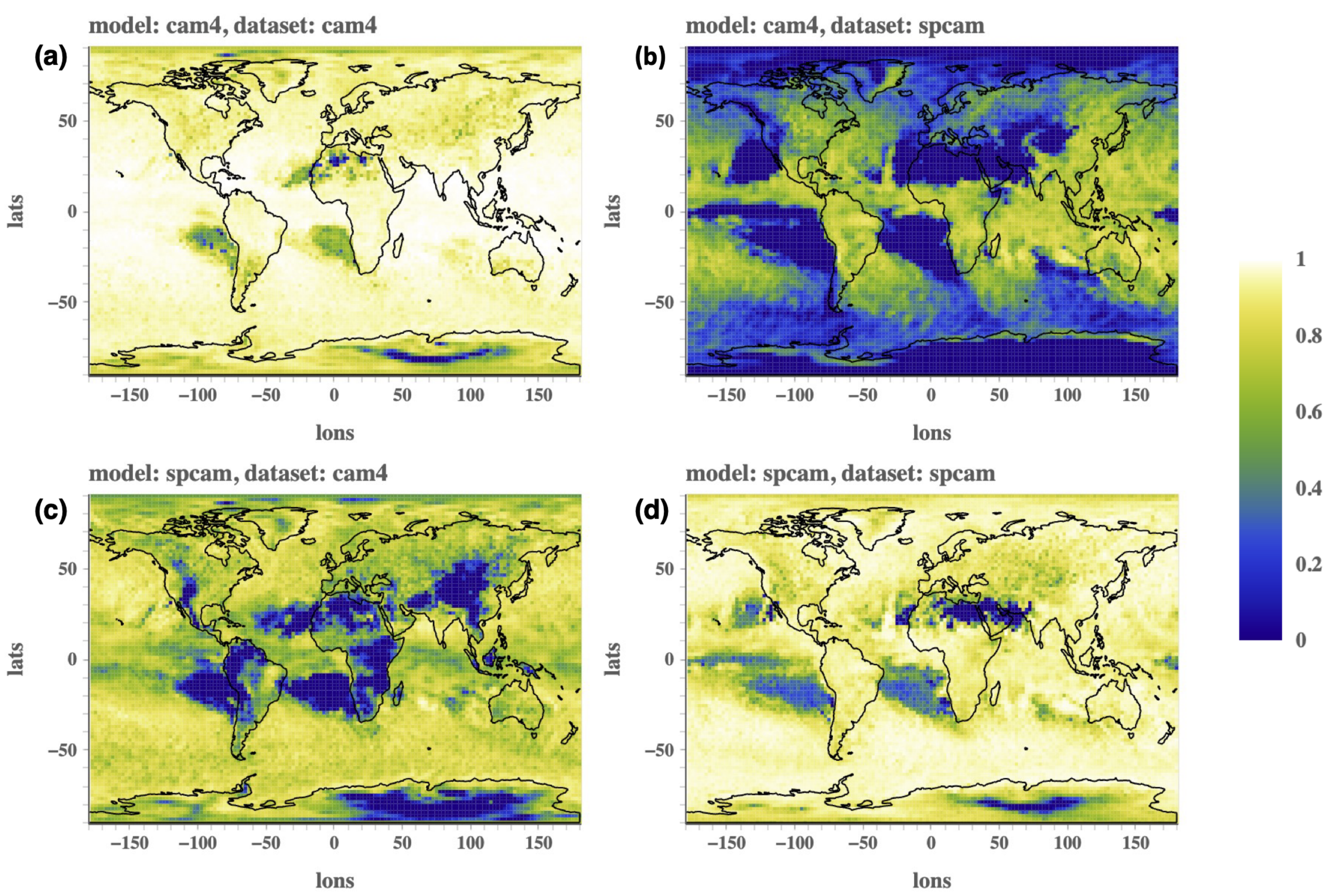

FIG2:Model transferability to another dataset evaluated on predicting the scalar output precipitation variable. Performance of (a) CAM4 surrogate on CAM4 test data, (b) CAM4 surrogate on SPCAM data, (c) SPCAM surrogate on CAM4 test data, and (d) SPCAM surrogate on SPCAM test data.

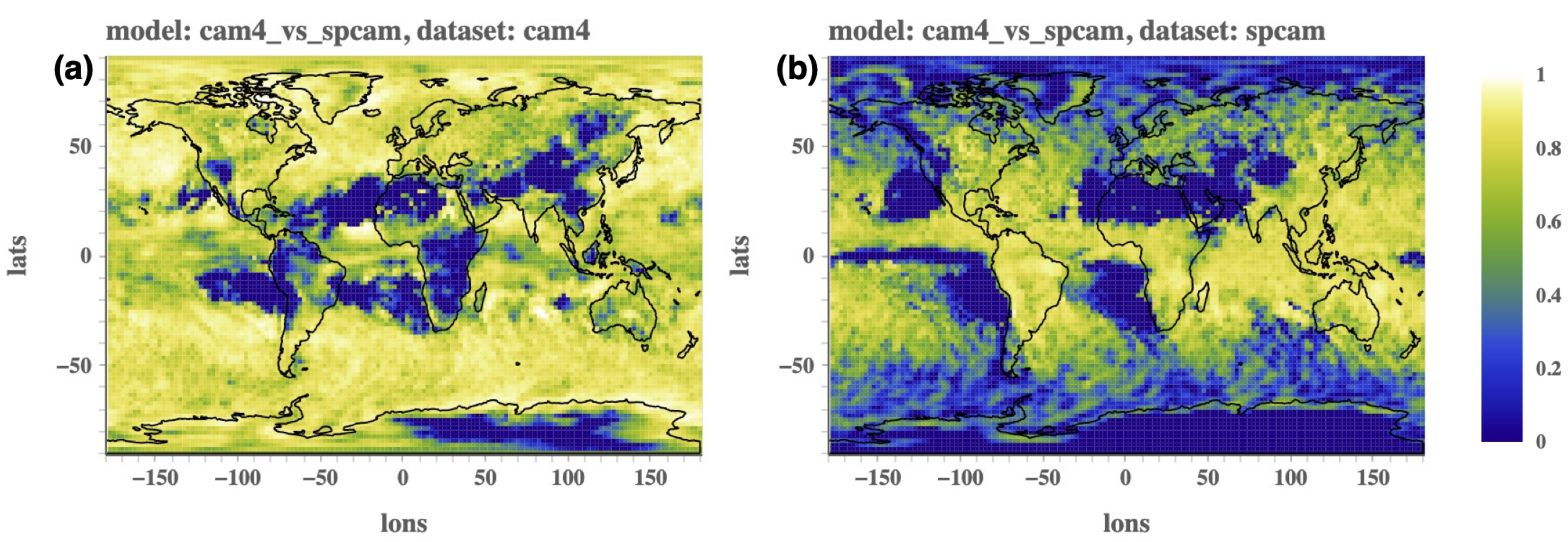

FIG3:Model Disagreement between a CAM4 and SPCAM trained model on two datasets.

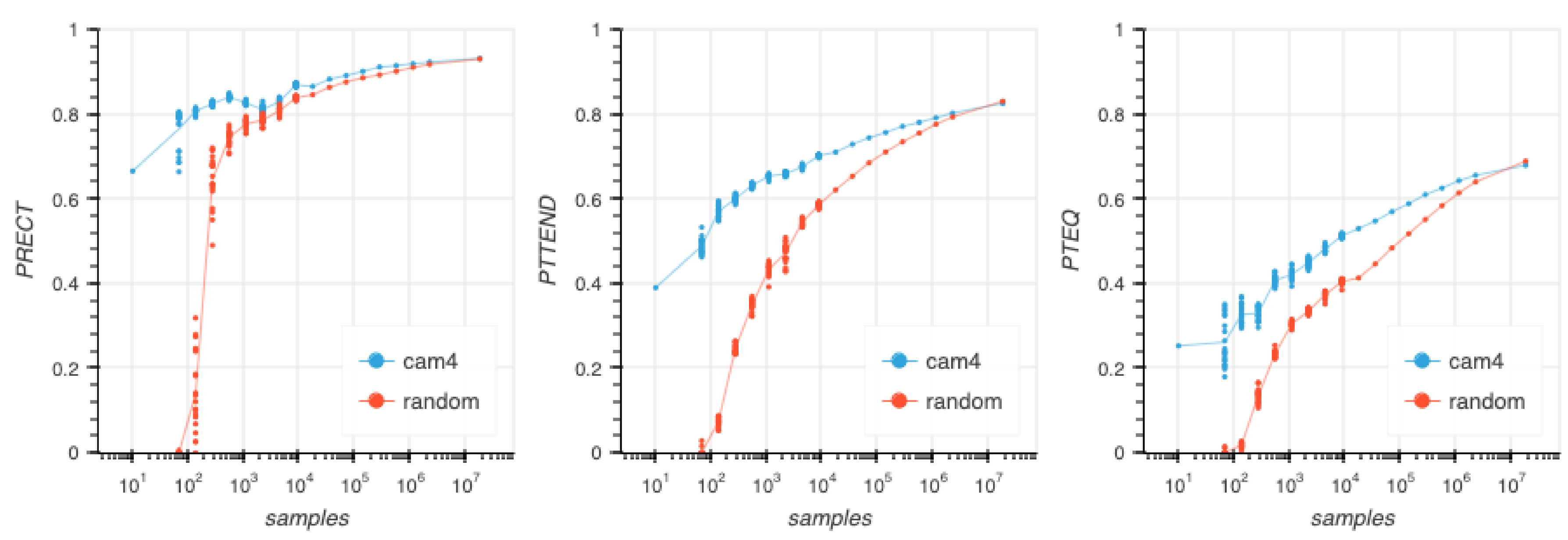

FIG4:The skill of each model as a function dataset size it was trained (or fine-tuned on) for precipitation, temperature tendency and moisture tendency.

Conclusions

- Both CAM4-trained and SPCAM-trained surrogates perform exceptionally when evaluated on the same parent data source, but the SPCAM-trained model generalizes better than the CAM4-trained surrogate. The model disagreement is much greater when evaluated on SPCAM data.
- We show strong gains in the ability to emulate SPCAM when fine-tuning a CAM4-trained surrogate on varying amounts of SPCAM data. Future work involves running the fine-tuned SPCAM emulator in a hybrid framework using TorchClim (https://doi.org/10.5194/gmd-17-5459-2024).