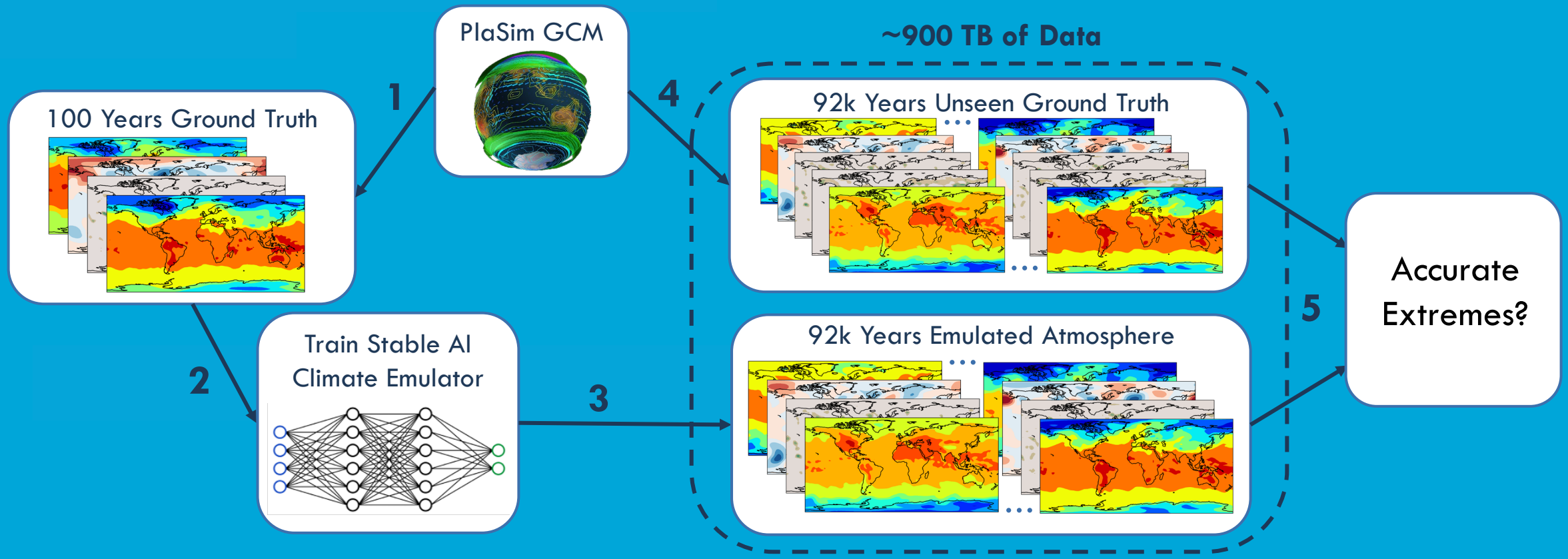
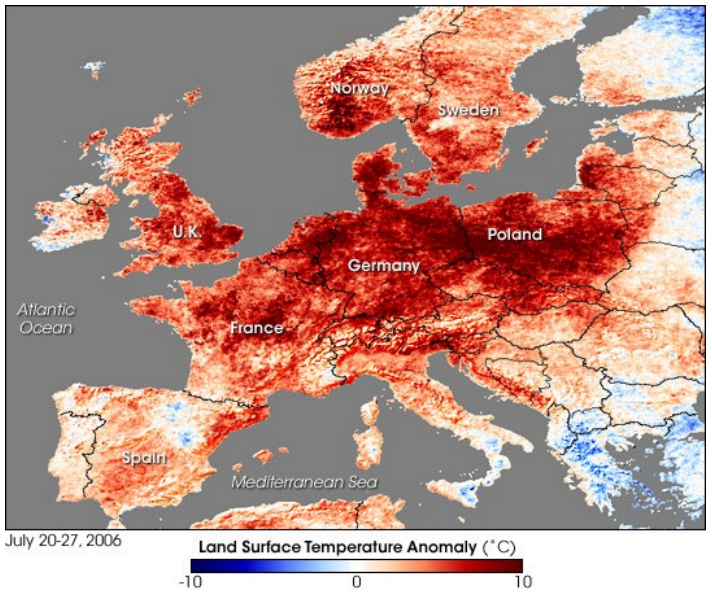
Can AI Climate Emulators Quantify the Statistics of the Rarest Unseen Weather Extremes?



Alex Wikner* (UChicago), Amaury Lancelin* (LMD & RTE), Troy Arcomano (AI2), Karan Jakhar (UChicago), Pedram Hassanzadeh (UChicago), Dhruvit Patel (UChicago), Freddy Bouchet (LMD, CNRS)

The few most extreme weather events have more impact than all the others, but we lack data to study those



- **Historical records** are **too short** to study the rarest events (e.g., 50 years of data vs. events with > 100-year return periods).
- **Climate models** are powerful but biased, and high-resolution ones are often too **computationally expensive** for rare event analysis.
- Statistical techniques like **Extreme Value Theory** are useful tools but exhibit **large uncertainties** and **do not sample** actual events

AI emulators have the potential to substantially reduce extreme weather risk uncertainty

- AI emulators can remain stable for many decades
- Computational efficiency allows emulators to produce very many samples, potentially of high-impact extremes
- However, emulators may be inaccurate for events with return periods longer than the training

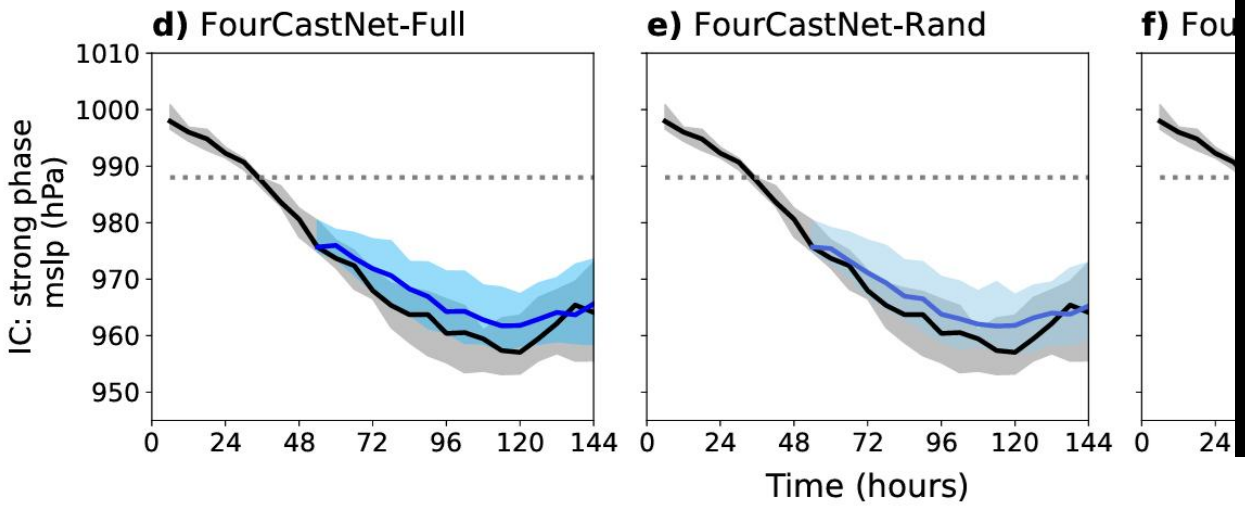
AI climate emulators

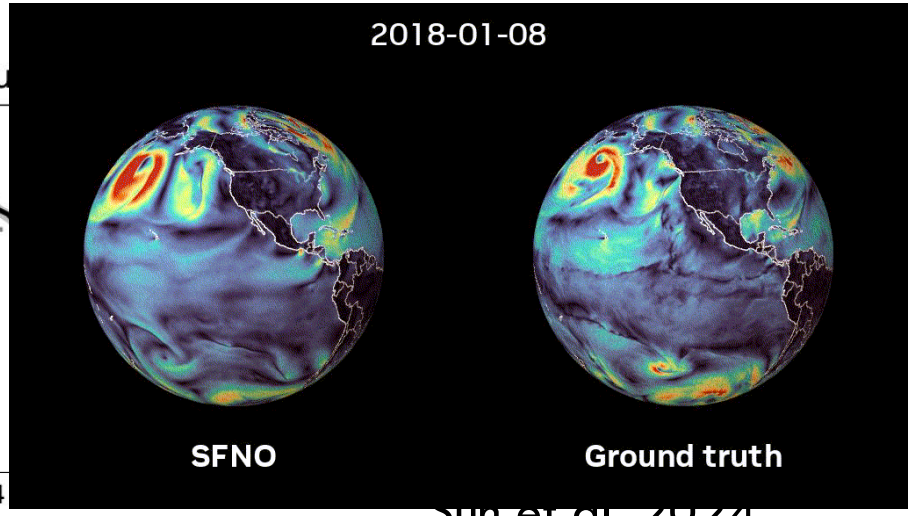
autoregressive

$$\boldsymbol{X}(t+\Delta t) = \mathbf{NN}(\boldsymbol{X}(t), f(t), \boldsymbol{\theta})$$

$$\boldsymbol{f} = \begin{bmatrix} \text{TOA solar radiation} \\ \text{SST} \\ \text{land, ice etc} \end{bmatrix}$$

boundary-value solver





Credits: Nvidia

Sun et al. 2024

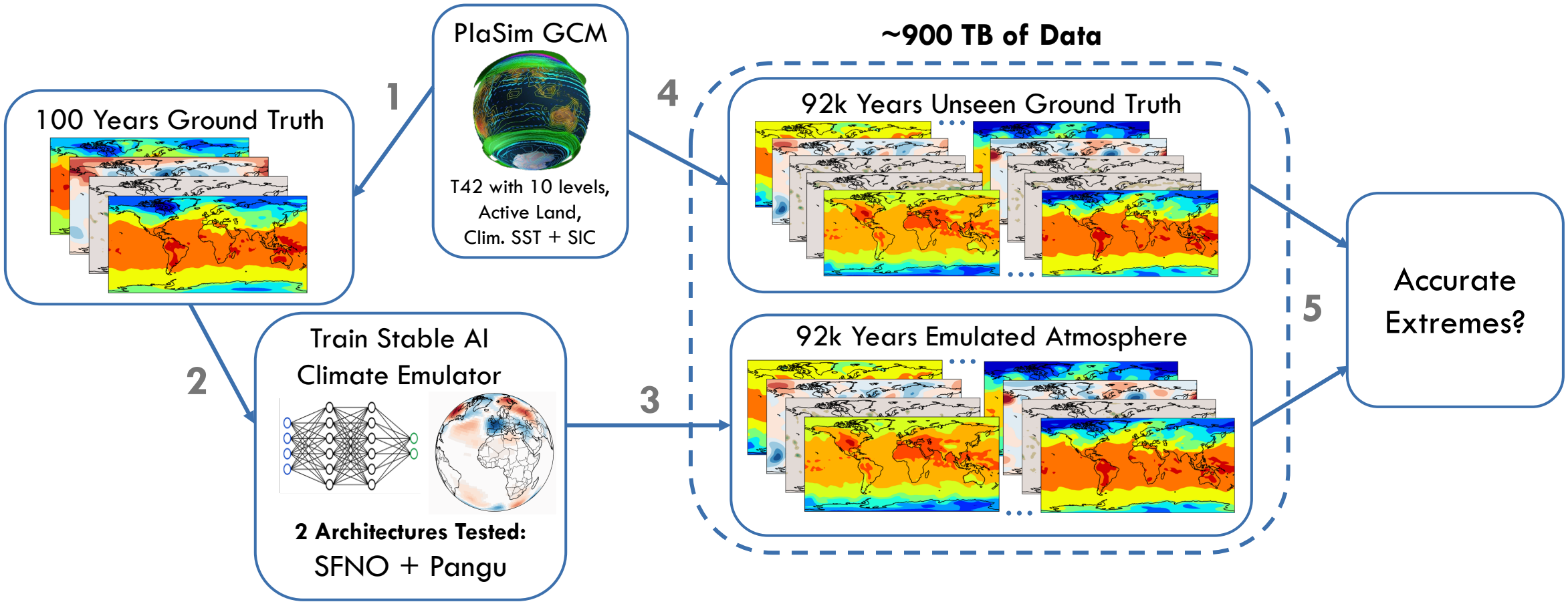
Key questions

Can AI climate emulators learn from short-duration training data such that:

1. They emulate very rare events with return periods longer than the training data?
2. Are they useful compared to traditional statistical methods: unbiased estimators of rare probabilities?*
3. Very rare events in the true system have similar dynamics to very rare events produced by the emulator?*

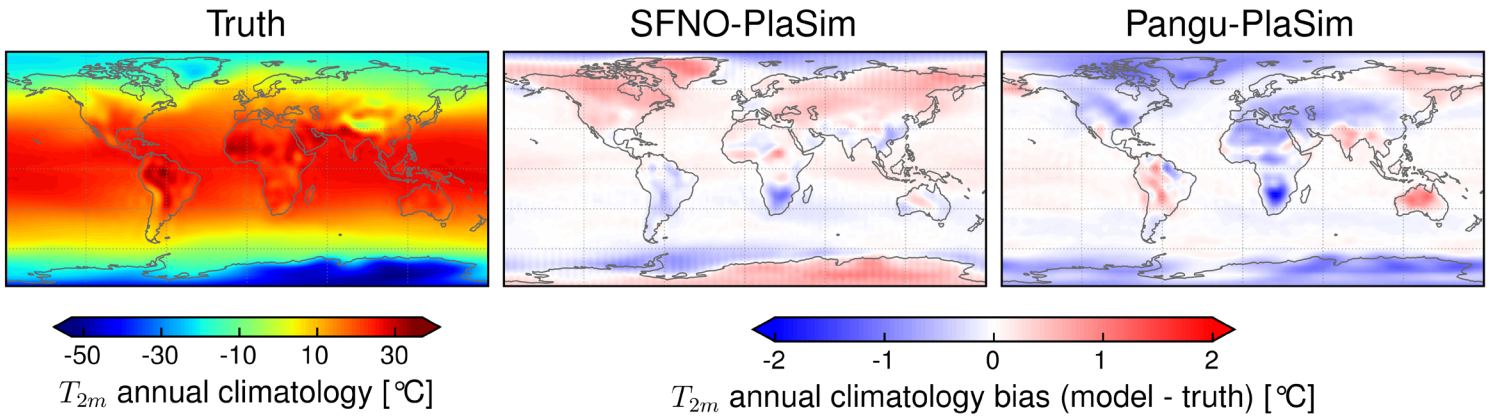
* To validate 2 and 3, we need a very long ground truth data set

We evaluate emulated rare event return times with a very long ground truth data set



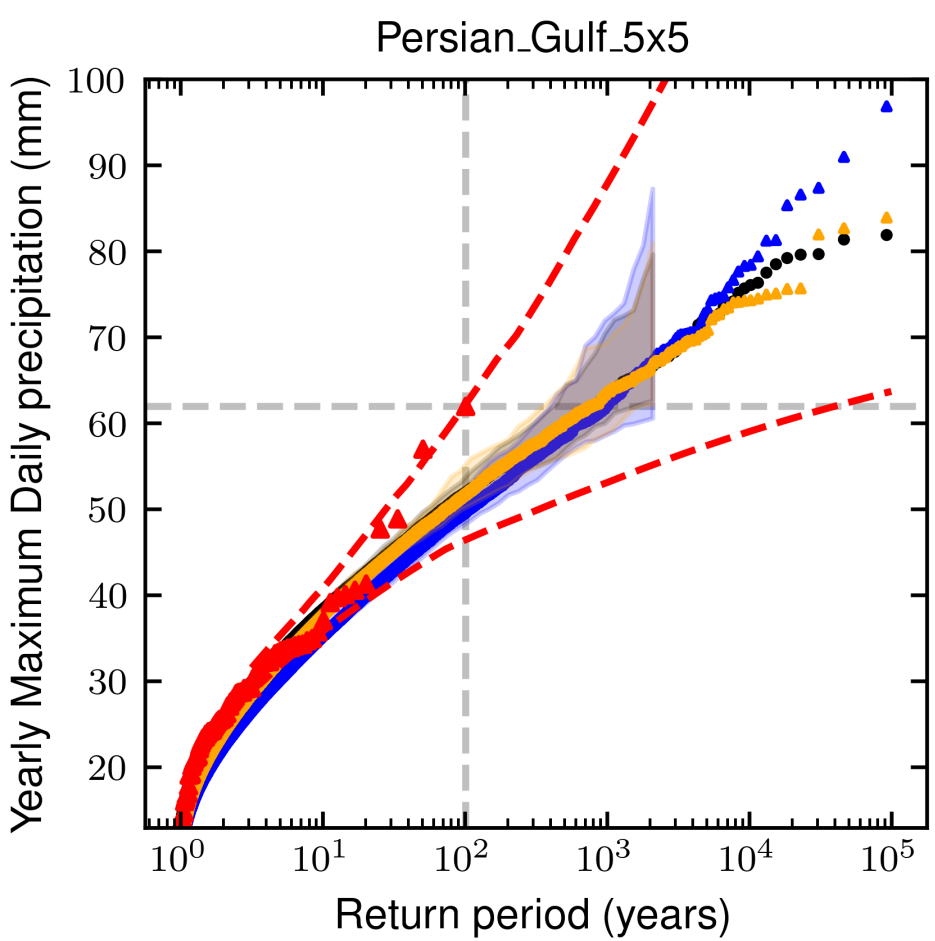
Emulators are long-term stable and have a correct mean and variability

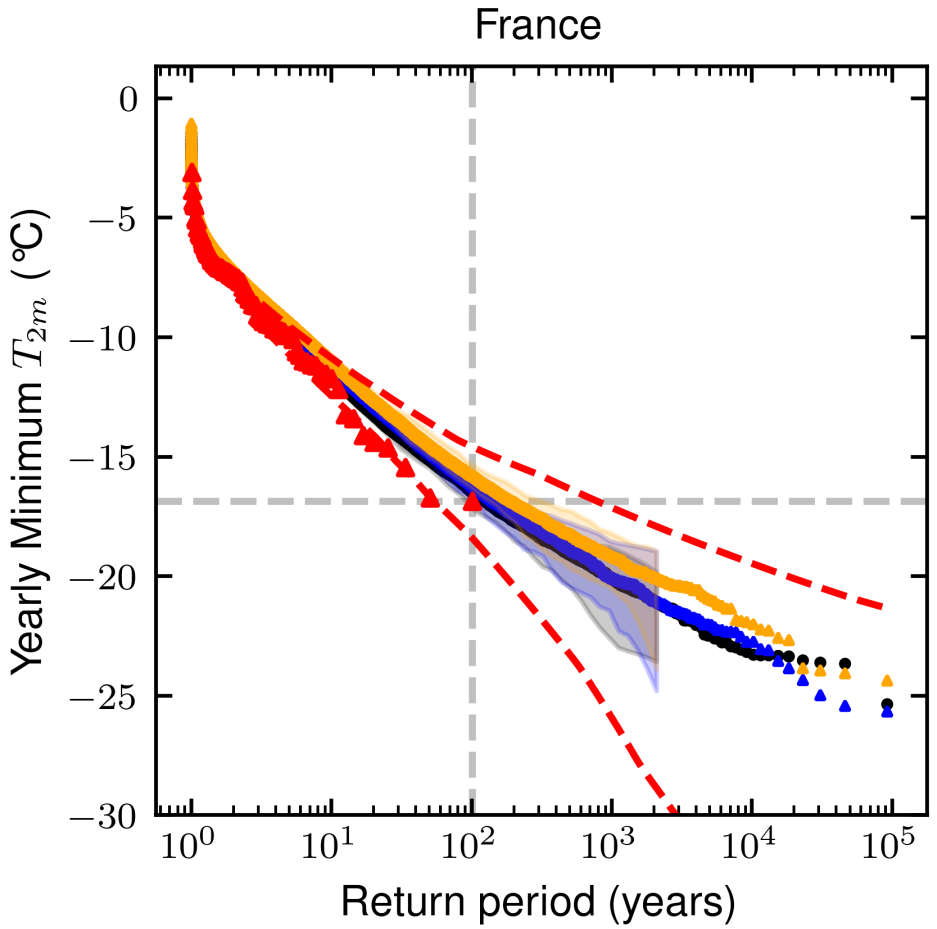
- Mean biases of a few percents



- Both emulators get the right variability modes

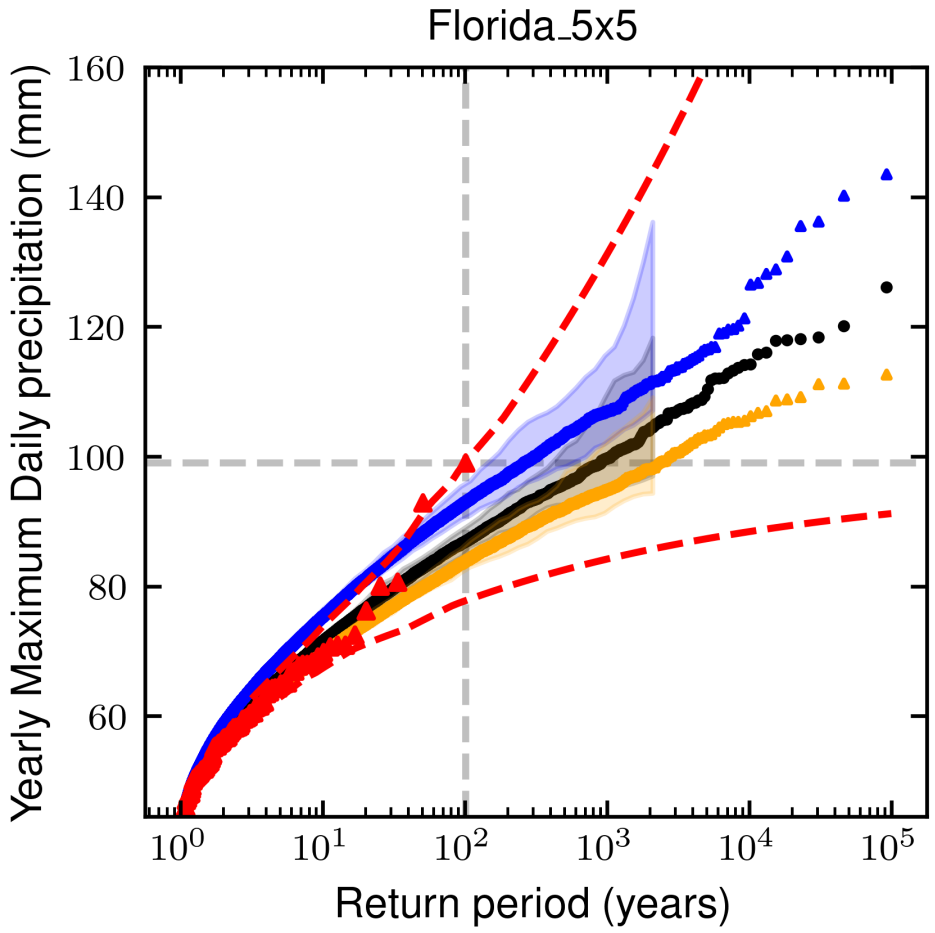
Emulators extrapolate well beyond training data…

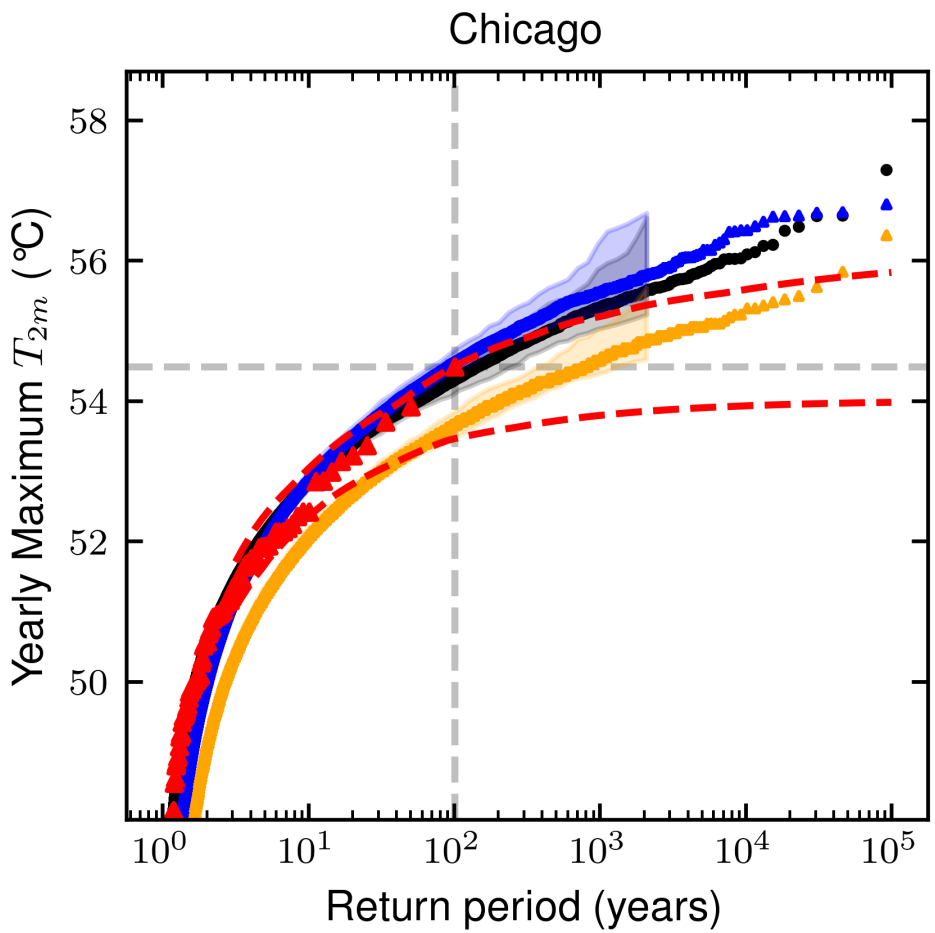




● Ground truth ● SFNO-PlaSim ● Pangu-PlaSim ▲ Training set GEV Fit

...but could have inaccurate return periods due to biases

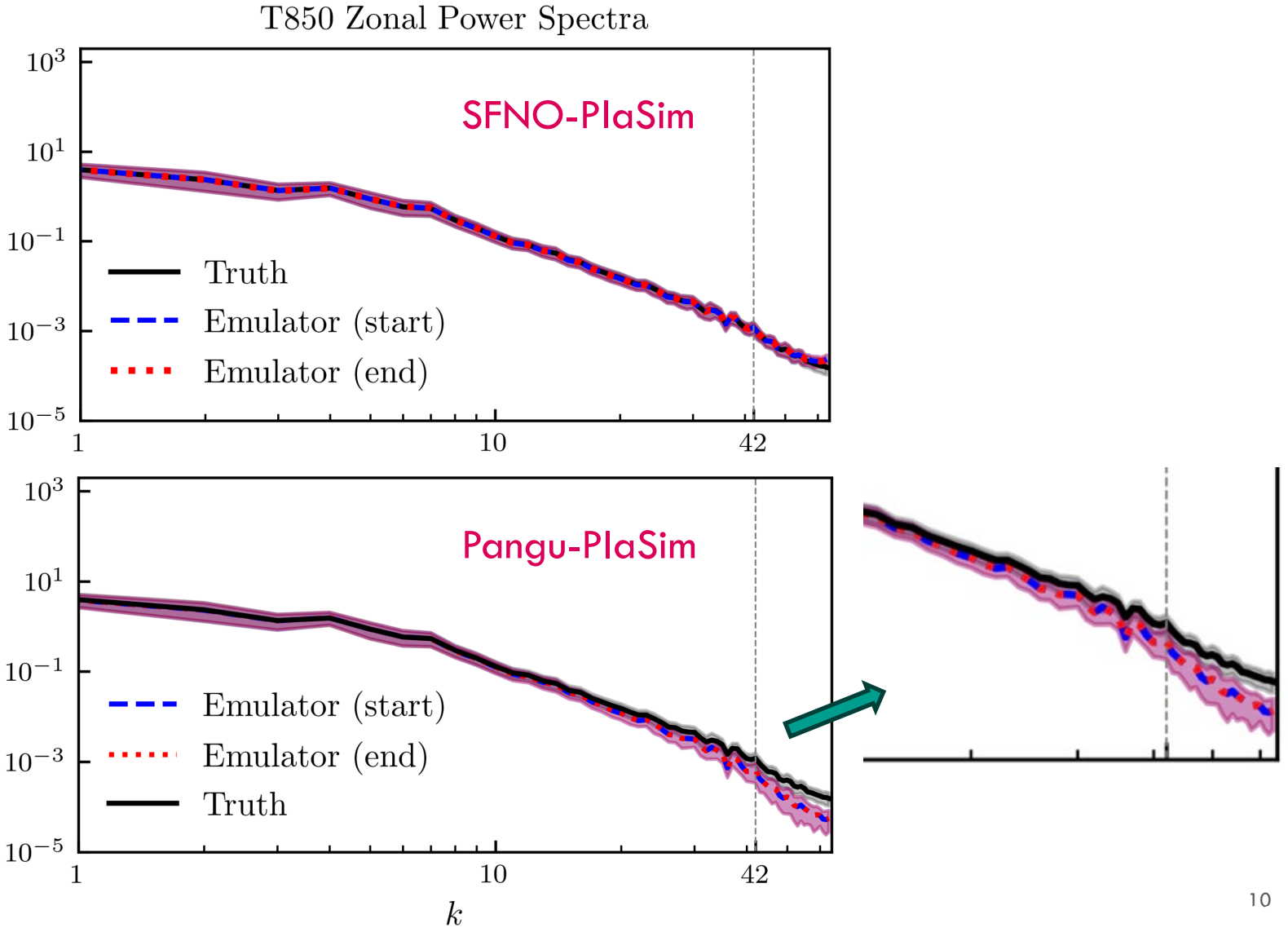




● Ground truth ● SFNO-PlaSim ● Pangu-PlaSim ▲ Training set GEV Fit

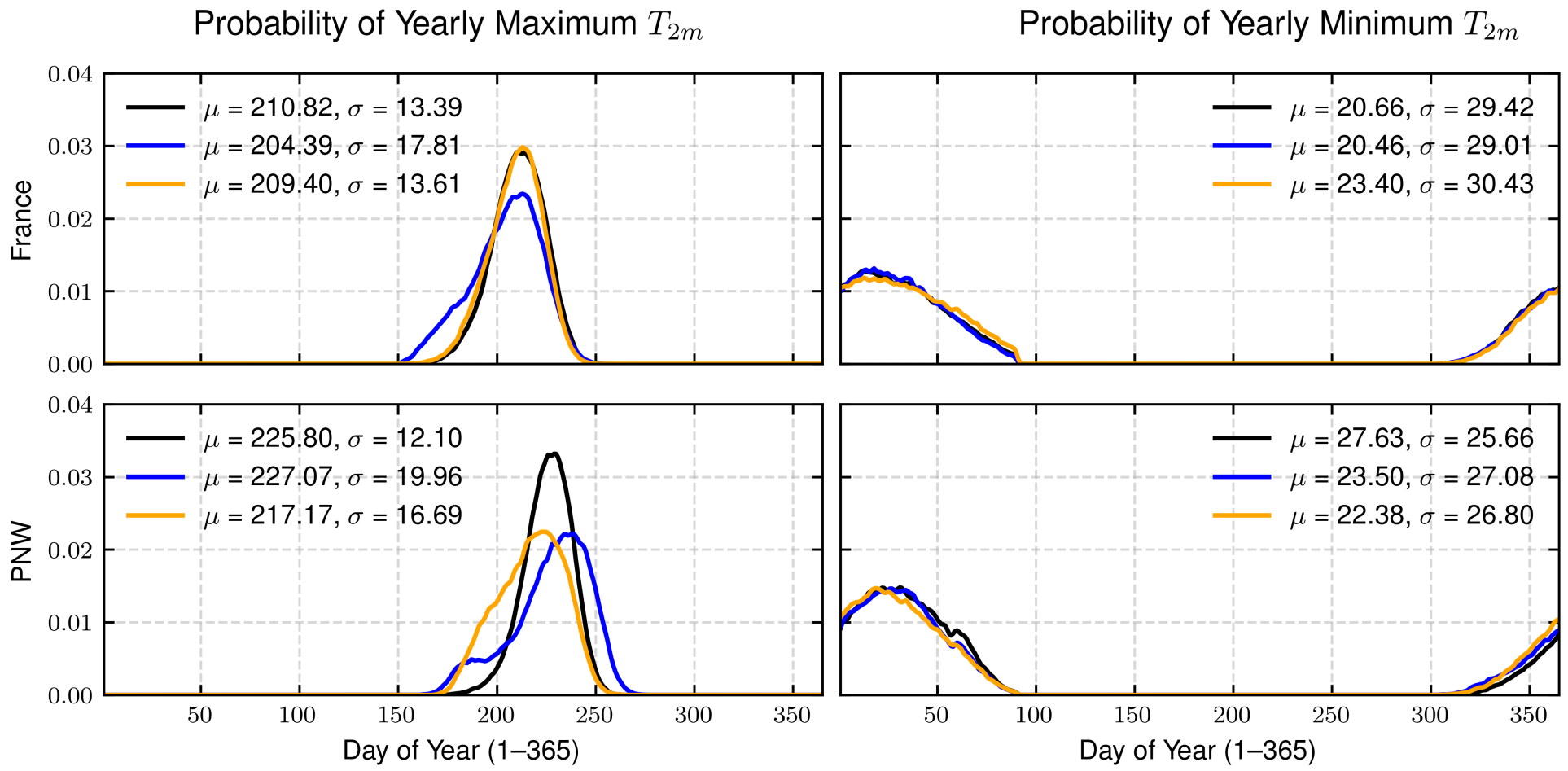
Return period biases have multiple sources, including spectral bias and missing land variables

Spectral biais explains the consistent underestimation of extremes by Pangu-PlaSim



Return period biases have multiple sources, including spectral bias and missing land variables

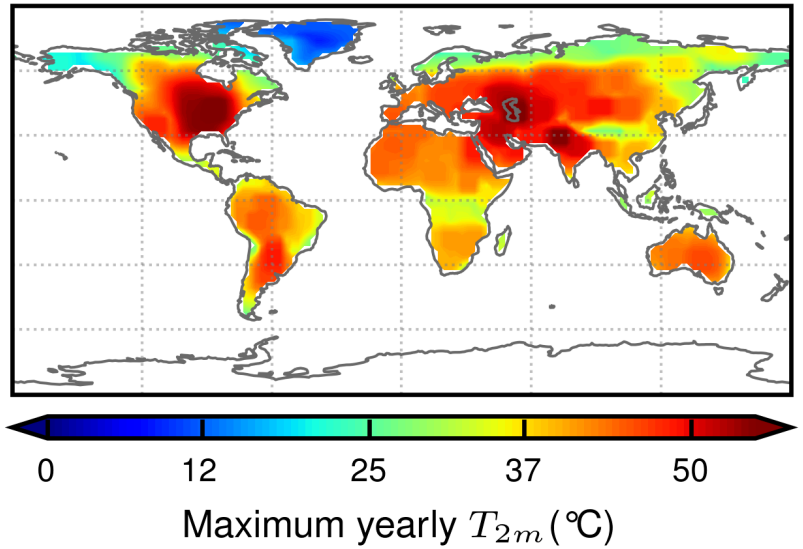
Missing soil moisture seems to create seasonal biaises for the two emulators



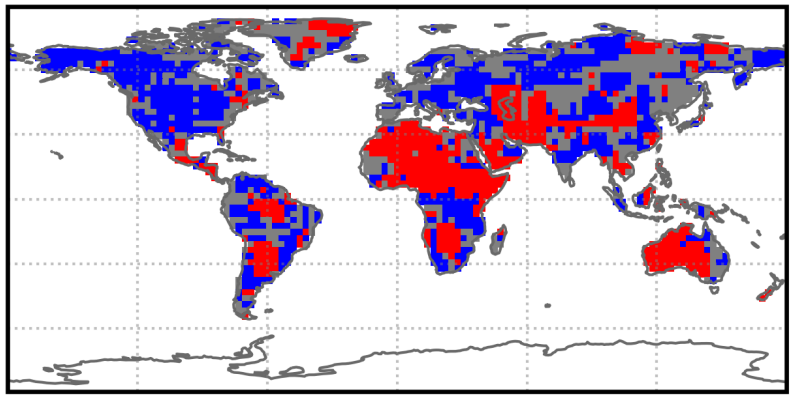
Systematic global analysis shows emulator can outperform EVT

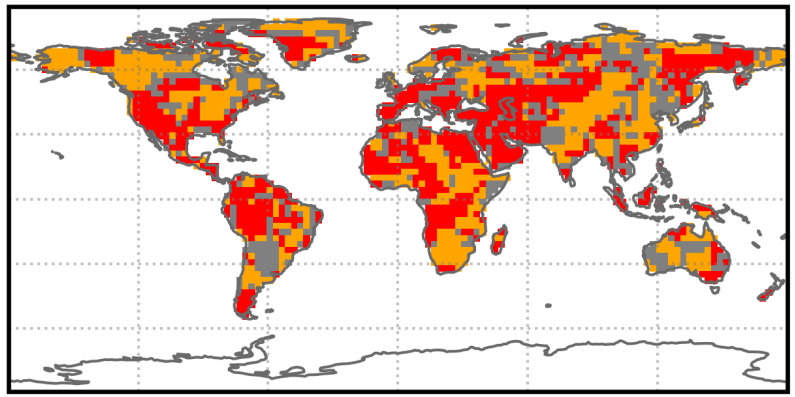
2000-year return period heatwaves events

Ground truth



Best mean return value



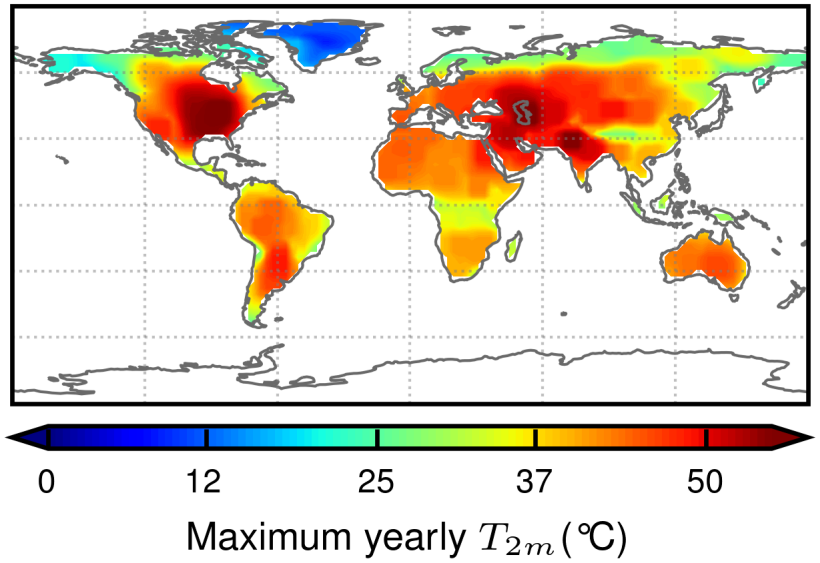


Simple debiasing techniques really helps!

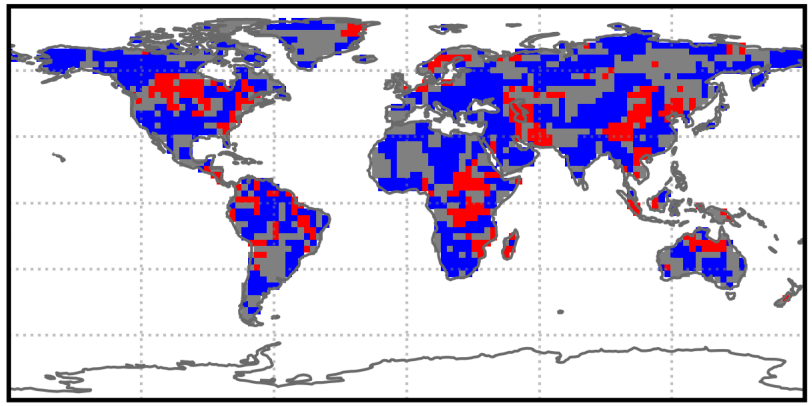
Debiasing with Mean and Standard Deviation:

$$A_{debiased} = \frac{\sigma_A^{train}}{\sigma_A^{emul}} \left(A - \mu_A^{emul} \right) + \mu_A^{train}$$

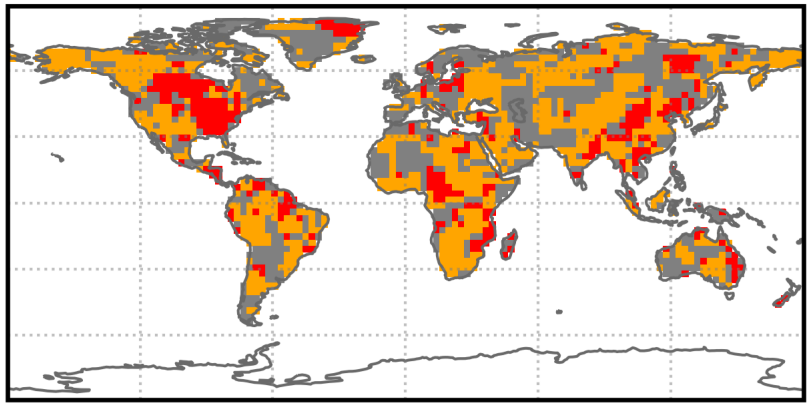
Ground truth



Best mean return value



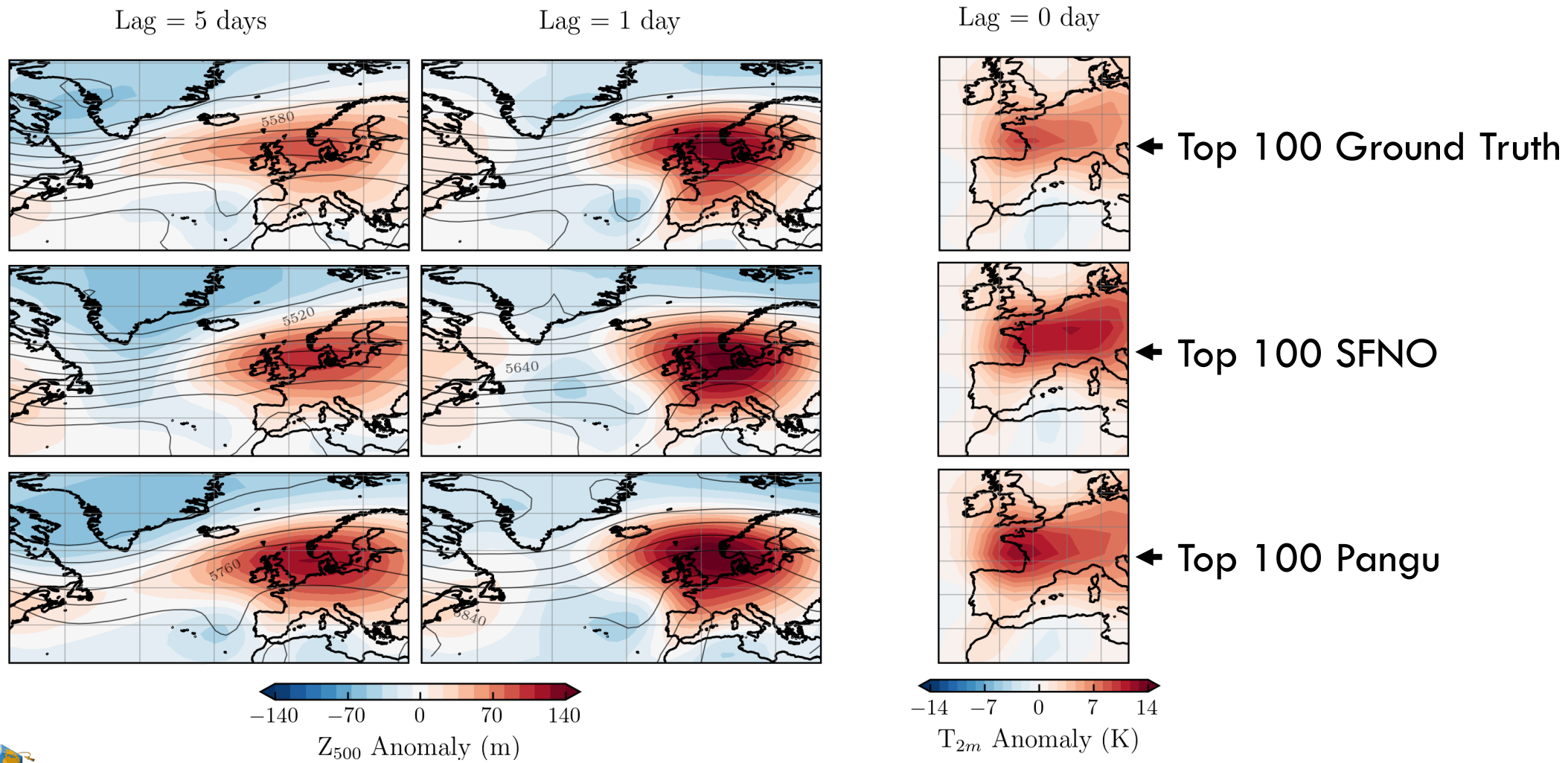
Training GEV fit (16%)
SFNO-PlaSim (49%)
None (35%)



Training GEV fit (17%)
Pangu-PlaSim (46%)
None (37%)

Emulators produce dynamically consistent extremes

Composite of Maximum Yearly T2m over France



Key takeaways

- AI climate emulators **trained on 100 years** of GCM simulation produce very rare events with return periods **much longer** than the training duration
- Emulated rare events seems **dynamically similar** to the ground truth
- Exact return periods of rare events are affected by **emulator biases**, in particular for events impacted by **missing variables**
- Even **simple debiasing** techniques make emulators **more accurate** than EVT

Future Steps

- Scaling to **higher resolution** and **coupled-model (atm + ocean + land)** to solve issue of missing variables
- Use (AI) **Rare Event Algorithms** to obtain additional extremes to retrain and validate them with **physically generated** data

AI Rare-event algorithm Pre-print

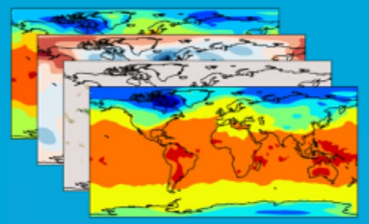
Thank you for your attention!

Back up slides

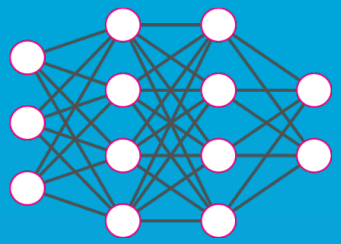
Presented at MLESM 2025

Coupling AI Emulators and Rare Event Algorithms to Sample Extreme Weather Events

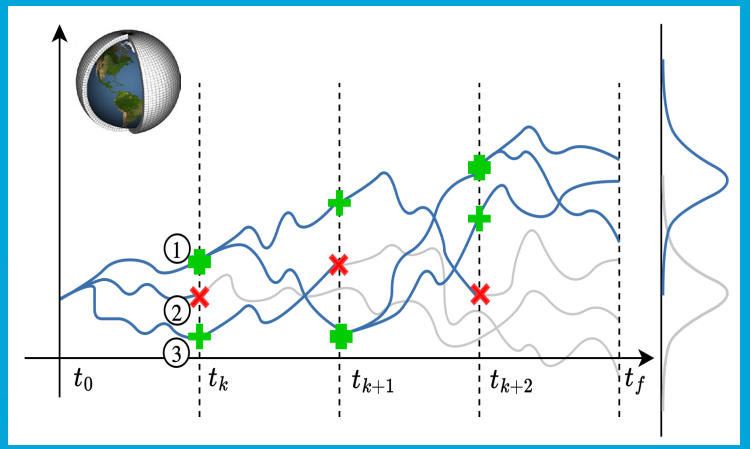
Data from GCM

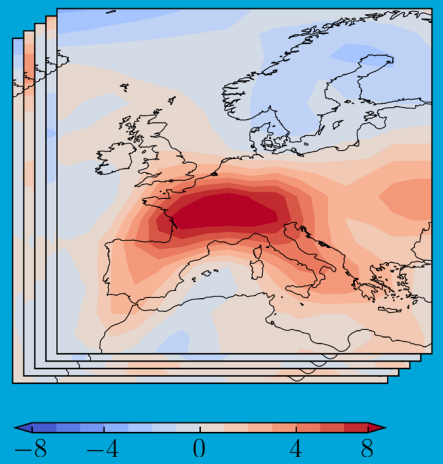


Train AI Climate emulator



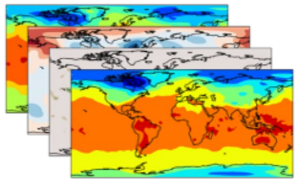
Use the **emulator** to drive **Rare Event Sampling** in **GCM**





Amaury Lancelin*, Alex Wikner* (UChicago), Pedram Hassanzadeh (UChicago), Dorian Abbot (UChicago), Jonathan Weare (Courant, NYU), Freddy Bouchet (LMD, CNRS), Laurent Dubus (RTE)

PlaSim GCM

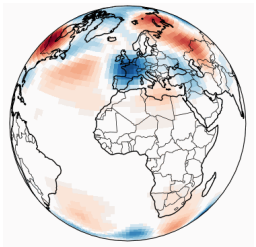


T42 with 10 levels,
Active Land,
Clim. SST + SIC

100-years of data

Train AI Climate Emulator

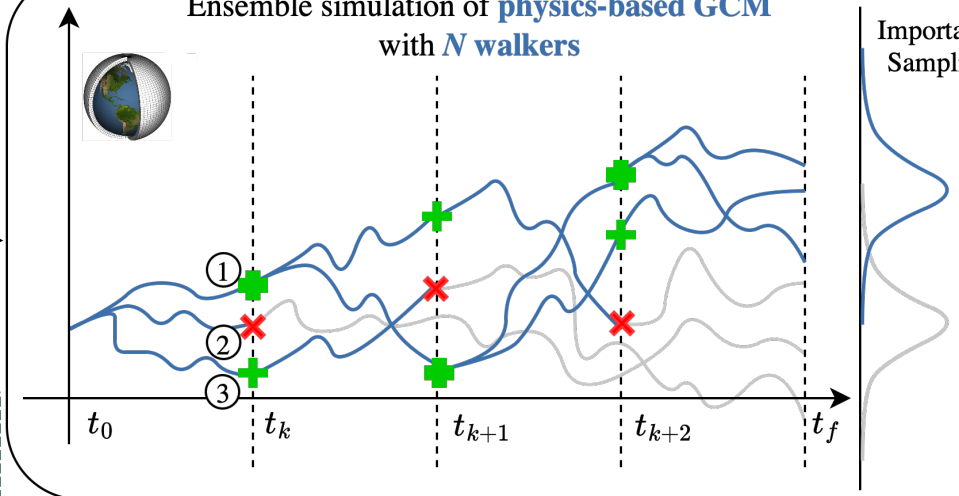
Vision-Transformer based



Long-term stable and with weather skill up to 15-days

At each resampling step t_k

Ensemble simulation of **physics-based GCM** with ***N* walkers**



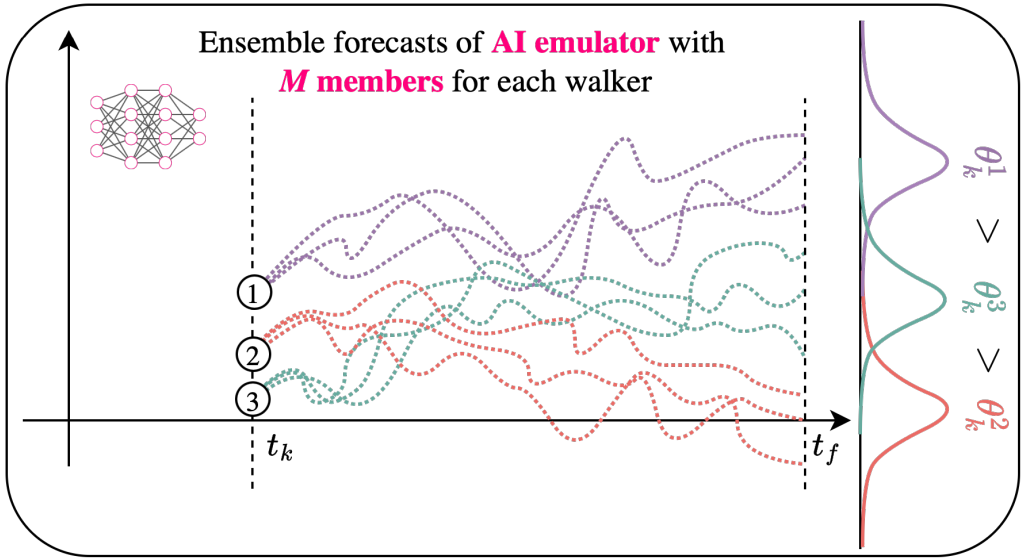
State vector $\{X^i_{t_k}\}_i$

Score function $\{\theta^i_k\}_i$

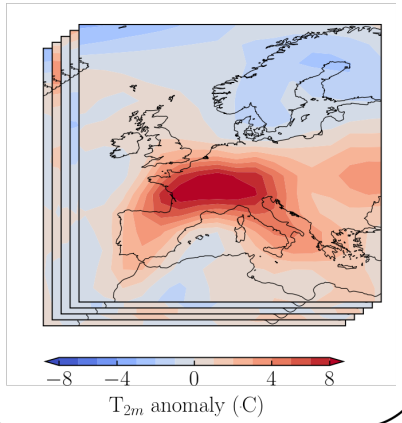
Each walker is duplicated based on the **weight** w^i_k

$$w^i_k \propto e^{C\theta^i_k}$$

Ensemble forecasts of **AI emulator** with ***M* members** for each walker



We generate a large catalog of extreme events



Validate with **50k-years** ground truth

We compute unbiased probabilities of extremely rare events