Huge Ensembles of Weather Extremes using the Fourier Forecasting Neural Network

NERSC

William D. Collins
and Ankur Mahesh
Berkeley Lab and UC Berkeley
NERSC
NVIDIA

”Must-have” use case for Huge Ensembles for Extremes

Proposal: Generate statistics on simulated LLHIs that could have occurred under historical conditions, as well as their drivers, by generating

HENS: Huge Ensembles of 10^N members,

where $N \geq 4$ required to converge statistics

The ensemble will consist of short (2-week long) hindcasts.

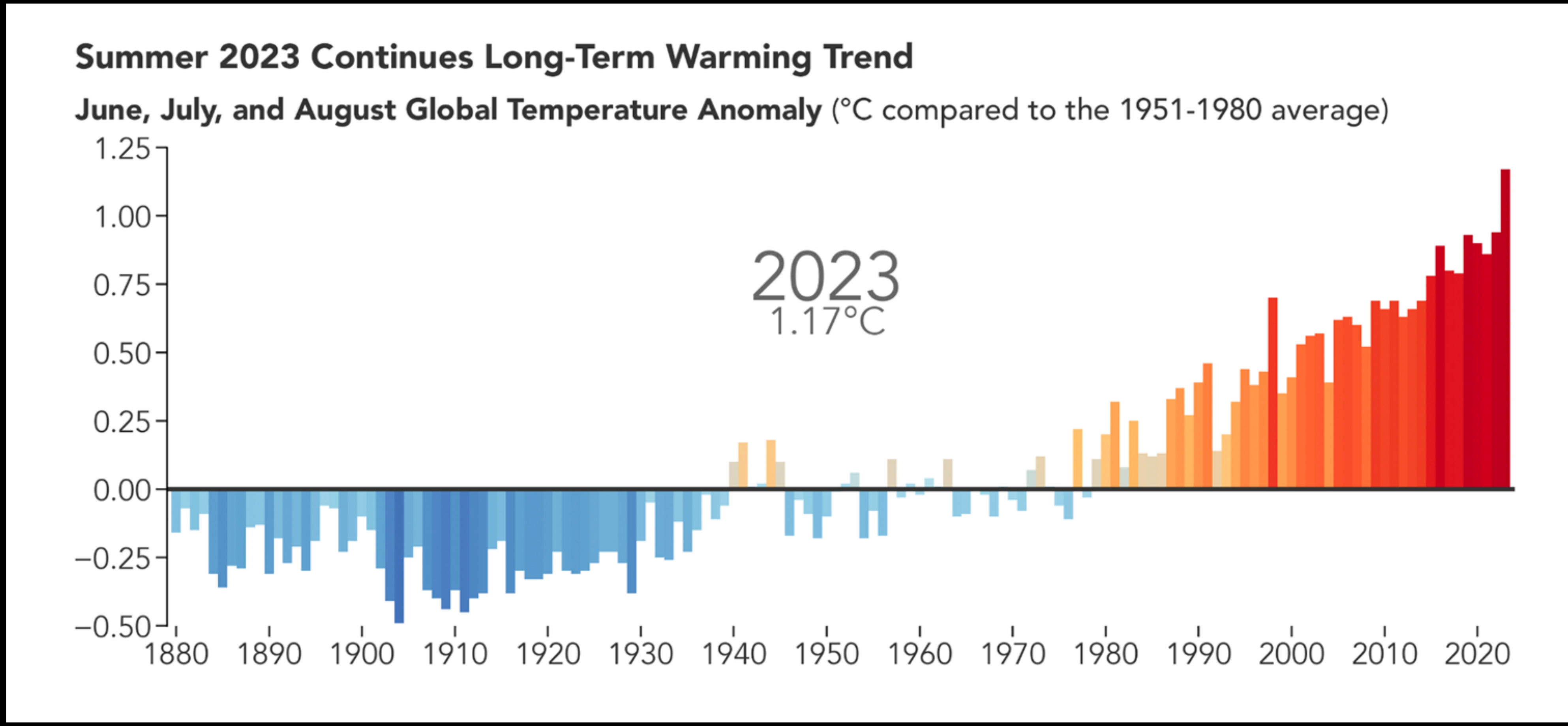
Hypothesis: Ergodicity of climate system means we can ”trade” increasing ensemble size with increasing length of sampling time.

Huge Ensembles (HENS) for Summer 2023

Constructing HENS: We construct ensembles with FourCastNet using the same ensembling techniques as operational weather centers

Validating HENS: We validate these ensembles on extremes using the same techniques as NWP

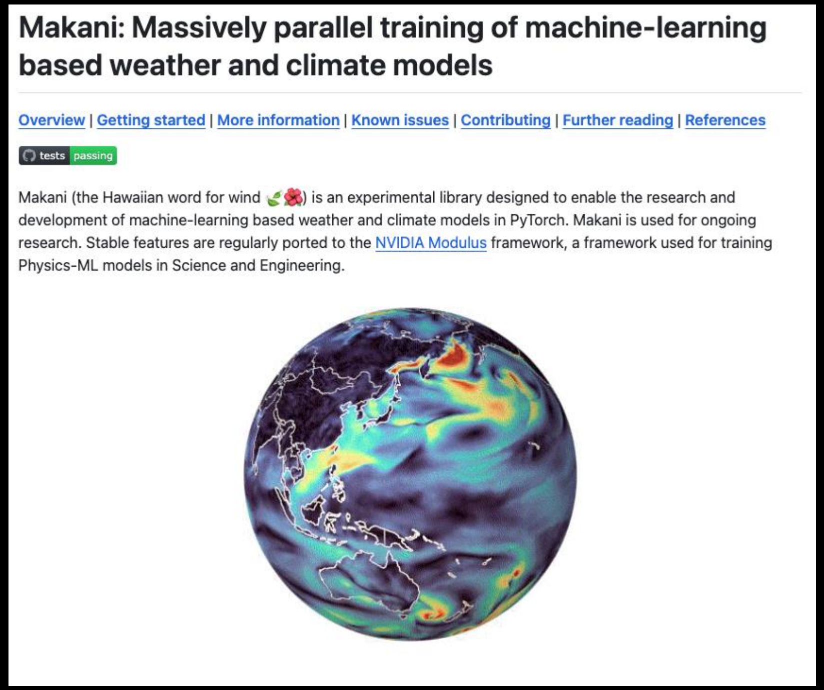
LLHIs in HENS: Summer 2023 was the hottest summer on record. We will study and quantify near-miss LLHIs in ultra-large counterfactuals of summer 2023.



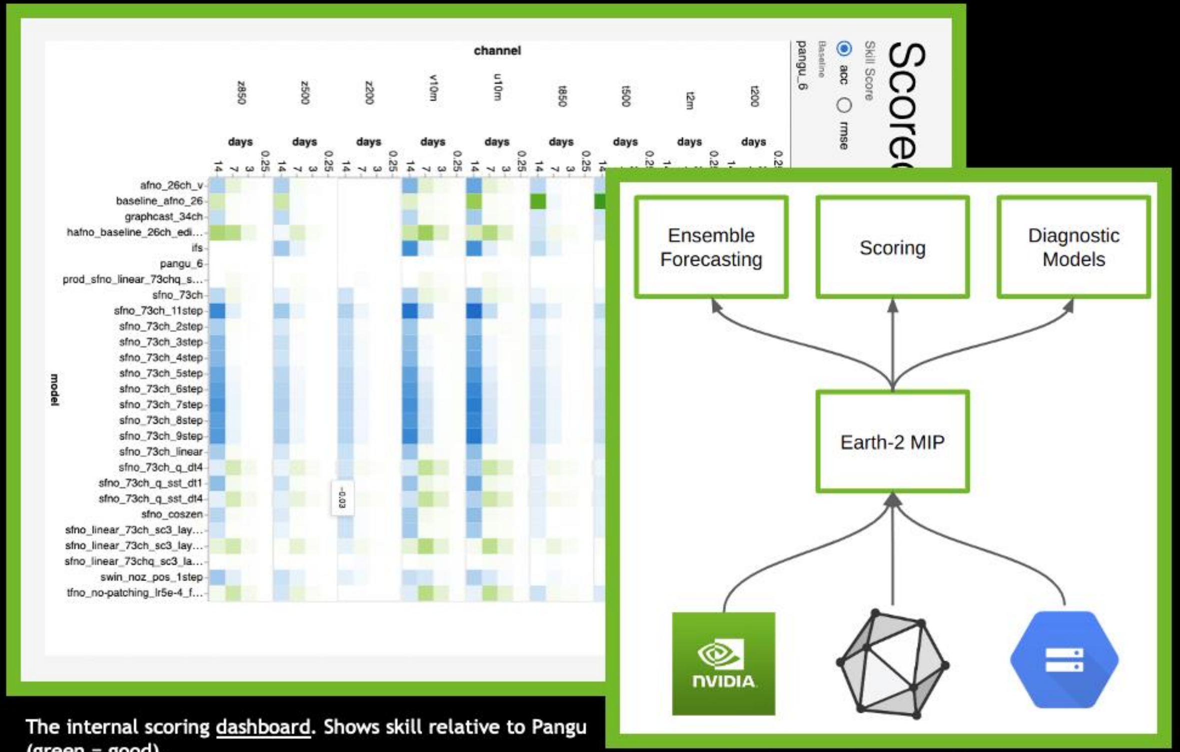
Source: NASA Earth Observatory

FourCastNet: an Open-Source AI-Driven Digital Twin

Modulus-Makani & Earth2-MIP Repositories



https://github.com/NVIDIA/modulus-makani



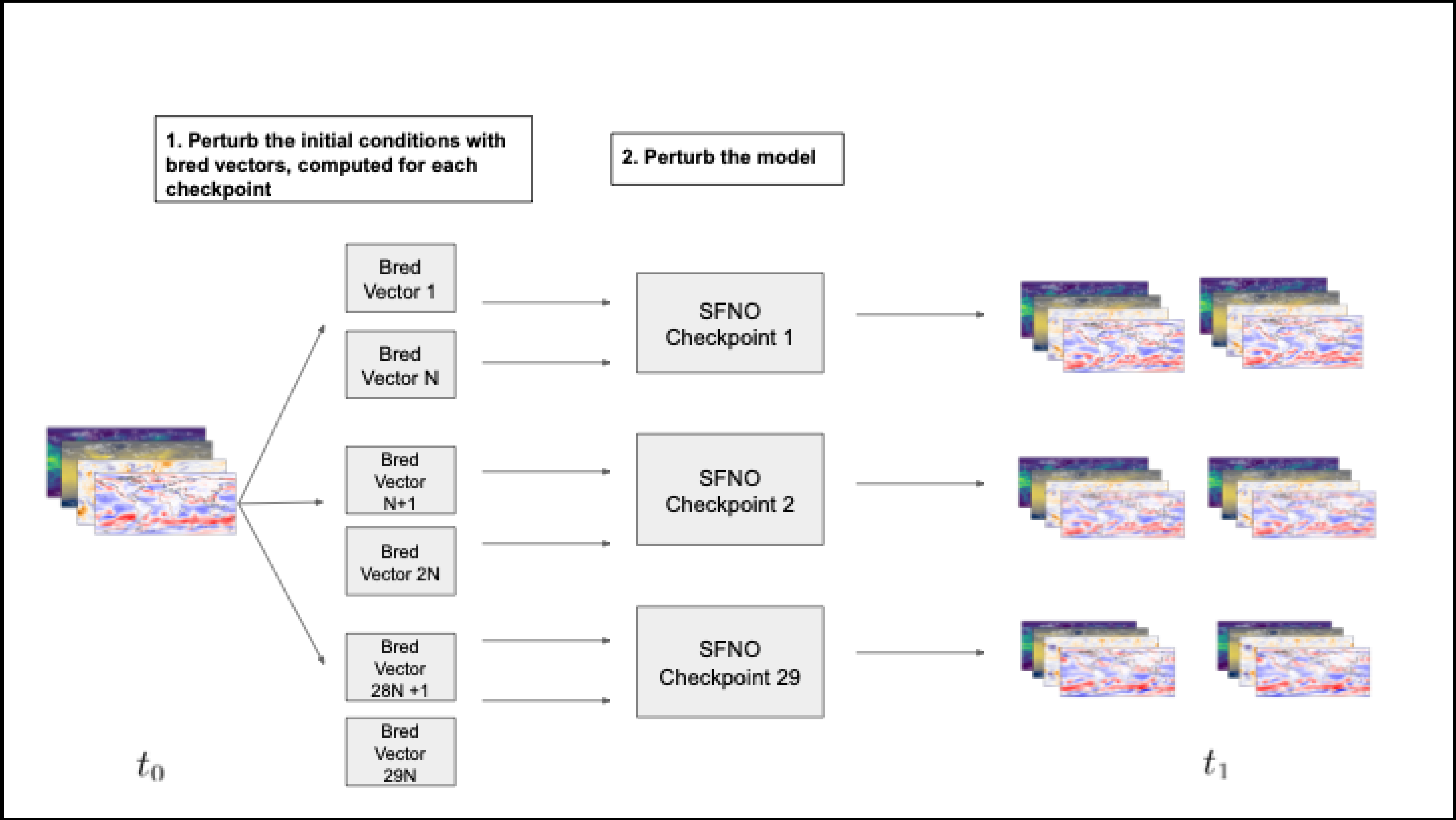
The internal scoring dashboard. Shows skill relative to Pangu (green = good).
Public version features a reduced set of models.

Multiple models, multiple sources

https://github.com/NVIDIA/earth2mip

- Scope — Global
- Model Type — Full-Atmosphere AI Surrogate
- Architecture — Fourier Neural Operator
- Resolution — 25km, 6-hourly
- State Variables — Temperature, wind, pressure, humidity
- Training Data — ERA5 Reanalysis
- Inference Time — 3 sec (2-week forecast)
- Calibration — IC + Bayesian model uncertainty
- Speedup /NWP — O(10,000 – 100,000)
- Power Savings — O(10,000)
- Max Rollout — Years
- Project Type — Open-source

Constructing Huge Ensembles with FourCastNet

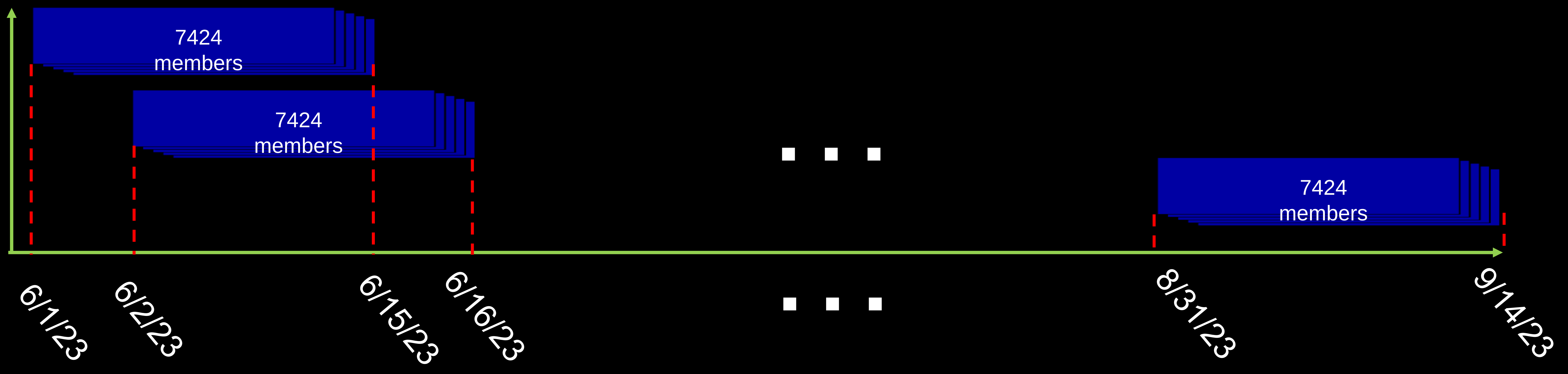


Parameters and Contents of HENS

The parameters of HENS are:

Number of models	Number of initial conditions
29	256

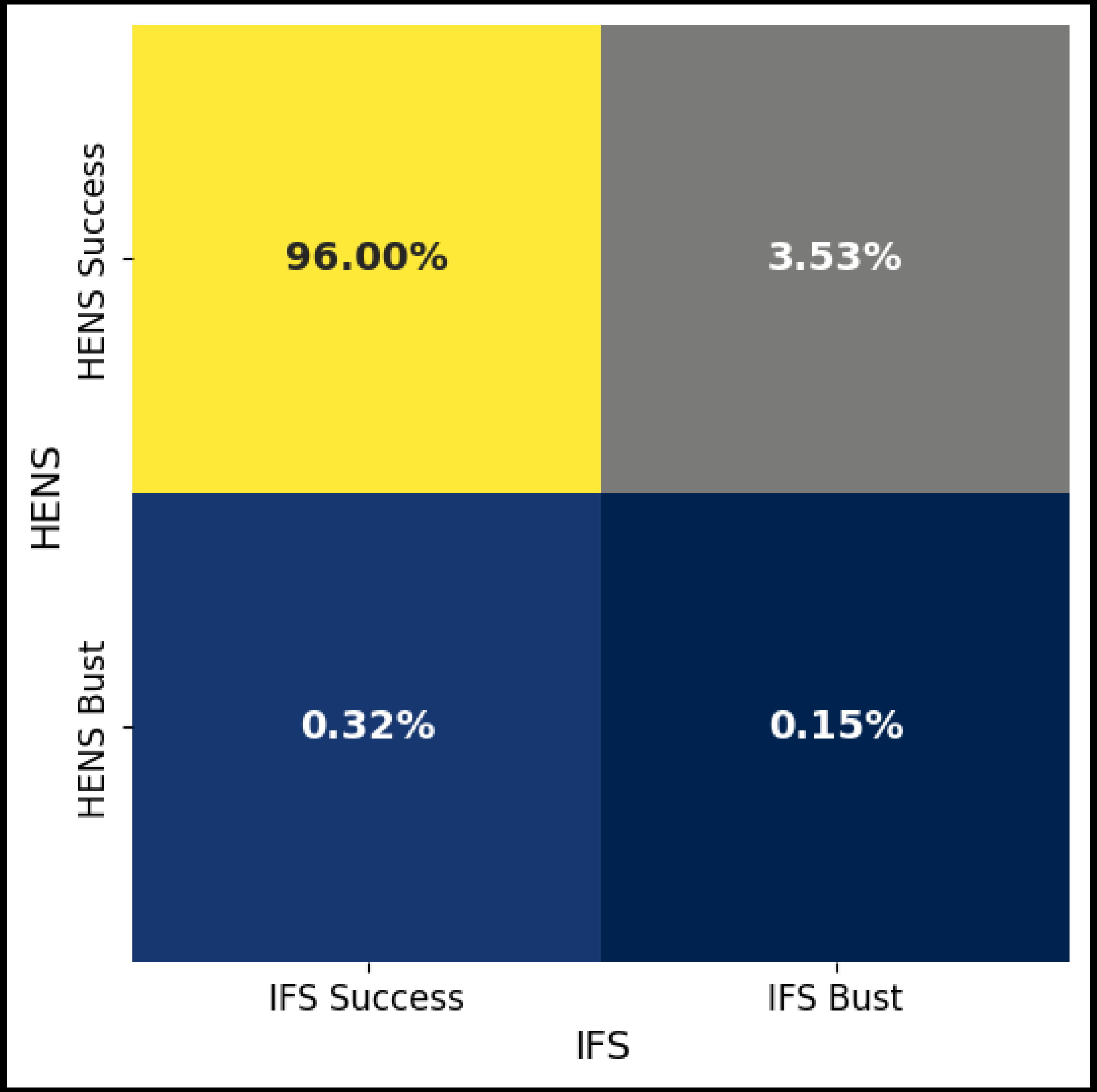
HENS consists of 29*256=7424 15-day hindcasts started from each summer day:

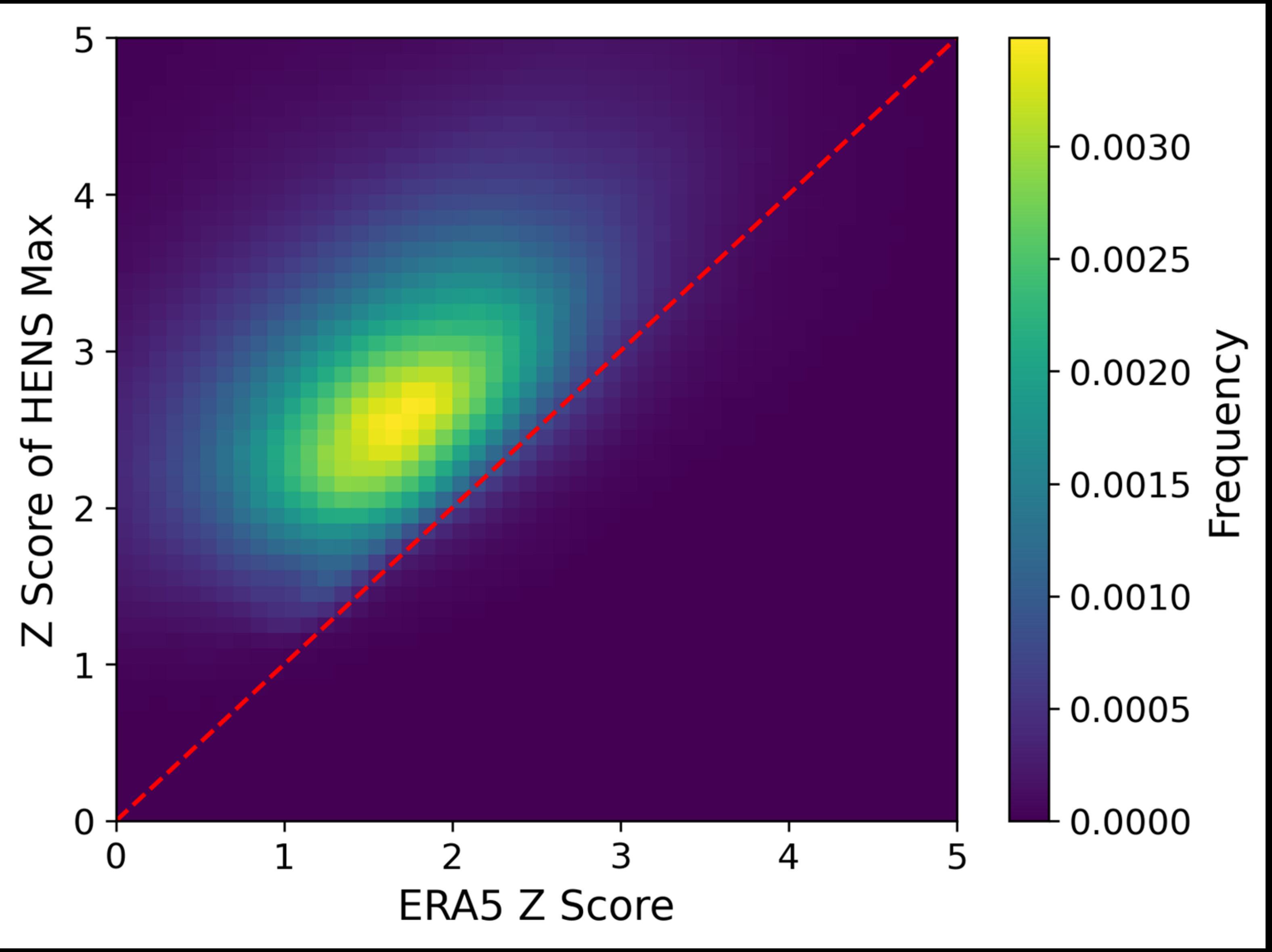


HENS captures extremes outside range of IFS ensemble

For the ensembles initialized throughout summer, at a 10 day lead time (240-258 hours), **the HENS ensemble range includes extremes missed by the IFS ensemble range.**

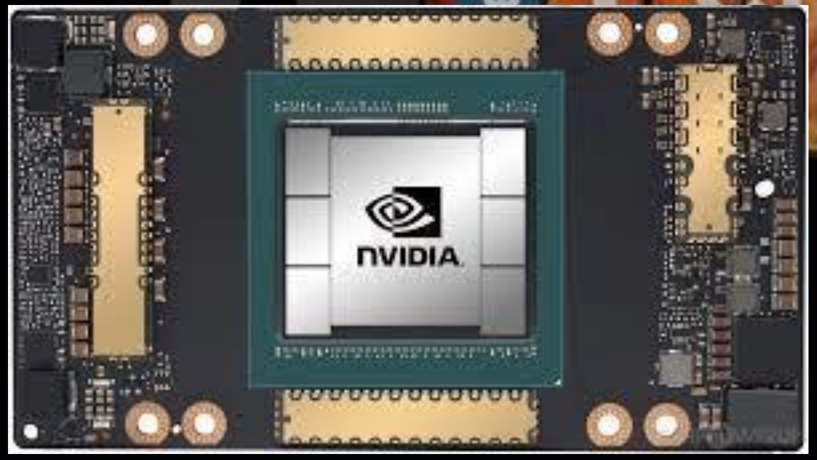
- For these IFS misses, the maximum HENS member is greater than ERA5.
- Most of the distribution is above the 1:1 line. (right)





What are the current and unique strengths and foundational capabilities of DOE for this topic?

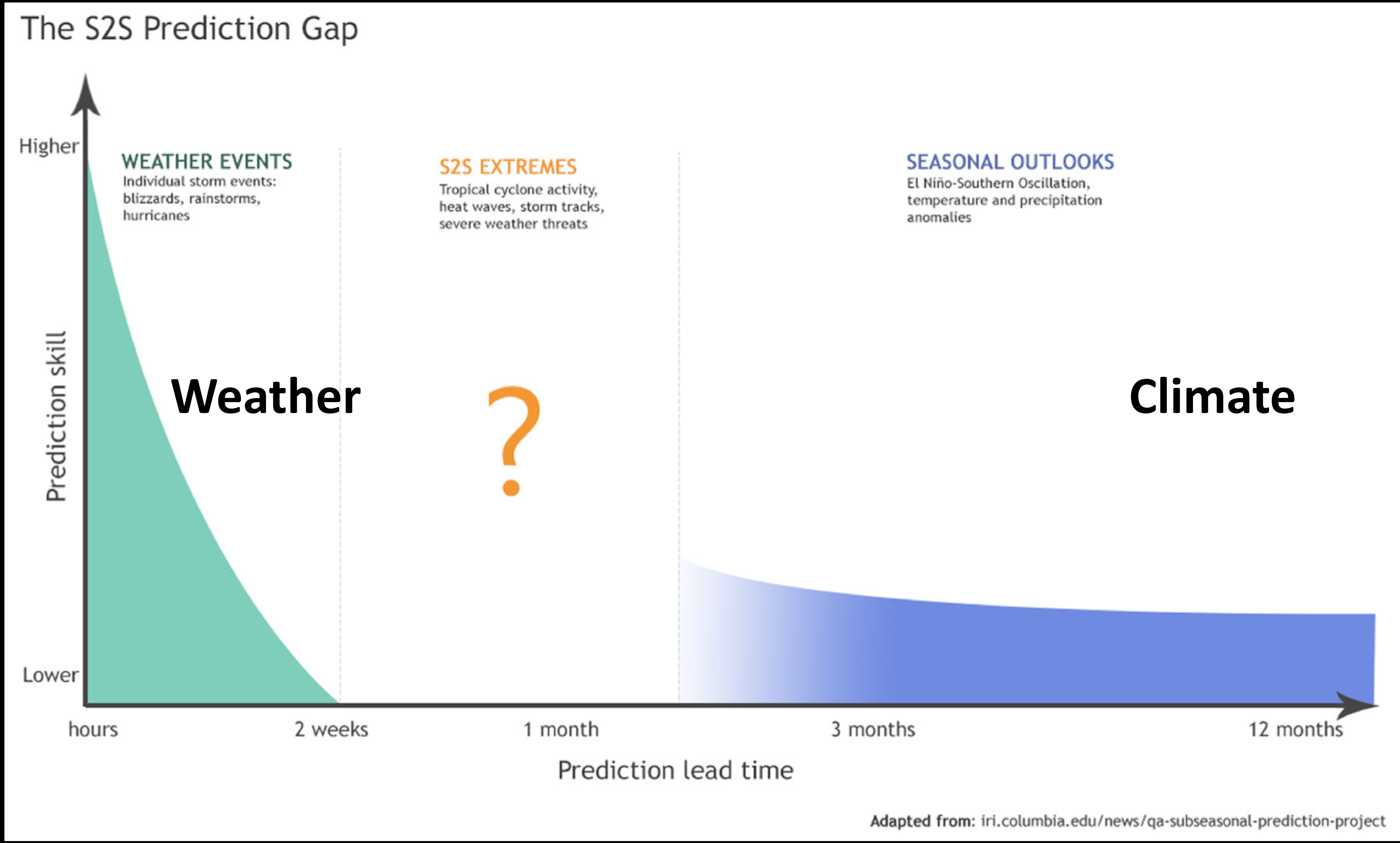
Enabling Technology: DOE’s GPU Exascale Systems



2022: ~1.7 exaFLOPS/s (OLCF)

What are the grand challenges in advancing the research on this topic?

Filling the gap between weather (short-term) and climate (long-term).



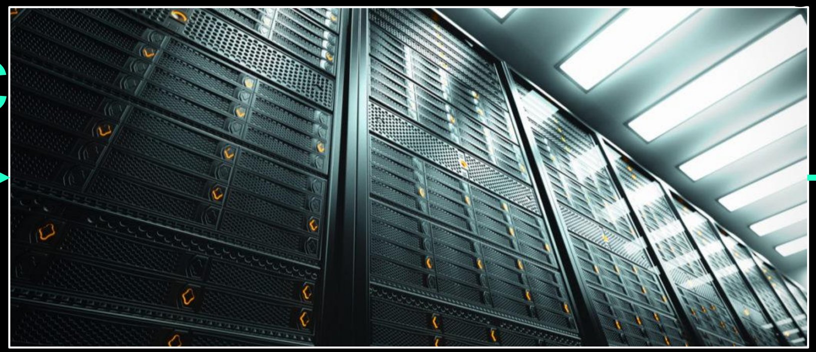
What are the gaps in research / infrastructure / coordination that prevent advances?

GPU-enabled emulator speed stresses storage rates

Perlmutter Supercomputer

8 GB/sec →

Perlmutter All-Flash Scratch



10 GB/sec →

Project Disk

Parameter	Value
# of nodes / GPUs	64 / 256
Run time for 1 start day	~45 minutes
Run time for 1 season	~69 hours ≅ 3 days
GPU-hours	69 * 256 = 17664 hrs = 2 yrs

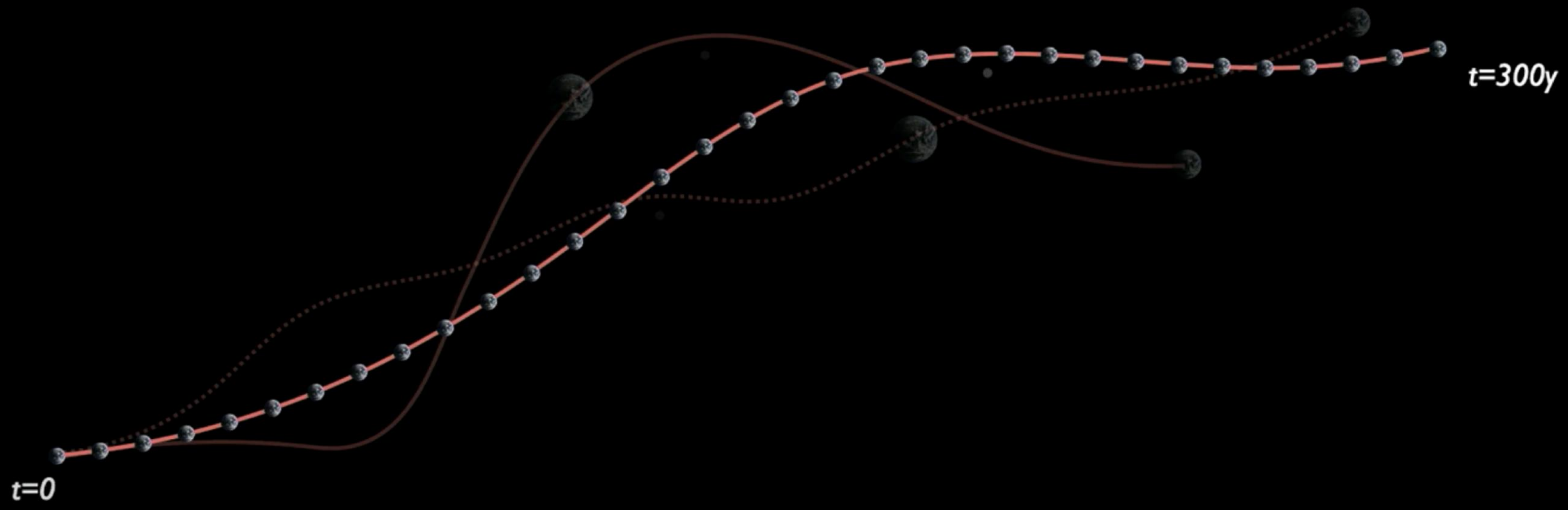
Model Field	Meaning
t2m	2-meter surface air temperature
t2d	2-meter surface dewpoint temperature
tcwv	Total column water vapor
t850	Air temperature at 850 hPa
q850	Specific humidity at 850 hPa
z500	Geopotential height at 850 hPa
msl	Mean sea level pressure

What opportunities exist to overcome each of those gaps?

Switch to exchanging regeneration methods, not data?

AI nimbly generates details between "checkpoints" saved only infrequently from physics-based climate simulations

-- Bjorn Stevens, GTC 2021



What role could other agencies play in facilitating our science?

Out of 15 agencies in USGCRP, NSF is a prime target.

U.S. Global Change Research Program

Agencies

Department of Agriculture	Department of Commerce	Department of Defense
Department of Energy	Department of Health and Human Services	Department of Homeland Security
Department of Housing and Urban Development	Department of the Interior	Department of State

Department of Transportation	Environmental Protection Agency	National Aeronautics & Space Administration
National Science Foundation	Smithsonian Institution	U.S. Agency for International Development

NCAR
LEAP STC

What are reasonable 2–5 year, 5–10 year, and long-term goals for addressing these grand challenges?

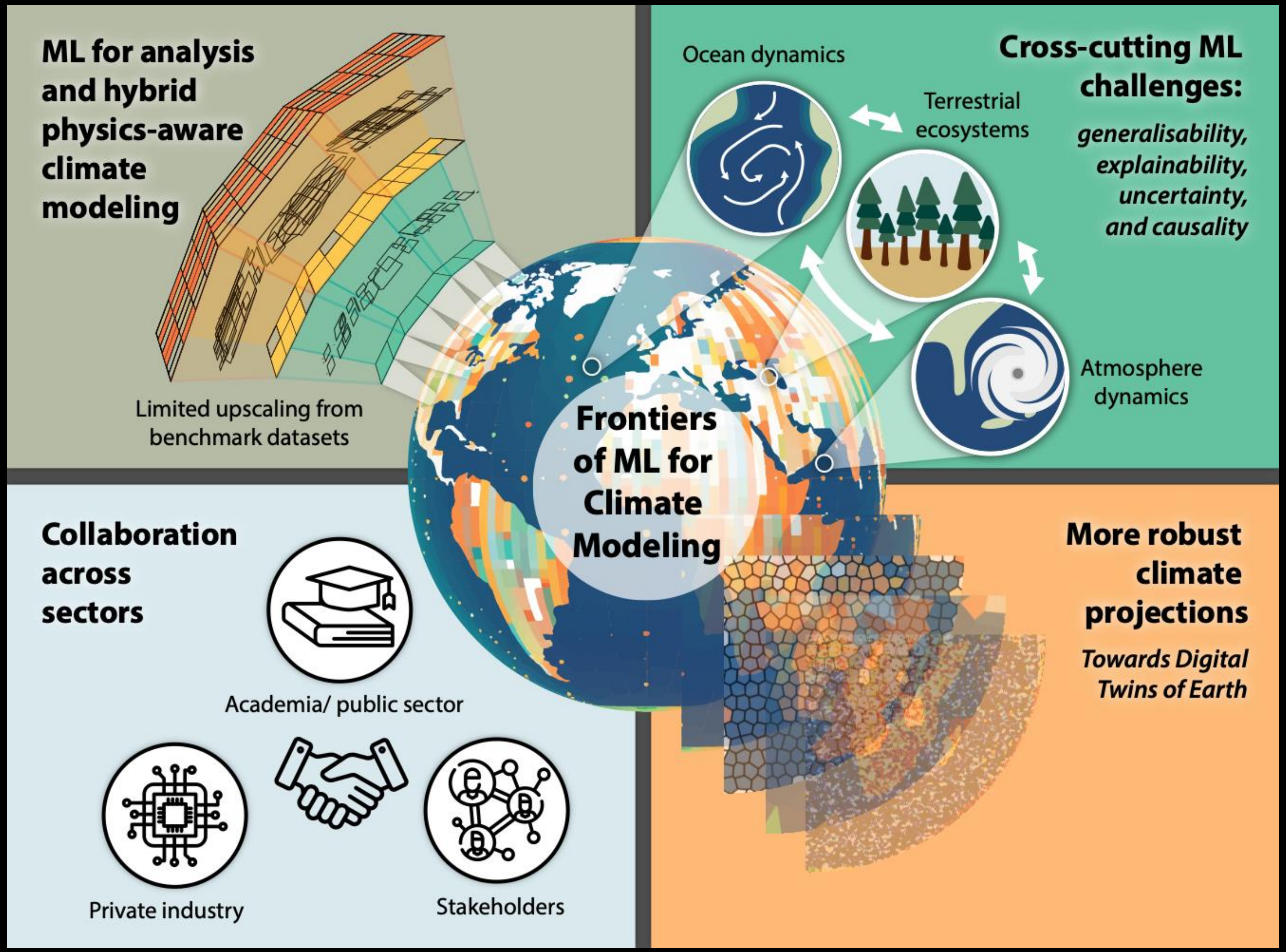
Some suggestions for goals

2-5 year goals:

1. 1st hybrid ESM simulations → CMIP7
2. 1st diagnostics for emulator physics
3. 1st formal intercomparisons of emulators & hybrid ESMs

5-10 year goals:

1. Establishment of parallel PCMDI for ML/AI?
2. Extension of data-driven emulators to IAMs
3. Revisiting the Charney report in ML/AI era



Conferences led by CASCADE4

Machine Learning for Actionable Climate Science

Exploiting Machine Learning to Enhance Earth System Modeling and Analysis Across Scales

Gordon Research Conference

June 22 - 27, 2025 • Bryant University

Exploiting Machine Learning to Enhance Earth System Modeling and Analysis Across Scales

This GRC conference will explore how to best push the frontiers of ML beyond state-of-the-art approaches, especially in

1. developing hybrid Earth system models with greater fidelity,
2. providing capabilities for climate extremes through large ensembles with emulators as well as enhancing detection and attribution methods, and
3. advancing climate model analysis and benchmarking.

This interdisciplinary conference brings together ML and climate scientists, as well as the private sector, to accelerate progress towards actionable climate science.

Organizers

Veronika Eyring
Chair

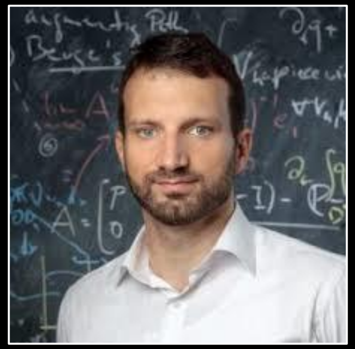
William Collins
Co-Chair

Pierre Gentine
Vice Chair

Laure Zanna
Co-Vice Chair

Chris Lofholm
Conference Operations Associate

Thanks to all collaborators!



Boris Bonev

Noah Brenowitz

Yair Cohen

Peter Harrington

Karthik Kashinath

Thorsten Kurth

Ankur Mahesh

Josh North

Travis O’Brien

David Pruitt

Mark Risser

Shashank Subramanian

Jared Willard

Mike Pritchard

Acknowledgements:

This research was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and by the Regional and Global Model Analysis Program area within the Earth and Environmental Systems Modeling Program.

Acknowledgements:

This research was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and by the Regional and Global Model Analysis Program area within the Earth and Environmental Systems Modeling Program.

The research used resources of the National Energy Research Scientific Computing Center (NERSC), also supported by the Office of Science of the U.S. Department of Energy, under Contract No. DE-AC02-05CH11231.

Questions?

HENS Part I

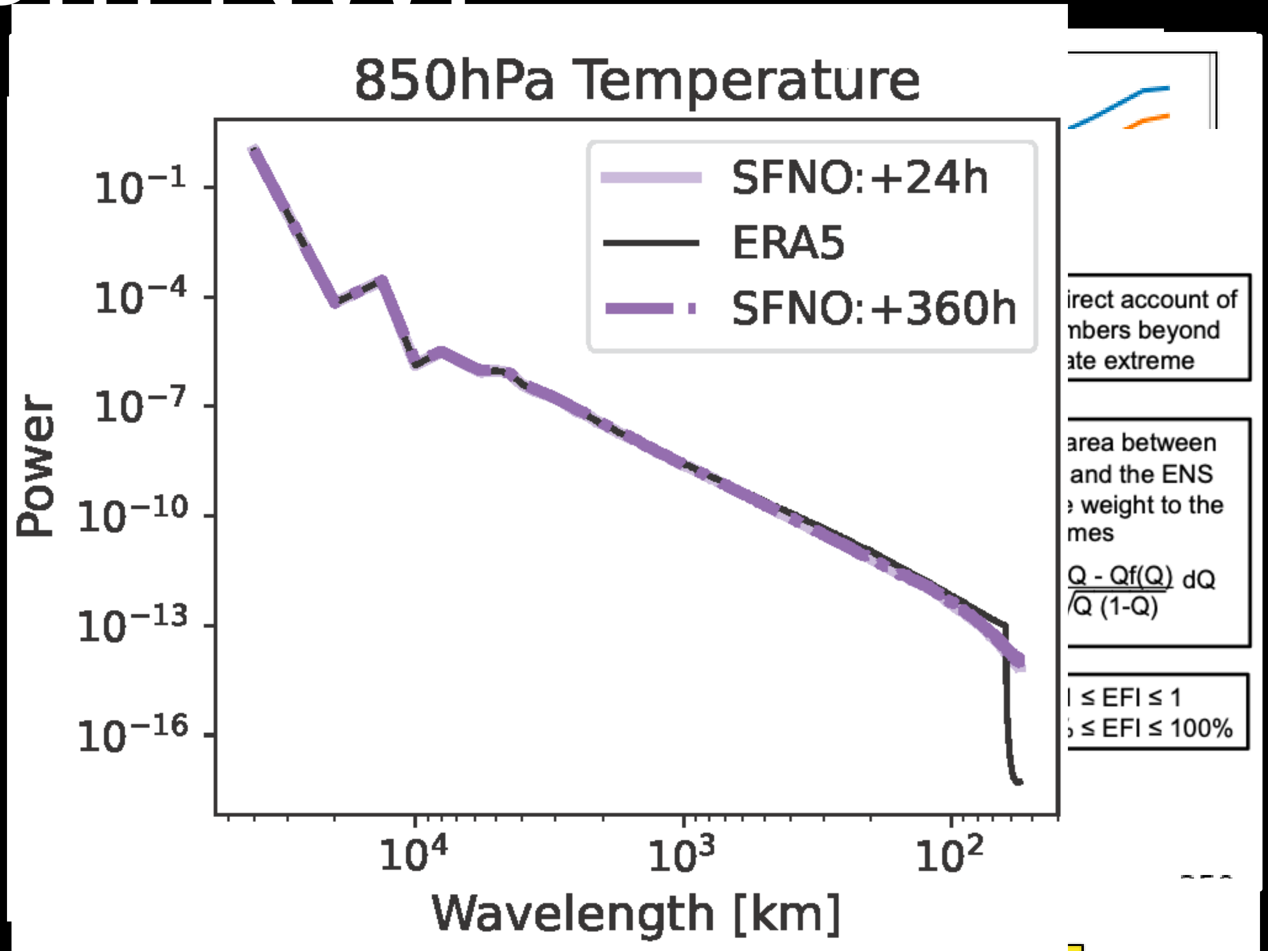
HENS Part II

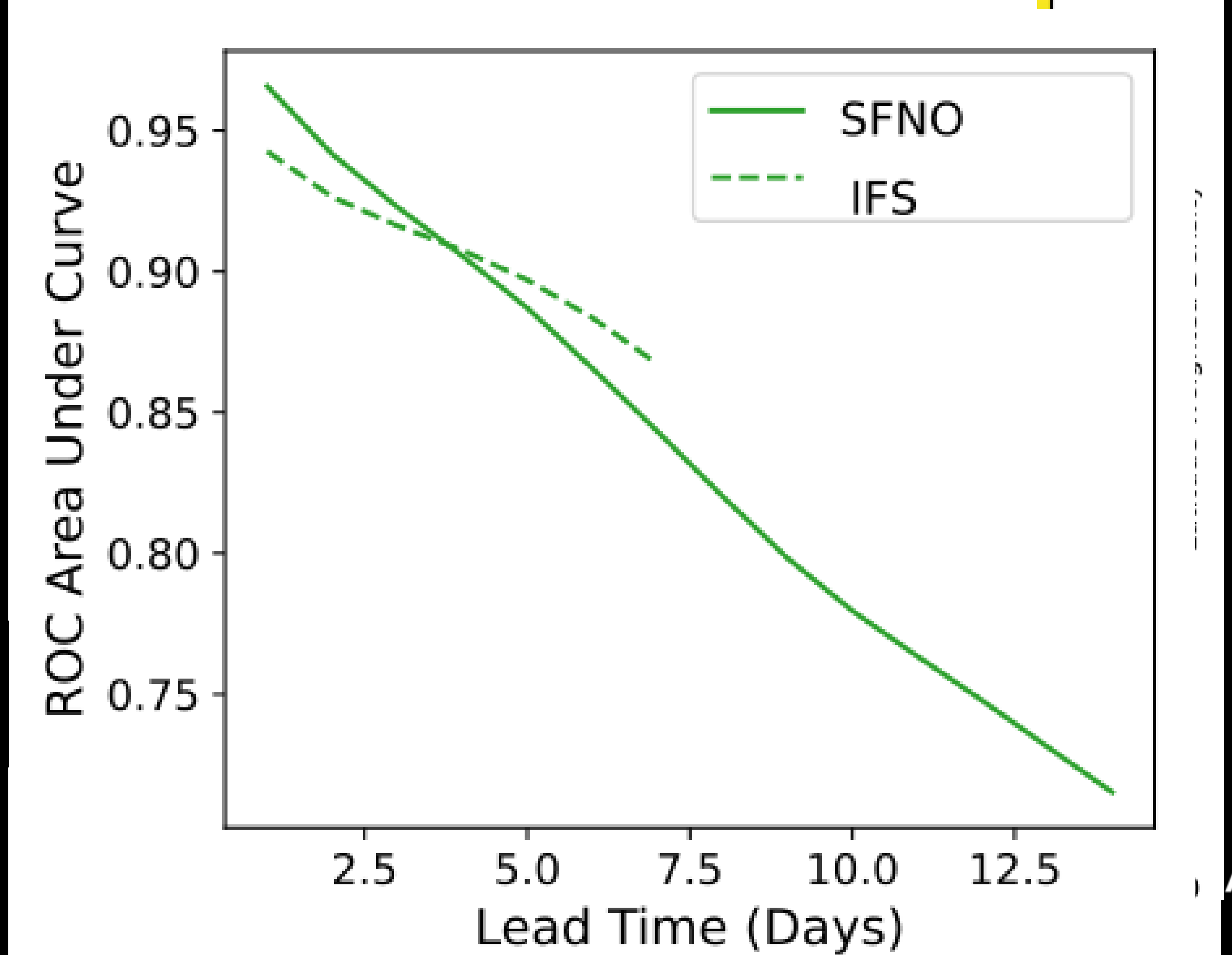
BACKUP SLIDES

Validation of HENS using Metrics from NWP

We validate FourCastNet using the same metrics used by operational weather centers.

1. **Can the forecast distinguish between extremes and non-extremes?**
2. **Do forecast probabilities of extremes match their observed occurrence?**
3. **Does the ensemble spread match its skill?**
4. **Does the distance between ensemble and climatology match numerical models?**
5. **Are the power spectra realistic?**





Model Variants are Statistically Indistinguishable

- Denote the HENS output by

$$\{X_{ij} : i = 1, \ldots, 29;\ j = 1, \ldots, 256\}, \text{where}$$
$$i = 1, \ldots 29 \text{ indexes the individual model variants and}$$
$$j = 1, \ldots, 256 \text{ indexes the ensemble members for each varia}$$

- Assume X_{ij} arises from a Gaussian with variant-specific mean:

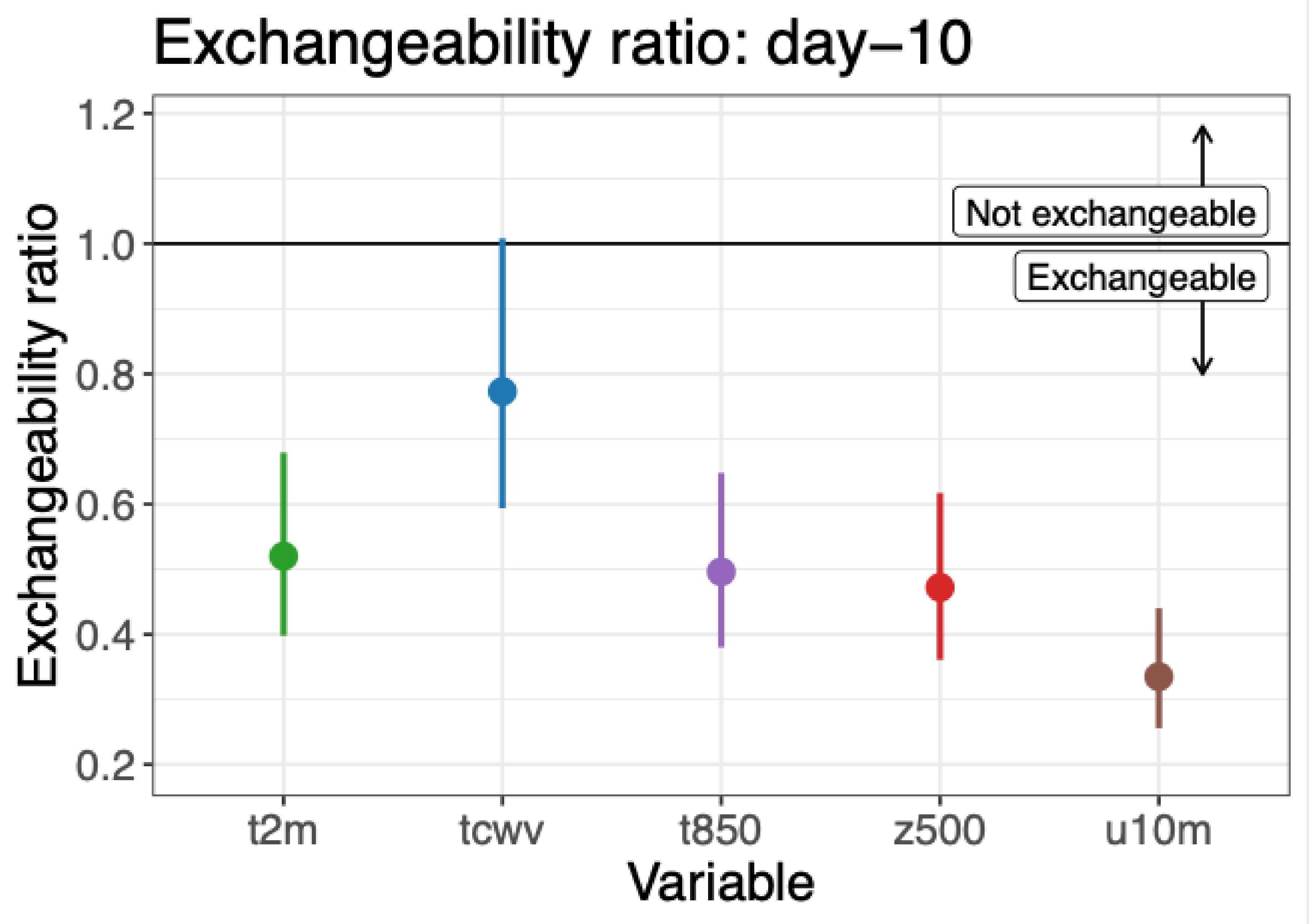
$$X_{ij} \sim N(m_i, \tau^2)$$

- Assume variant-specific mean arises from a different Gaussian:

$$m_i \sim N(M, \sigma^2)$$

- Define an "exchangeability ratio" $R = \sigma/\tau$.
- The possible values of R mean:

$$\boxed{R = \frac{\sigma}{\tau} \begin{cases} < 1 & \text{models are } \textbf{interchangeable} \\ > 1 & \text{models are } \textbf{different} \end{cases}}$$



Information on Extremes Gained from Huge Ensembles

- Define **information gain** for n random variables to be

$$G_n = \max_{i=1,\dots,n} \frac{|X_i - \overline{X}_n|}{S_n}\Bigg|_{\overline{X}=0,S_n=1} = \max_{i=1,\dots,n} |X_i|$$

- We seek expected information gain $E[G_n]$ as a function of n.
- The cumulative distribution function of G_n is

$$\begin{aligned} P(G_n \leq x) &= \Big(P(|X_i| \leq x)\Big)^n \\ P(|X_i| \leq x) &= 2P(X_i \leq x) - 1 \equiv 2\Phi(x) - 1 \end{aligned}$$

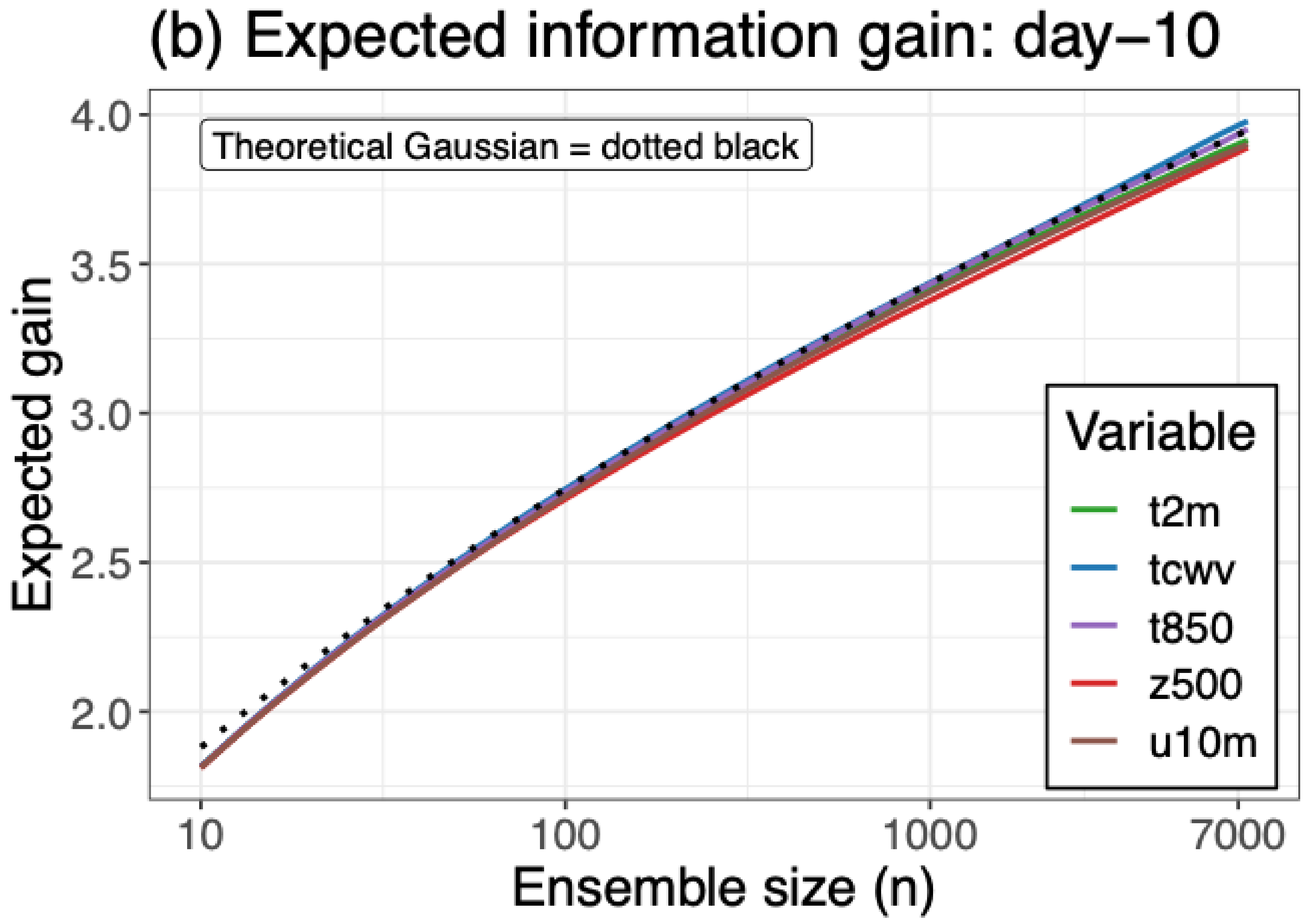
where $\Phi(\cdot)$ is the CDF of $N(0,1)$.
- Hence, for Gaussian data, the CDF of G_n is

$$P(G_n \leq x) = \Big(2\Phi(x) - 1\Big)^n.$$

- The expected gain $E[G_n]$ is

$$\boxed{E[G_n] = \int_0^\infty x \frac{d\,P(G_n \leq x)}{d\,x} dx},$$

which can be solved with numerical integration.



Theory for Large-Sample Behavior of Extreme Statistics

- Assume that HENS arises from a Normal distribution, i.e.,

$$X_i \stackrel{\text{IID}}{\sim} N(\mu, \sigma^2),$$

- Statistical theory yields the uncertainty in the sample mean as

$$\boxed{SD[\overline{X}_n] = \frac{\sigma}{\sqrt{n}}}.$$

- Similarly, the uncertainty in the sample standard deviation is:

$$\boxed{SD[S_n] \xrightarrow[n\to\infty]{} \frac{\sigma}{\sqrt{n}}}$$

- The uncertainty in the $100\alpha^{\text{th}}$ sample percentile from a sample of size n, denoted $X_n(\alpha)$, is

$$\boxed{SD[X_n(\alpha)] \approx \frac{1}{\sqrt{n}} f(\alpha)}.$$

Large-Sample Behavior of Extreme Statistics