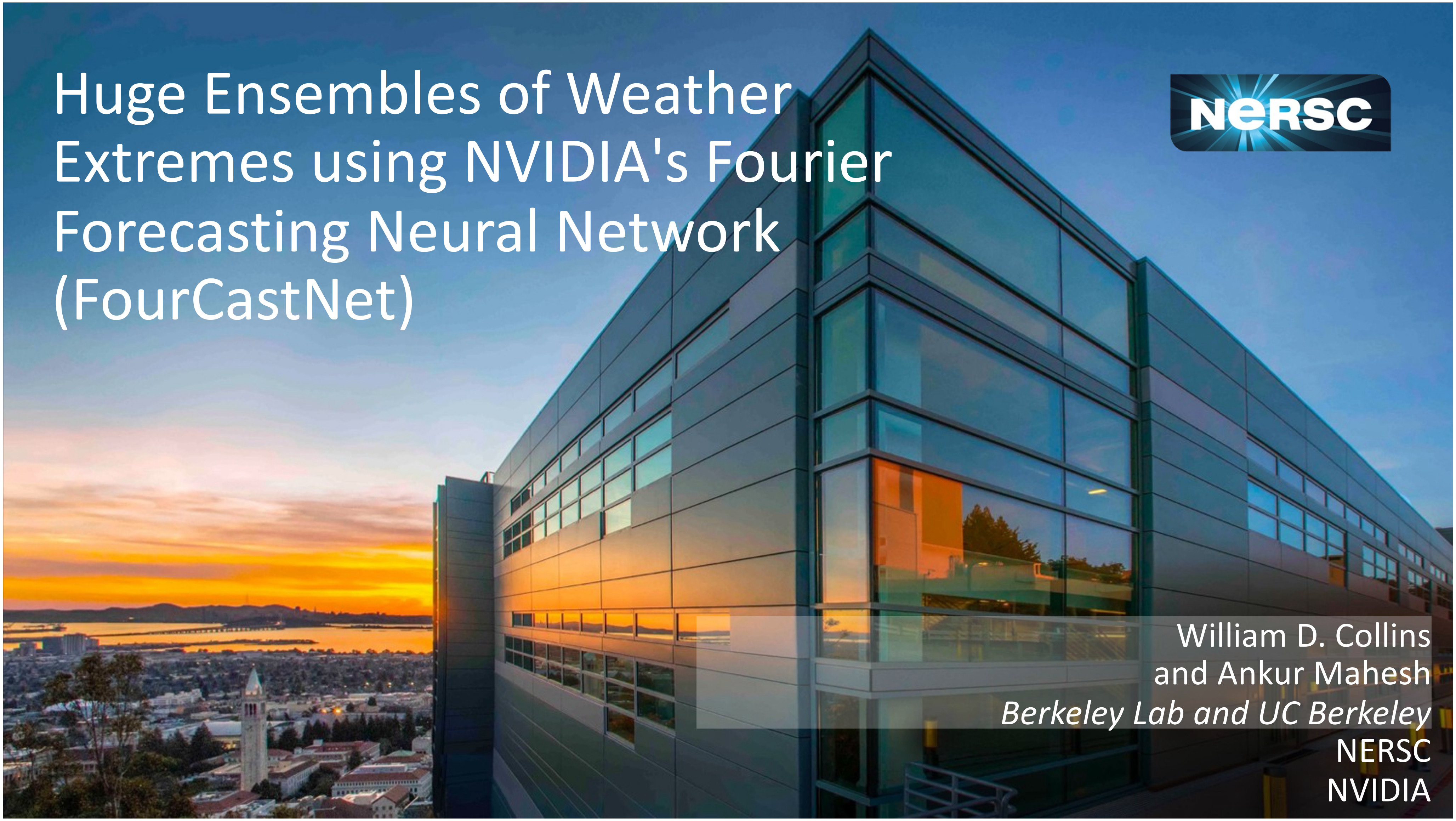
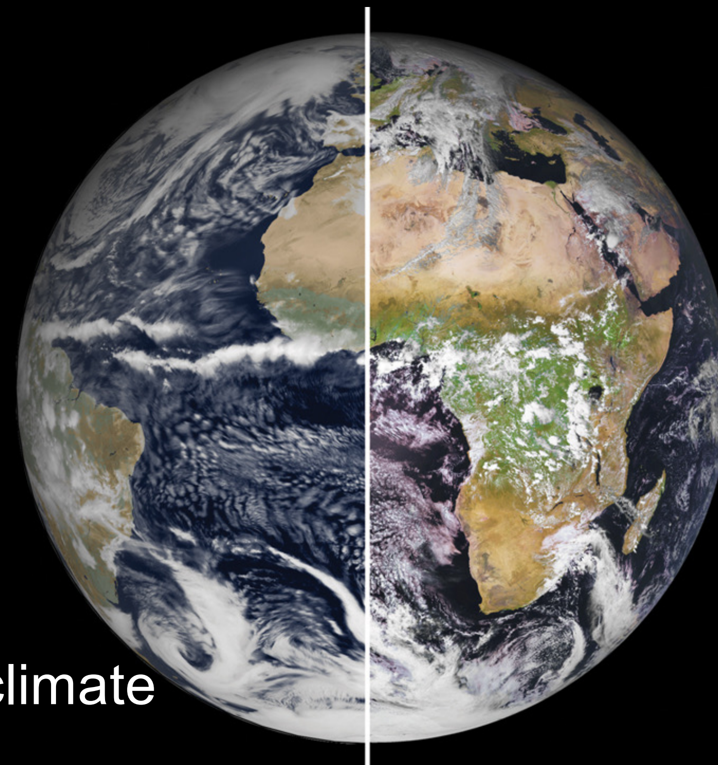


Huge Ensembles of Weather Extremes using NVIDIA's Fourier Forecasting Neural Network (FourCastNet)

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

Outline

- Low-likelihood High Impact Extremes (LLHIs)
- Background: data-driven forecasting
- FourCastNet model
    - Forecast results on extremes
    - Architecture & training strategy
- Ongoing/future work
    - Construction of Huge Ensembles
    - Operationalizing data-driven weather & climate Prediction



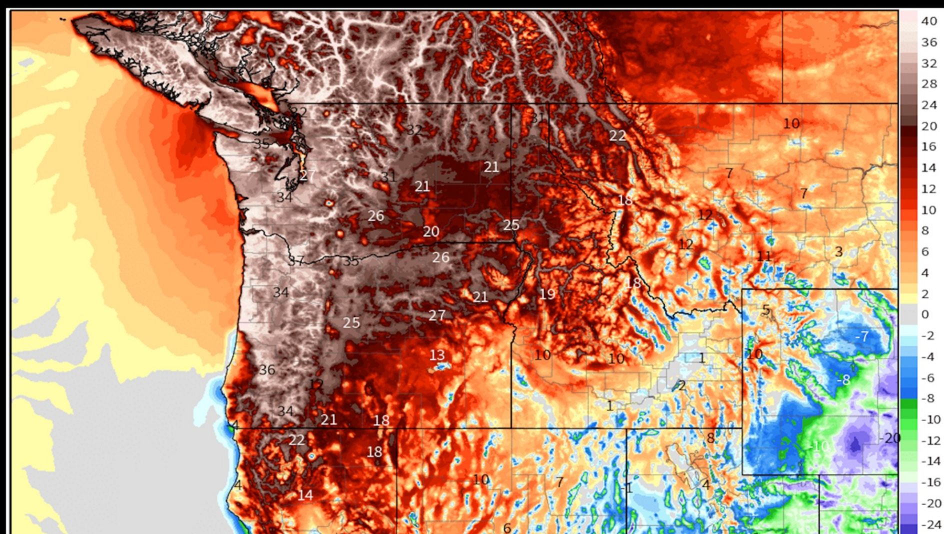
Low-Likelihood High Impact Extremes

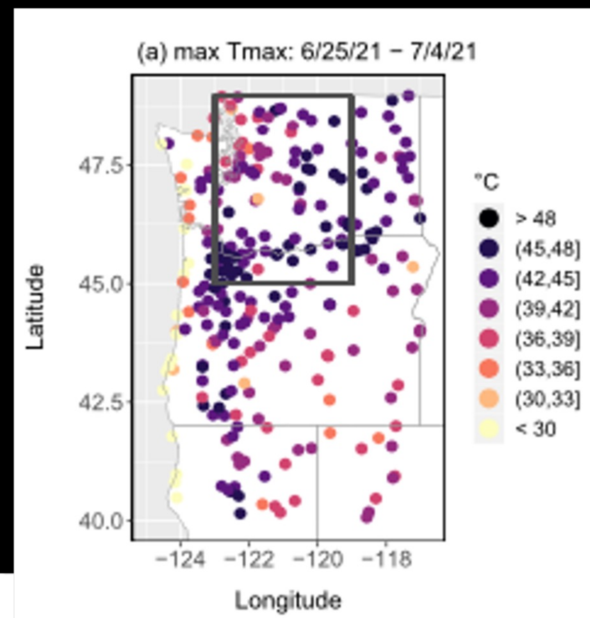
IPCC AR6:
… In summary, the future occurrence of LLHI events linked to climate extremes is generally associated with low confidence, but cannot be excluded, especially at global warming levels above 4°C.

Types of LLHIs to be investigated

- *Heatwaves*
- *Atmosphere rivers*
- *Tropical cyclones and hurricanes*
- *Extreme downpours and flooding*
- *Co-occurring extreme events*
- *Western US hydroclimate extreme events (droughts, floods, mountain snow, wildfires)*

Low-Likelihood High Impact Extremes

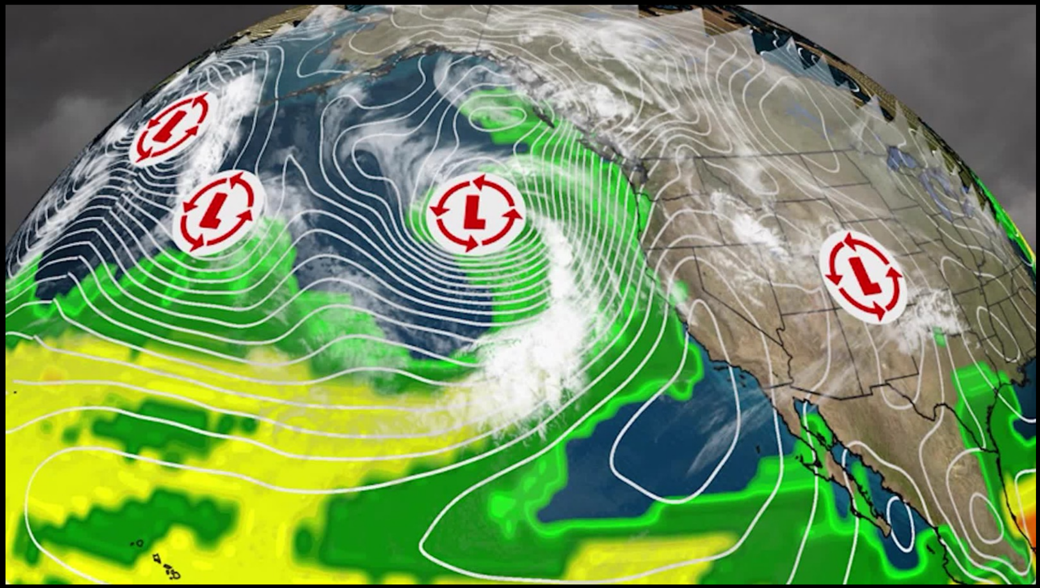




Anthropogenic Contributions to the 2021 Pacific Northwest Heatwave

Emily Bercos-Hickey[1], Travis A. O'Brien[1,2], Michael F. Wehner[3], Likun Zhang[1,4], Christina M. Patricola[1,5], Huanping Huang[1,6], and Mark D. Risser[1]

Low-Likelihood High Impact Extremes



Multiple ARs landing in California Dec-Jan, 2023

Why LLHIs? From the IPCC AR6 WG1 SPM….

- "**Low-likelihood outcomes**, *such as ice-sheet collapse, abrupt ocean circulation changes, some compound extreme events, and warming substantially larger than the assessed very likely range of future warming,* **cannot be ruled out and are part of risk assessment.**"
- "**Low-likelihood, high-warming outcomes are associated with potentially very large impacts,** *such as through more intense and more frequent heatwaves and heavy precipitation, and high risks for human and ecological systems, particularly for high GHG emissions scenarios.*"
- "*Low-likelihood, high-impact outcomes could occur at global and regional scales even for global warming within the very likely range for a given GHG emissions scenario.* **The probability of low-likelihood, high-impact outcomes increases with higher global warming levels (high confidence).** "

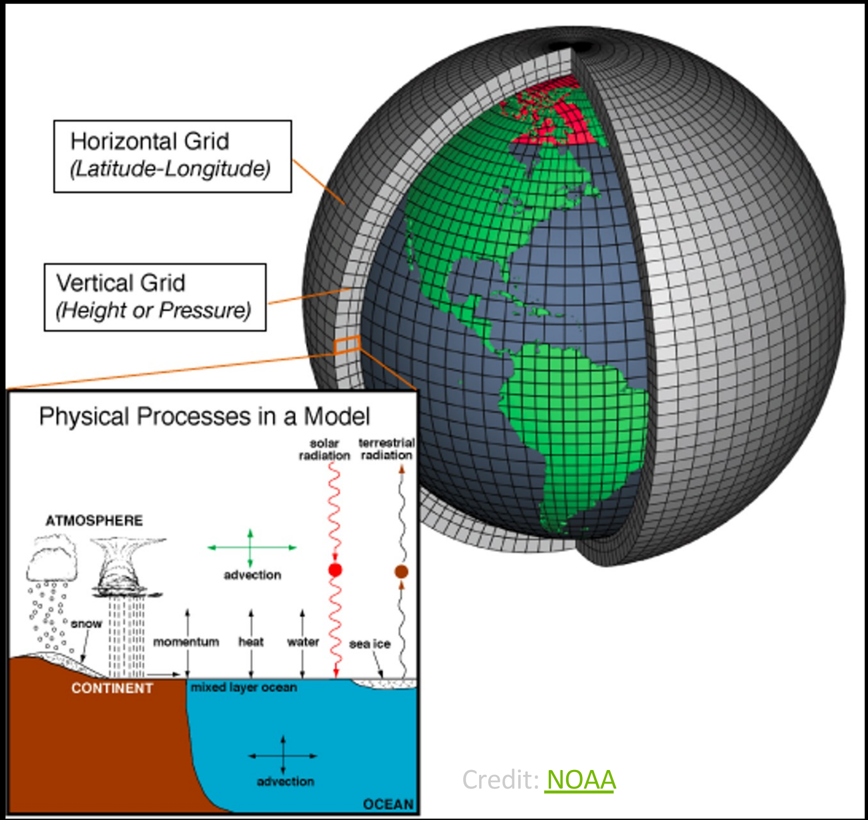
Numerical Weather Prediction (NWP)

- Urgency under climate change:
  - Reducing uncertainties, predicting extremes, disaster mitigation, etc.…
- Entire earth system: complex phenomena across wide range of physical scales

- Traditional NWP consumes substantial HPC resources
  - Dedicated machines running physical models + data assimilation 24/7

Ingredients of NWP

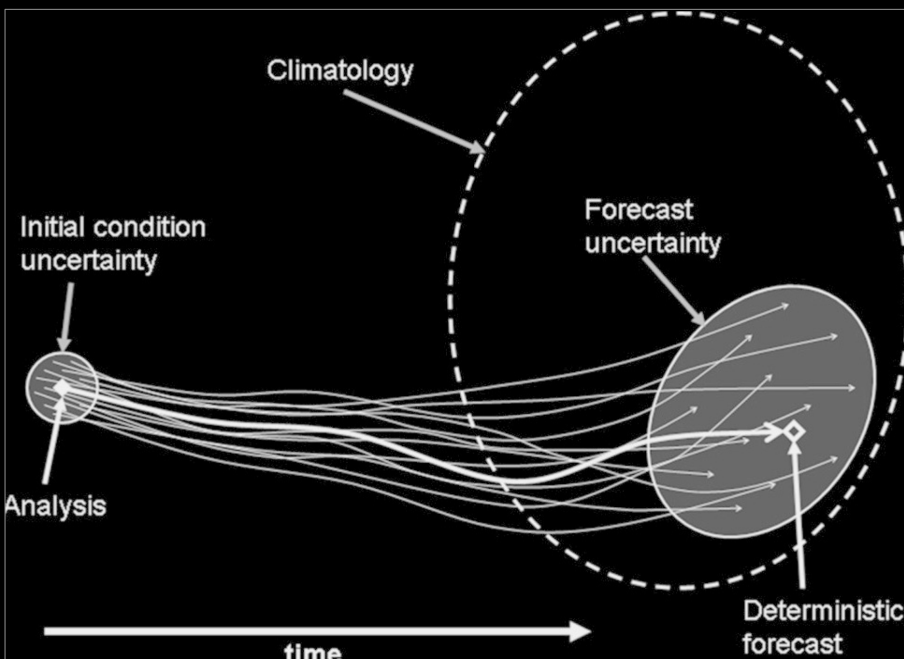
- PDEs w/ hundreds of variables, complex parameterizations for subgrid and multi–physics processes
- High resolution grids (computational cost scales as ~fourth power)
- Large ensembles for UQ and long-term forecasting

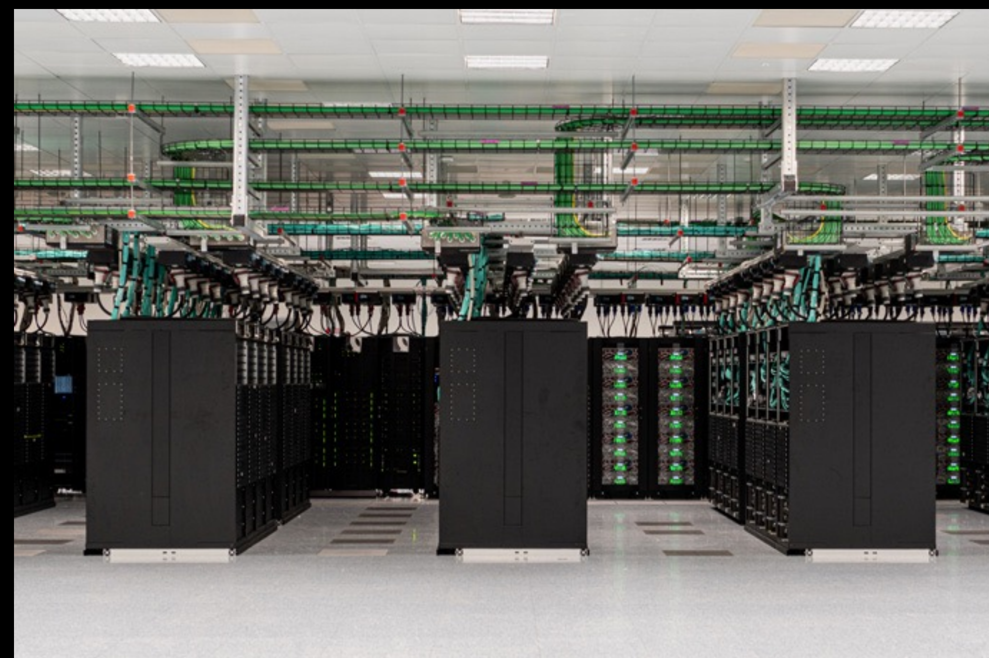


Credit: [NOAA]()

Deterministic Numerical Weather Prediction (NWP)

Solving equations of motion for an incompressible fluid on a rotating sphere

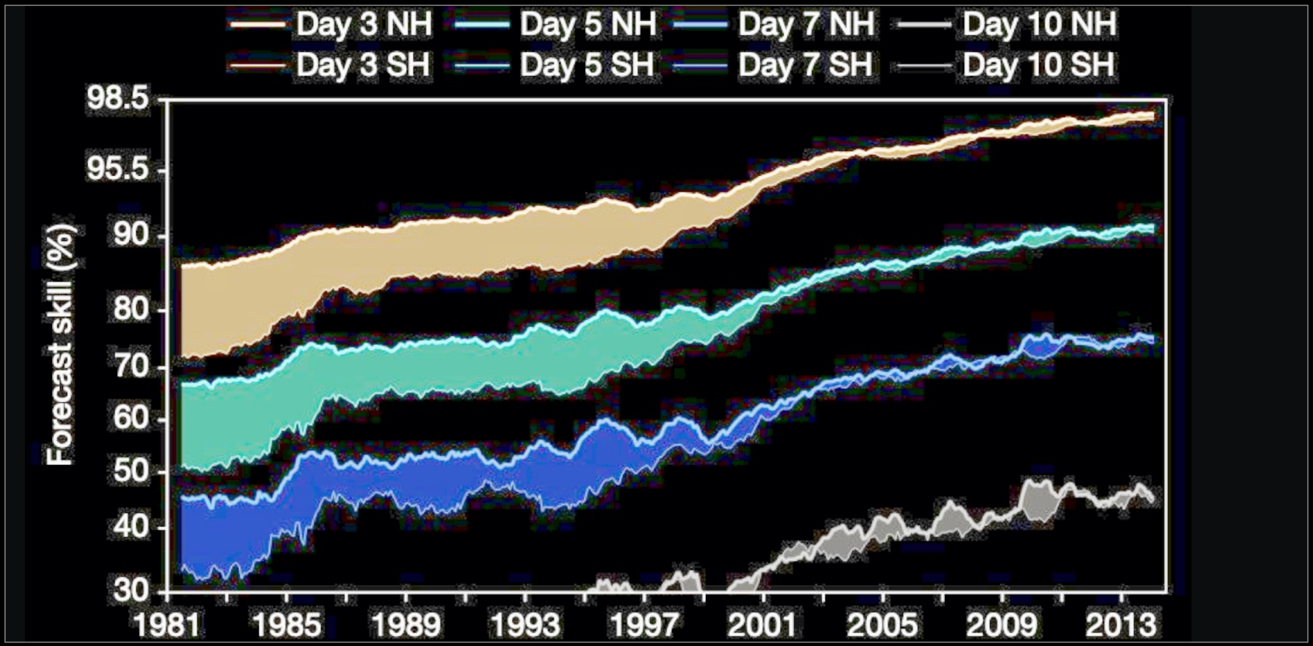




HPC-intensive: European Center for Medium Range Weather Forecasting

The slow but steady evolution of classical NWP

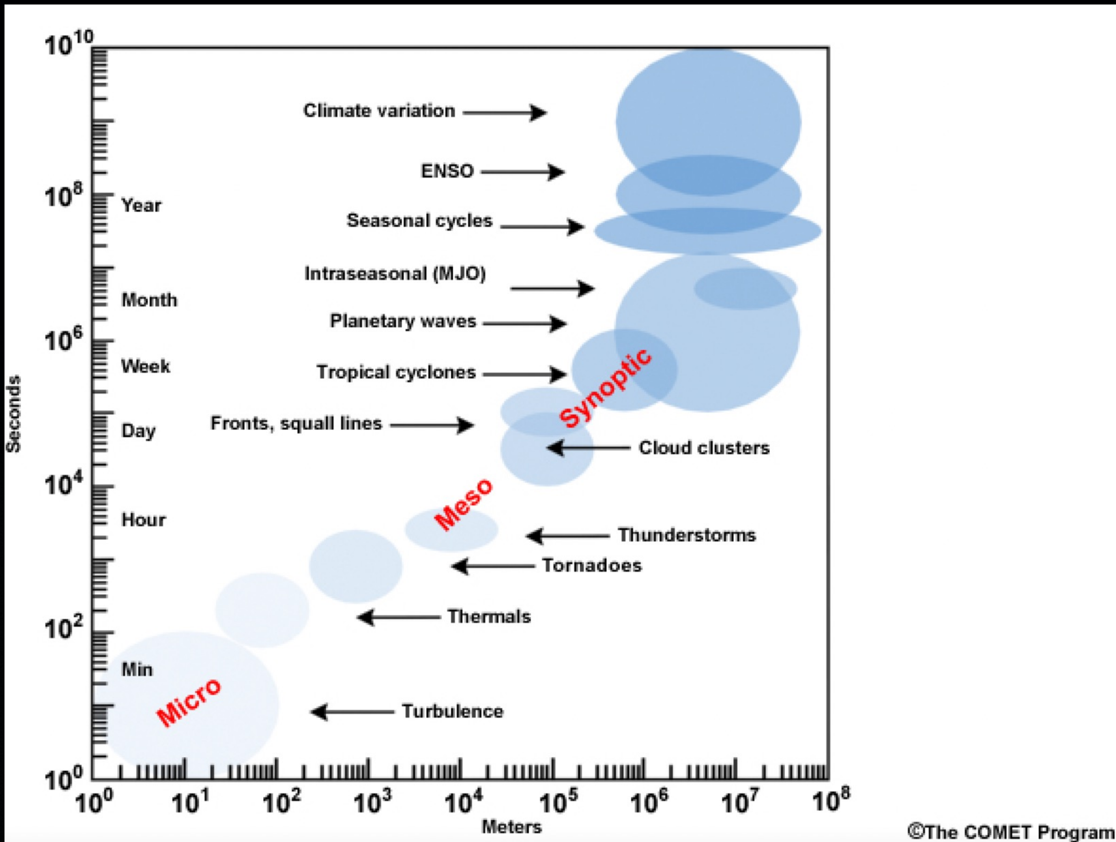
Data assimilation and other advances have slowly revolutionized the quality & accuracy over decades

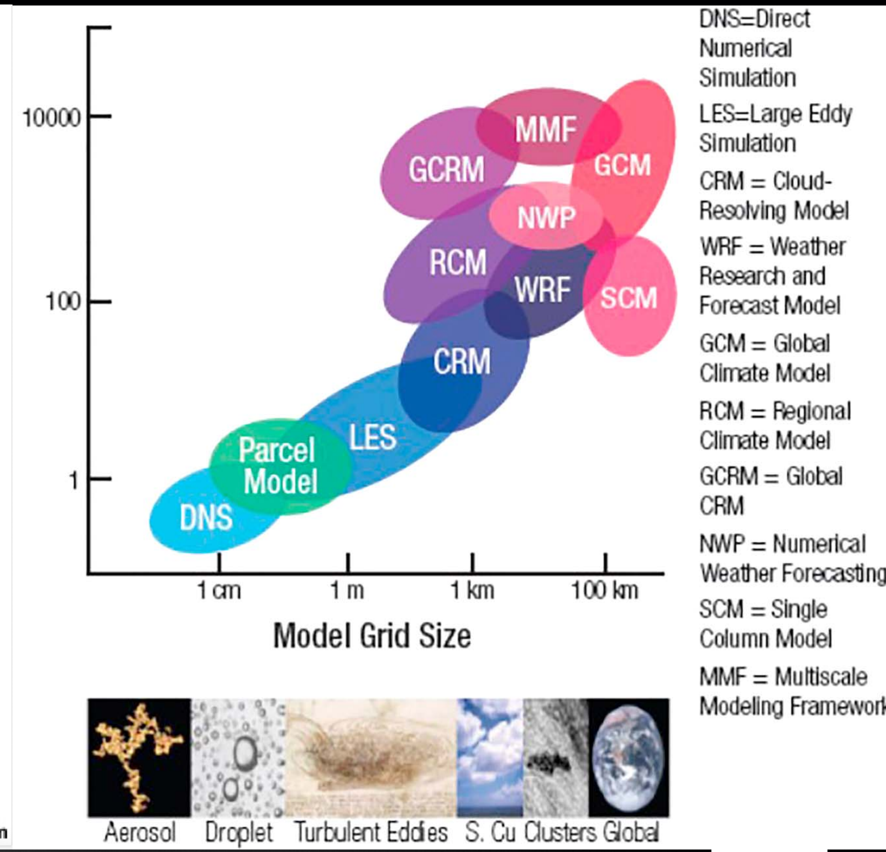


Bauer et al., *Nature*, 2015

Caveat: These advances are the foundation for assimilated state estimates that ML approaches rely on.

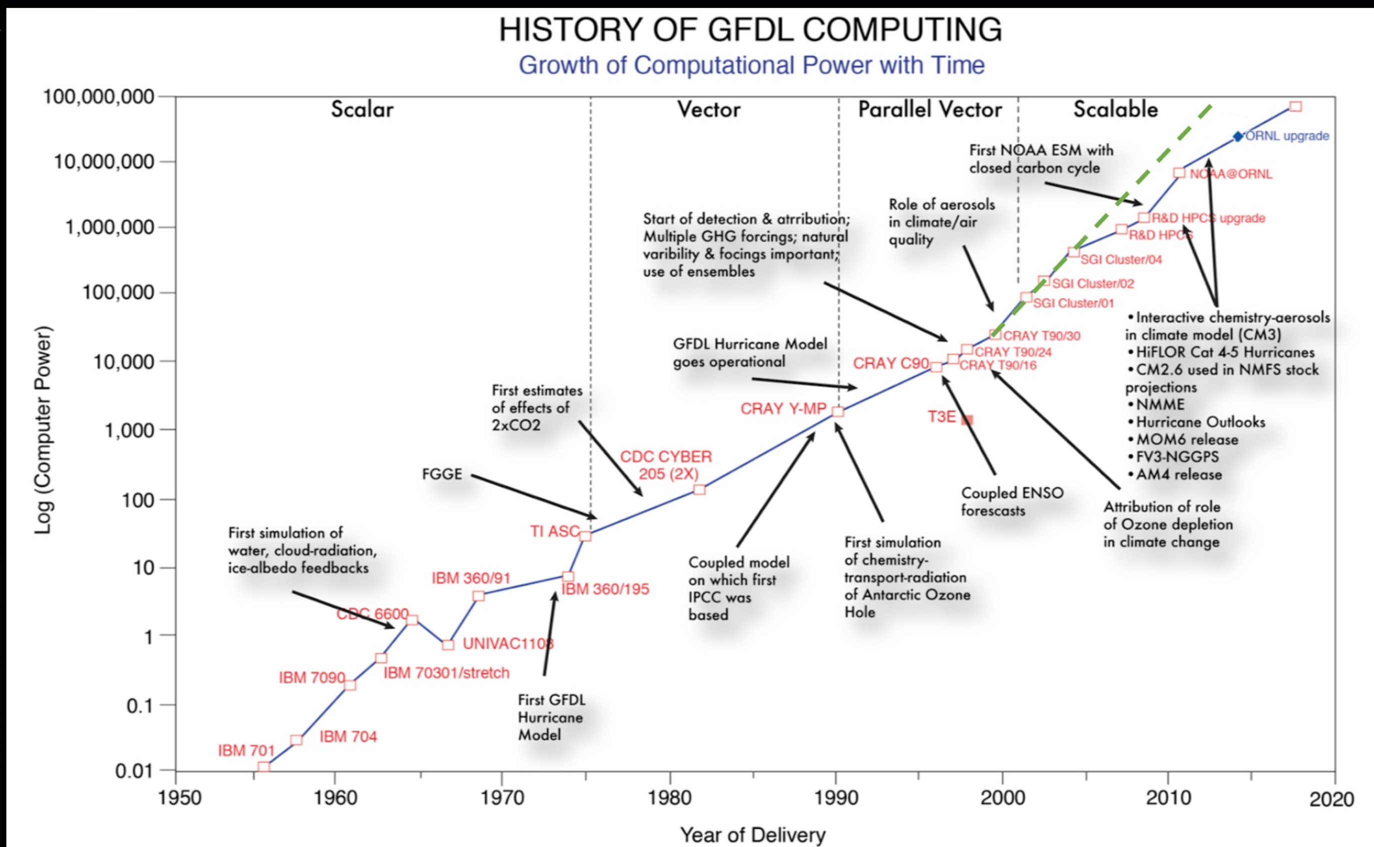
Climate Phenomena and Models are Inherently Multiscale





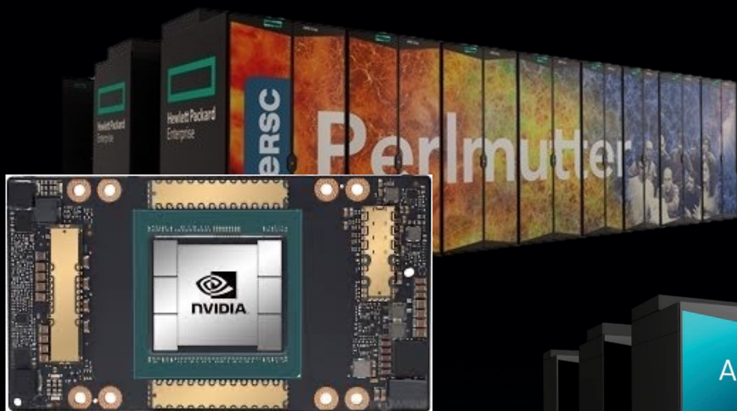
Faster computers have bridged 4 orders of magnitude, out of 16

X 10 Billion



V. Balaji / GFDL

Enabling Technology: DOE's GPU Exascale Systems

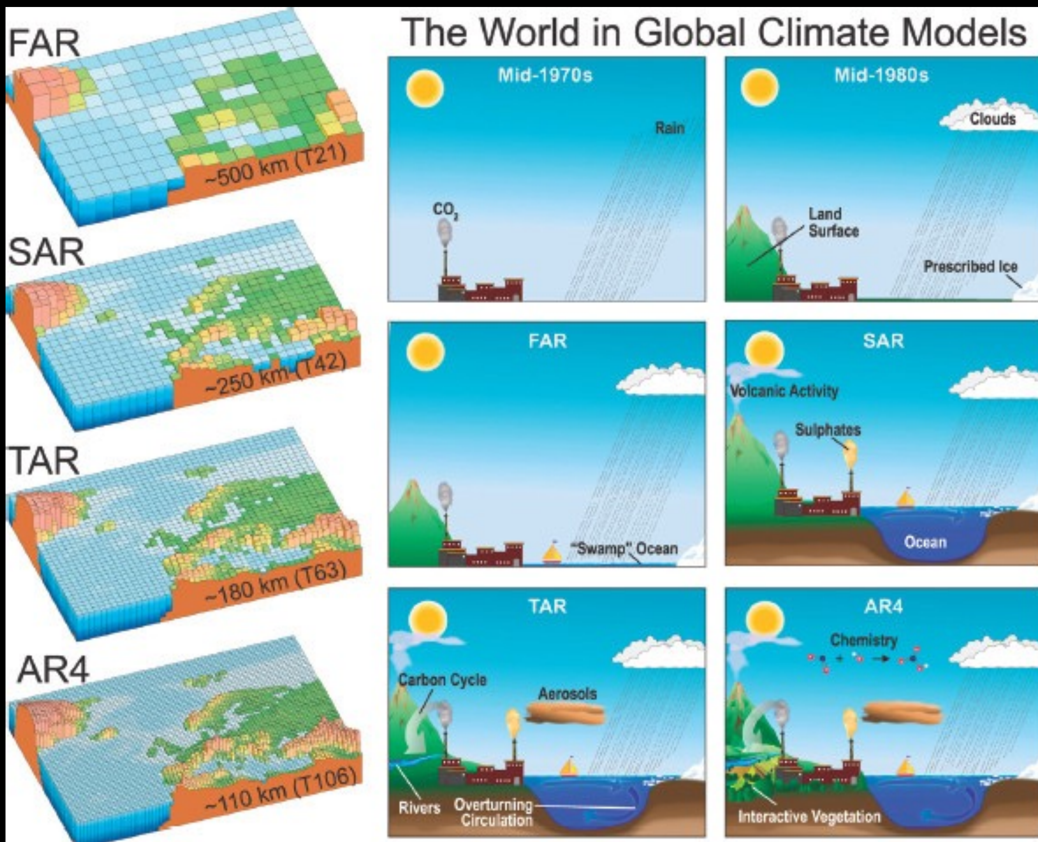


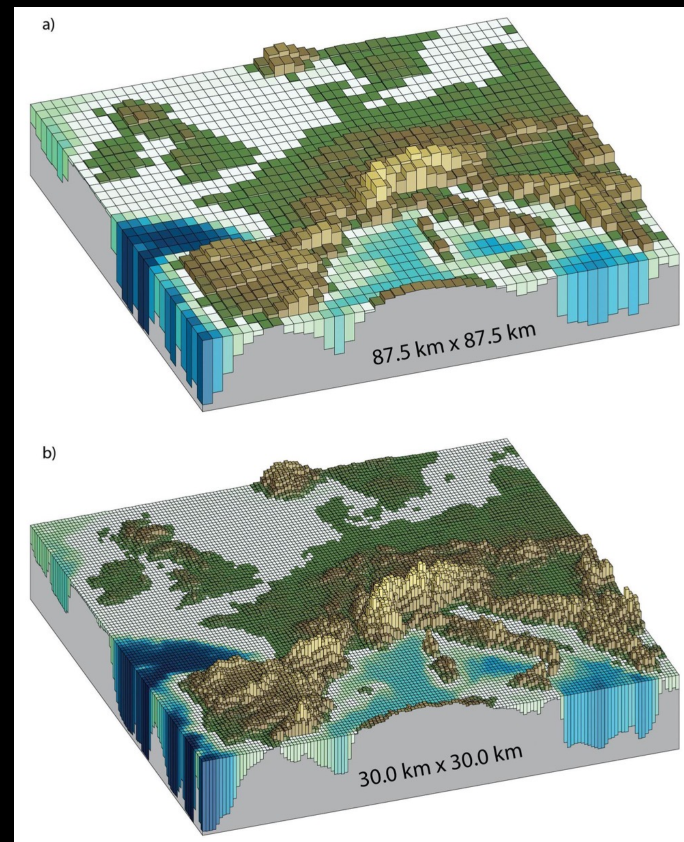
2021: 100 petaFLOPS/s (NERSC)

2023: ~1 exaFLOPS/s (ALCF)

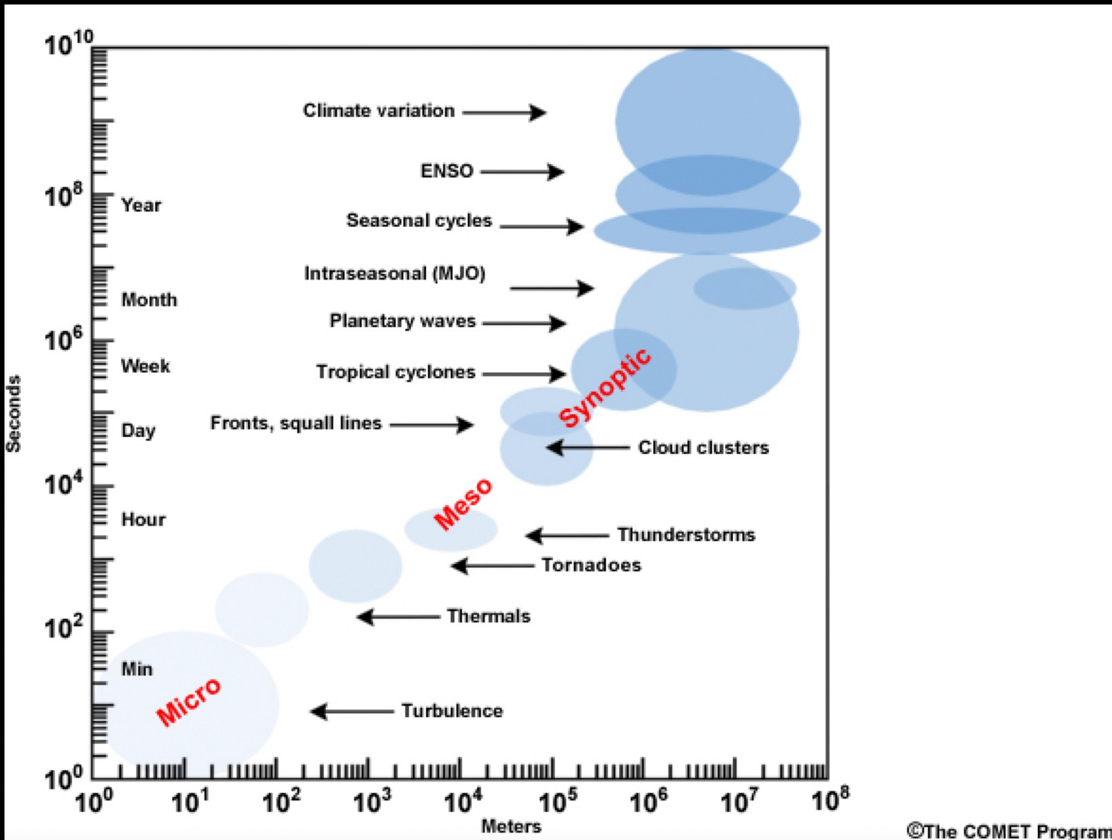
2022: ~1.7 exaFLOPS/s (OLCF)

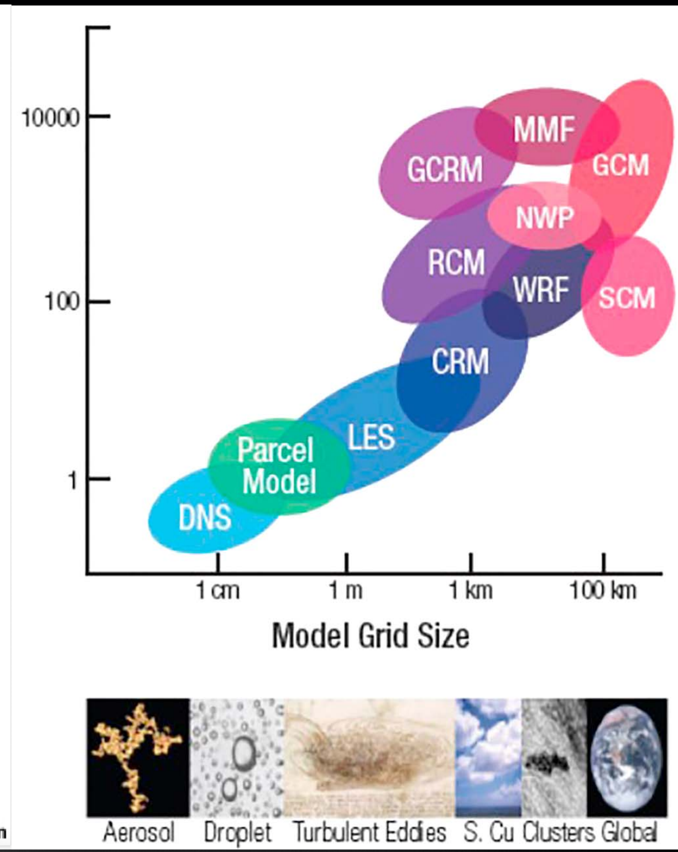
Gains in Resolution for Climate Modeling





Climate Phenomena and Models are Inherently Multiscale

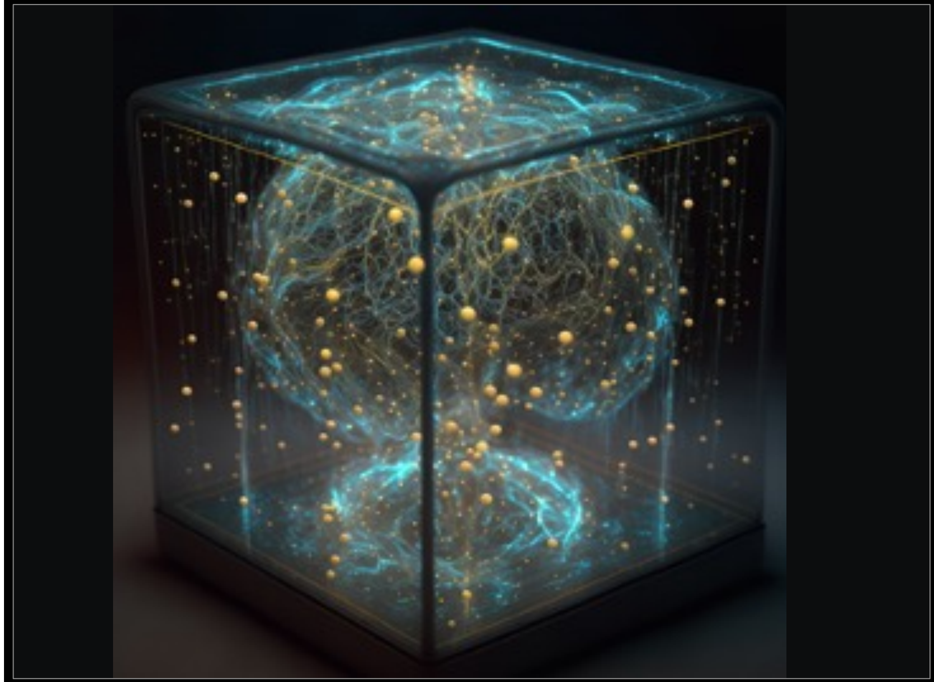




DNS=Direct Numerical Simulation
LES=Large Eddy Simulation
CRM = Cloud-Resolving Model
WRF = Weather Research and Forecast Model
GCM = Global Climate Model
RCM = Regional Climate Model
GCRM = Global CRM
NWP = Numerical Weather Forecasting
SCM = Single Column Model
MMF = Multiscale Modeling Framework

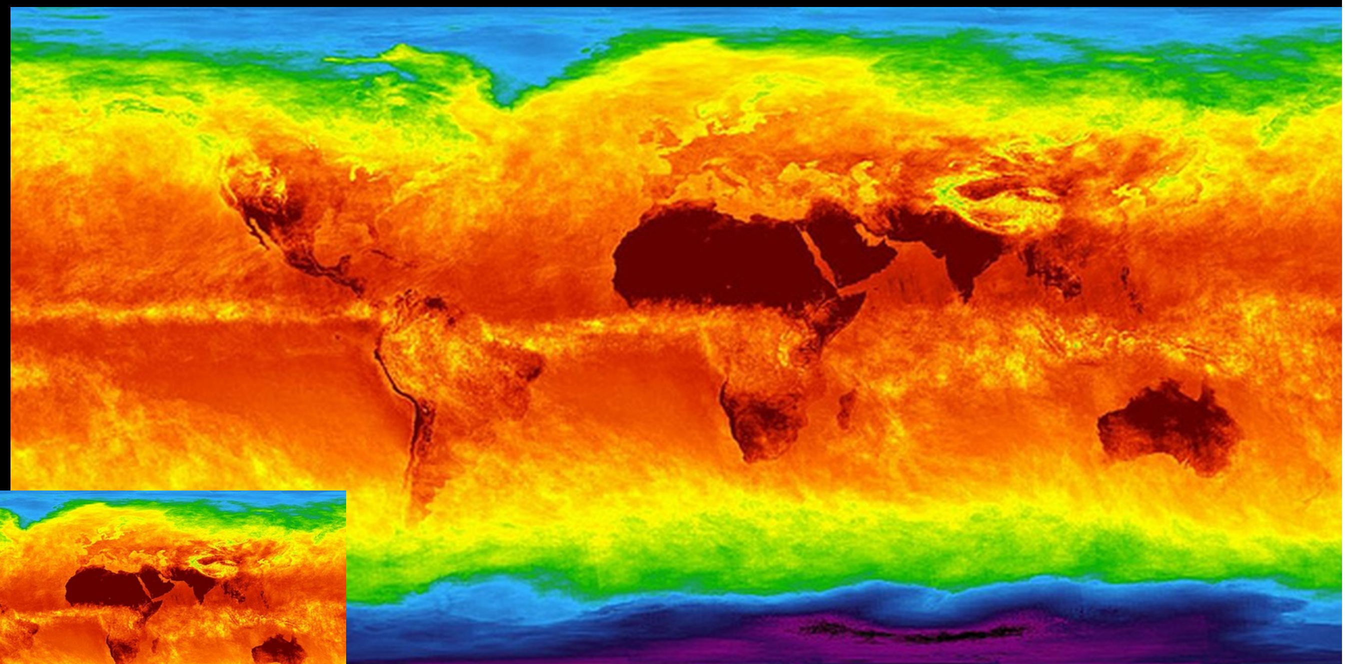
Fully Data-Driven Weather Prediction (DDWP)

A modern alternative, intriguingly complementary

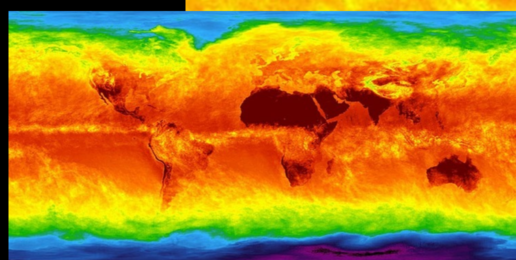


- Mining the data assimilated states from operational NWP
- # training samples = length of satellite record (~ 15k days)
- Can be stood up by small teams within tech companies.
- Is producing skill gains rapidly.

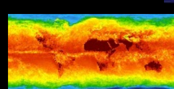
FourCastNet: DDWP at very high resolution



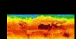
FourCastNet, Pathak et al. (2022), 0.25°, ~1,000,000 Pixels, ViT+AFNO



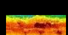
GNN, Keisler et al. (2022), 1°, 64,000 Pixels, Graph Neural Networks



DLWP, Weyn et al. (2020). 2°, 16K pixels, Deep CNN on Cubesphere/(2021) ResNet



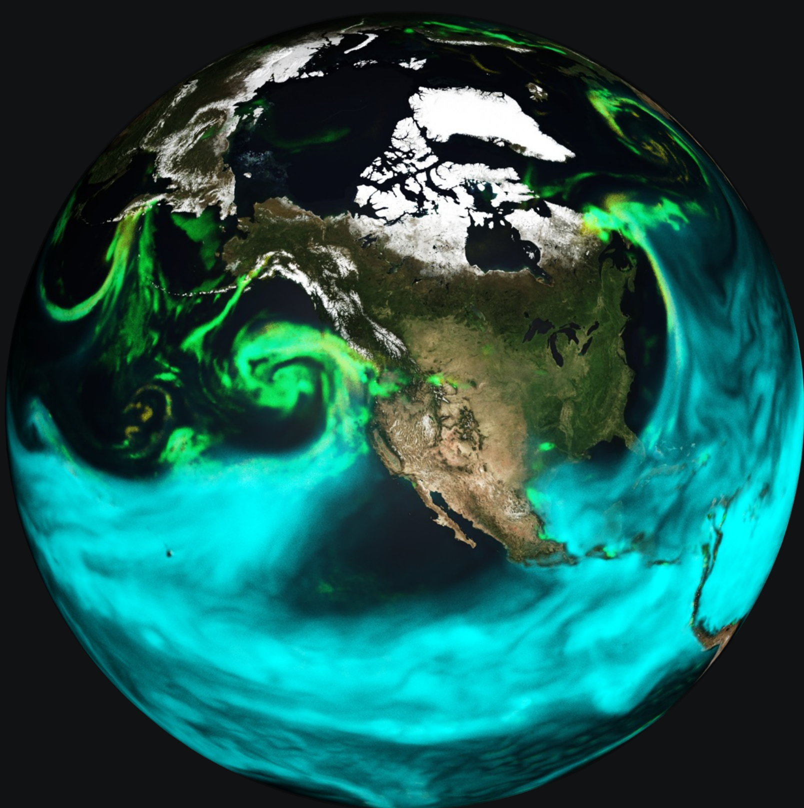
Weyn et al. (2019), 2.5° N.H only, 72x36, 2.6k pixels, ConvLSTM



WeatherBench, Rasp et al. (2020). 5.625°, 64x32, 2K pixels, CNN

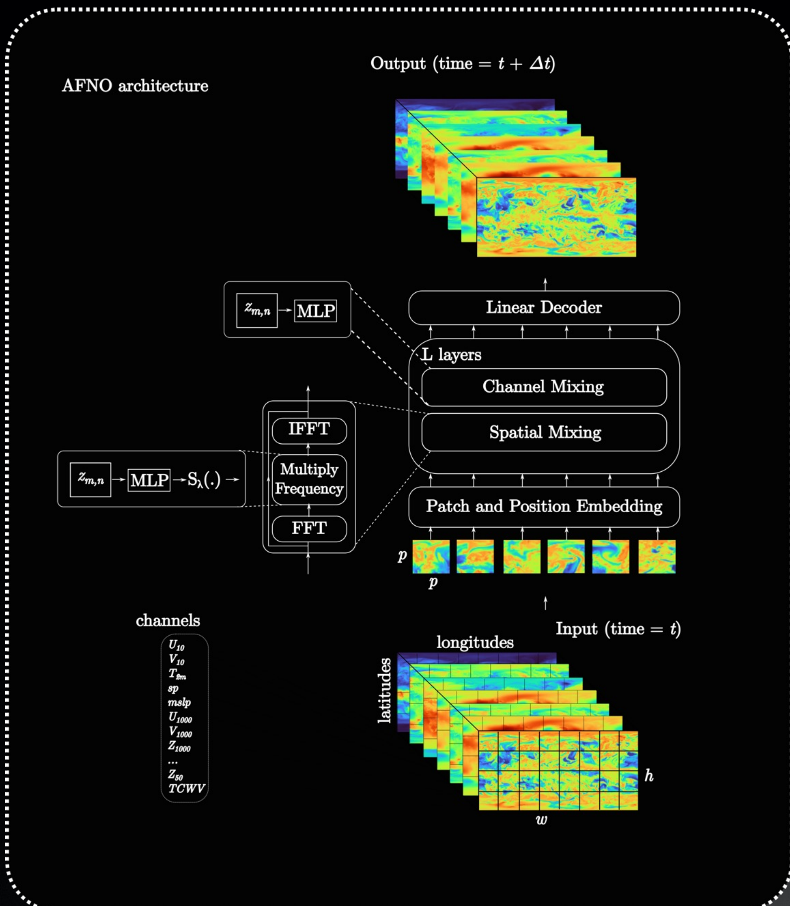
Deuben & Bauer (2018), 6° , 60x30, 1.8K pixels, MLP

FourCastNet pushing the frontier of AI-Driven Digital Twins



Scope	Global
Model Type	Full-Atmosphere AI Surrogate
Architecture	Fourier Neural Operator
Resolution	25km, 6-hourly
Training Data	ERA5 Reanalysis
Initial Condition	ERA5 / GFS / UFS
Training Time	1000 GPU-hrs
Inference Time	3 sec (2-week forecast)
Calibration	IC + Bayesian model uncertainty
Speedup vs NWP	O(10,000 – 100,000)
Power Savings	O(10,000)
Max Stable Rollout	Years
Project Type	Open-source

FourCastNet (Fourier foreCasting Network)



- Purely data-driven ML surrogate weather model
- Fourier transform for global convolution
- Learns solution operator, mesh and resolution invariant
- Training data: ERA5 reanalysis
- Autoregressive time interval: 6 hours
- 73 state variables selected:
  - Temperatures, winds, geopotential & humidity (surface & 12 vertical levels)
  - Surface pressure, column water vapor, …

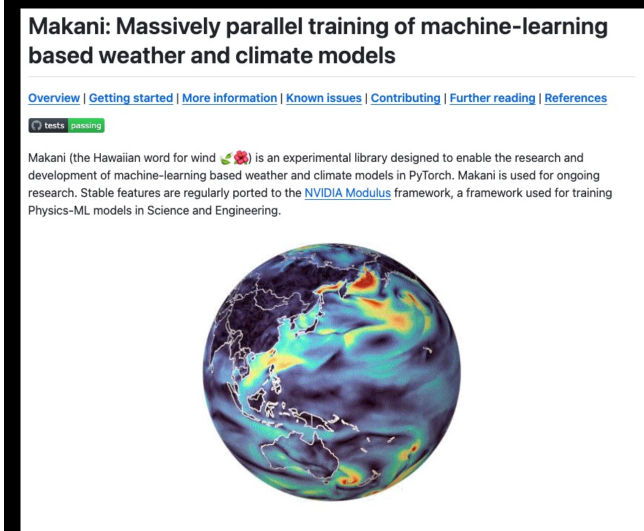
Extending to include radiation processes, vapor transport, clouds

Training set: 1979 to 2015
Validation set: 2016, 2017
Held out: 2018 onwards

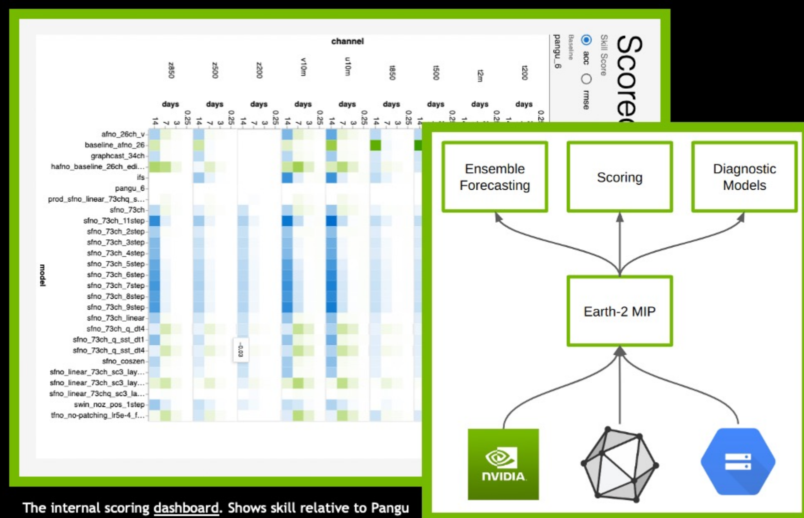
Pathak et al. (2022), FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators, arXiv:2202.11214

FourCastNet and other DDWP models available online

Modulus-Makani, Earth2-MIP, ECMWF AI Lab



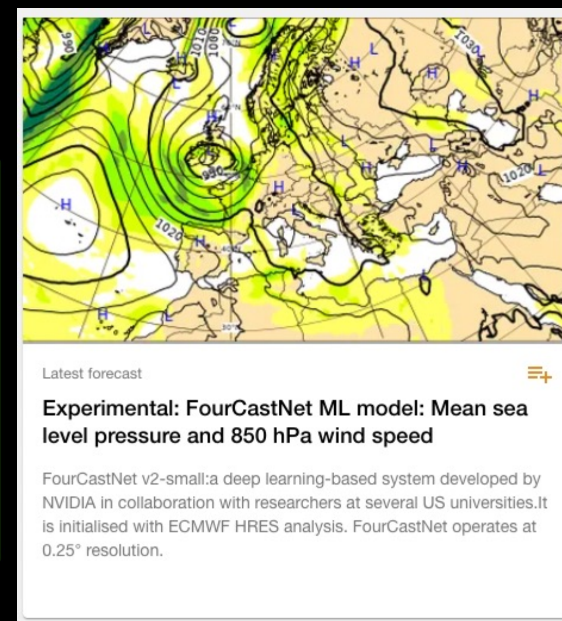
https://github.com/NVIDIA/modulus-makani



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

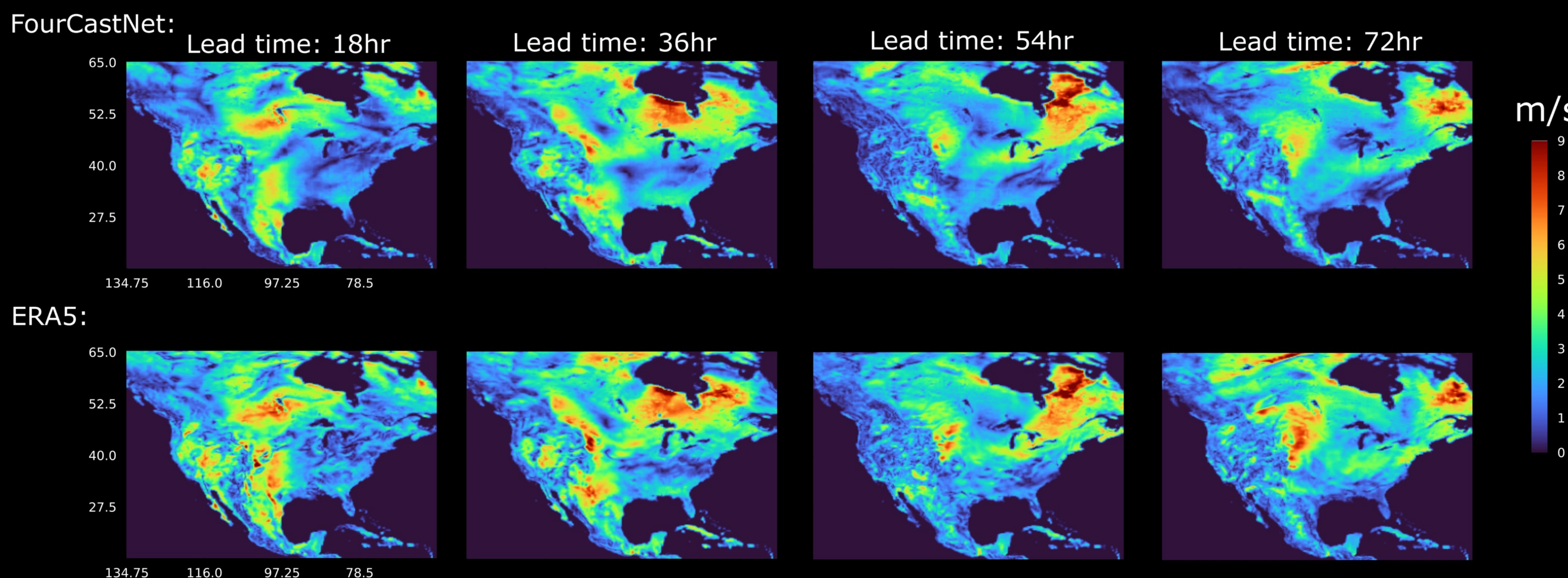


ECMWF | Charts

- Open sourcing and rapid training are indispensable for academic partnerships
- Our collaboration have actively contributed to these code bases.

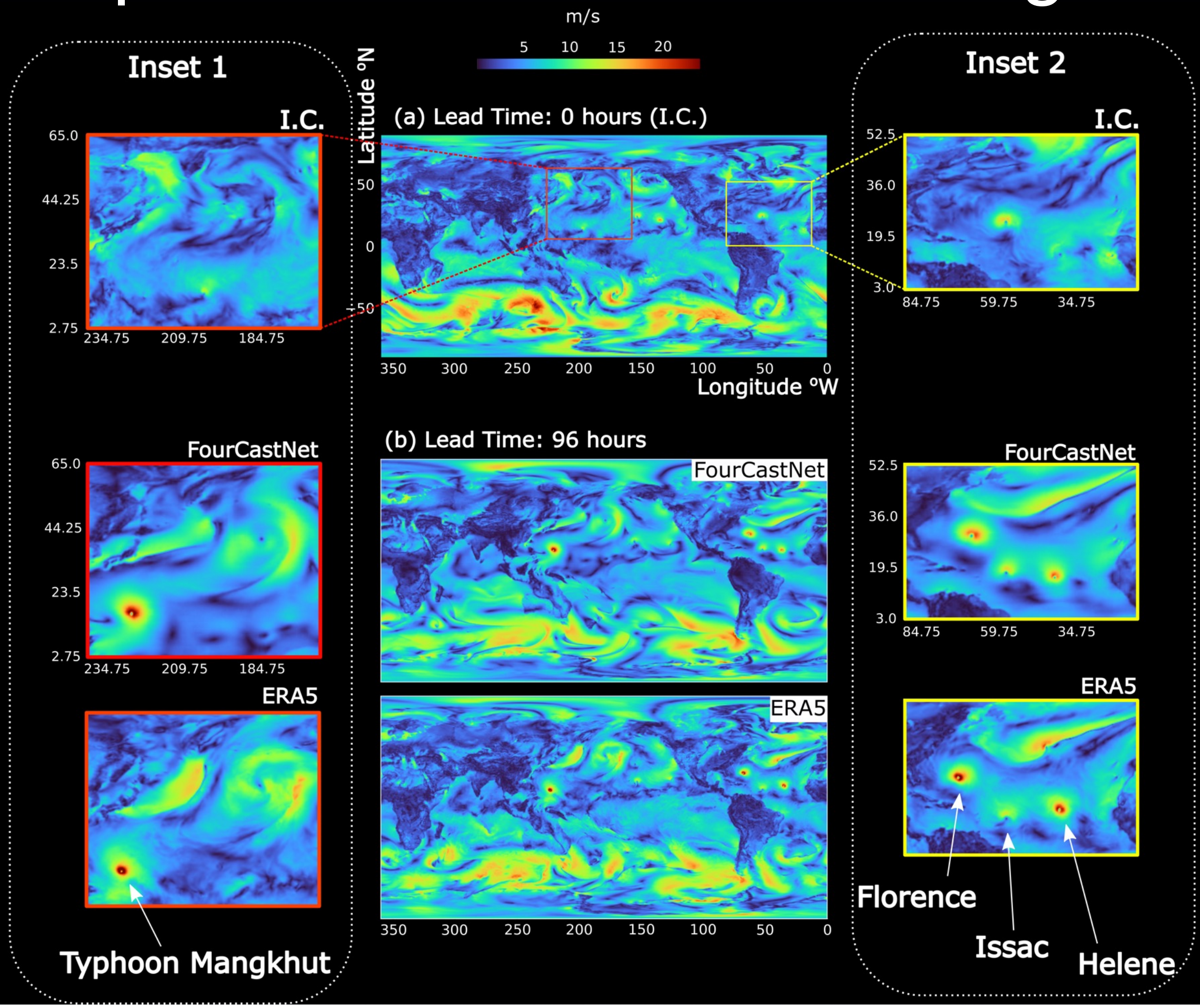
NVIDIA

FourCastNet predicts many atmospheric fields accurately

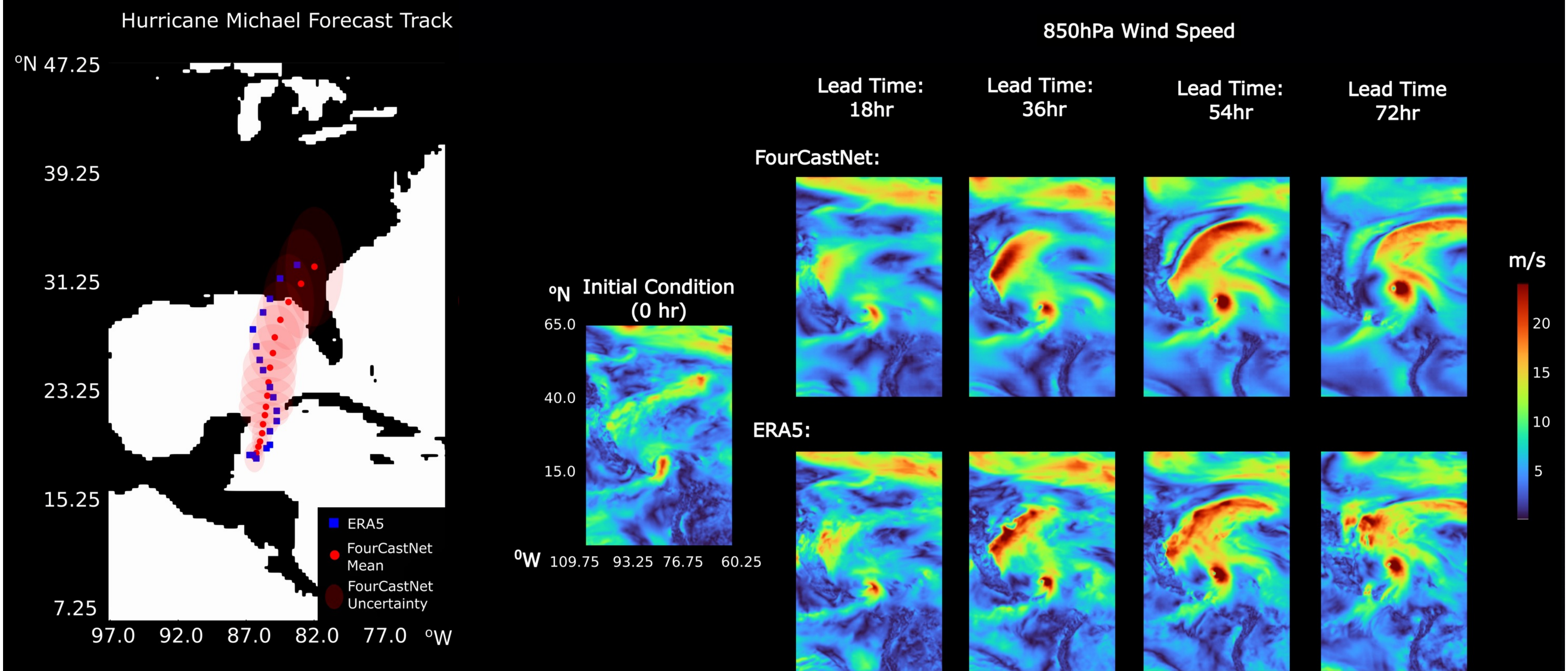


FCN has impressive skill on forecasting extremes

Including tropical cyclones, extra-tropical cyclones, and atmospheric rivers.

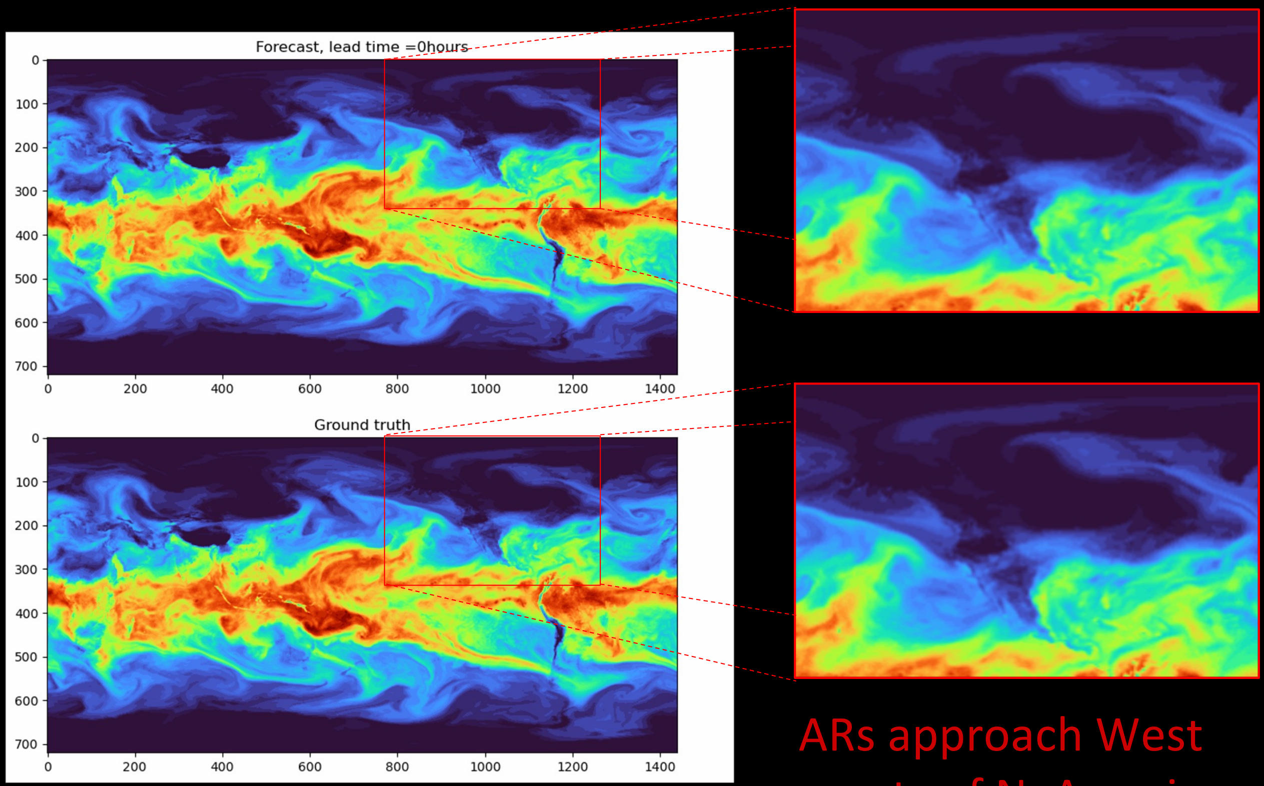


FourCastNet predicts hurricane paths and intensities



FCN accurately emulates Atmospheric Rivers

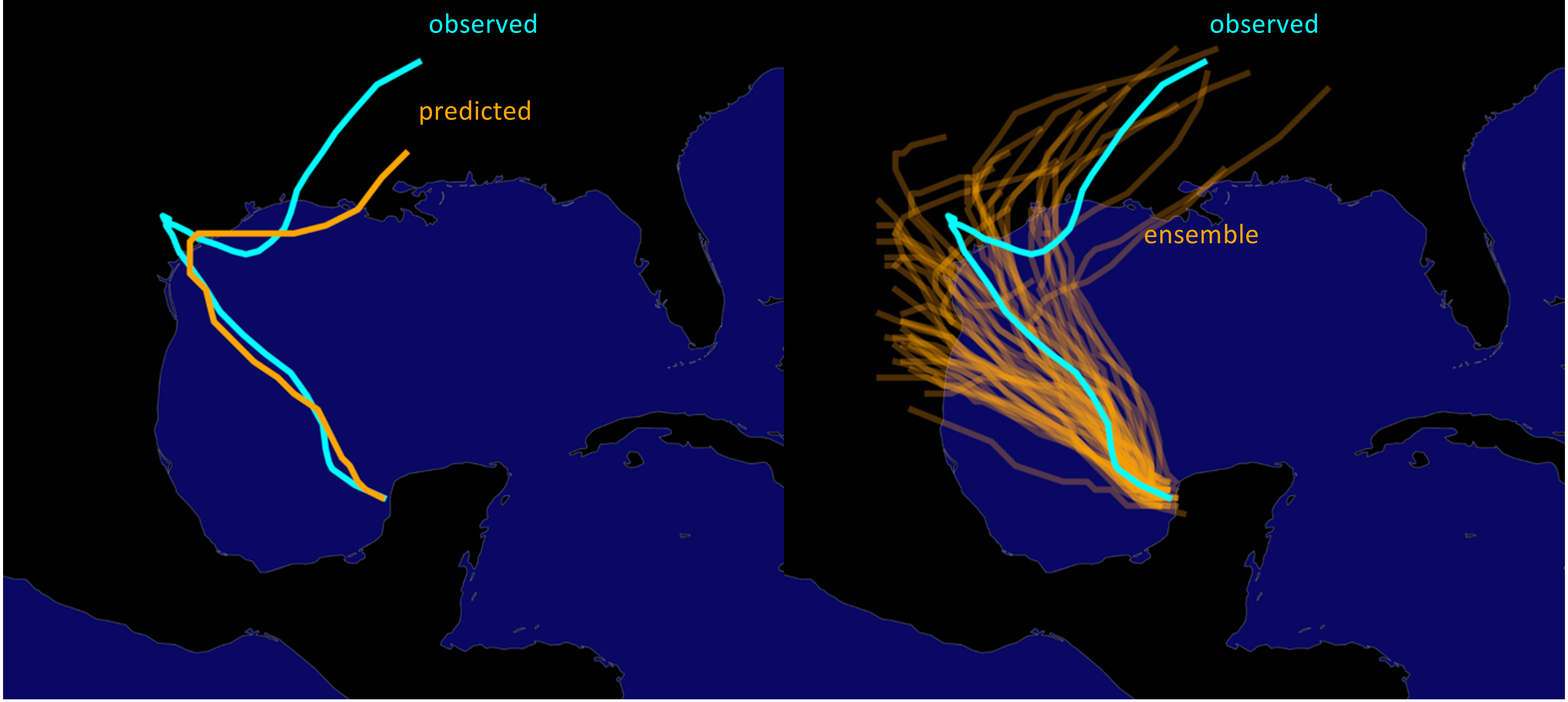
- Backbone forecasts moisture vars well (e.g., total column water vapor has ACC>0.6 out to 10 days)
- Key to capturing atmospheric river dynamics



ARs approach West coasts of N. America, Europe

FourCastNet’s speed enables massive ensembles.

To capture low-likelihood high-impact extreme events more accurately – far into the long tails of distributions



For simulating LLHIs, DDWP is a “killer app”

Characterizing LLHIs requires examination of many realizations of them.

The only way to produce a sufficient number of realizations is via simulation.

Given the low frequency of LLHIs, the only way to conduct these simulations is with an emulation of NWP/ESM codes that runs orders of magnitude faster.

This is why DDWP is valuable for characterizing rare climate extremes.

”Must-have” use case for DDWP 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.

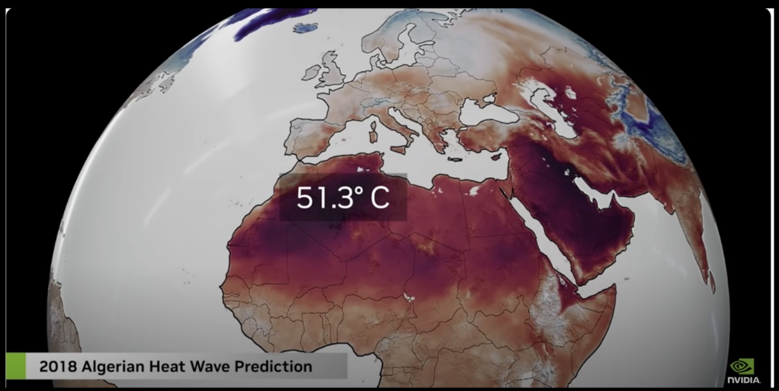
Hypothesis: Ergodicity of climate system means we can ”trade” increasing ensemble size with increasing length of sampling time.

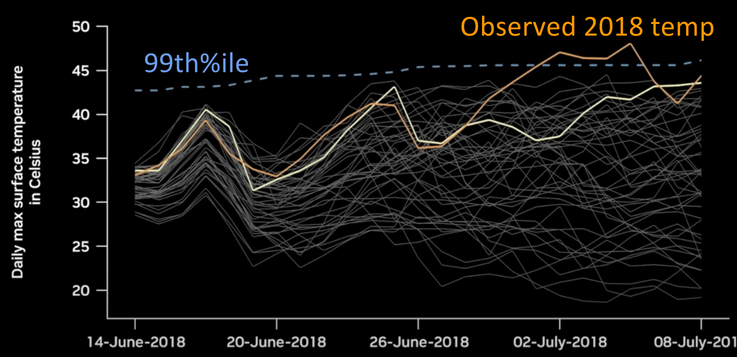
Data-Driven Forecasts from FourCastNet

FourCastNet is a machine learning architecture used for weather and climate prediction.

It can produce a 2-week, 0.25 degree forecast in 0.5 seconds. It's ~4 orders of magnitude faster than traditional models.

This enables the creation of **huge** ensembles (1,000-10,000 members). This ensemble size is vital for capturing the long-tails of climate distributions.





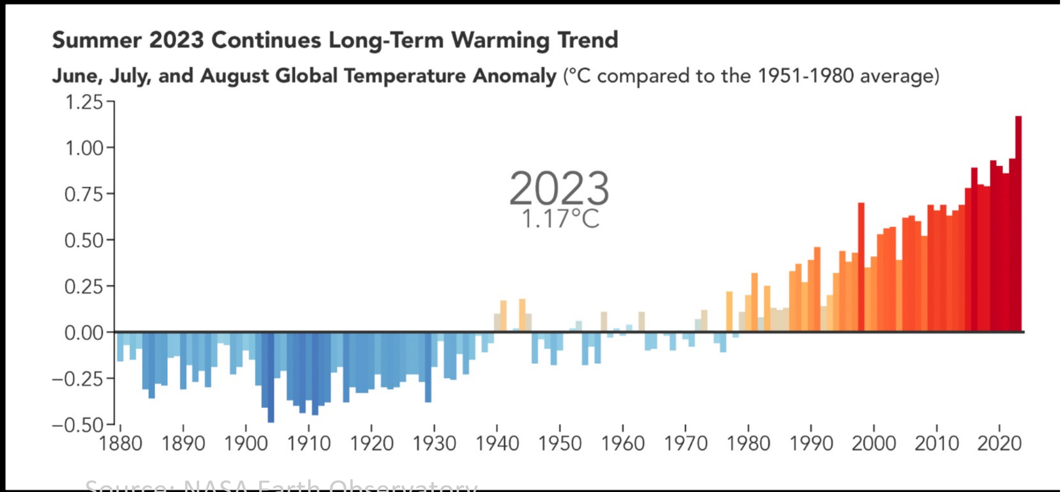
NVIDIA: https://www.youtube.com/watch?v=FUUT6IrQjo4

Huge Ensembles (HENS)

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

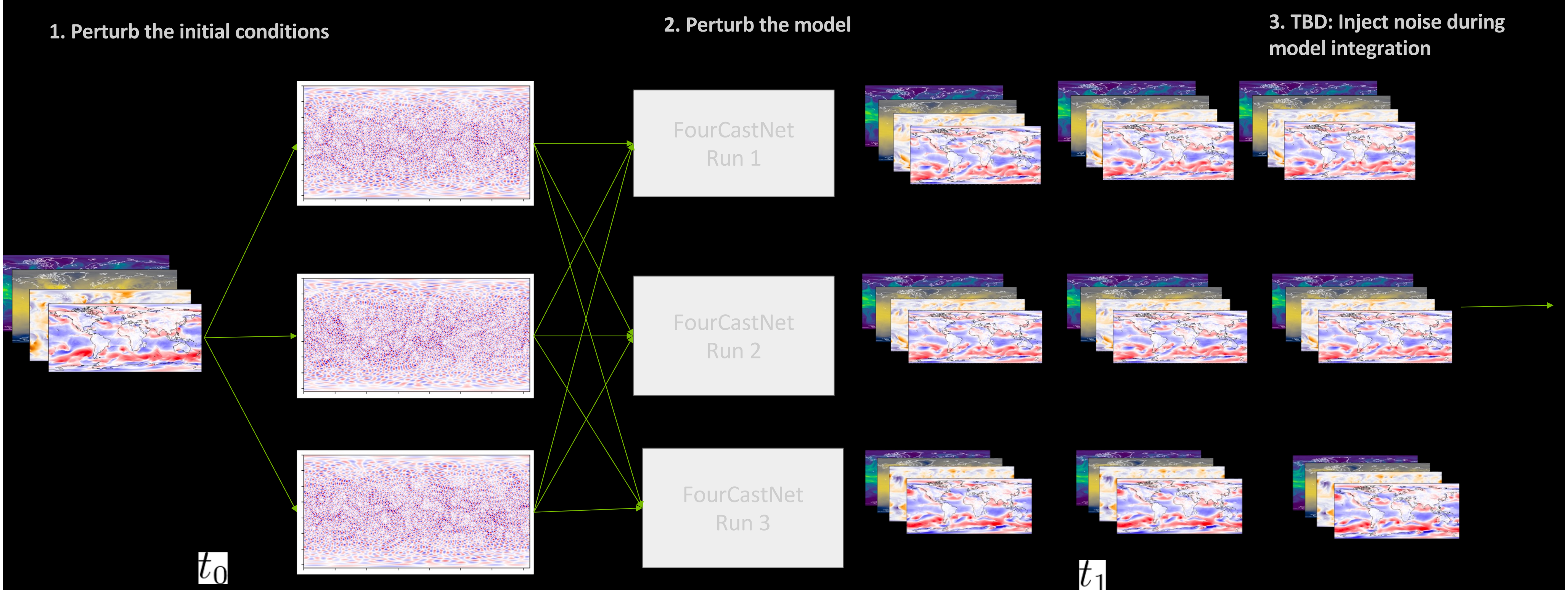
Validating HENS: We validate these ensembles on extremes using the same techniques as NWP

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

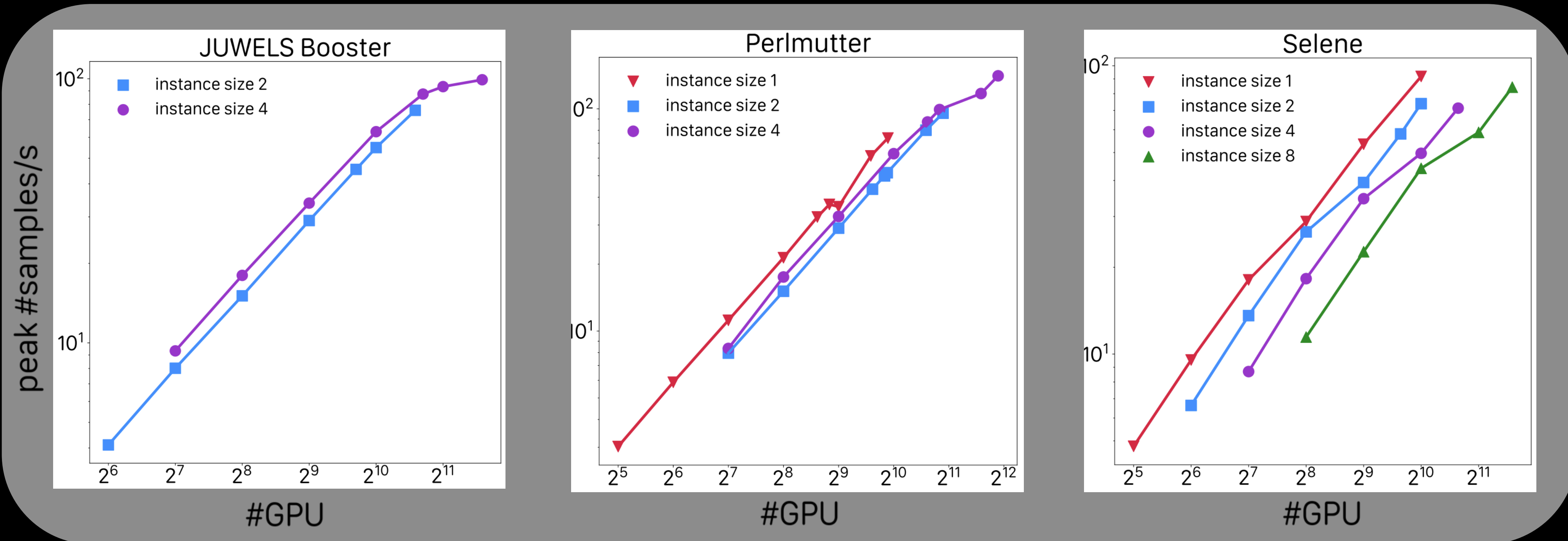
Constructing Huge Ensembles with FourCastNet



NVIDIA

FCN scales efficiently up to ~ 4000 GPUs on 3 HPC systems

NVIDIA's full-stack AI + HPC expertise shines thru... training on large amounts of the world's petabytes of past weather data.

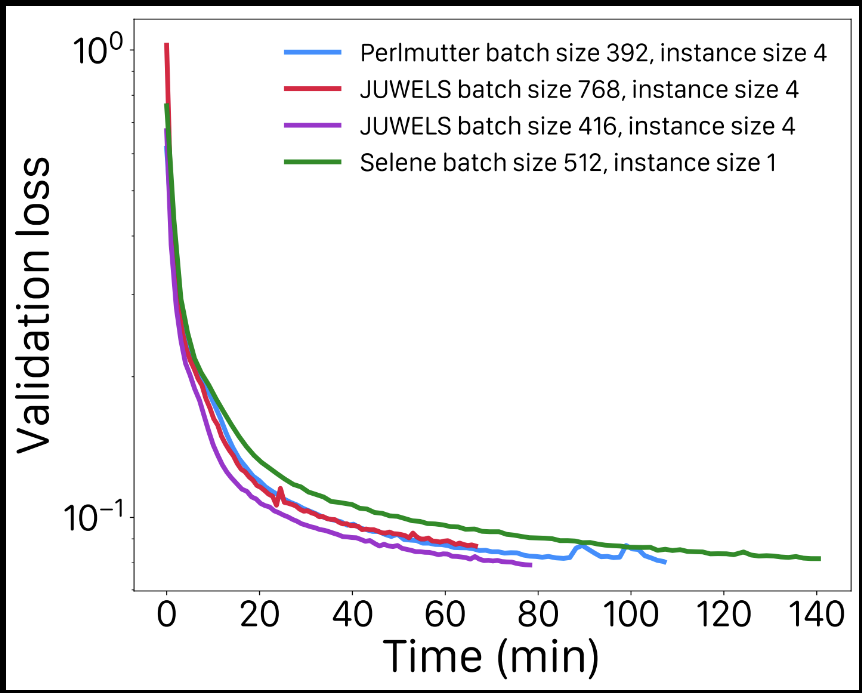


Peak performance is 140.8 petaFLOPS in mixed precision (averaged over a full epoch). Time to solution decreased from 24+ hours to 67 minutes with model and data parallelism

FourCastNet: Accelerating Global High-Resolution Weather Forecasting using Adaptive Fourier Neural Operators, Kurth et al. (2022), https://arxiv.org/abs/2208.05419

Another key scaling property: training in <1 day

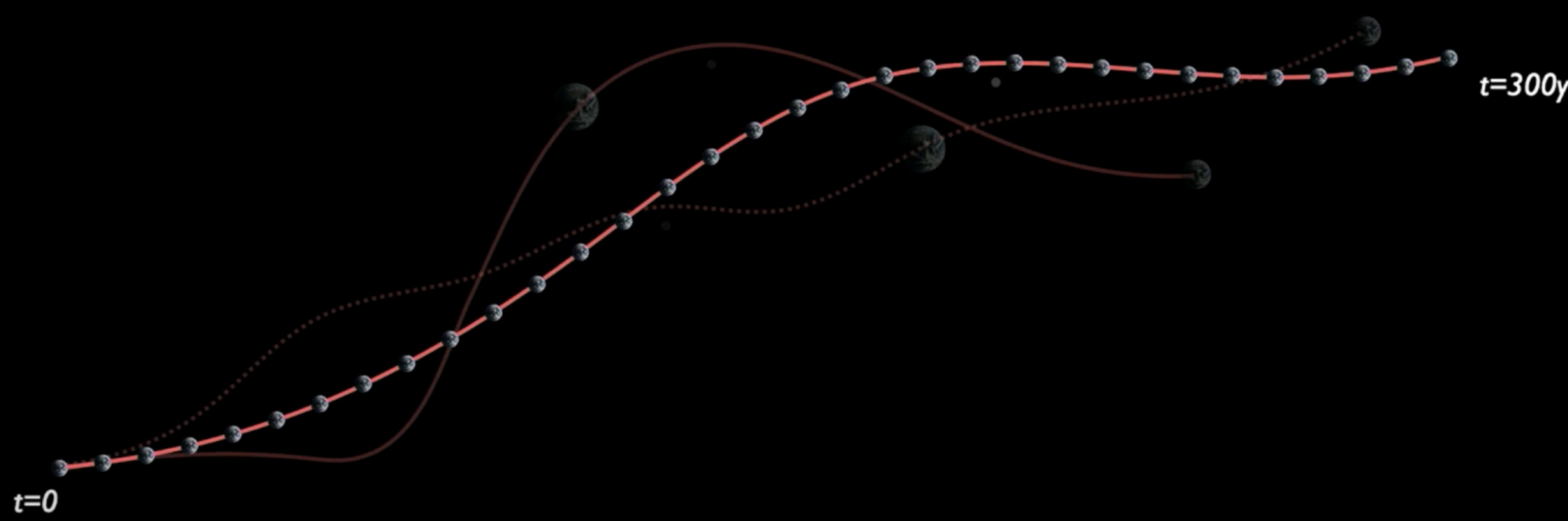
Model and data parallelism reduces training time from ~ 24 hours to 67.4 minutes



Given Future Data, FCN's Speed Yields Fast Tethering

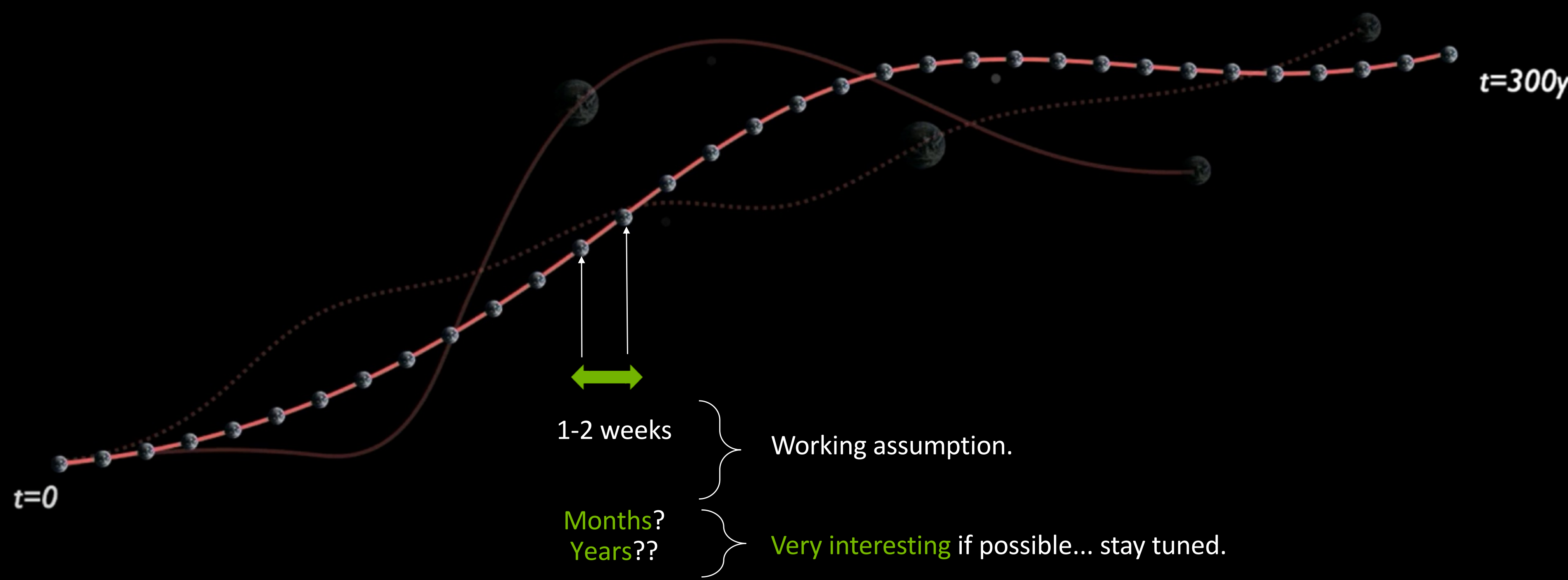
AI nimbly generates details between "checkpoints" saved only infrequently from physics-based climate simulations

-- Bjorn Stevens, GTC 2021



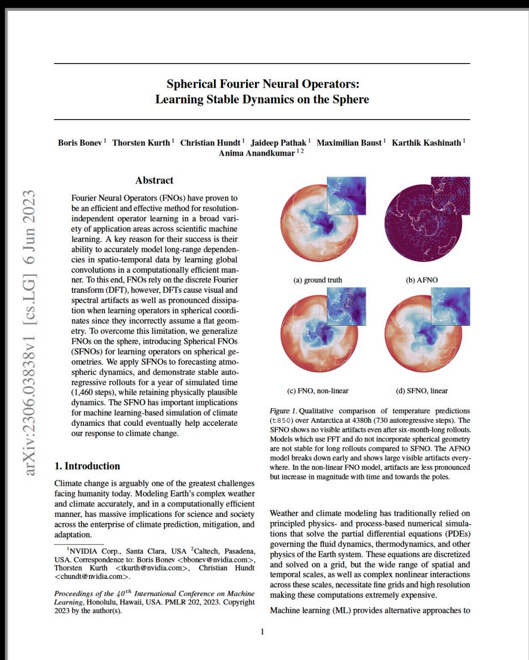
Game changer: Regenerate data rather than moving it

For how long can full-AI models like FourCastNet be trusted to "tether" between climate checkpoints?



Spherical harmonics enable stable, high-fidelity long rollouts.

Respecting spherical geometry key to maintaining robustness for auto-regressive rollout.



arXiv:2306.03838v1 [cs.LG] 6 Jun 2023

Spherical Fourier Neural Operators: Learning Stable Dynamics on the Sphere

Boris Bonev [1] Thorsten Kurth [1] Christian Hundt [1] Jaideep Pathak [1] Maximilian Baust [1] Karthik Kashinath [1] Anima Anandkumar [1 2]

Abstract

Fourier Neural Operators (FNOs) have proven to be an efficient and effective method for resolution-independent operator learning in a broad variety of application areas across scientific machine learning. A key reason for their success is their ability to accurately model long-range dependencies in spatio-temporal data by learning global convolutions in a computationally efficient manner. To this end, FNOs rely on the discrete Fourier transform (DFT), however, DFTs cause visual and spectral artifacts as well as pronounced dissipation when learning operators in spherical coordinates since they incorrectly assume a flat geometry. To overcome this limitation, we generalize FNOs on the sphere, introducing Spherical FNOs (SFNOs) for learning operators on spherical geometries. We apply SFNOs to forecasting atmospheric dynamics, and demonstrate stable autoregressive rollouts for a year of simulated time (1,460 steps), while retaining physically plausible dynamics. The SFNO has important implications for machine learning-based simulation of climate dynamics that could eventually help accelerate our response to climate change.

(a) ground truth (b) AFNO

(c) FNO, non-linear (d) SFNO, linear

Figure 1. Qualitative comparison of temperature predictions (`t850`) over Antarctica at 4380h (730 autoregressive steps). The SFNO shows no visible artifacts even after six-month-long rollouts. Models which use FFT and do not incorporate spherical geometry are not stable for long rollouts compared to SFNO. The AFNO model breaks down early and shows large visible artifacts everywhere. In the non-linear FNO model, artifacts are less pronounced but increase in magnitude with time and towards the poles.

1. Introduction

Climate change is arguably one of the greatest challenges facing humanity today. Modeling Earth's complex weather and climate accurately, and in a computationally efficient manner, has massive implications for science and society across the enterprise of climate prediction, mitigation, and adaptation.

Weather and climate modeling has traditionally relied on principled physics- and process-based numerical simulations that solve the partial differential equations (PDEs) governing the fluid dynamics, thermodynamics, and other physics of the Earth system. These equations are discretized and solved on a grid, but the wide range of spatial and temporal scales, as well as complex nonlinear interactions across these scales, necessitate fine grids and high resolution making these computations extremely expensive.

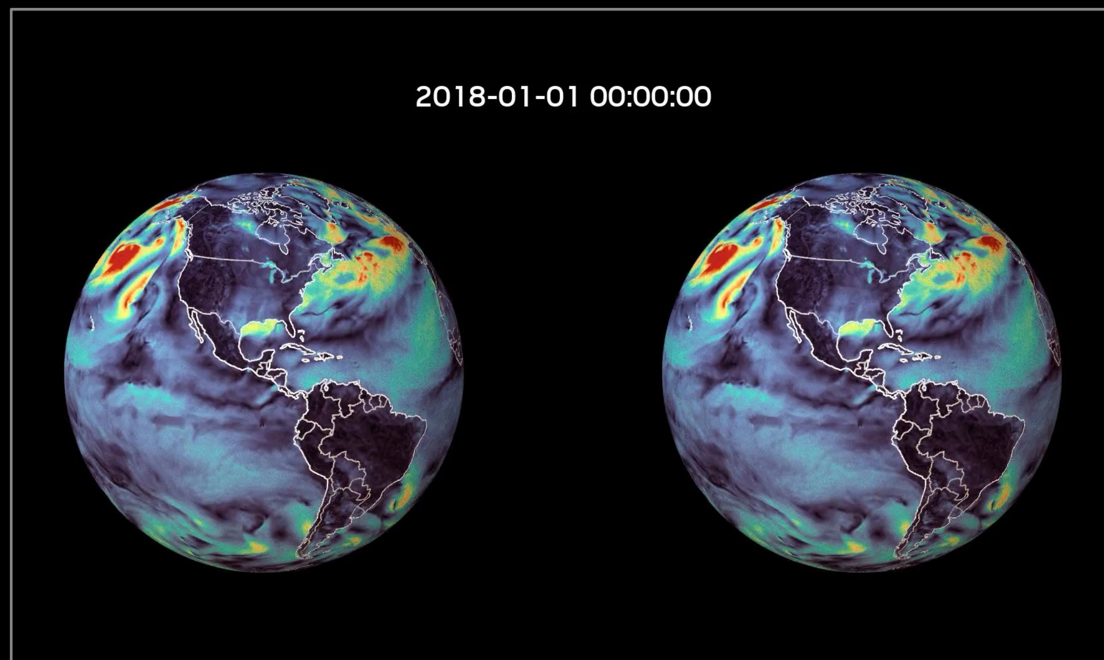
Machine learning (ML) provides alternative approaches to

[1]NVIDIA Corp., Santa Clara, USA [2]Caltech, Pasadena, USA. Correspondence to: Boris Bonev <bbonev@nvidia.com>, Thorsten Kurth <tkurth@nvidia.com>, Christian Hundt <chundt@nvidia.com>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

1

Bonev et al., ICML, 2023



* Stable one-year rollout (1460 autoregressive steps) computed in 13 minutes on a single NVIDIA RTX A6000

As of January 2024, our newest spherical FNO is making convincing S2S predictions and stable up to decades.

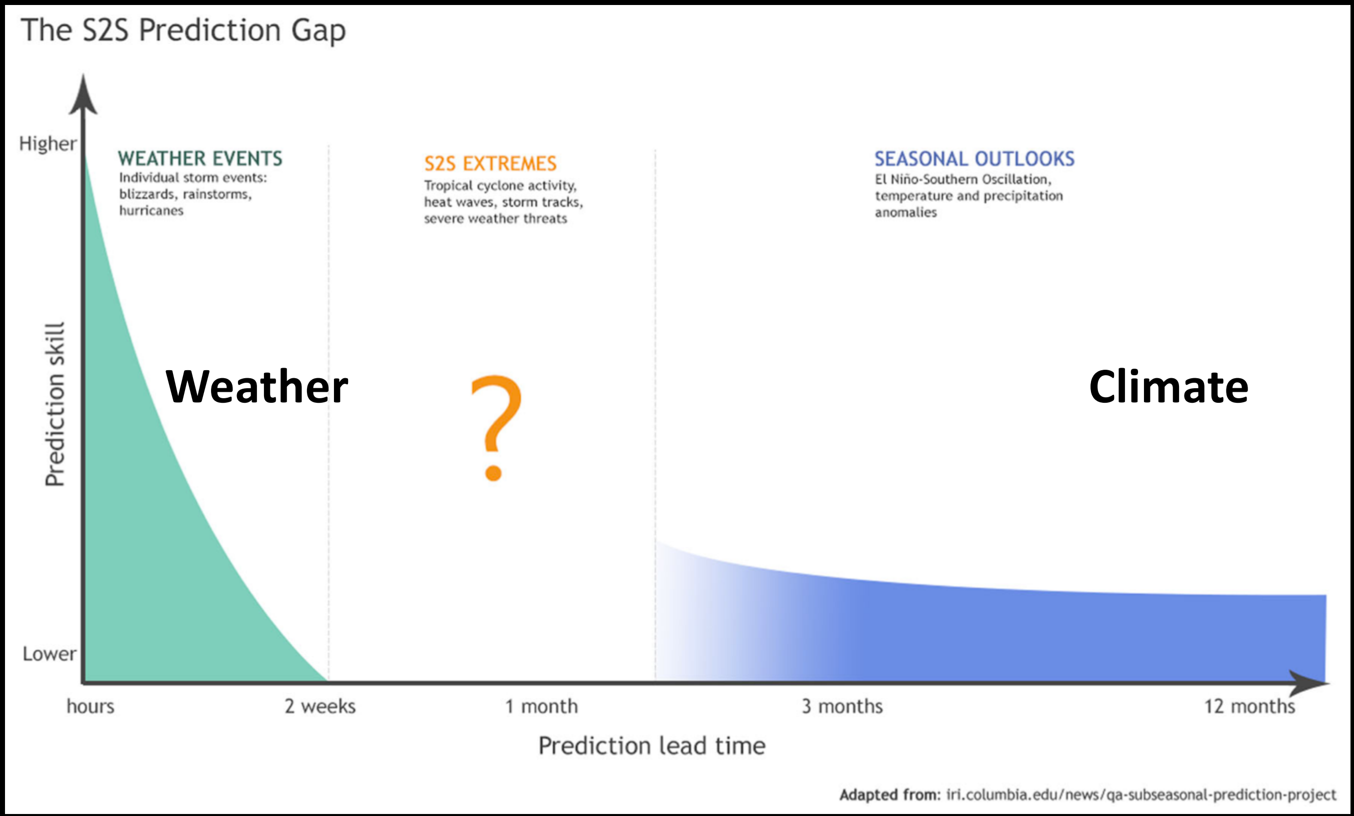
While not yet ready for real-world S2S forecasting this is a community milestone.

harmonics

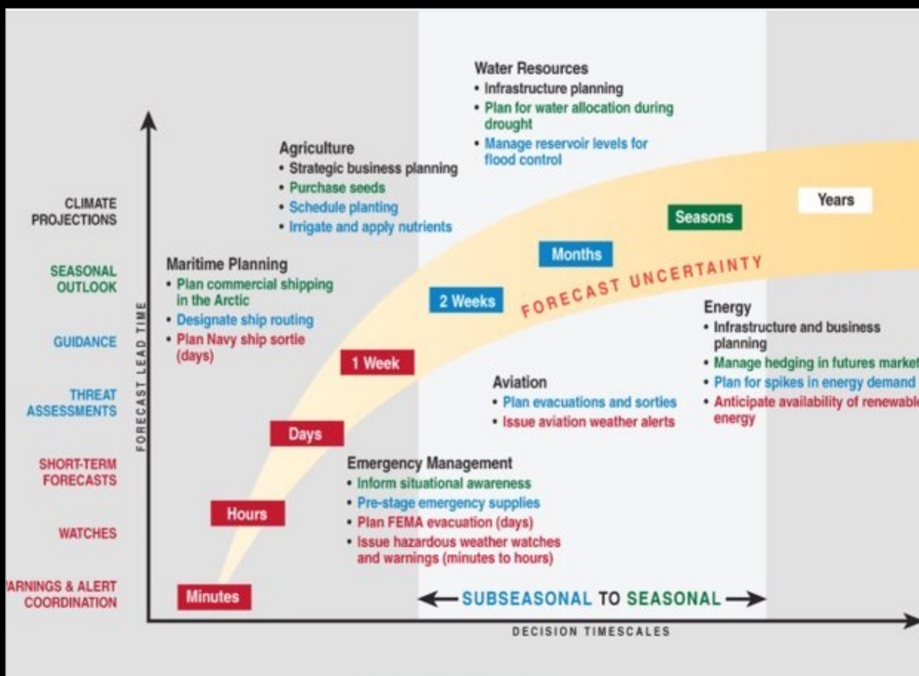
https://github.com/NVIDIA/torch-harmonics

NVIDIA

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



A critical timescale with immense value: sub-seasonal ensemble forecasting



- Iterating on & resolving current pathologies.
- In partnership with operational NWP centers:
  - ECMWF & NOAA
- Exploring data assimilation, optimal initialization.
- Generative post-processing.

Source: National Academies, "Next-generation Earth System Prediction", https://nap.nationalacademies.org/catalog/21873/next-generation-earth-system-prediction-strategies-for-subseasonal-to-seasonal

Towards stable, accurate, high-fidelity rollouts for decadal climate

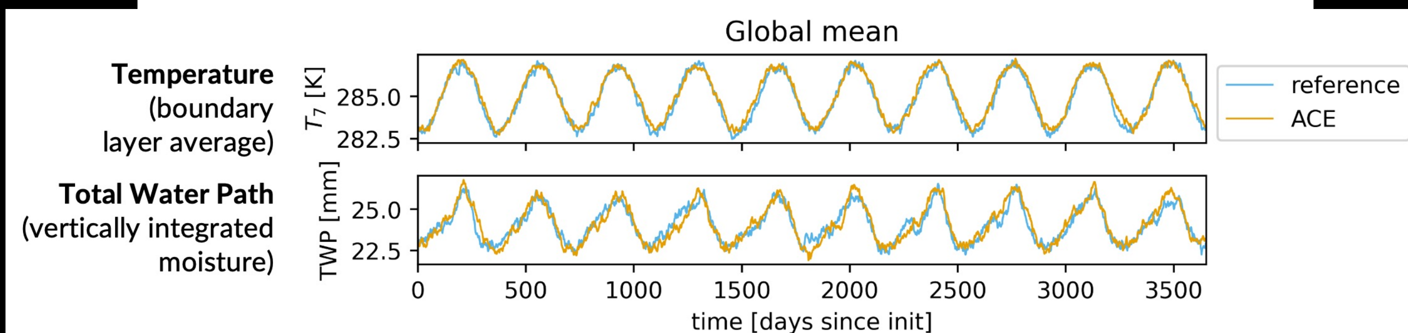
Enabled by spherical harmonics, equivariance, and physical constraints

Physics > Atmospheric and Oceanic Physics

[Submitted on 3 Oct 2023 (v1), last revised 6 Dec 2023 (this version, v2)]

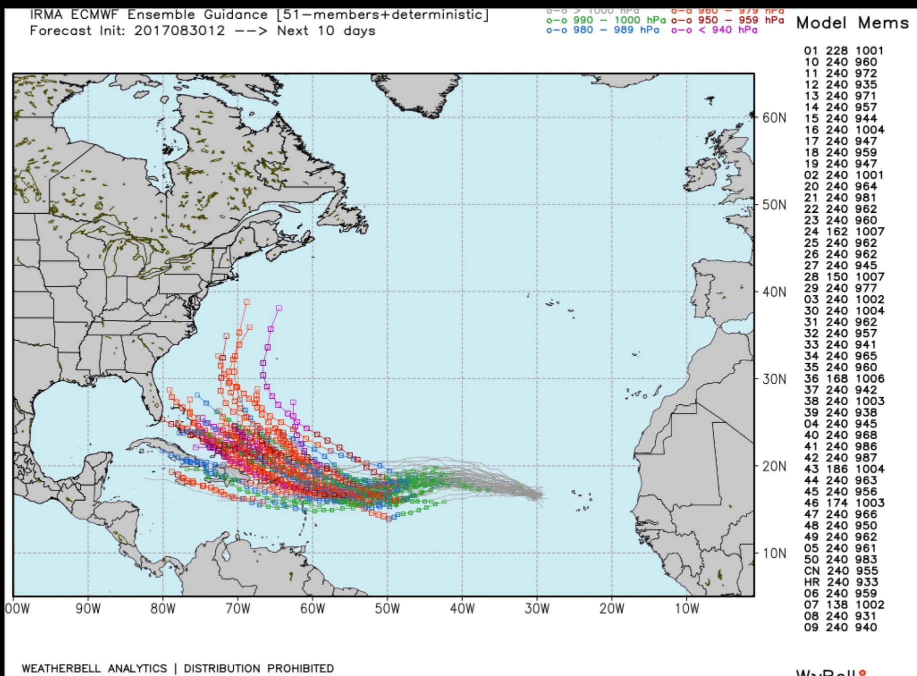
ACE: A fast, skillful learned global atmospheric model for climate prediction

Oliver Watt-Meyer, Gideon Dresdner, Jeremy McGibbon, Spencer K. Clark, Brian Henn, James Duncan, Noah D. Brenowitz, Karthik Kashinath, Michael S. Pritchard, Boris Bonev, Matthew E. Peters, Christopher S. Bretherton



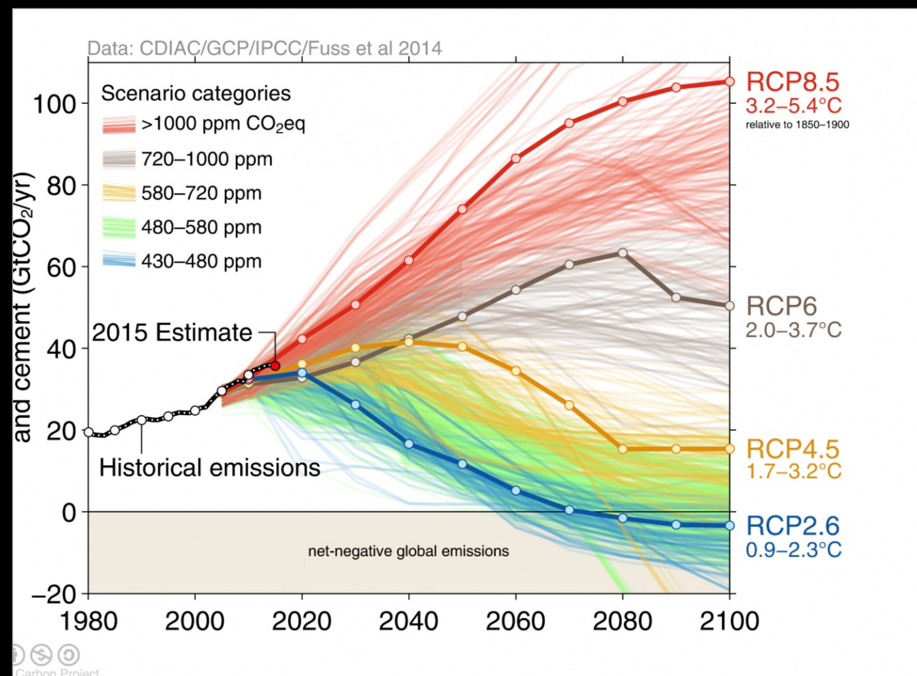
FCN will sample future extreme-weather for decades

For more accurate statistics, to capture rare extremes, and identify significant forecast-bifurcations



Sample Weather Ensembles Over Time

ECMWF



Sample across many possible climate futures

IPCC 2014

Towards kilometer-scale emulation

To meet the needs of society:
(i) predicting impacts at scales that matter
(ii)interacting with data at low-latency

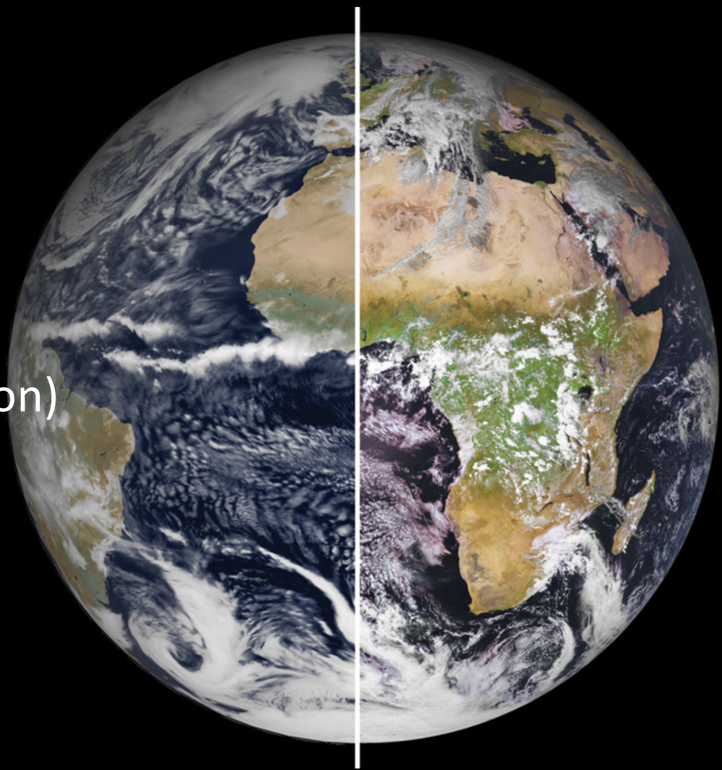
Model: FourCastNet ~ 2.5B Parameter FCN (5km Resolution)

System: 4K H100

Training: Full State Vector, 10 Years > 5 PB > ~3 Days

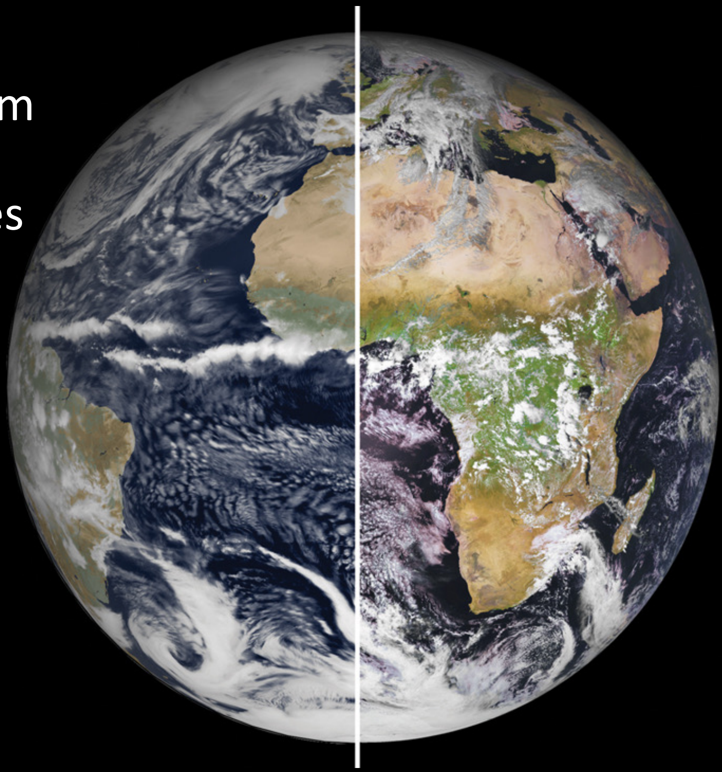
Inference: 30 Days, 1000-member ensemble > ~1 Hour

O(1,000)X Speed-Up Versus Simulation



Next Steps and Future Directions

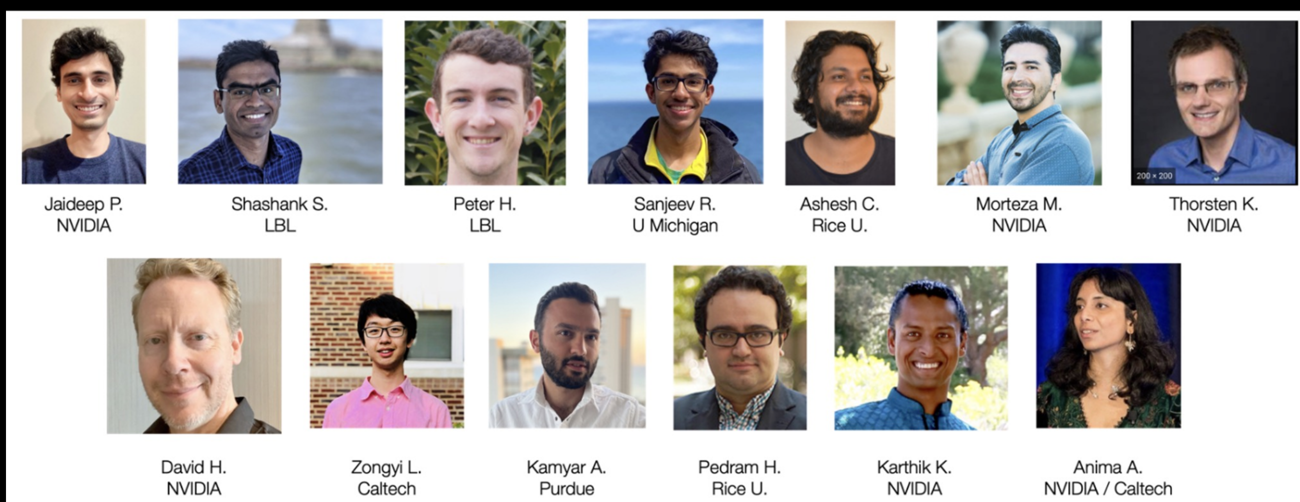
- Extending emulators to all components of the Earth system
- Developing extrapolation methods to “no analog” climates
- Introducing data-driven climate prediction to the IPCC (Intergovernmental Panel on Climate Change)
- Integrating data-driven weather prediction with existing operational forecasting
- Demonstrating the real-world benefits from emulation



NVIDIA

Thanks to all collaborators!

Ankur Mahesh



Mike Pritchard

Data-Driven Models Revolutionize Environmental Prediction

At the dawn of an entirely new way to simulate the global atmosphere

Questions?

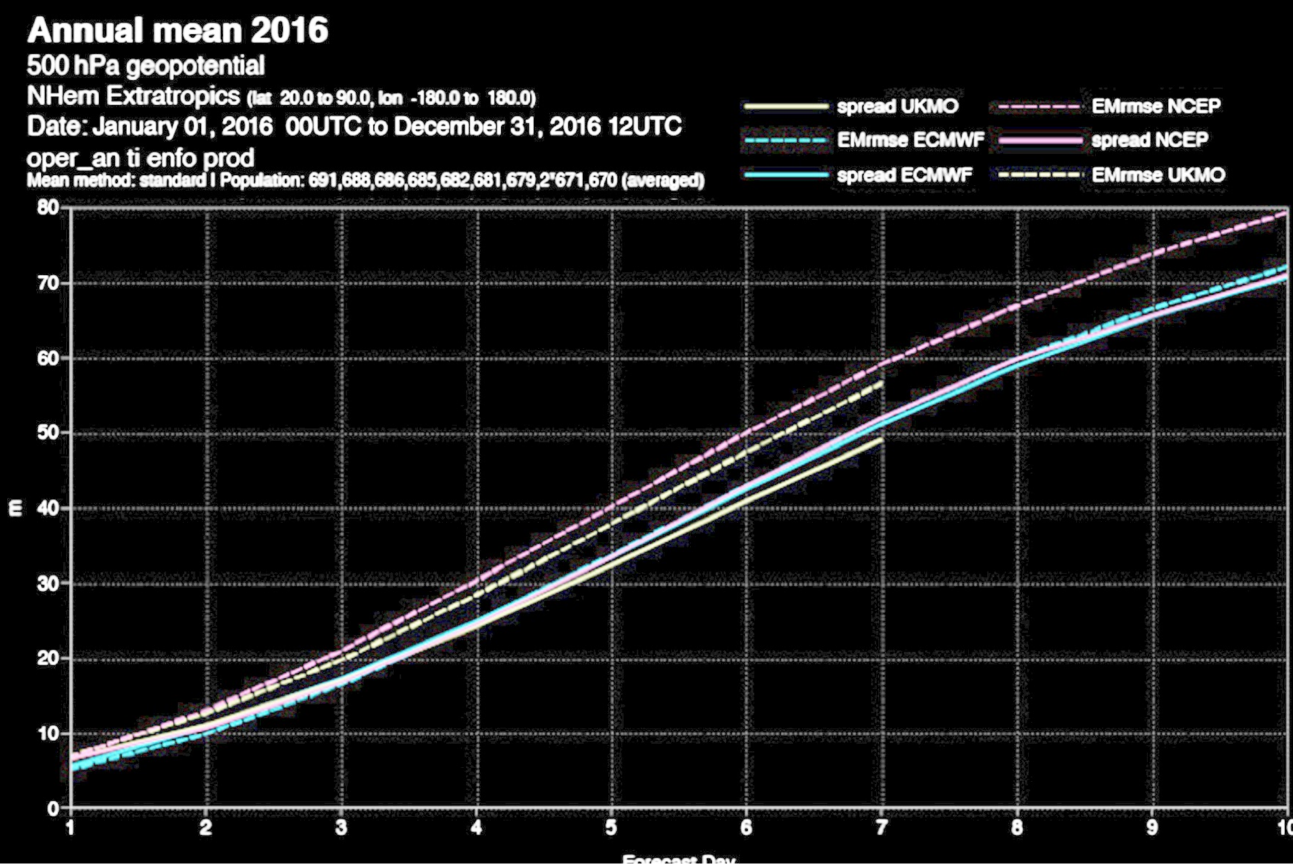
Feedback?

Paths Forward?

BACKUP SLIDES

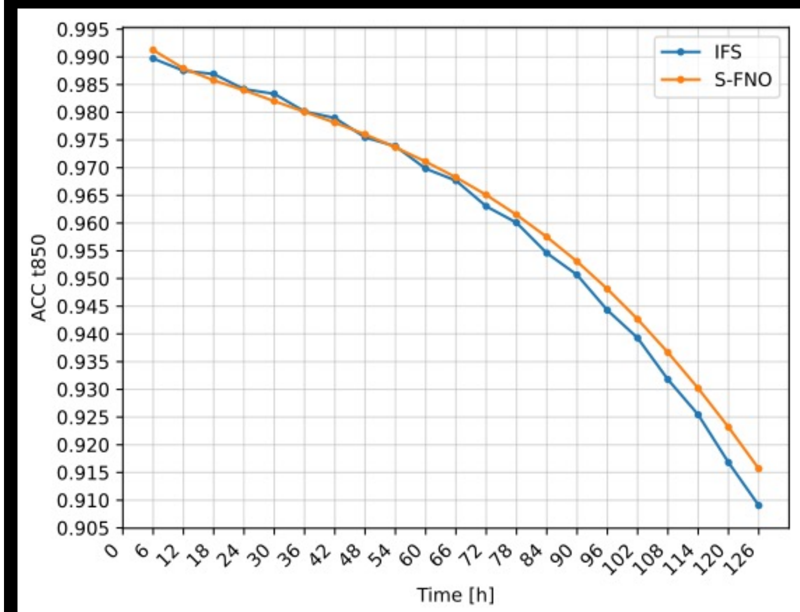
The spread of an ensemble prediction should grow at the same rate as its error

The European Center for Medium Range Weather Forecasting (ECMWF) is respected for its well-calibrated ensembles

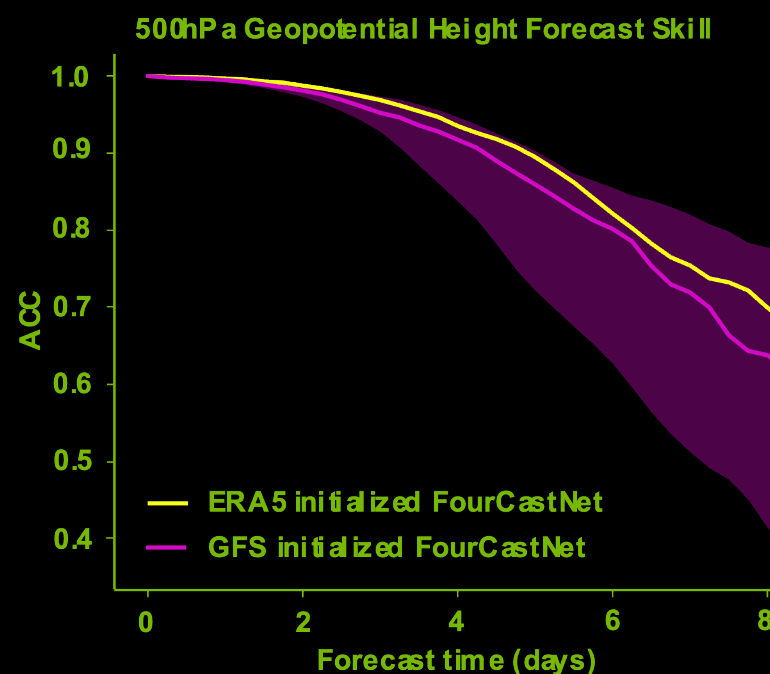


FCN medium-range forecast skill comparable to IFS

And can be initialized with GFS initial conditions, proving zero-shot transfer learning.



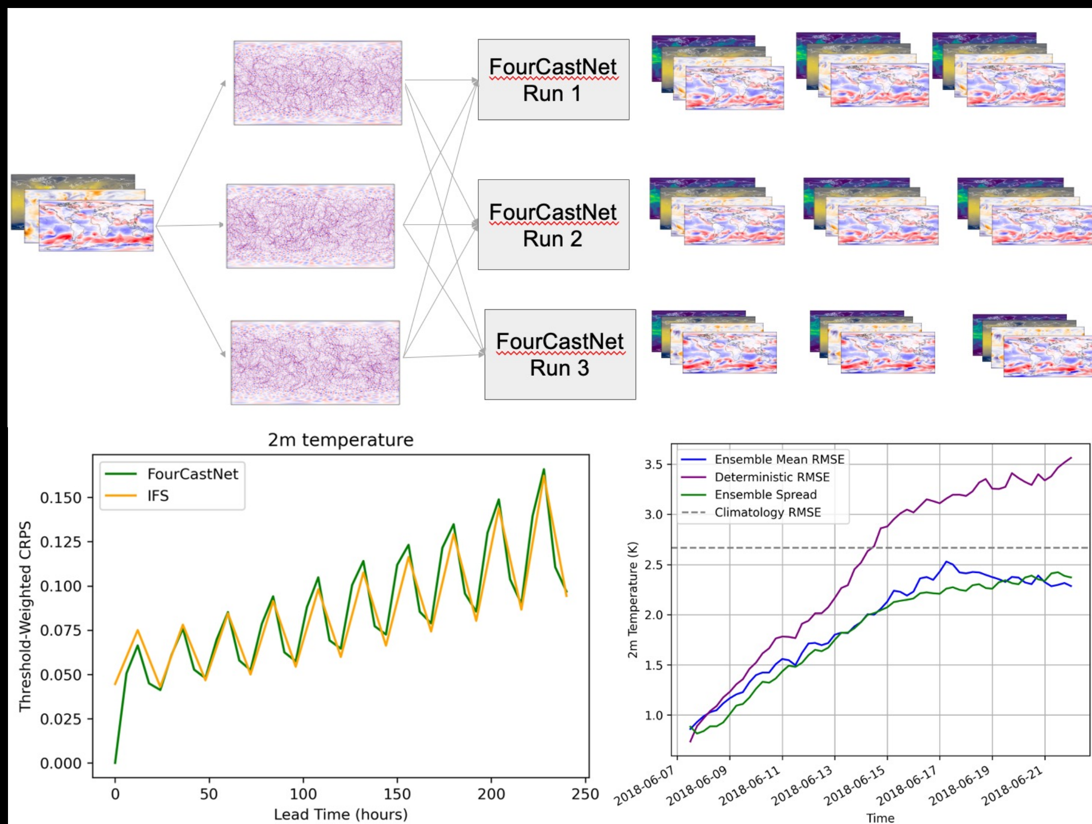
t2m	3 days 00:00:00	0.02
	7 days 00:00:00	-0.00
	14 days 00:00:00	0.00
t850	3 days 00:00:00	-0.03
	7 days 00:00:00	-0.04
	14 days 00:00:00	-0.03
u10m	3 days 00:00:00	-0.09
	7 days 00:00:00	-0.06
	14 days 00:00:00	-0.06
v10m	3 days 00:00:00	-0.09
	7 days 00:00:00	-0.06
	14 days 00:00:00	-0.05
z500	3 days 00:00:00	0.09
	7 days 00:00:00	0.01
	14 days 00:00:00	-0.02



Modeled on the ECMWF scorecard
Green is better, Blue is worse.

FourCastNet passes key tests for extreme forecasts

FCN's ensembles calibrated using initial condition uncertainty and model uncertainty drawing on NWP wisdom



Ankur Mahesh