NVIDIA

Earth-2: Towards kilometer-scale Digital Twins for Weather and Climate

Karthik Kashinath, Principal Engineer and Scientist, AI-HPC

Co-Lead, NVIDIA Earth-2 Initiative

Agenda

- NVIDIA's Earth-2 initiative: The Big Picture
- FourCastNet & latest accomplishments in Weather
- Towards kilometer-scale emulation
- Technology platforms

Realistic weather and climate simulation is a computational grand challenge

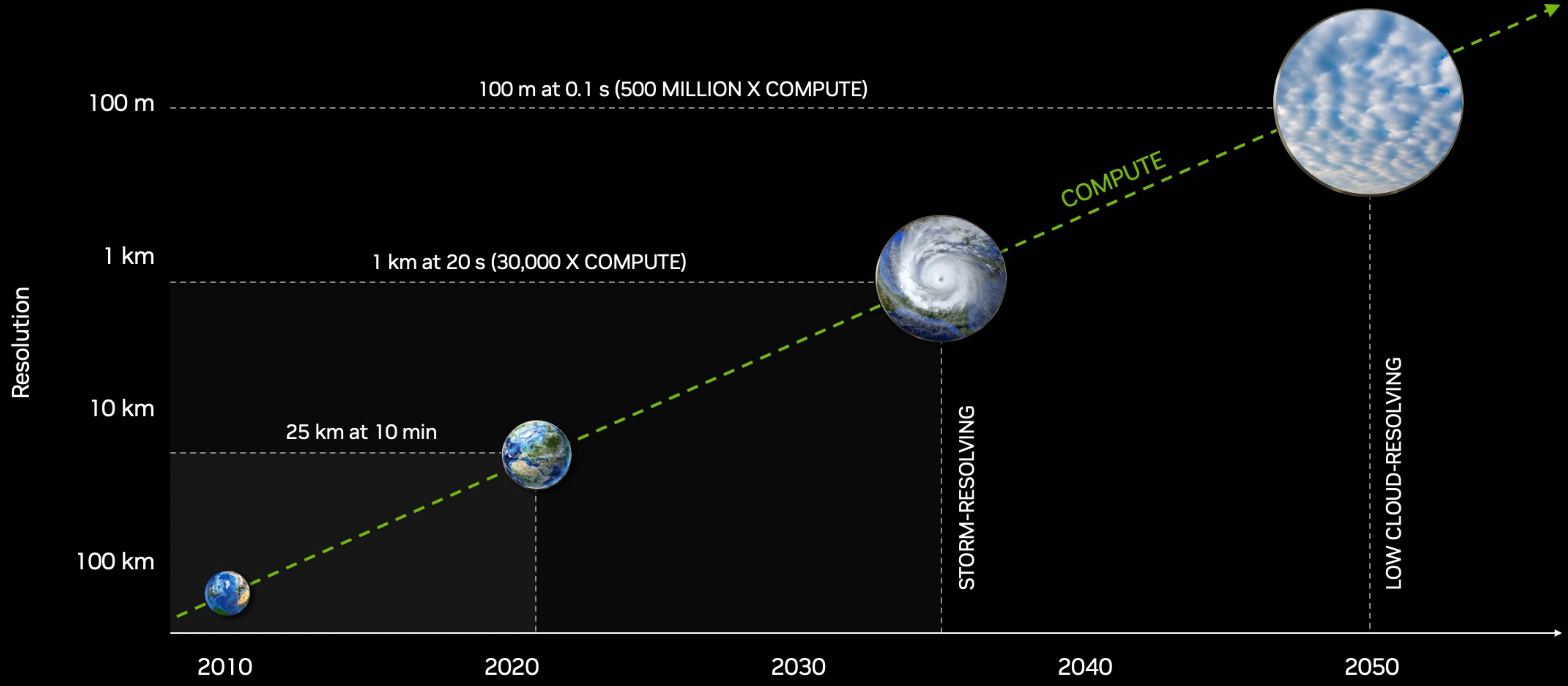
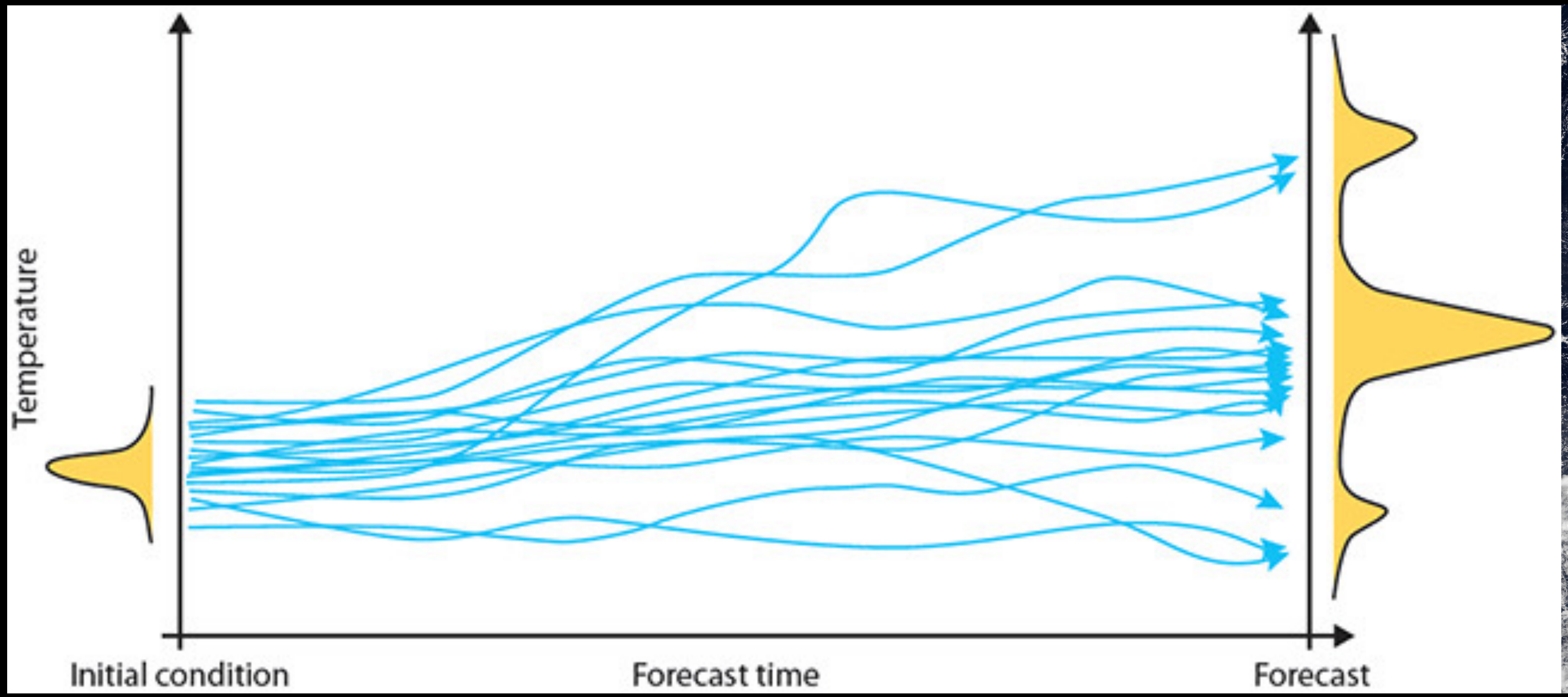
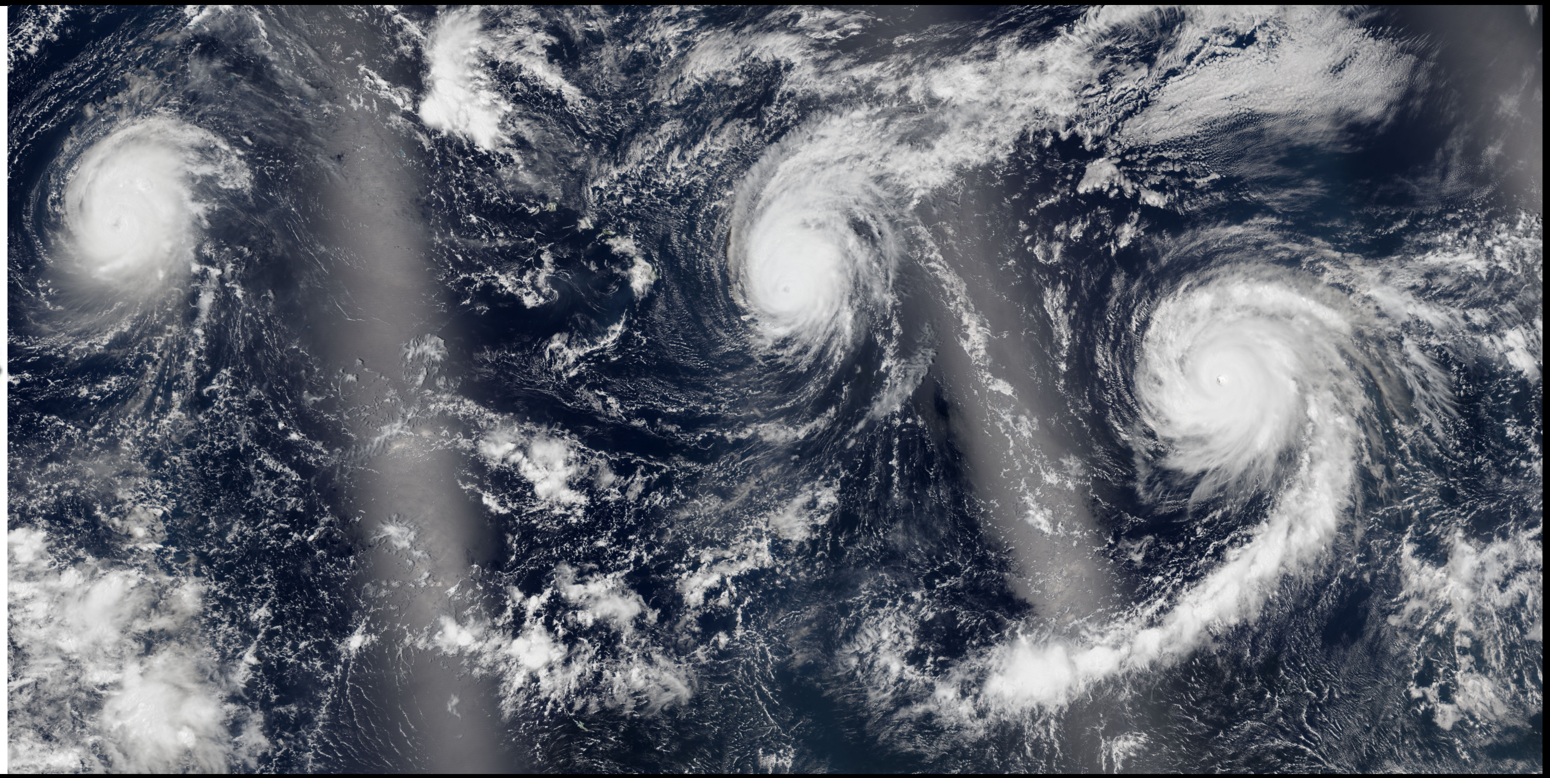


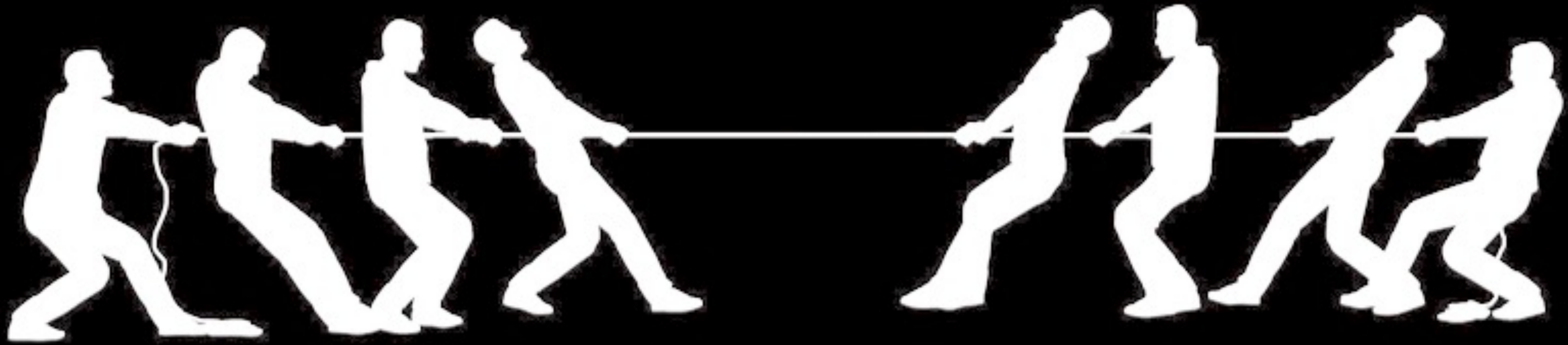
Figure adapted from: Schneider, T., Teixeira, J., Bretherton, C. *et al.* "*Climate goals and computing the future of clouds*". *Nature Climate Change* **7,** 3–5 (2017)

Throughout history, ensemble size of weather and climate predictions has been limited.

Under computational constraints, ensemble size trades off against horizontal resolution







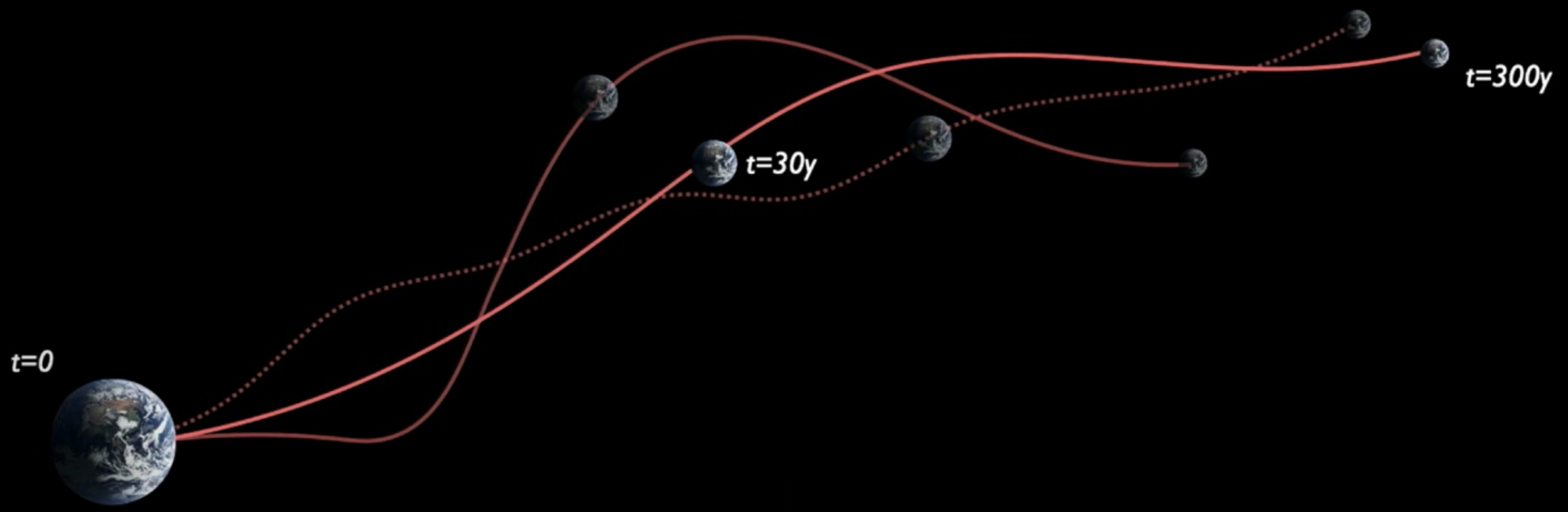
Ensemble size

Resolution

Exploding Data Volumes in High-Resolution Climate Prediction

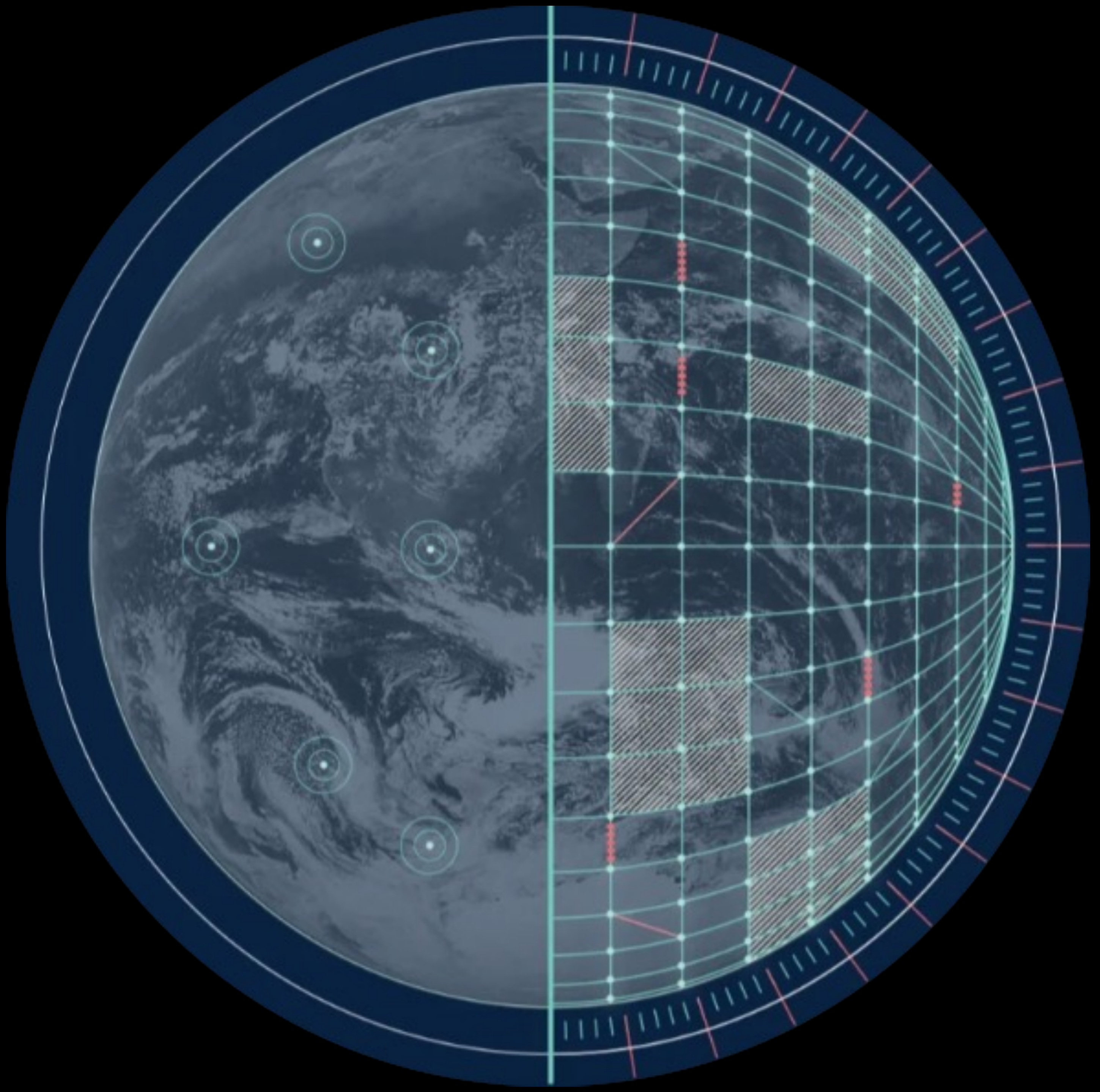
"We can compute km-scale predictions, but can't effectively extract information content, let alone interact with it"

-- Prof. Dr. Bjorn Stevens.



EARTH-2: A highly interactive weather and climate information system

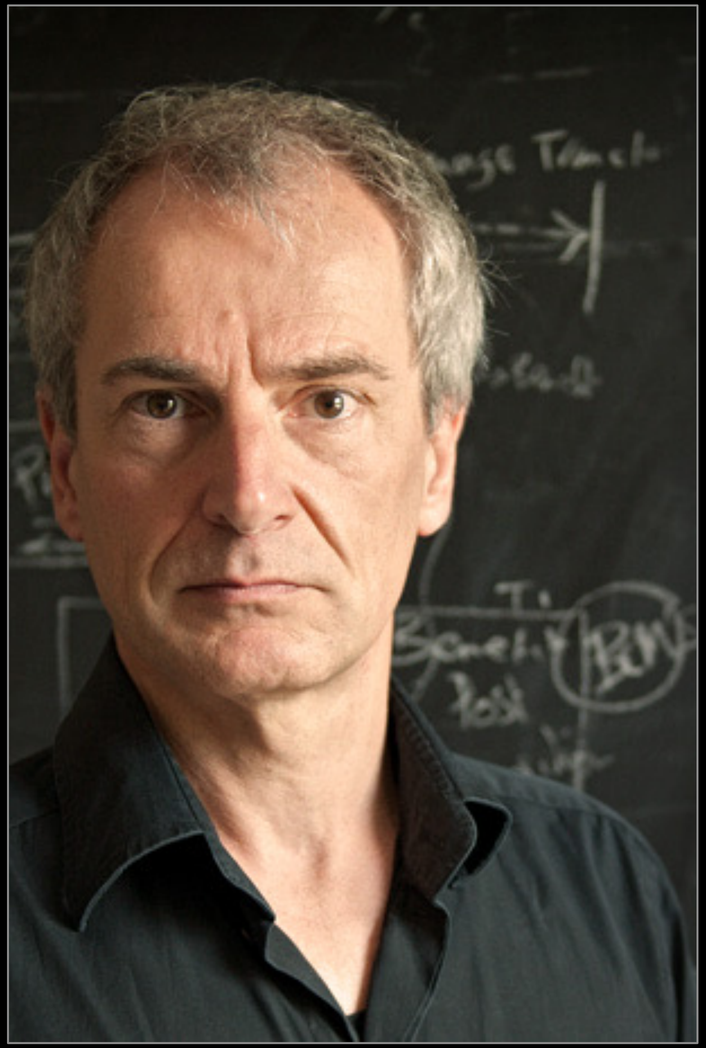
For serving society with next-gen weather and climate predictions.



Drawing on Destination Earth: Bauer, Stevens, Hazeleger. "*A Digital Twin of Earth for the Green Transition*". *Nature Climate Change* **11,** 80–83 (2021)

The vision of Earth-2 shaped by world-leading scientists

Thought-leading weather & climate visionaries, champions of interactive digital twins.



Peter Bauer

ECMWF

Peter Dueben

ECMWF

Bjorn Stevens

MPI-Hamburg

Francisco Doblas-Reyes

Barcelona Supercomputing Center

Thomas Schulthess

ETHZ and CSCS

Earth-2 is in Collaboration with International Weather and Climate Science

NVIDIA's AI, engineering & full-stack expertise complement research capacity in academia & government.

MAX-PLANCK-INSTITUT FÜR METEOROLOGIE

ECMWF
EUROPEAN CENTRE FOR MEDIUM RANGE WEATHER FORECASTS

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

NASA

NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE

DEPARTMENT OF ENERGY
UNITED STATES OF AMERICA

USMILE

Ai2

ARTIFICIAL INTELLIGENCE · ENVIRONMENTAL SCIENCE · RISK COMMUNICATION

LEAP

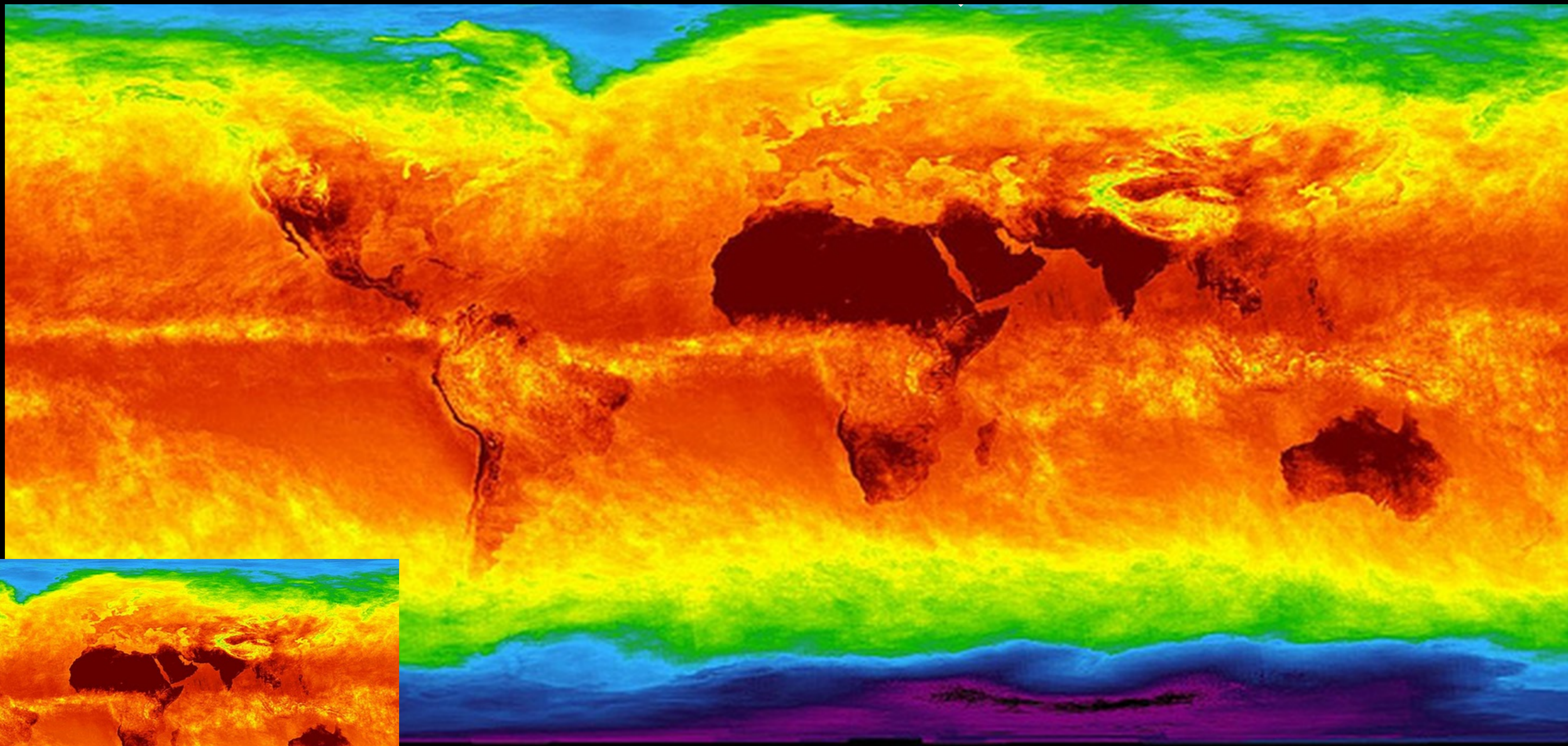
MeteoSwiss

Agenda

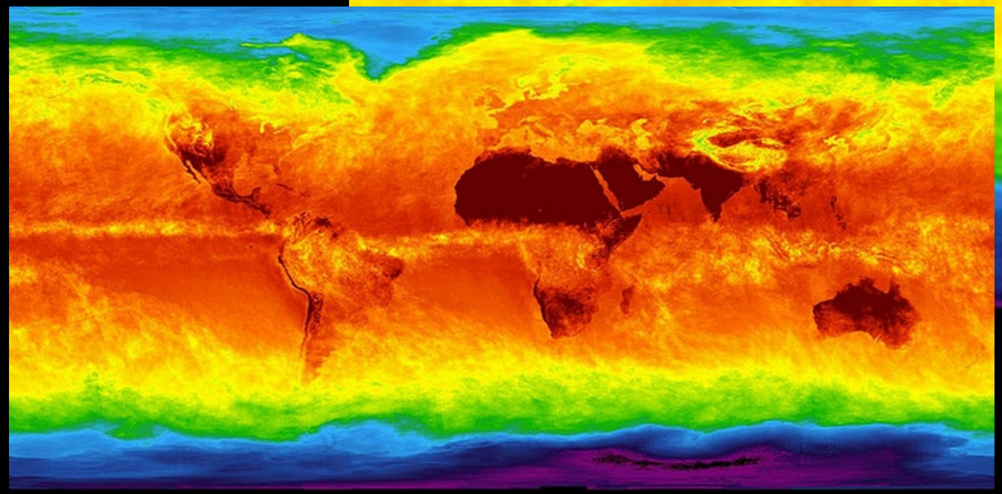
- NVIDIA's Earth-2 initiative: The Big Picture
- FourCastNet & latest accomplishments in Weather
- Towards kilometer-scale emulation
- Technology platforms

The *not-so-quiet* revolution in data-driven weather prediction!

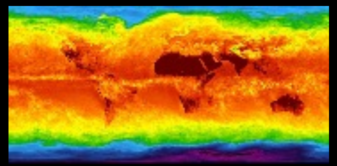
Staggering ~500x increase in resolution in 5 years



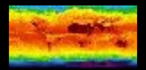
FourCastNet, Pathak et al. (2022), 0.25°, ~1,000,000 Pixels, ViT+AFNO



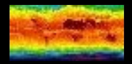
GNN, Keisler et al. (2022), 1°, 64,000 Pixels, Graph Neural Networks



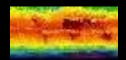
DLWP, Weyn et al. (2020). 2°, 16K pixels, Deep CNN on Cubesphere/(2021) ResNet



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



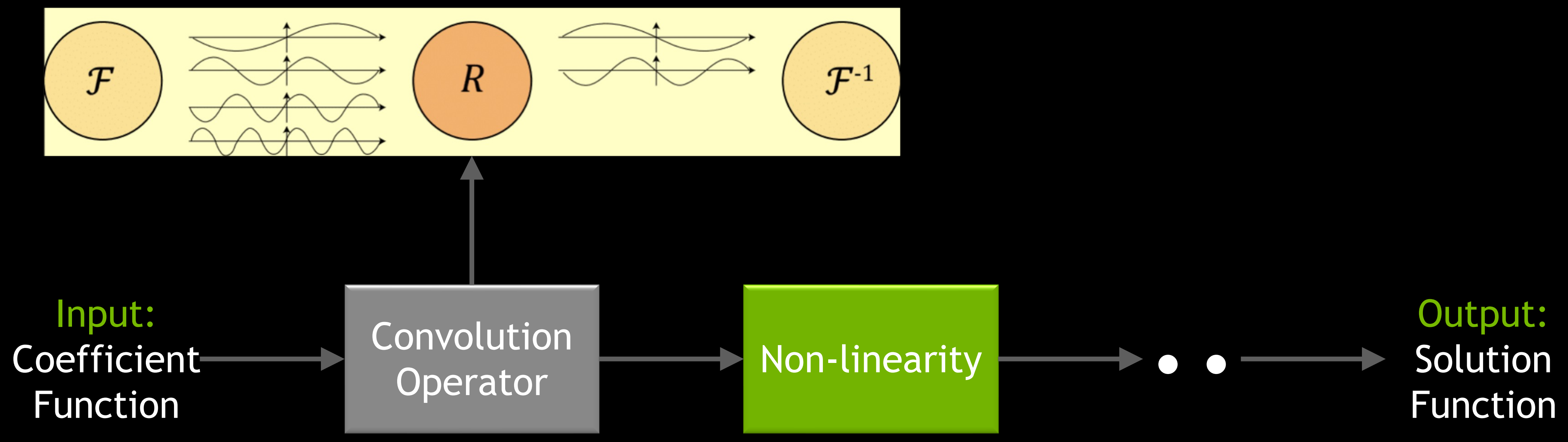
WeatherBench, Rasp et al. (2020). 5.625°, 64x32, 2K pixels, CNN



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

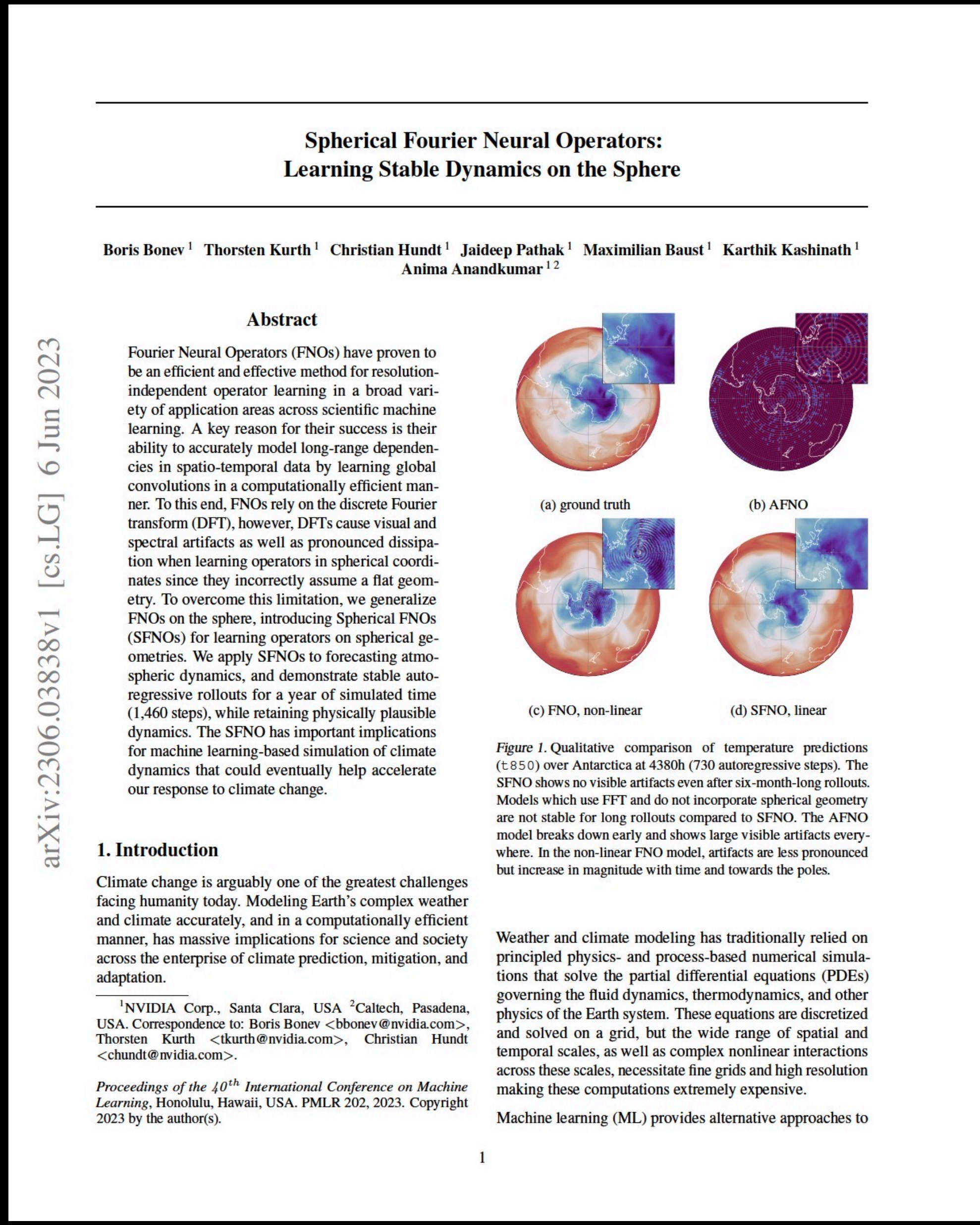
FourCastNet uses the Spherical Fourier Neural Operator (SFNO)

Spherical harmonics enable stable, high-fidelity long rollouts.



harmonics

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



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.

[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 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

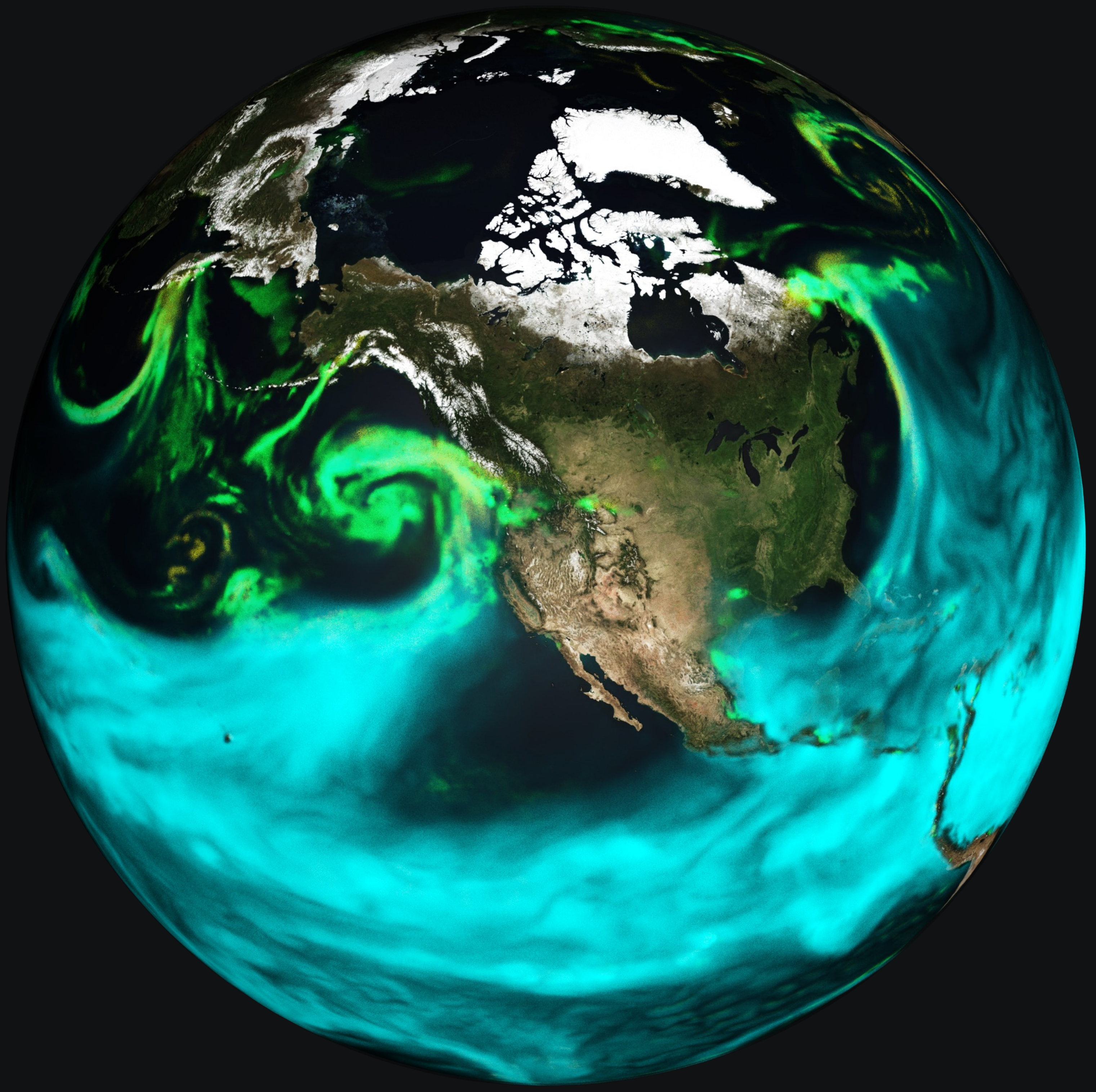
Weather and climate modeling has traditionally relied on principled physics- and process-based numerical simulations that solve the partial differential equations (PDEs) governing the fluid dynamics, thermodynamics, and other physics of the Earth system. These equations are discretized and solved on 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

- 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, …

ICML 2023 (https://arxiv.org/abs/2306.03838): Bonev et al., Spherical Fourier Neural Operators: Learning Stable Dynamics on the Sphere

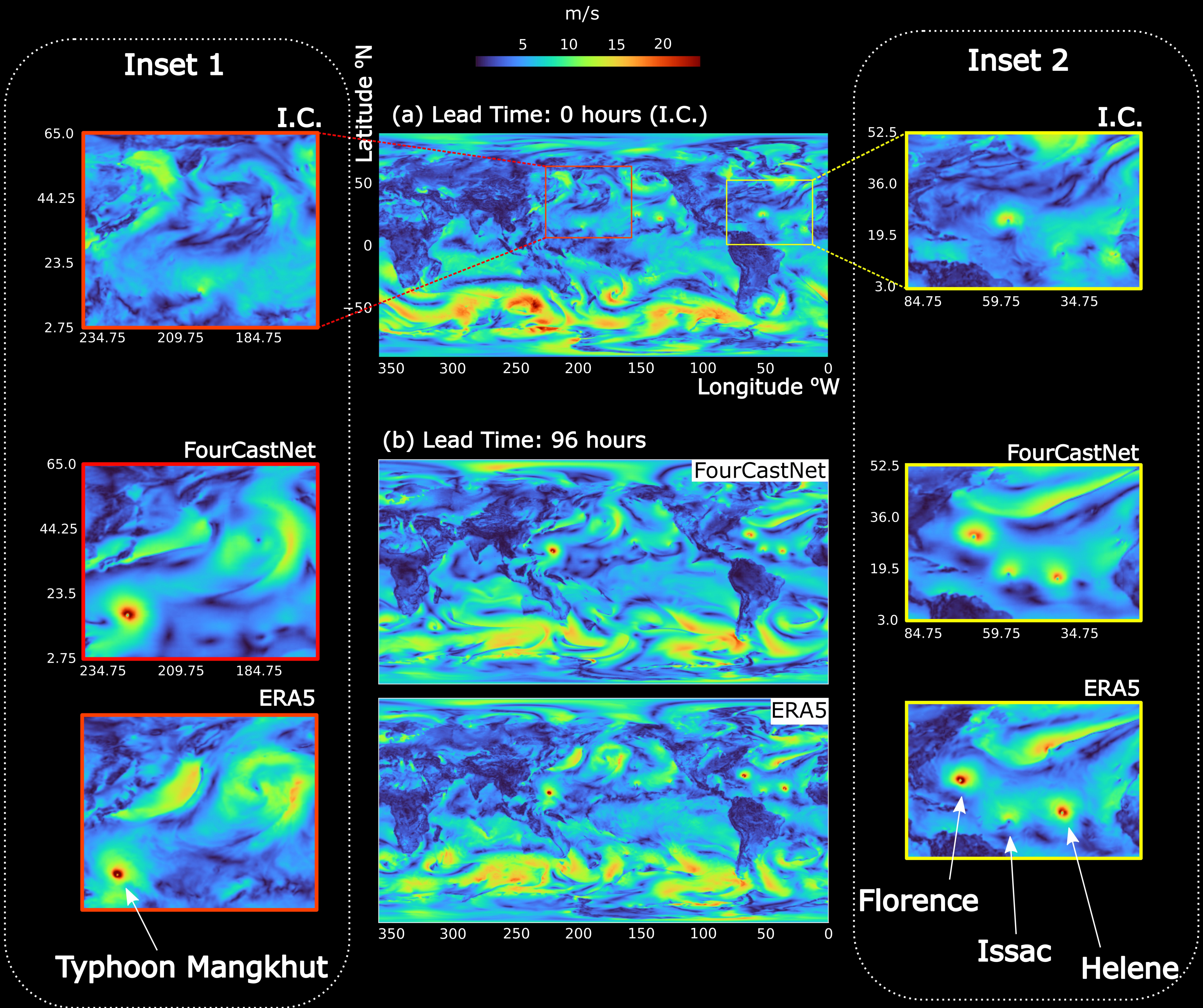


FourCastNet

Scope	Global
Model Type	Full-Atmosphere AI Surrogate
Architecture	Fourier Neural Operator
Resolution	25km, 6-hourly (up to 10km, 1-hourly)
Training Data	ERA5 Reanalysis
Initial Condition	ERA5 / GFS / UFS
Training Time	O(1000) GPU-hours (NVIDIA A100 80GB)
Inference Time	3 sec (10-day forecast)
Calibration	Initial condition + Bayesian model uncertainty
Speedup vs NWP	O(1,000 – 10,000)
Power Savings	O(1,000)
Max Stable Rollout	Years
Scalability	Over 90% up to 4000 GPUs, 140 petaFLOPS
Access	Open-source

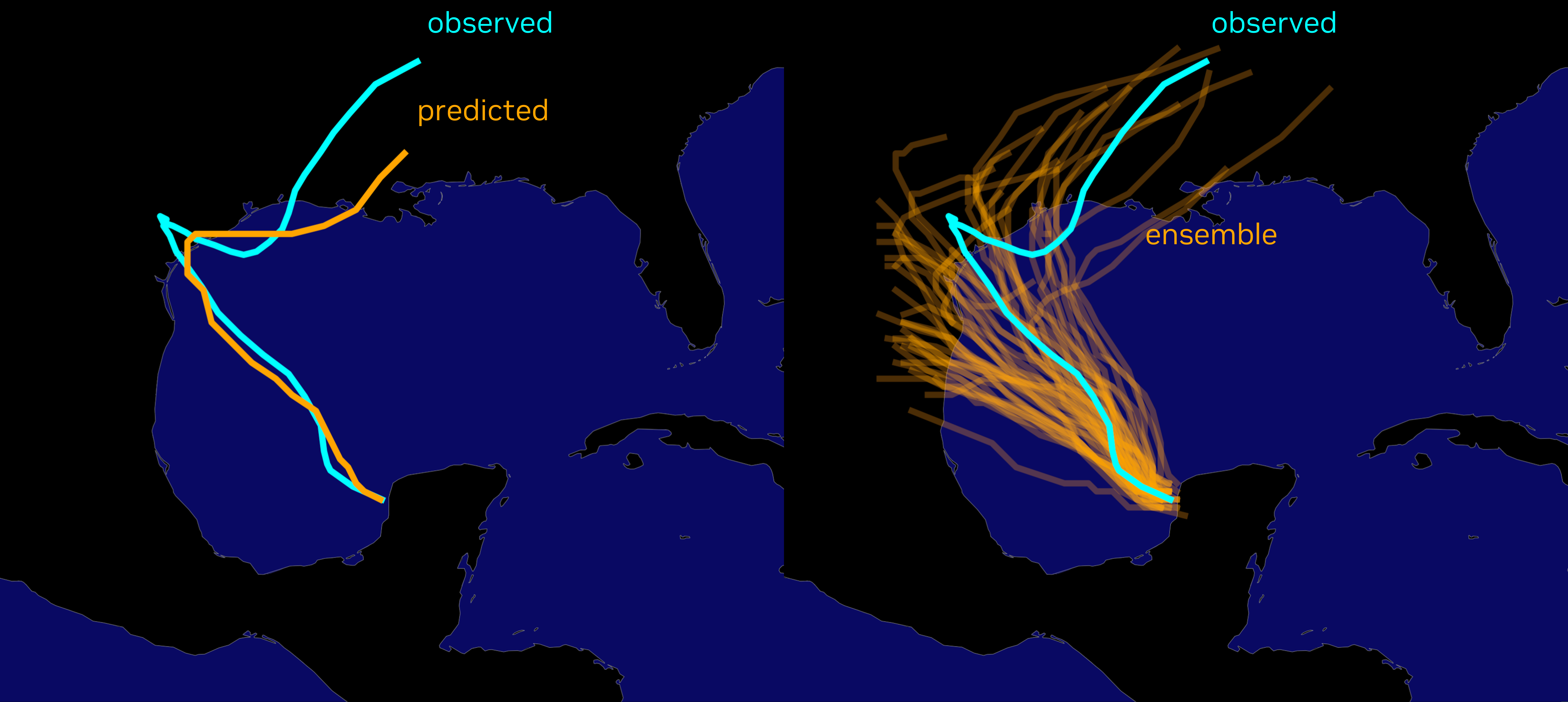
FCN forecasts extremes with high fidelity.

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



FCN's speed enables massive ensembles.

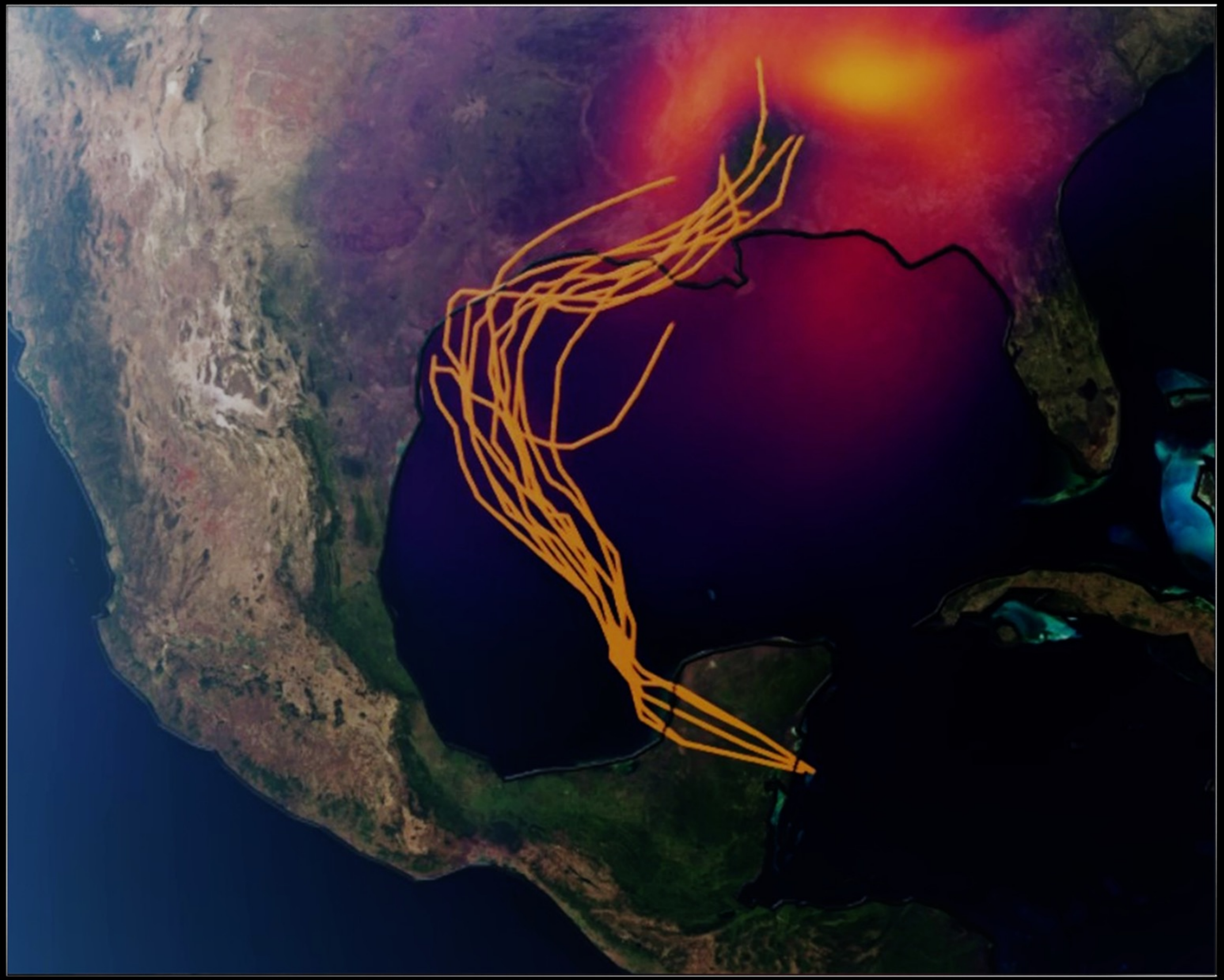
To capture low-likelihood high-impact extreme events more accurately – far into the long tails of distributions



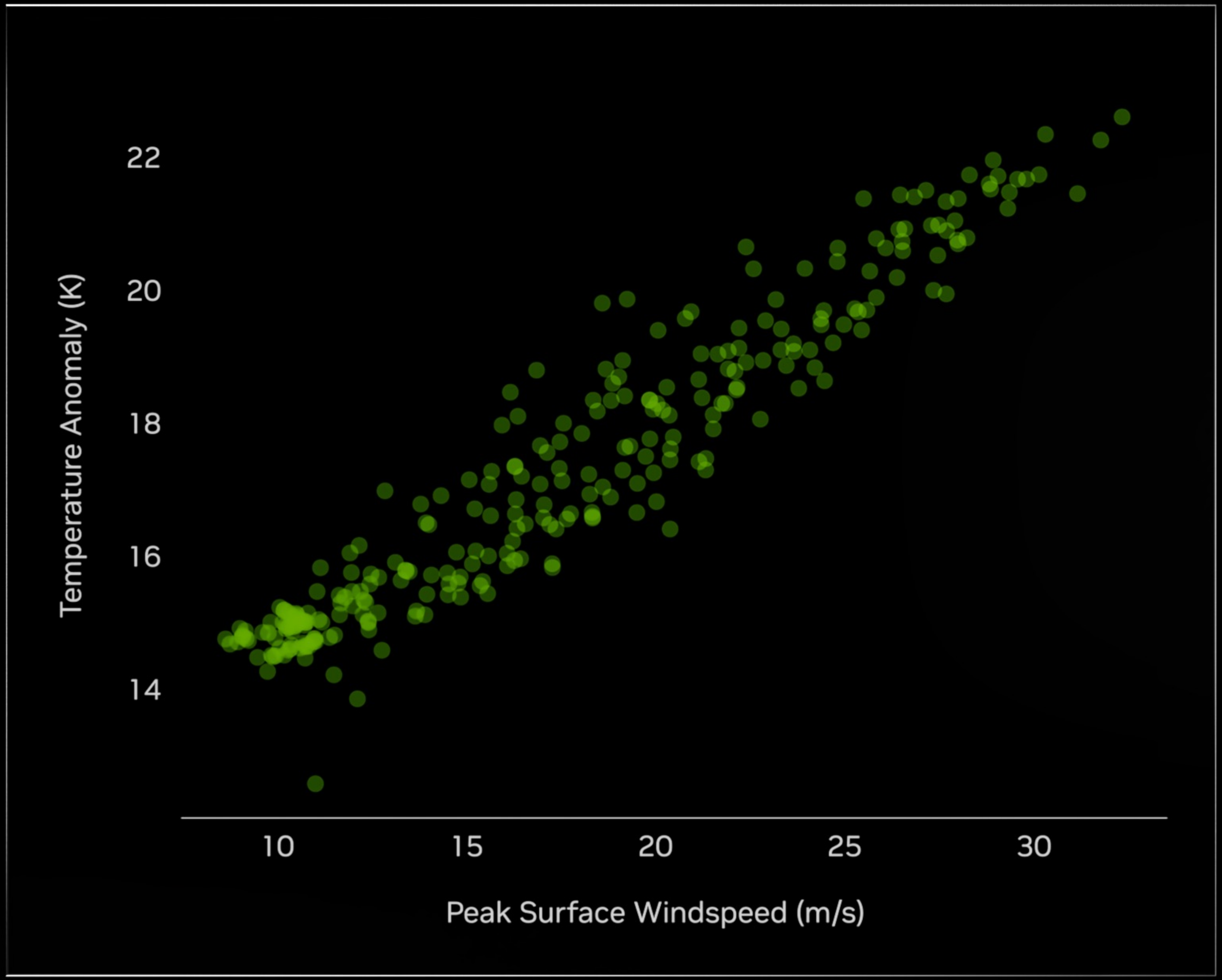
FourCastNet learns physics

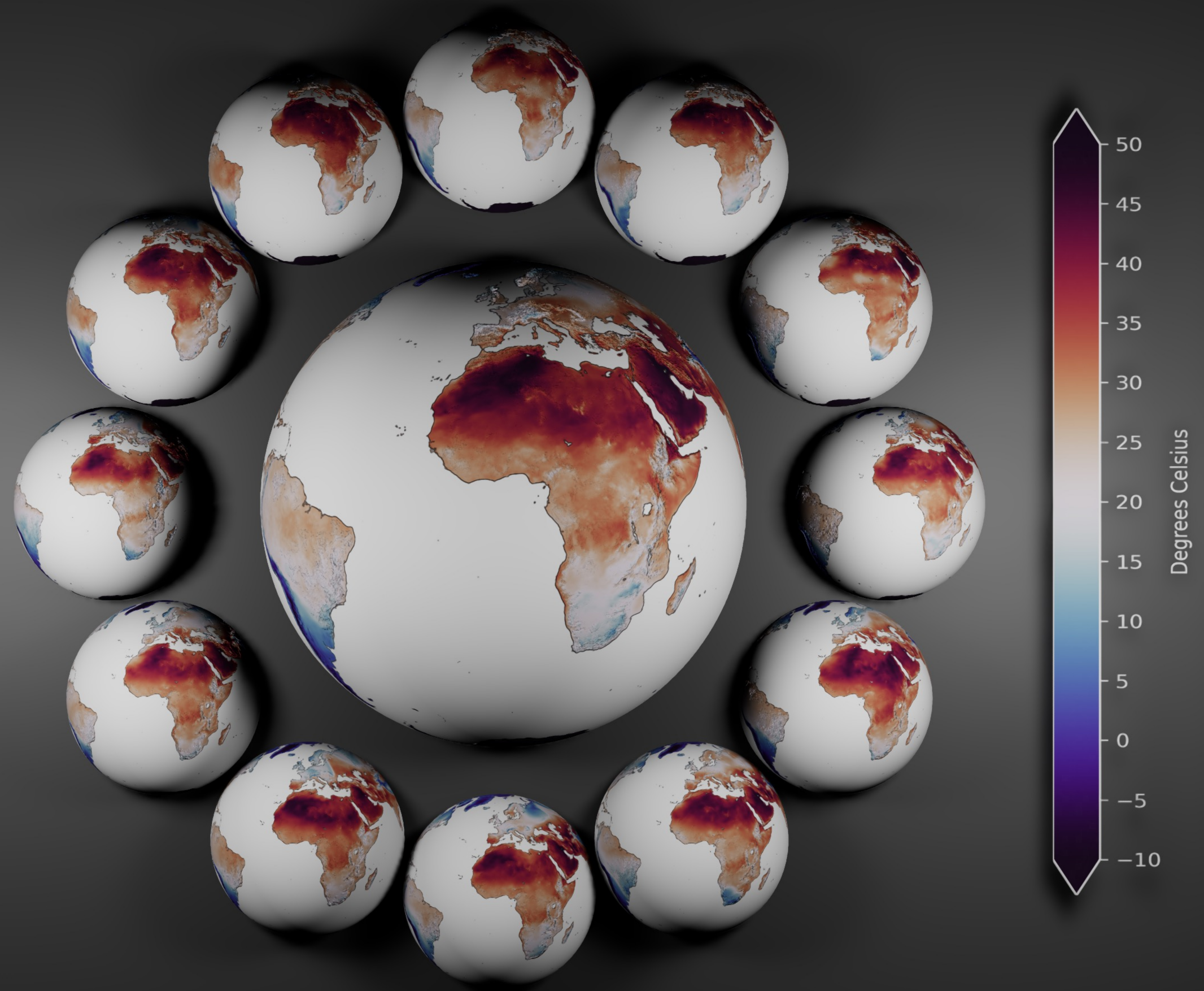
From data, without constraints

REALISTIC HURRICANE TRACKS OF HURRICANE HARVEY



THERMAL WIND STRUCTURE OBEY THEORY AND PHYSICS-BASED MODELS




50
45
40
35
30
25
20
15
10
5
0
−5
−10
Degrees Celsius

Agenda

- NVIDIA's Earth-2 initiative: The Big Picture
- FourCastNet & latest accomplishments in Weather
- Towards kilometer-scale emulation
- Technology platforms

Towards kilometer-scale emulation

To meet the needs of society:

(i) predicting impacts at scales that matter

(ii) interacting with data at low-latency

CHALLENGE

Exabytes of Data to Store

NEED

Full State Vector Interactivity

Any Region

Any Time Period

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

System: 4K H100

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

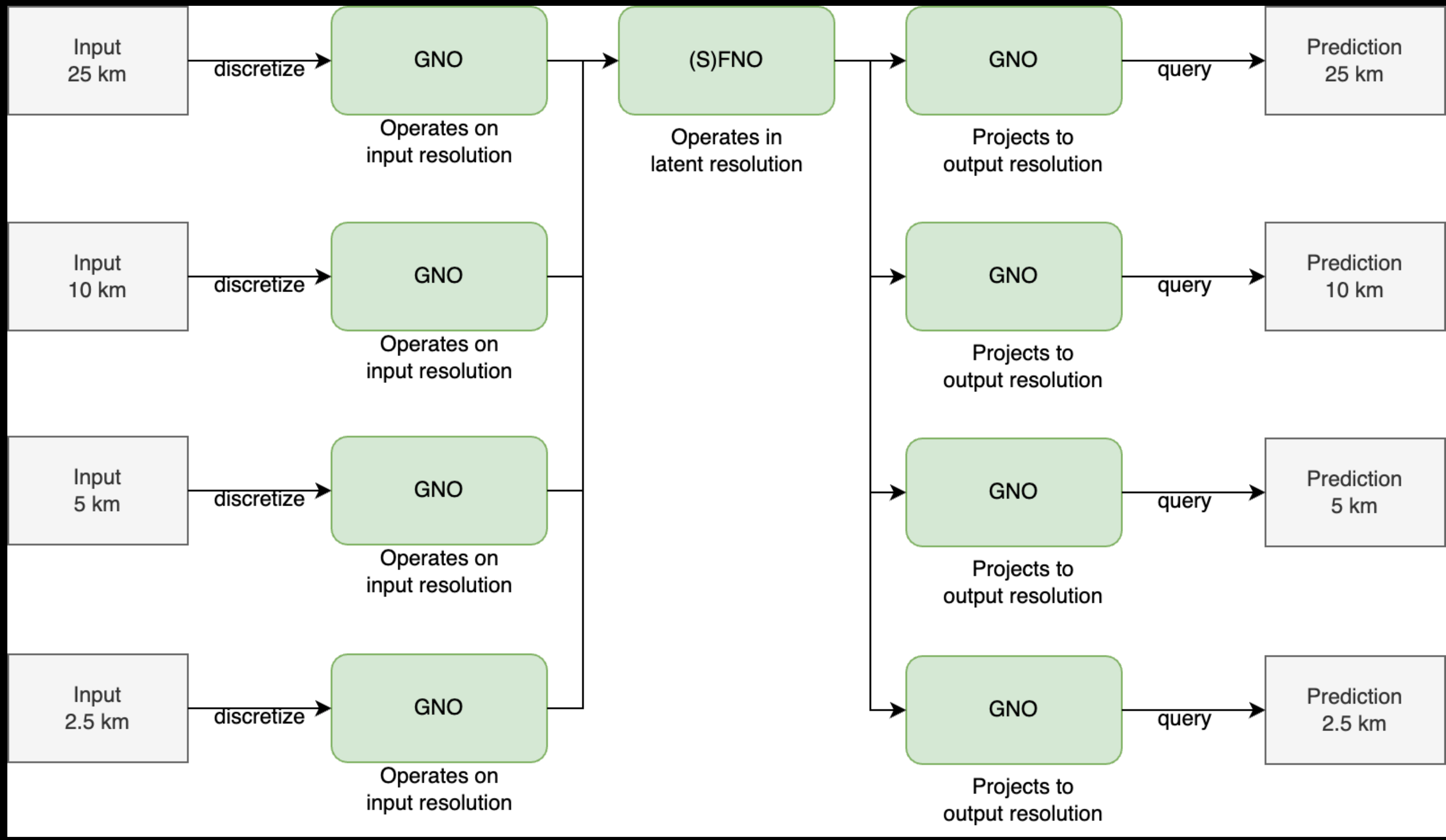
Inference: 30 Days, 1000-member ensemble > ~1 Hour

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

Towards kilometer-scale emulation

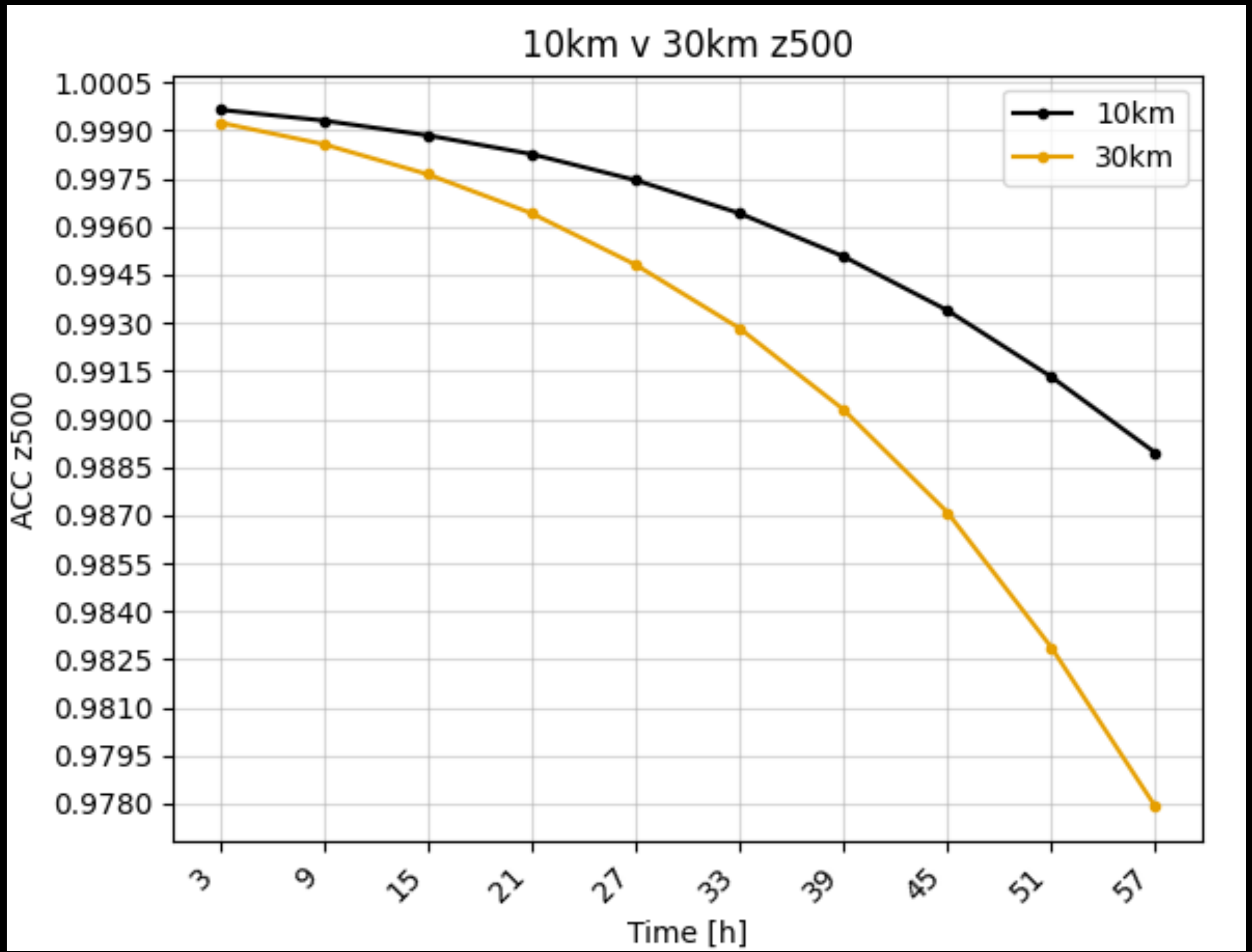
Full State Vector Interactivity – Any Region – Any Time Period

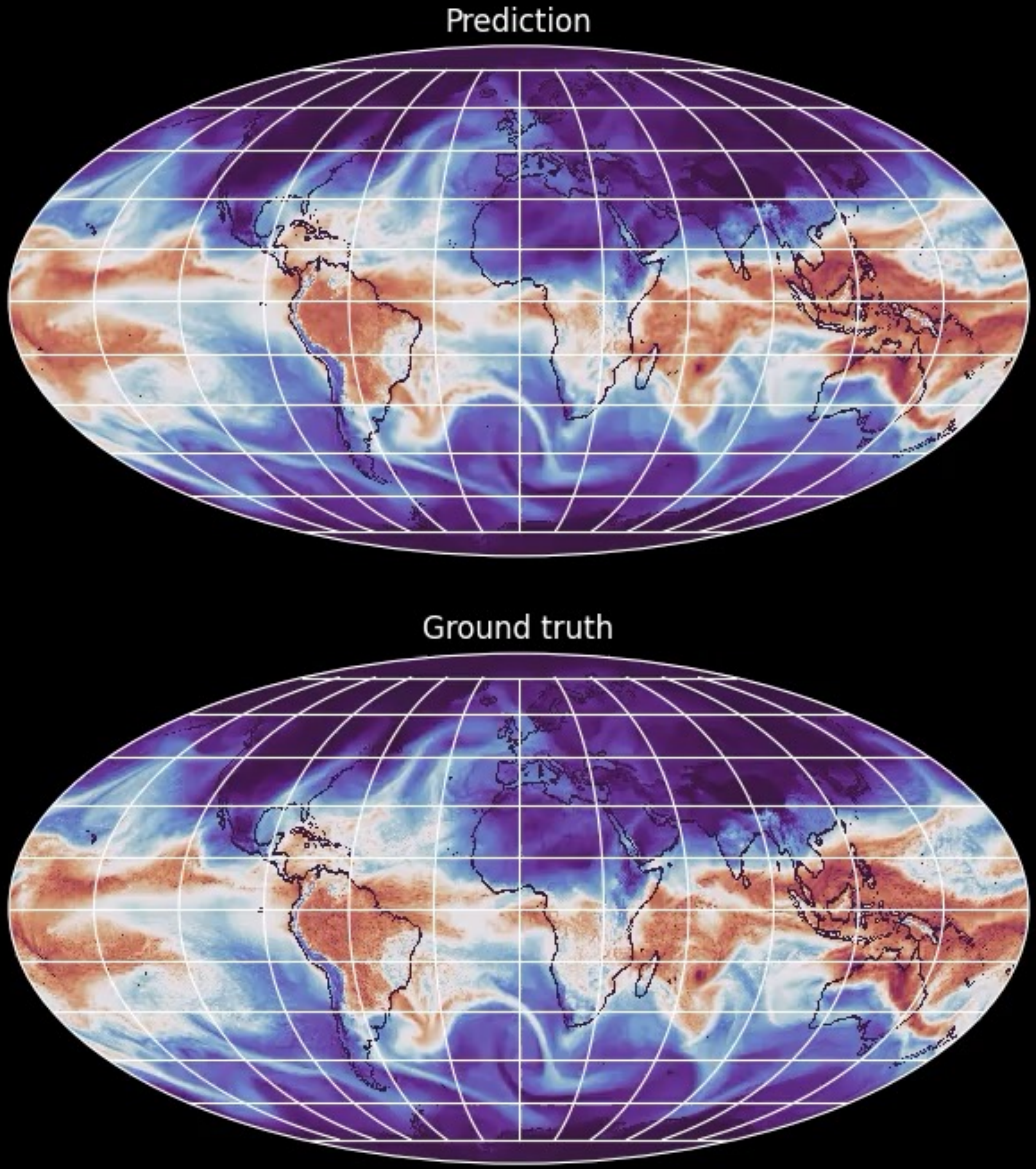
- Naively scaling to km resolution is computationally infeasible (>12k H100s needed to fit model)
- Need multi-resolution, progressive learning approaches to make this tractable
- Use 25km FCN as the backbone and learn a decoder to predict at 2.5km scale, progressively increasing resolution Graph Neural Operator to work on irregular data grids
    - Spherical Fourier Neural Operator for global integration
    - Graph Neural Operator to query at any spatial resolution



SFNO trained on 10km ICON simulations (nextGEMS cycle 2)

Higher resolution improves forecast skill





- Challenges:
  - 9x more compute trained from scratch (fine-tuning from 30km reduces cost by 75%)
  - 7x longer inference time (100-member ensemble in 15 minutes)
  - 2.5x more memory
  - Fine-scale blurring remains
- Limits of learning are yet to be seen!

Key challenges towards *global* km-scale emulation

Global auto-regressive ML forecasting

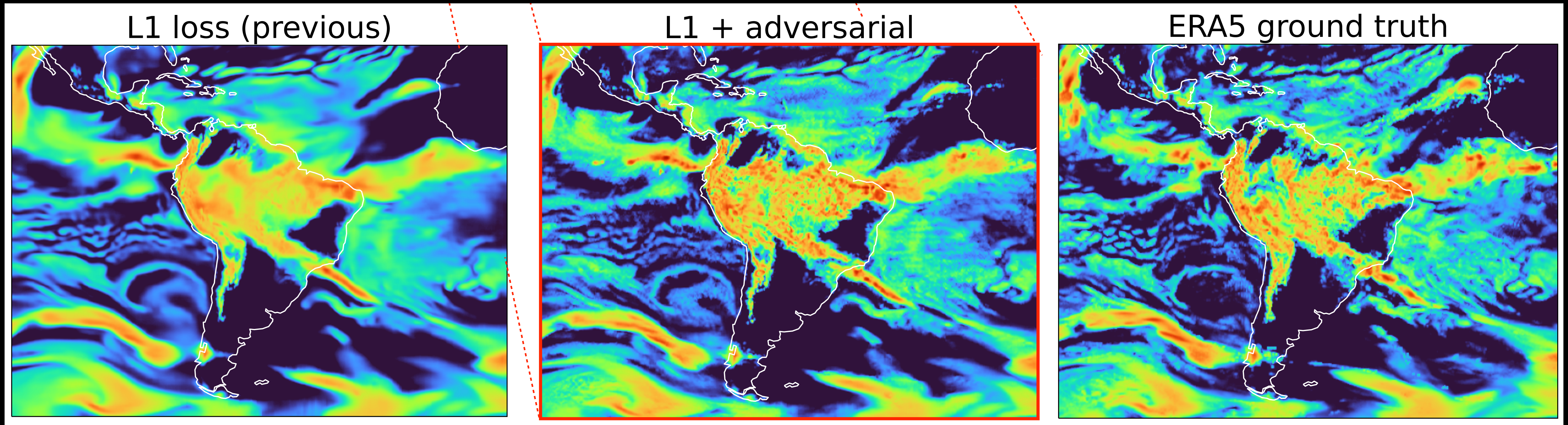
- ECMWF blog on AIFS, "*The IFS is unparalleled by ML models for the breadth of variables it predicts and its spatial resolution*".
- MSE training optimizes for the ensemble mean, especially with multi-step fine-tuning, resulting in fine-scale blurring and lower effective resolution than the training data.
- Ensembles are hard to calibrate.
- DDWP is dependent on data assimilation for training datasets and real-time initial conditions.
- Training directly on observations – multi-modal, sparse and unevenly distributed data.

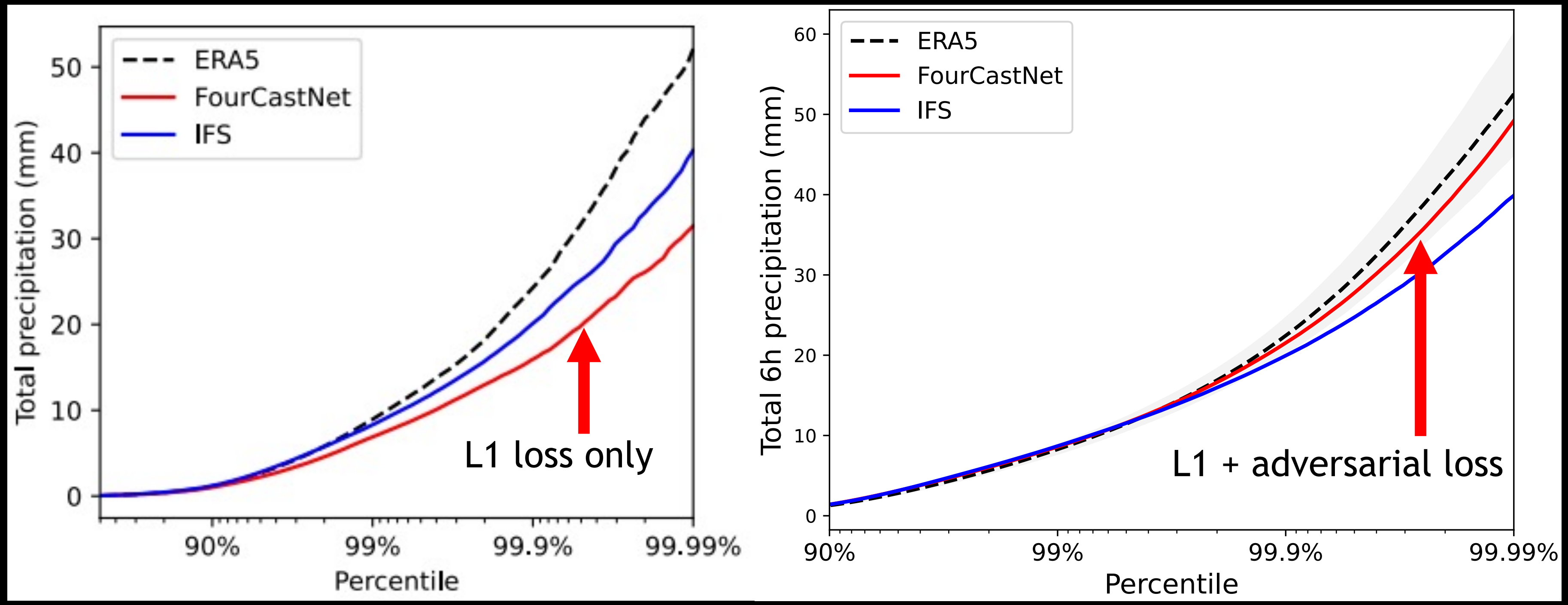
Global *km-scale* auto-regressive ML forecasting

- Model size and data should be scaled equally – *à la* Chinchilla (2022).
- Increased spatial resolution requires finer timesteps, but error accumulates with autoregressive rollout.
- Training SFNO on km-scale global data from scratch requires at least 12k H100s! Progressive SR fine-tuning could bring it down to 4k H100s.

Could Generative AI offer an alternative pathway for km-scale emulation?

Gen AI not new in data-driven weather and climate prediction, but the current gen AI revolution could be game-changing…

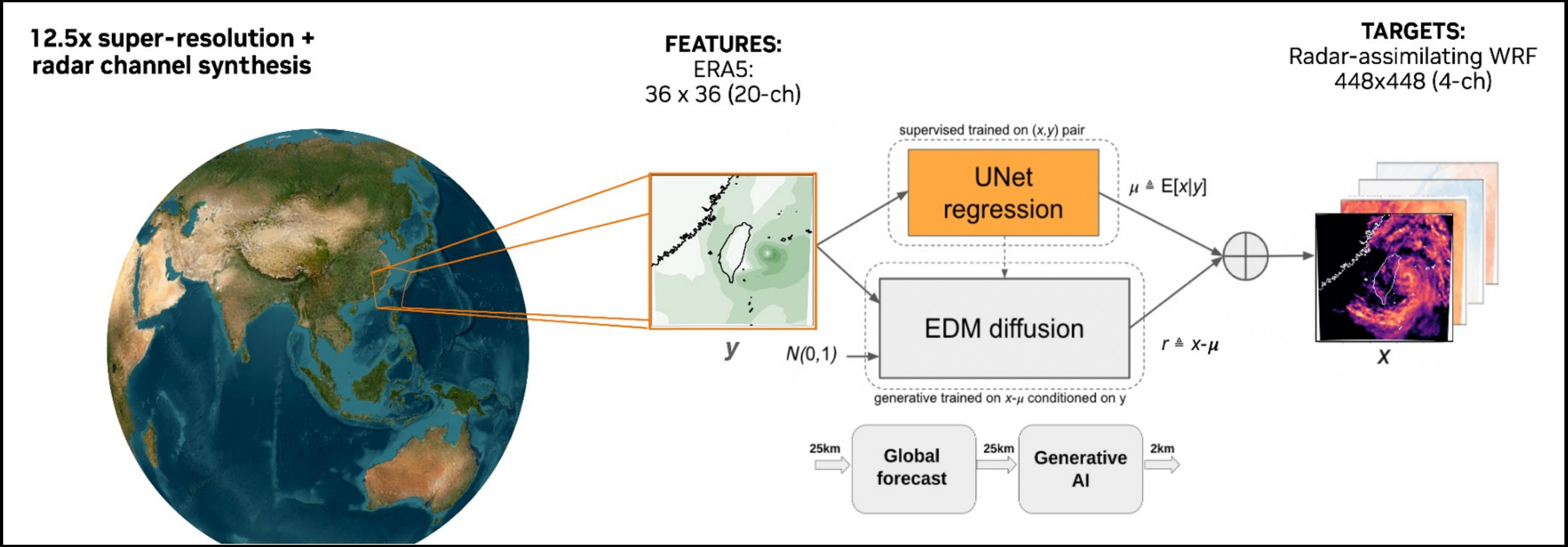




Duncan et al. (2022), Generative Modeling of High-resolution Precipitation Forecasts, arXiv:2210.12504

Generative diffusion modeling for regional km-scale forecasts

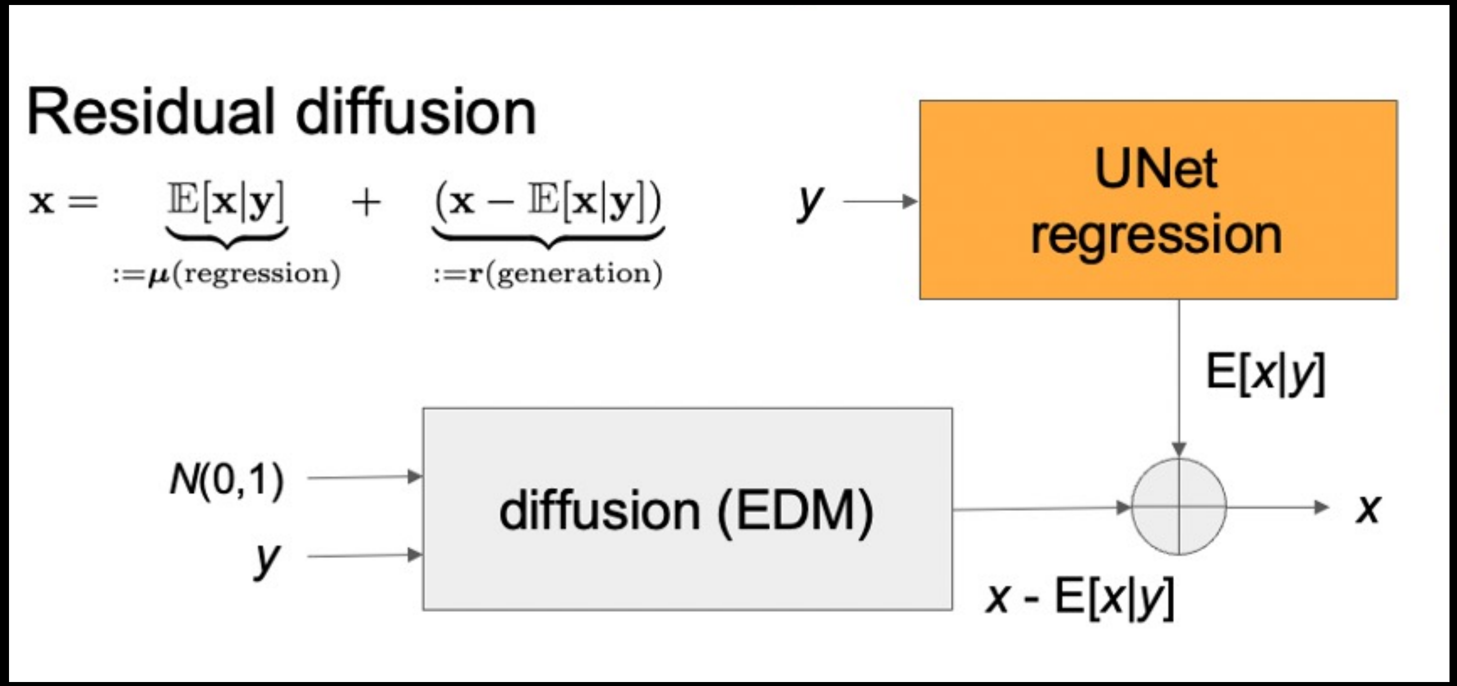
Tapping into extensive gen AI research and optimizations being developed

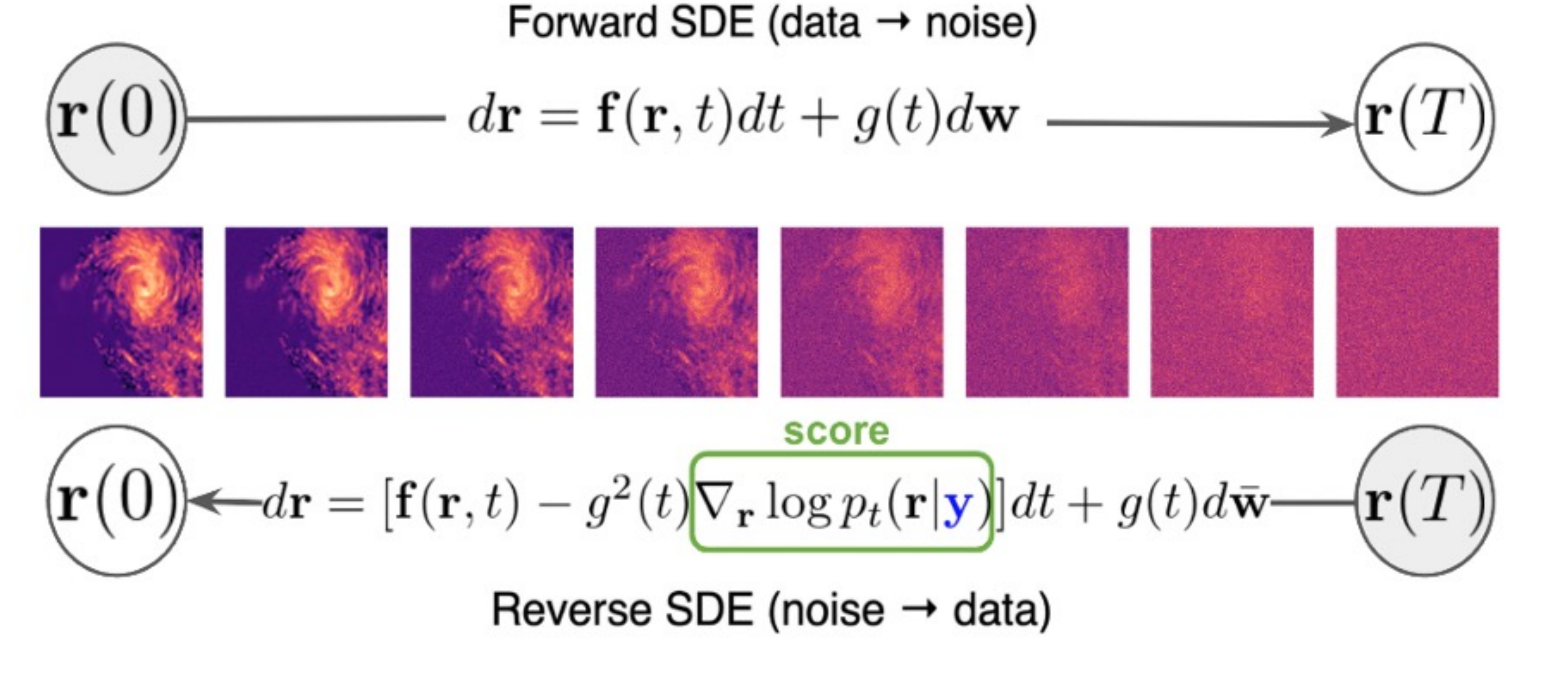


Mardani et al. (2023), Generative Residual Diffusion Modeling for Km-scale Atmospheric Downscaling, arXiv: 2309.15214

Generative diffusion modeling for regional km-scale forecasts

Tapping into extensive gen AI research and optimizations being developed



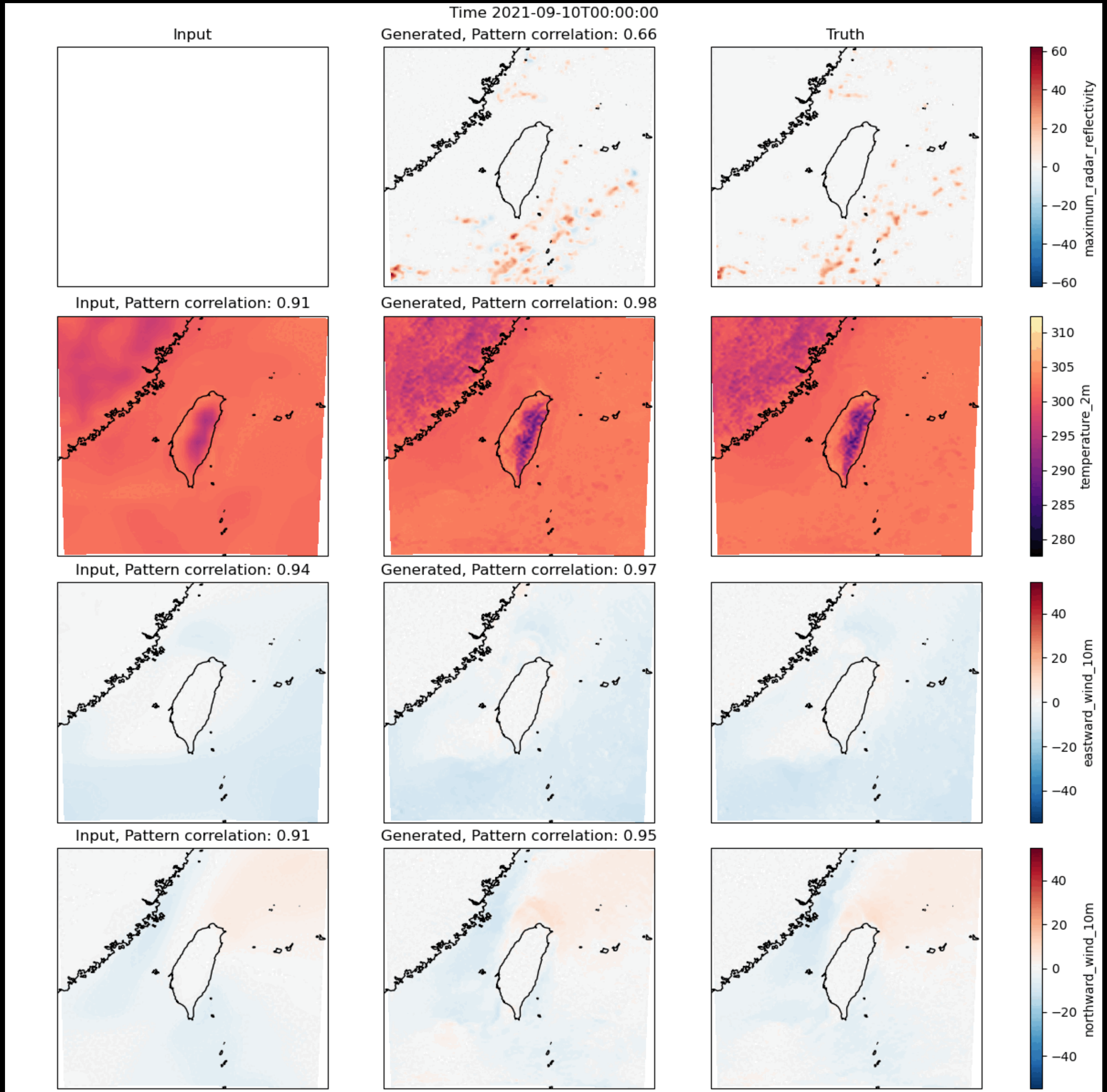


Mardani et al. (2023), Generative Residual Diffusion Modeling for Km-scale Atmospheric Downscaling, arXiv: 2309.15214

12.5x downscaling (25km → 2km) *and* channel synthesis

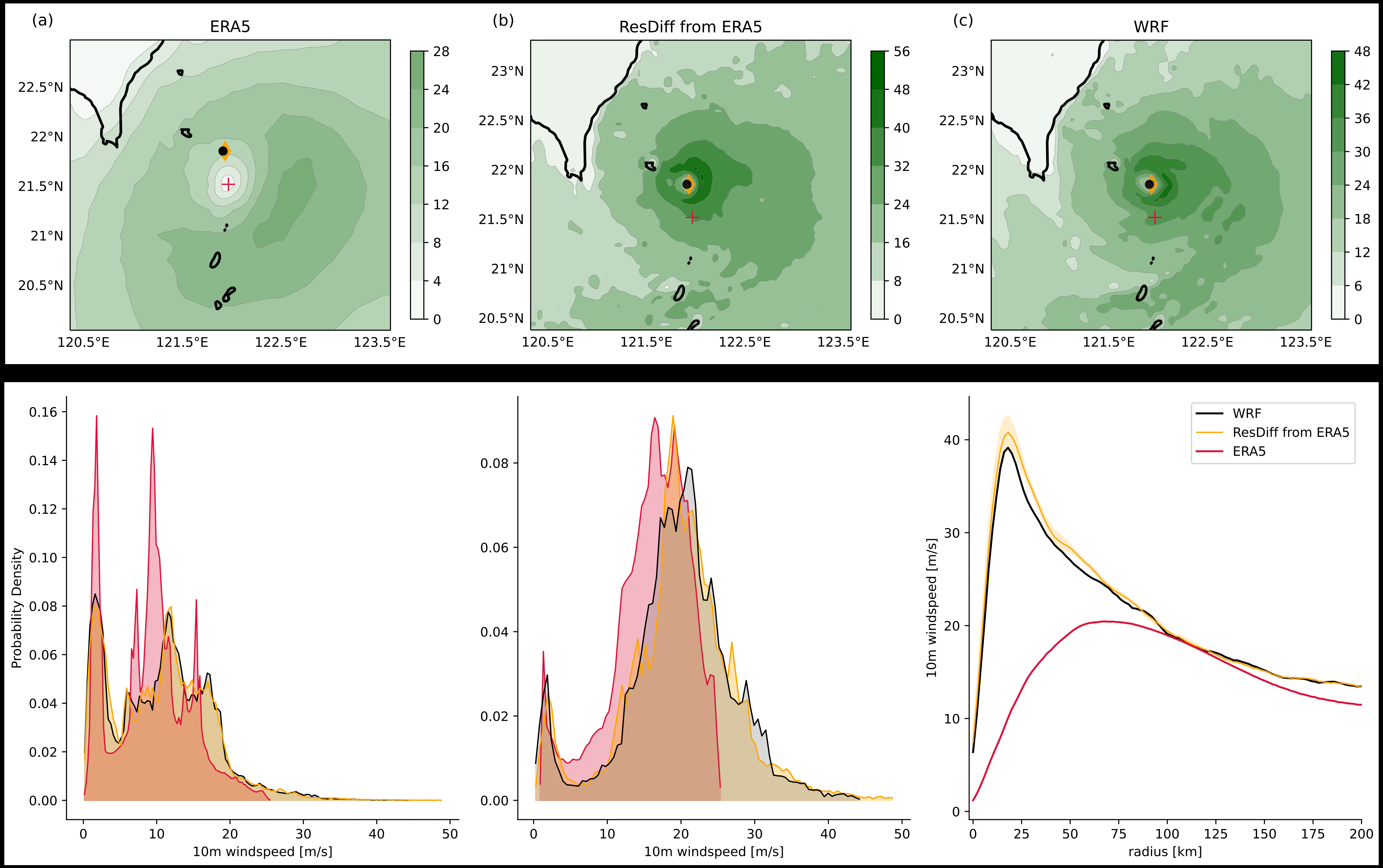
Stochastic prediction of radar reflectivity, temperature and 10m winds

FCN Forecast ***ResDiff*** ***WRF***



Tropical cyclones more intense and compact

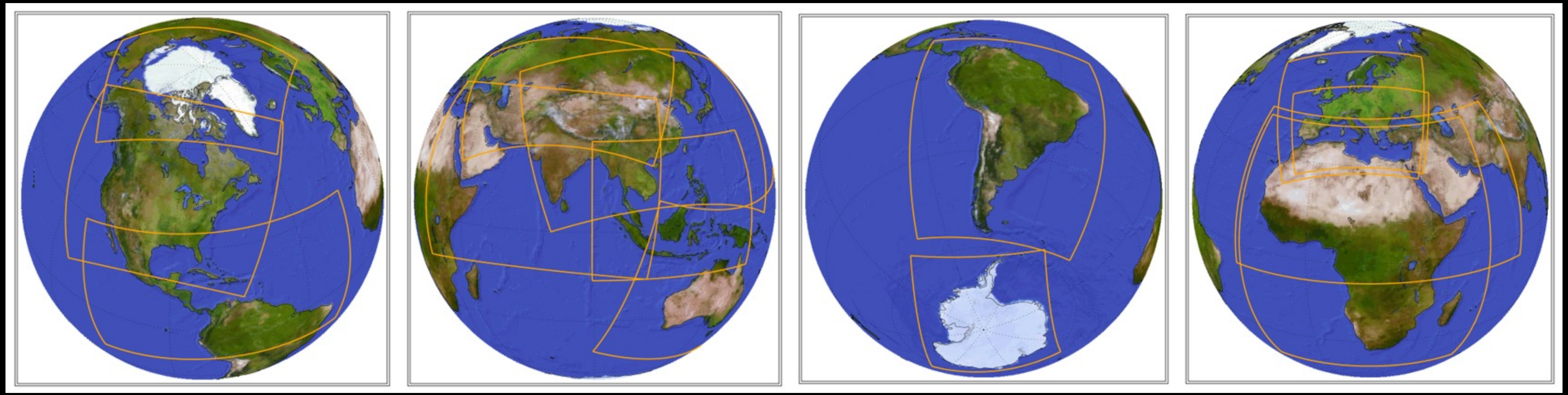
Not just super-resolution, but distribution shifts and (some) physics is learned



NVIDIA

A promising pathway to km-scale prediction…

Could ResDiff be scaled to multiple regions, perhaps the entire globe?



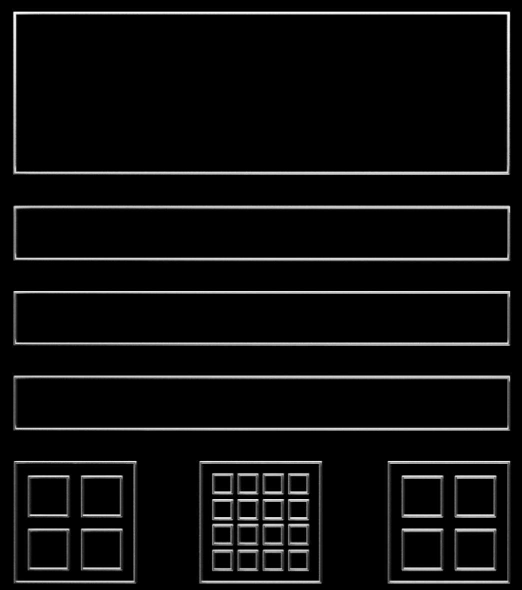
- 1000-member ensemble in 8 minutes. Massive ensembles *for free* (sampling from distribution)
- 200x data compression, 500x faster, and 2000x more energy efficient than a WRF simulation at 2km
- End-to-end ML solution when coupled with a global ML predictor
- Could be adapted for climate prediction by conditioning on hi-fidelity regional climate data, e.g., CORDEX

Agenda

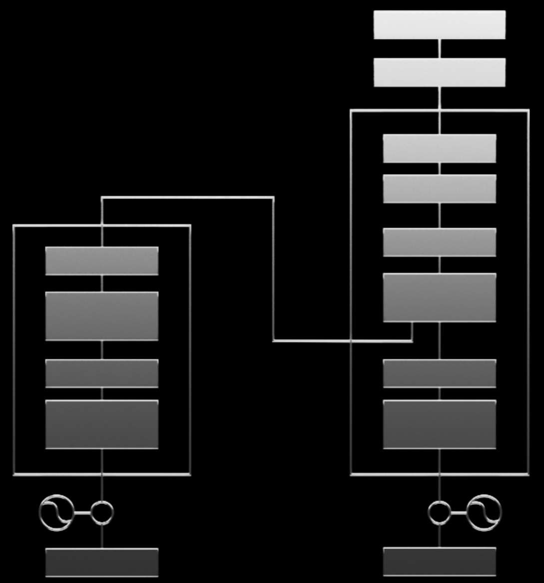
- NVIDIA's Earth-2 initiative: The Big Picture
- FourCastNet & latest accomplishments in Weather
- Towards kilometer-scale emulation
- Technology platforms

Earth-2 Platforms

NVIDIA
ACCELERATED COMPUTING



MODULUS PHYSICS-ML
FOURCASTNET



OMNIVERSE
DIGITAL TWINS

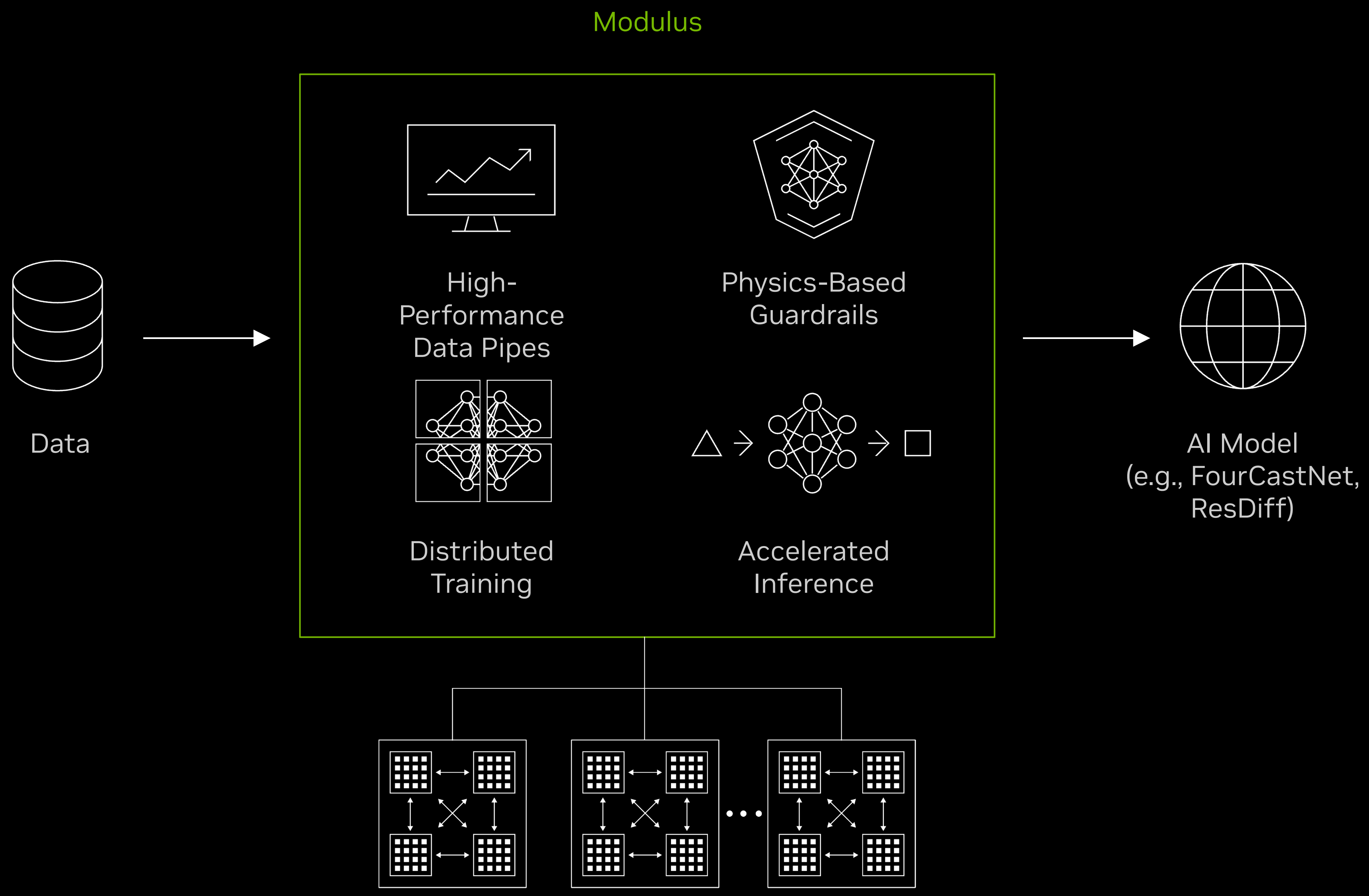
Accelerated Computing

3 supercomputers: Simulation, AI, Visualization

NVIDIA Modulus

Open-Source, Physics-ML Platform + Cloud-native end-to-end MLOps

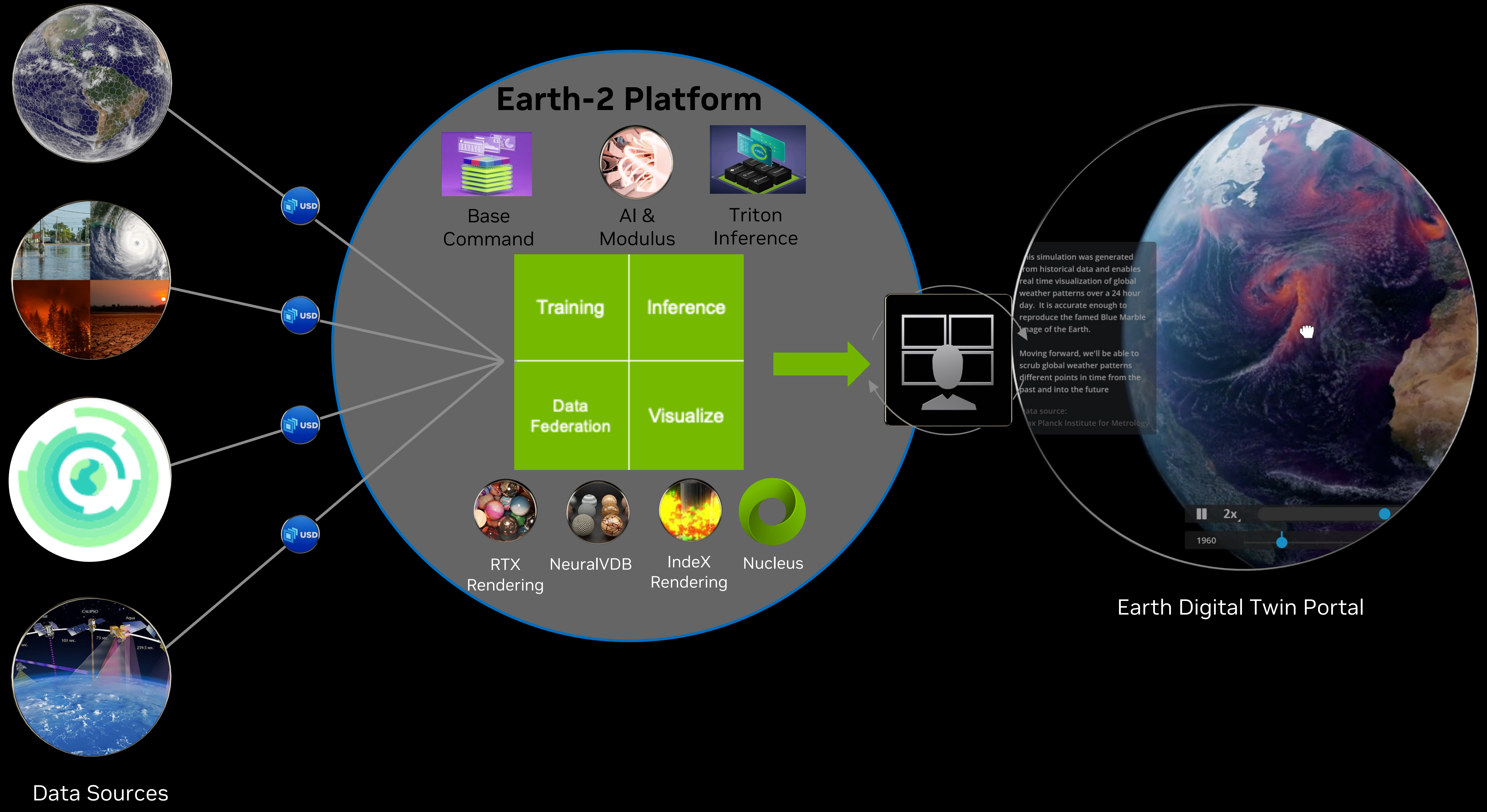
- End-to-end MLOPs pipeline for data ingest, processing, training, inference, deployment
- Optimized training and inference
- E2MIP for access to popular pre-trained models (a model zoo)
- E2MIP diagnostics and scorecards to verify, validate, inter-compare models
- Recipes for model development and fine-tuning for regional prediction, specific phenomena (cyclones, heat waves, etc.)



https://github.com/NVIDIA/modulus

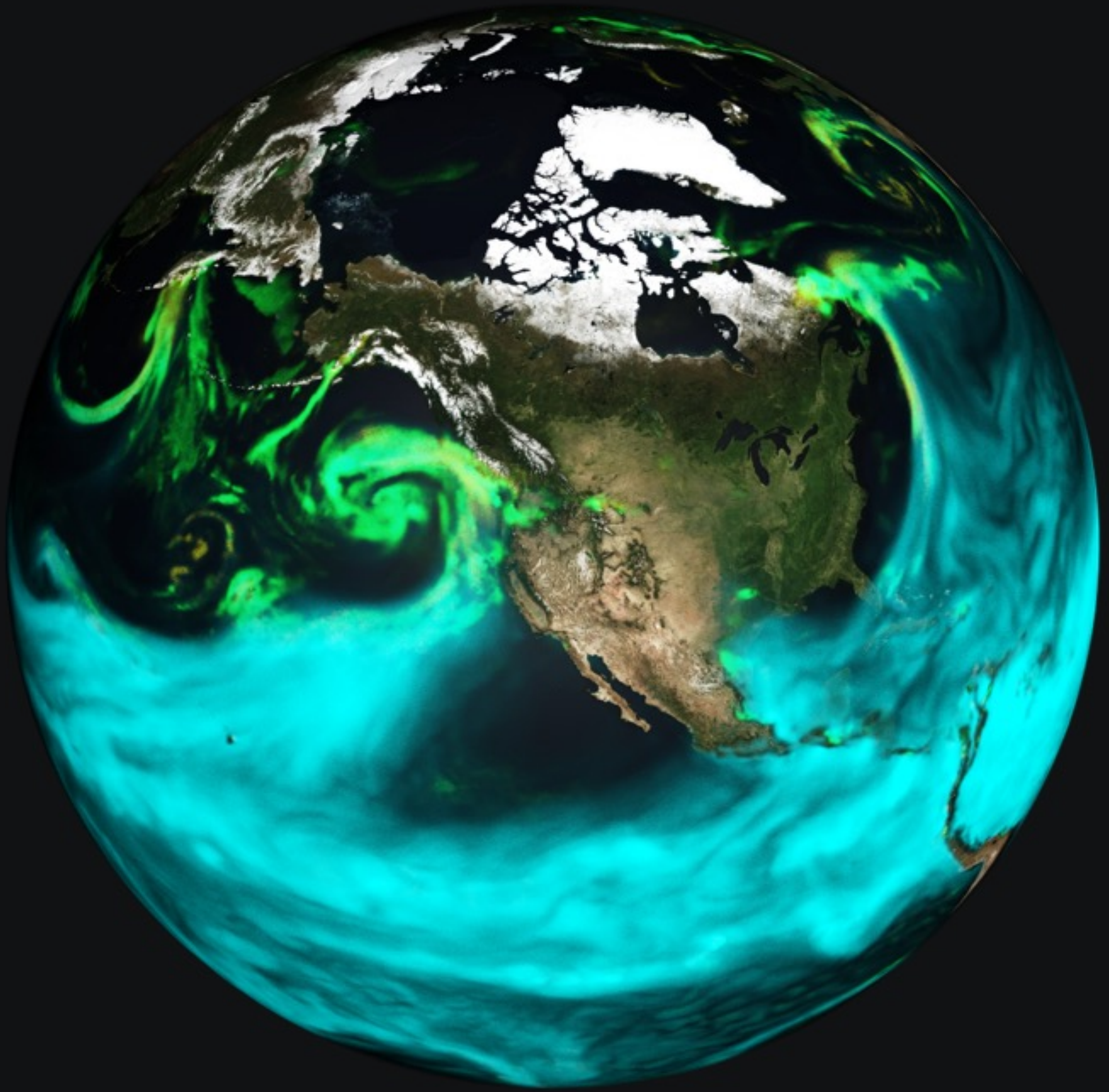
Earth-2 Digital Twin

Connecting Complex Simulations, Data and AI workflows, rendered in 3D



The Vision of Earth-2

Is Beginning to Take Shape



Acknowledging: Mike Pritchard, Anima Anandkumar, David Hall, Jaideep Pathak, Noah Brenowitz, Yair Cohen, Thorsten Kurth, Boris Bonev, Christian Hundt, Andre Graubner, Peter Messmer, Stan Posey, Akshay Subramaniam, Sanjay Choudhry, Farah Hariri, Niklas Roebler, Ram Cherukuri, Nicholas Geneva, Mathias Hummel, Christopher Lamb, Mike Houston, Kamyar Azizzadenesheli, Jean Kossaifi, Steffen Roemer, Marius Koch & David Appelhans, many more NV staff & our **generous external climate science advisors Bjorn Stevens, Peter Deuben, Peter Bauer, Nils Wedi, Thomas Schulthess, and Francisco Doblas-Reyes**.

NVIDIA