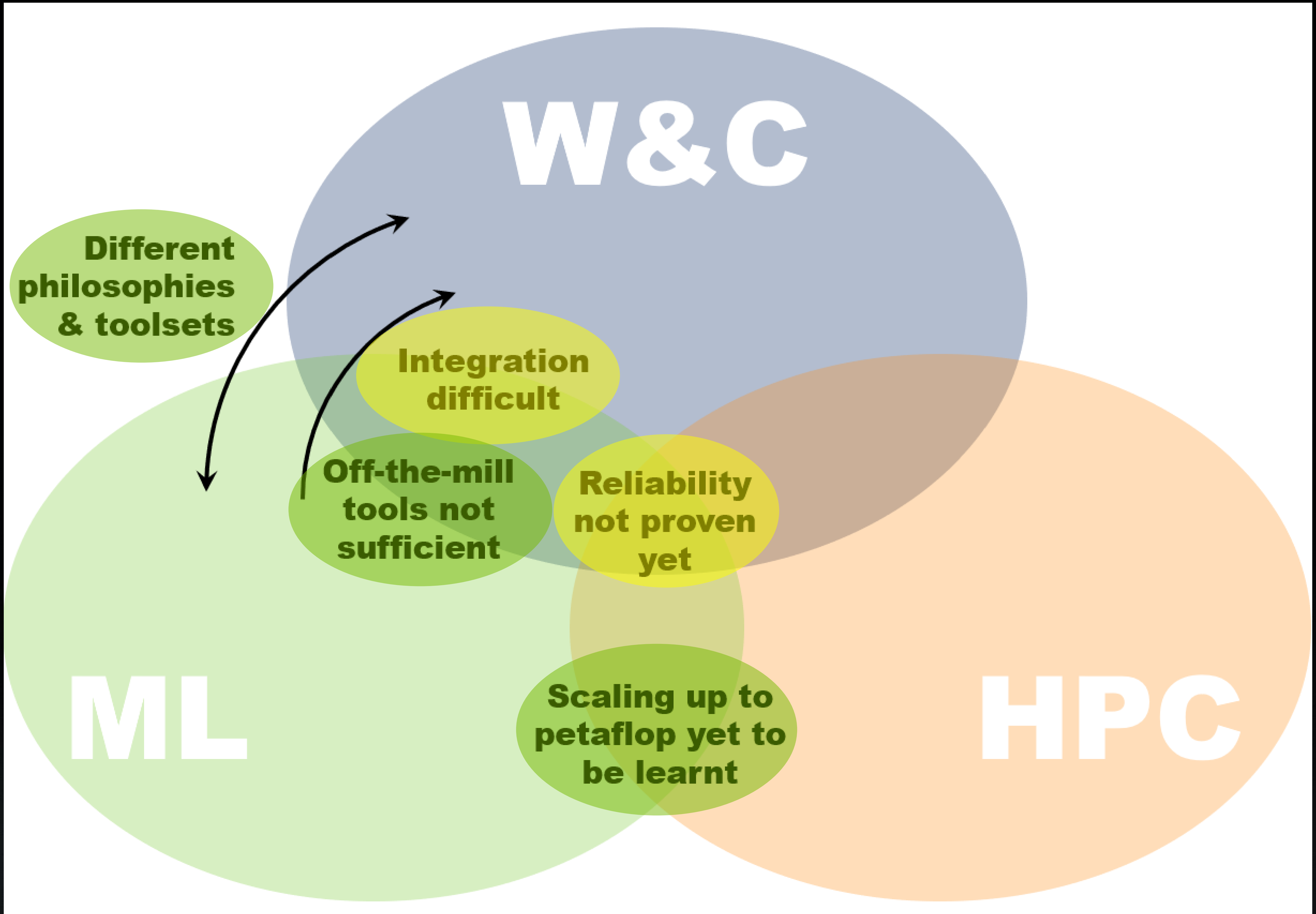
FOURCASTNET: A STEP TOWARDS BUILDING DIGITAL TWINS OF THE EARTH FOR NVIDIA'S EARTH-2 INITIATIVE

Karthik Kashinath, Senior ML Scientist-Technologist, AI-HPC, NVIDIA

EARTH DIGITAL TWIN
MACHINE LEARNING CHALLENGES AND APPROACHES



Source: Peter Dueben (ECMWF) on "Machine Learning for Weather and Climate Predictions" at the 7th ENES HPC Workshop 2022

EARTH DIGITAL TWIN
MACHINE LEARNING CHALLENGES AND APPROACHES

CHALLENGES

- Weather ⟶ Climate
- Extrapolation
- Physical consistency & causality
- Uncertainty quantification & Calibration
- Data fusion & assimilation
- Scale up & out
- …

APPROACHES

- Spatio-temporal modeling
- Super-resolution
- Segmentation
- Online learning
- Reinforcement Learning
- …

NVIDIA

FourCastNet

Global data-driven high-resolution Earth digital twin for Extreme Weather

J. Pathak
NVIDIA

S. Subramanian
LBL

P. Harrington
LBL

S. Raja
U. Michigan

A. Chattopadyay
Rice. U.

M. Mardani
NVIDIA

T. Kurth
NVIDIA

D. Hall
NVIDIA

Z. Li
Caltech

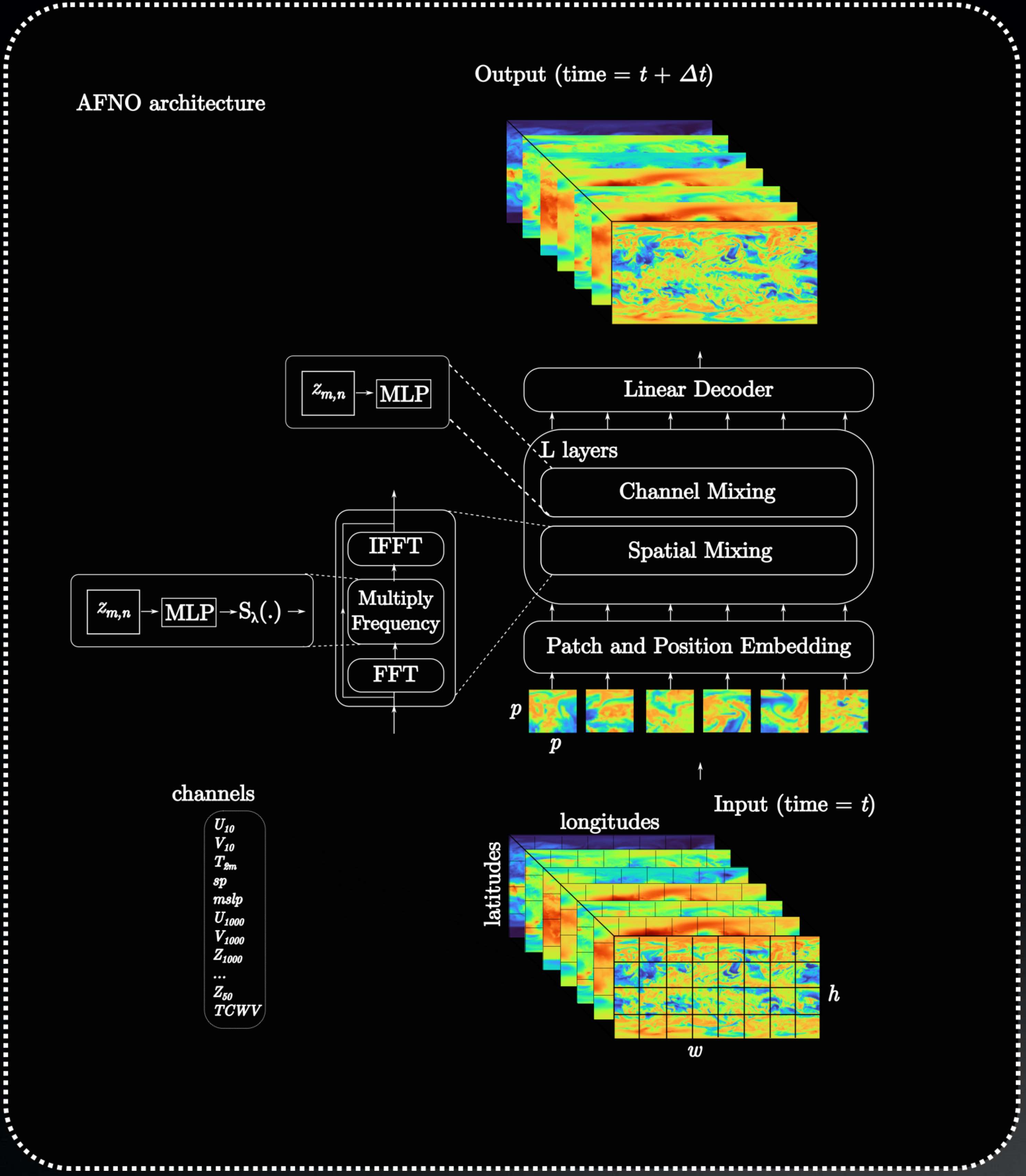
K. Azzizzadenesheli
Purdue

P. Hassanzadeh
Rice U.

K. Kashinath
NVIDIA

A. Anandkumar
Caltech/NVIDIA

FOURCASTNET (FOURIER FORECASTING NETWORK)



- Purely data-driven ML surrogate weather model
- Trained on ERA5 reanalysis data at the native resolution of 0.25 degrees

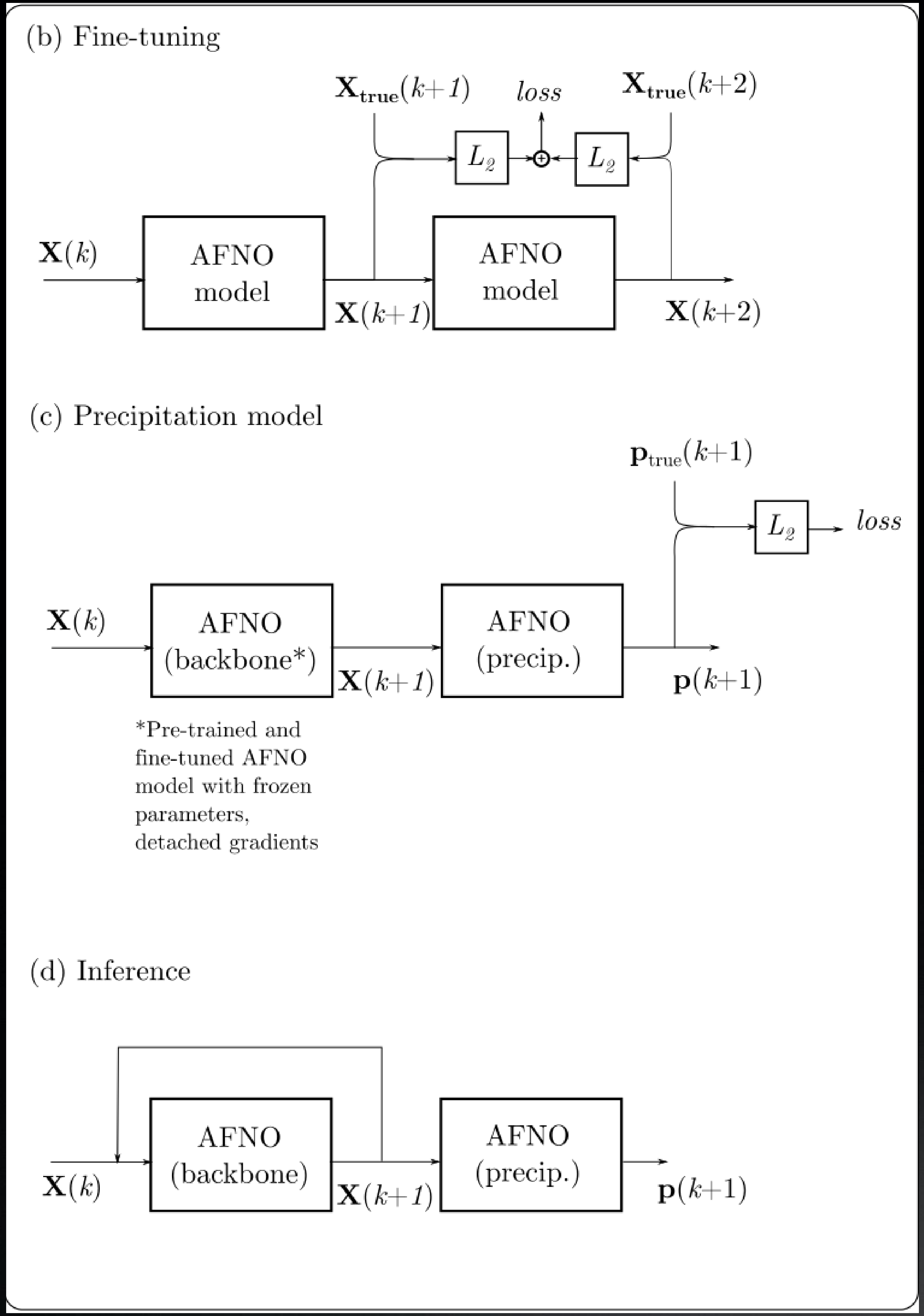
Vertical Level	Variables
Surface	$U_{10}, V_{10}, T_{2m}, sp, mslp$
$1000hPa$	U, V, Z
$850hPa$	T, U, V, Z, RH
$500hPa$	T, U, V, Z, RH
$50hPa$	Z
Integrated	$TCWV$

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 (FOURIER FORECASTING NETWORK)



- Purely data-driven ML surrogate weather model
- Trained on ERA5 reanalysis data at the native resolution of 0.25 degrees

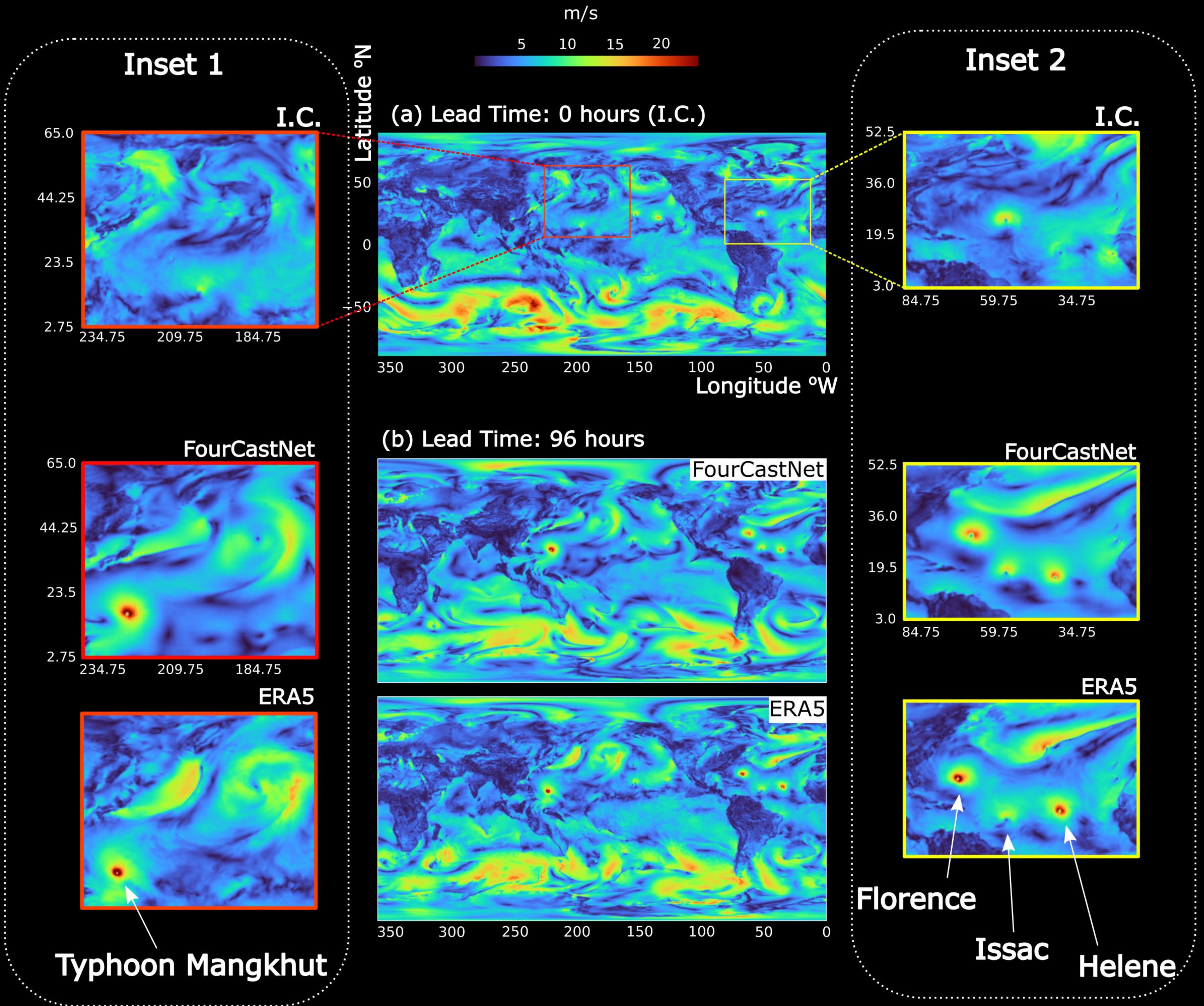
Vertical Level	Variables
Surface	$U_{10}, V_{10}, T_{2m}, sp, mslp$
$1000hPa$	U, V, Z
$850hPa$	T, U, V, Z, RH
$500hPa$	T, U, V, Z, RH
$50hPa$	Z
Integrated	$TCWV$

Extending to include radiation processes, vapor transport, clouds

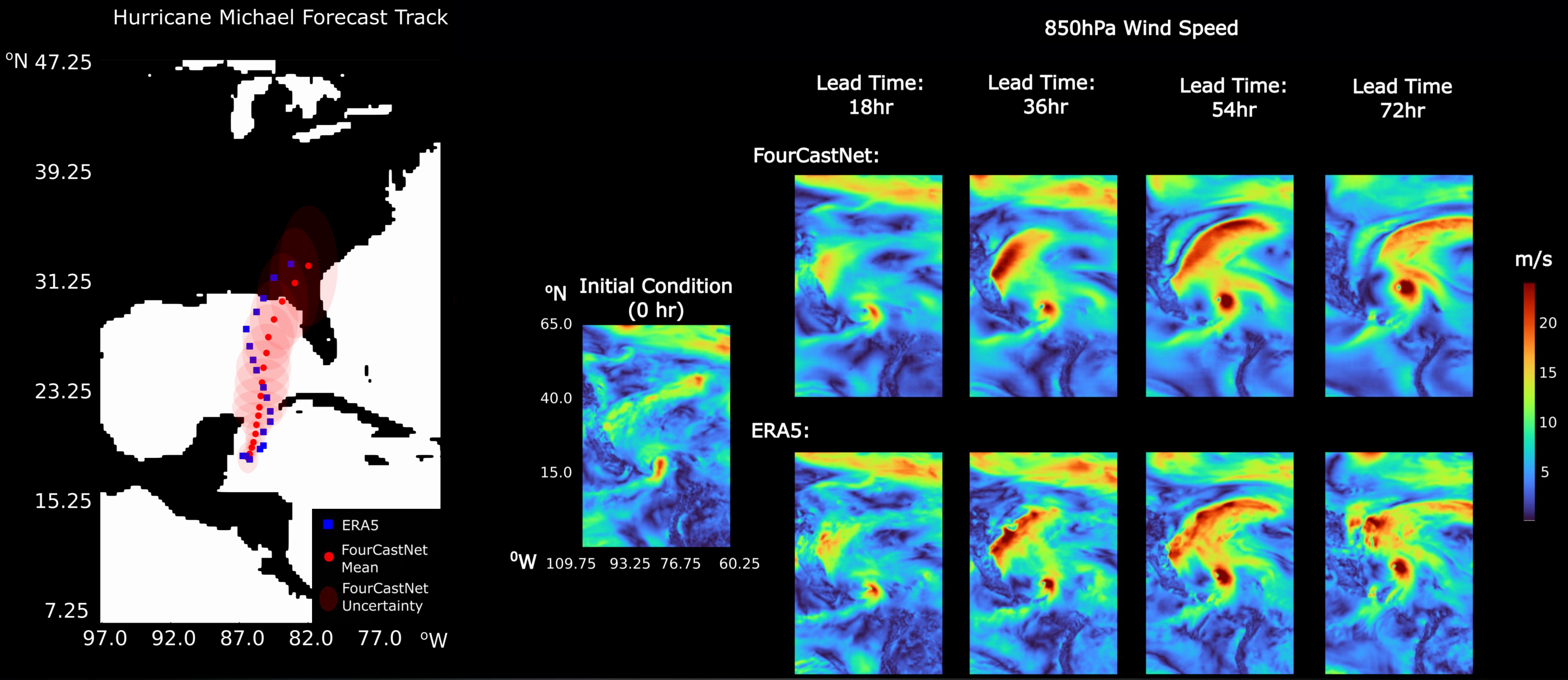
Training set: 1979 to 2015
Validation set: 2016, 2017
Held out: 2018 onwards

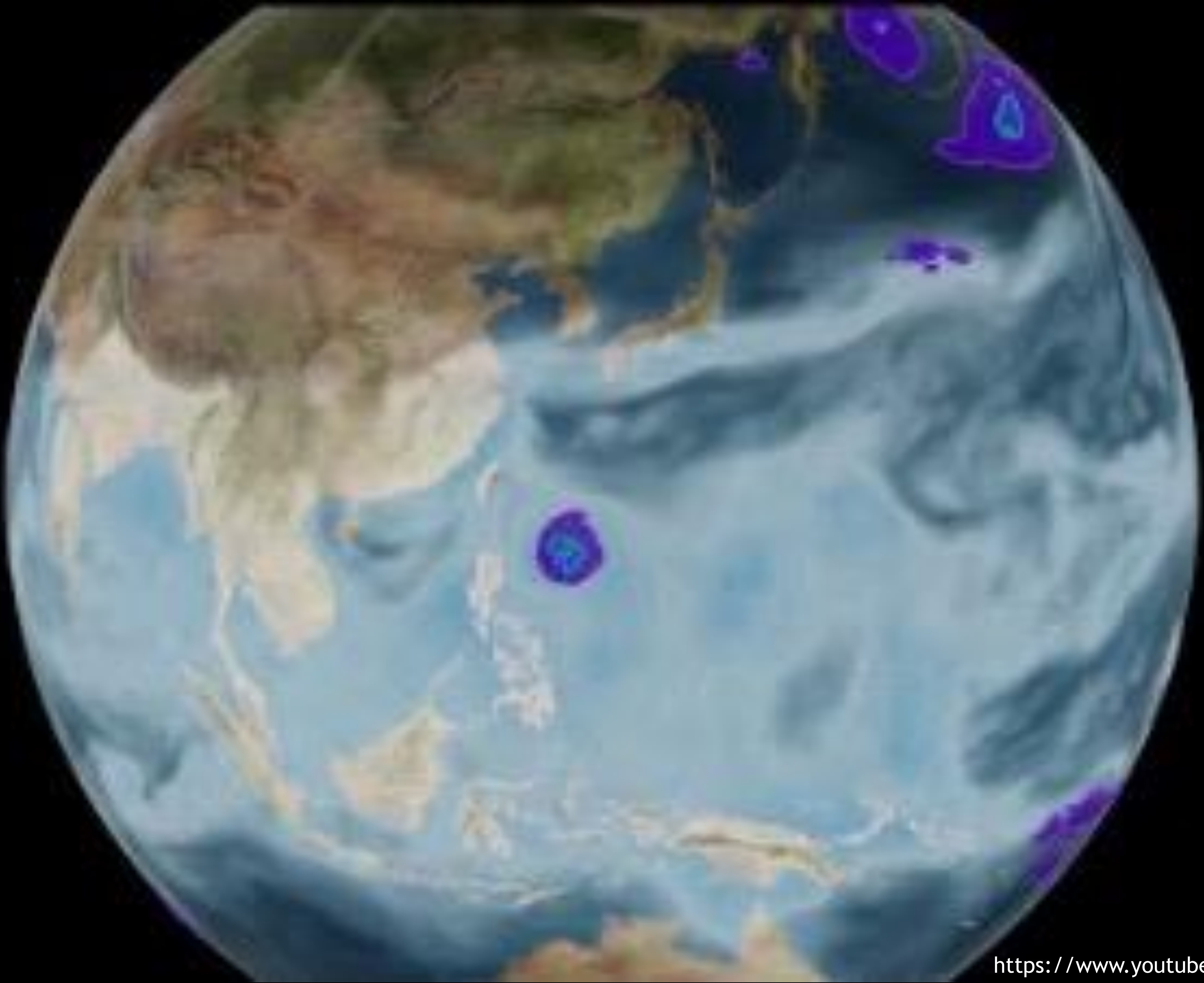
Pathak et al. (2022), FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators, arXiv:2202.11214

EXCELLENT SKILL ON FORECASTING SURFACE WINDS



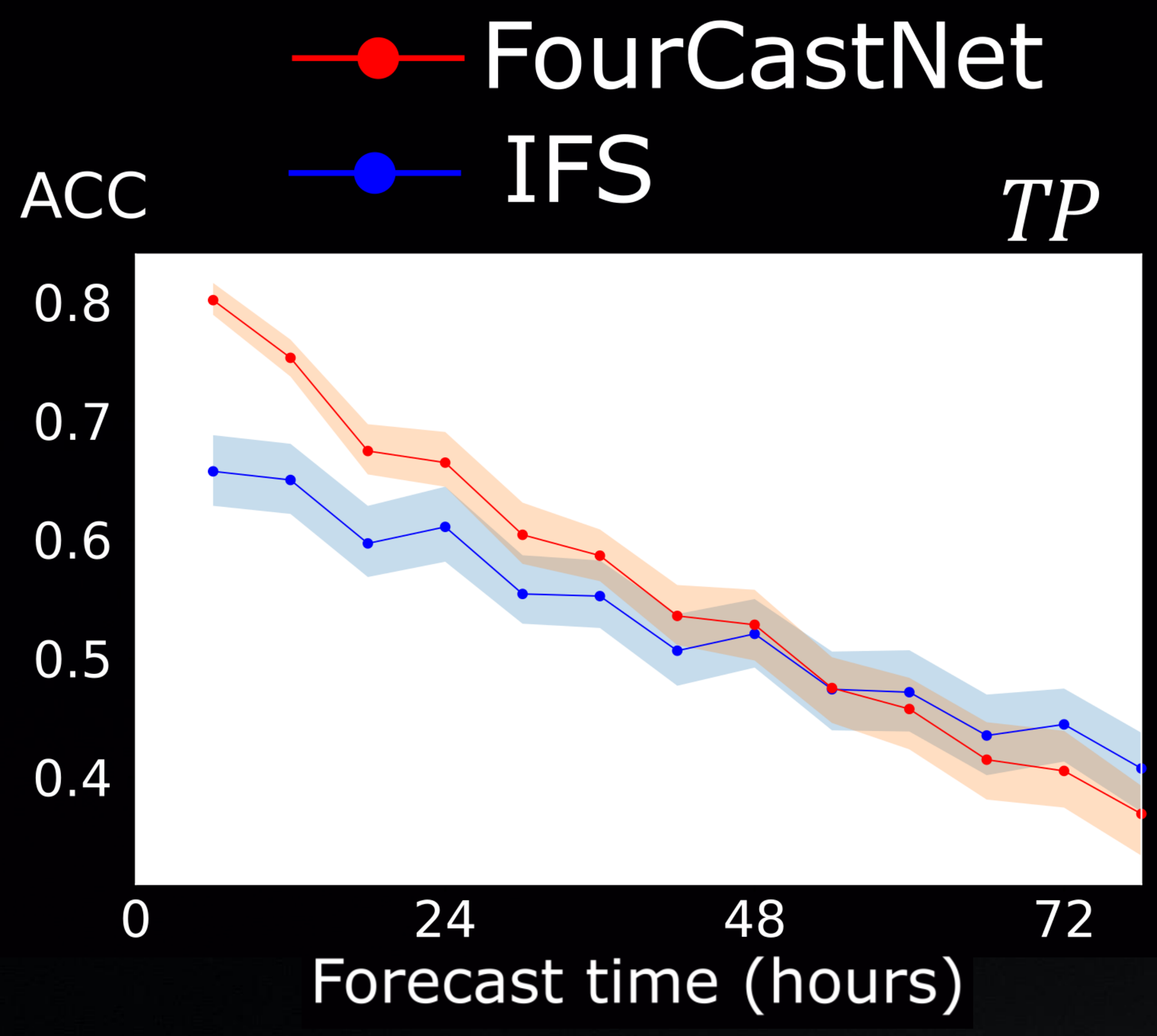
FOURCASTNET PREDICTS HURRICANE PATHS AND INTENSITIES

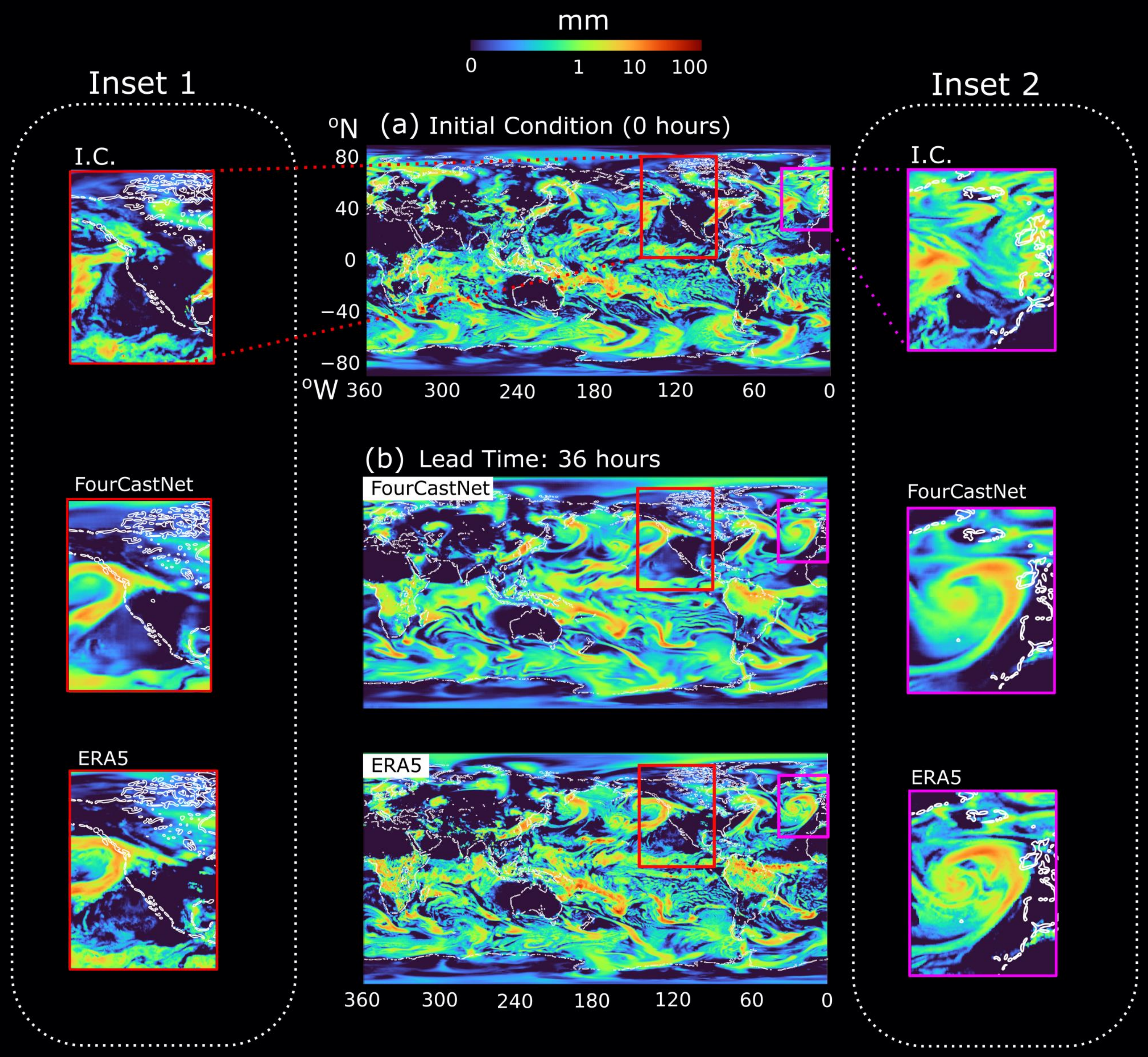




https://www.youtube.com/watch?v=JnGPxZ9glVk

UNPRECEDENTED SKILL ON PRECIPITATION FORECASTS

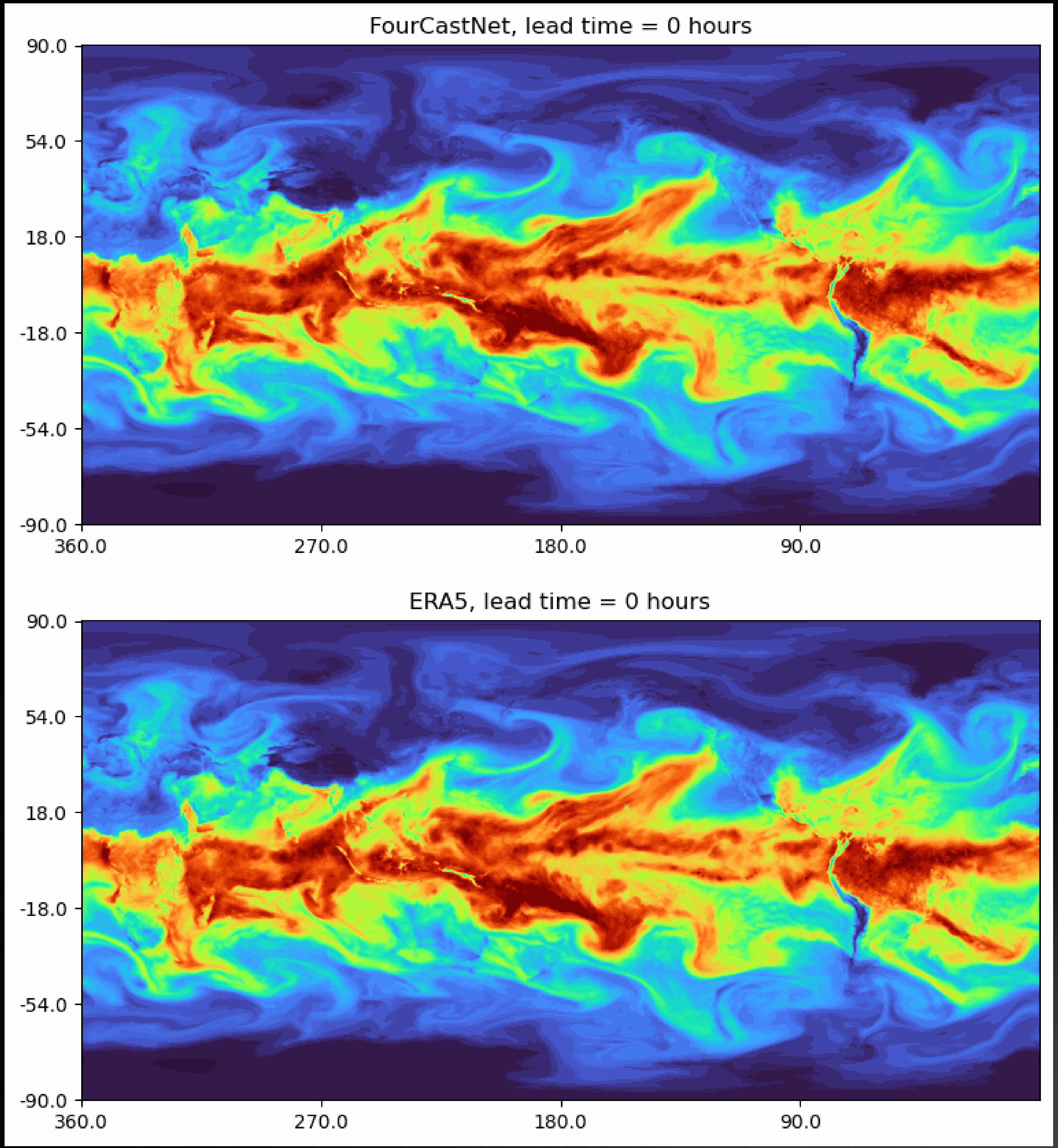


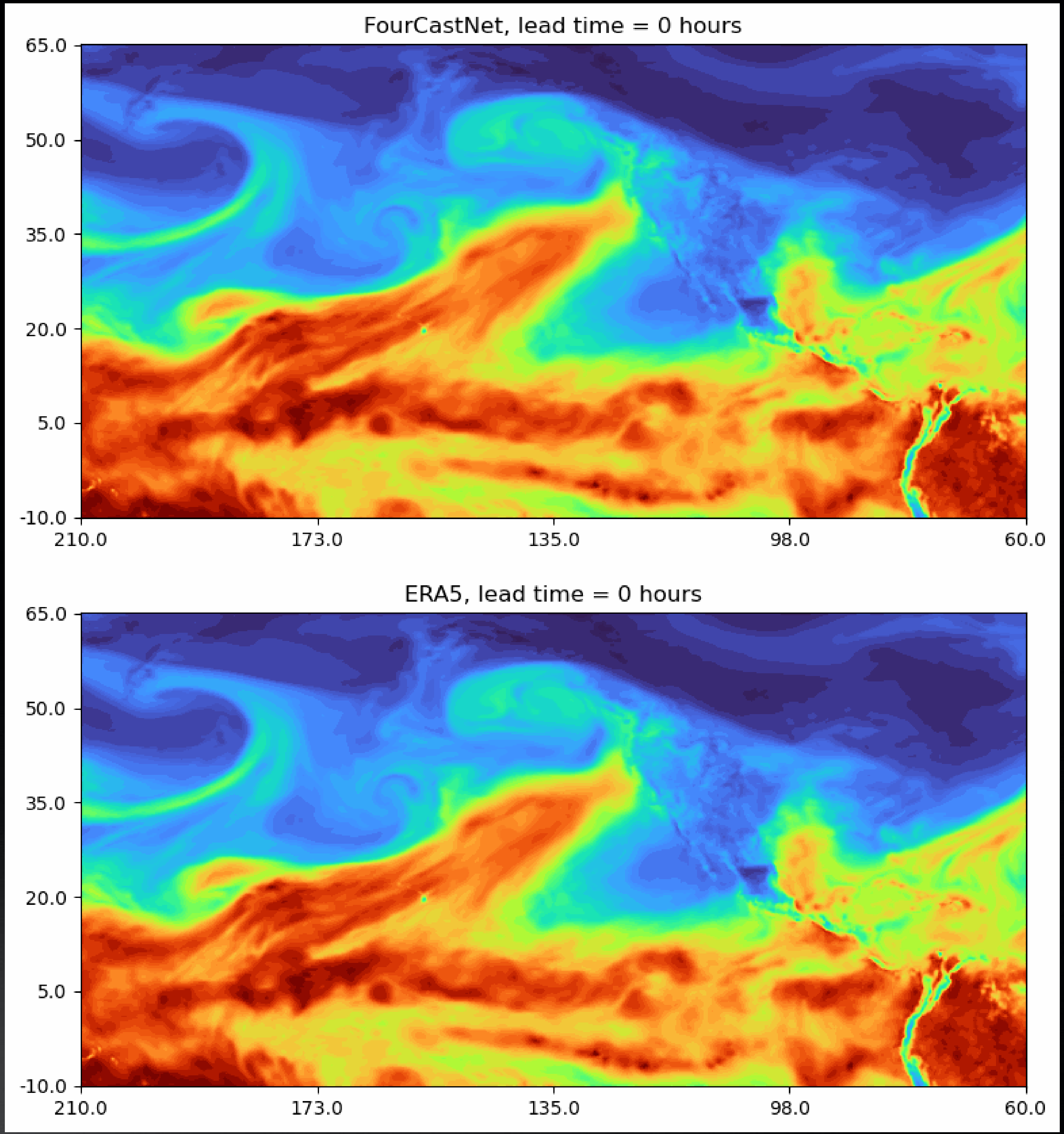


Note: Ground truth is ERA5, NOT observations

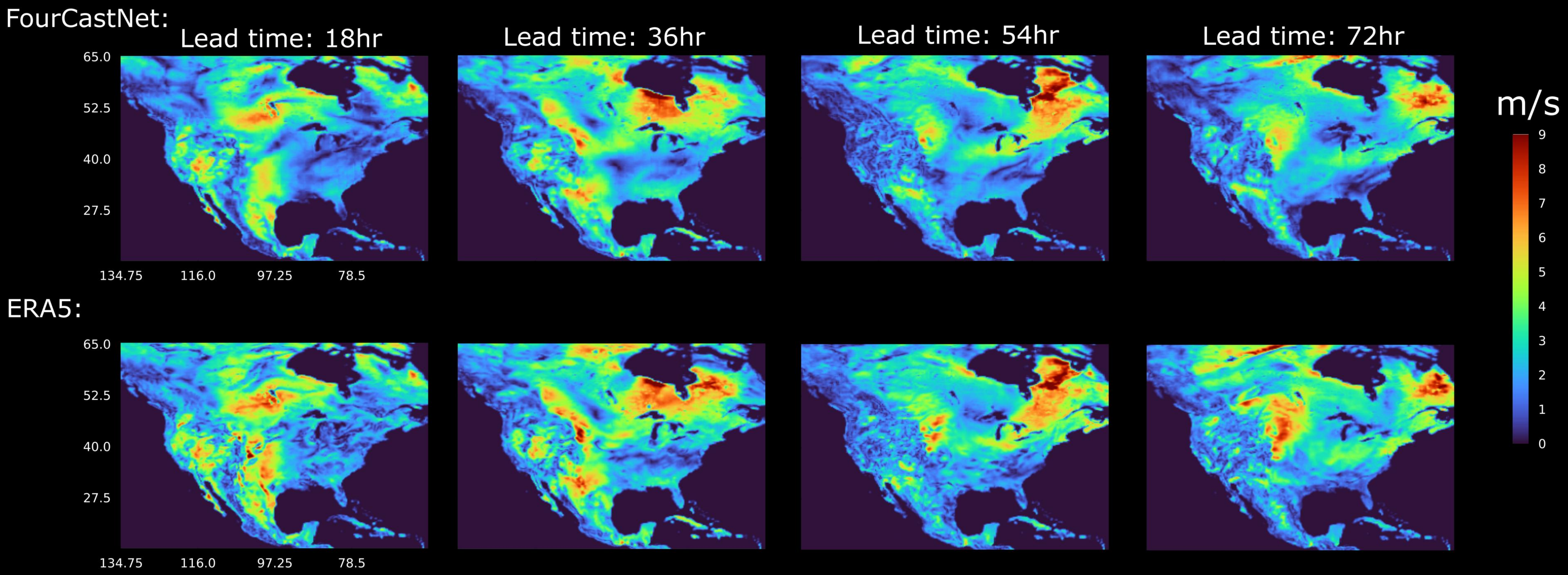
Pathak et al. (2022), FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators, arXiv:2202.11214

ACCURATE TOTAL COLUMN WATER VAPOR DYNAMICS

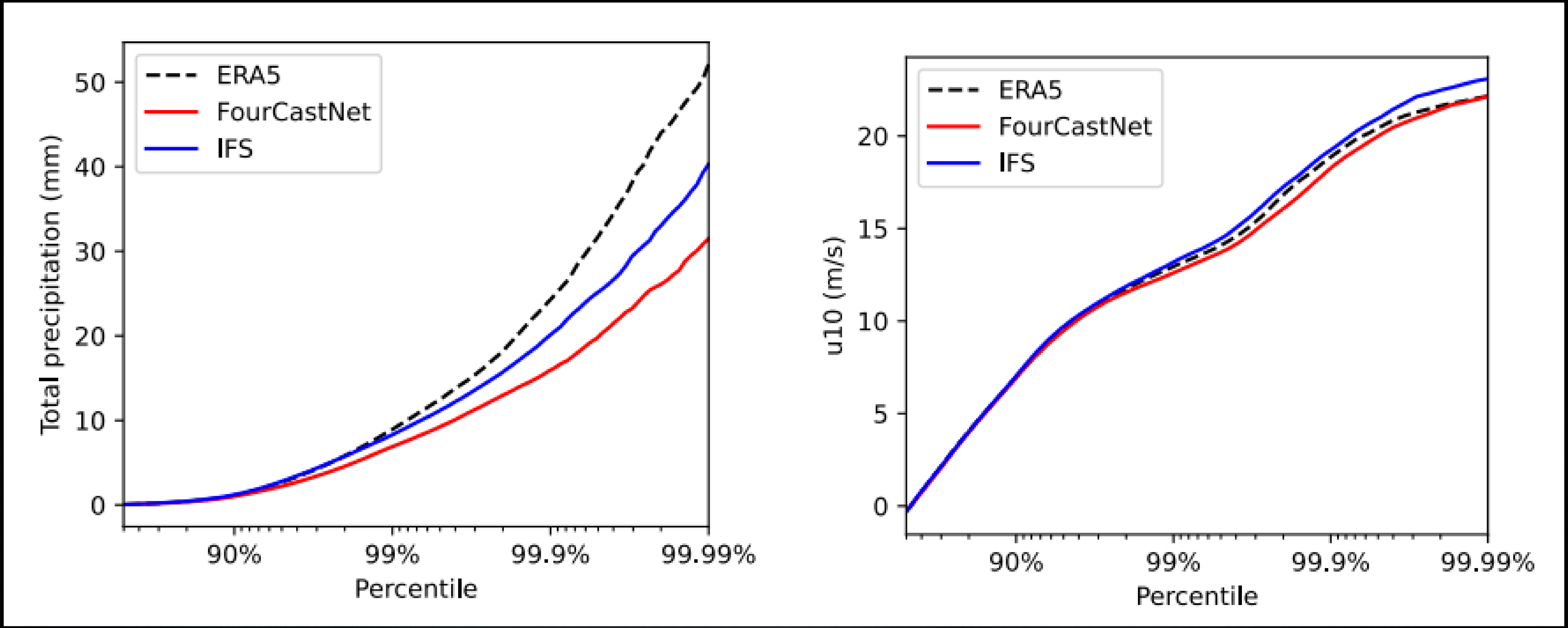




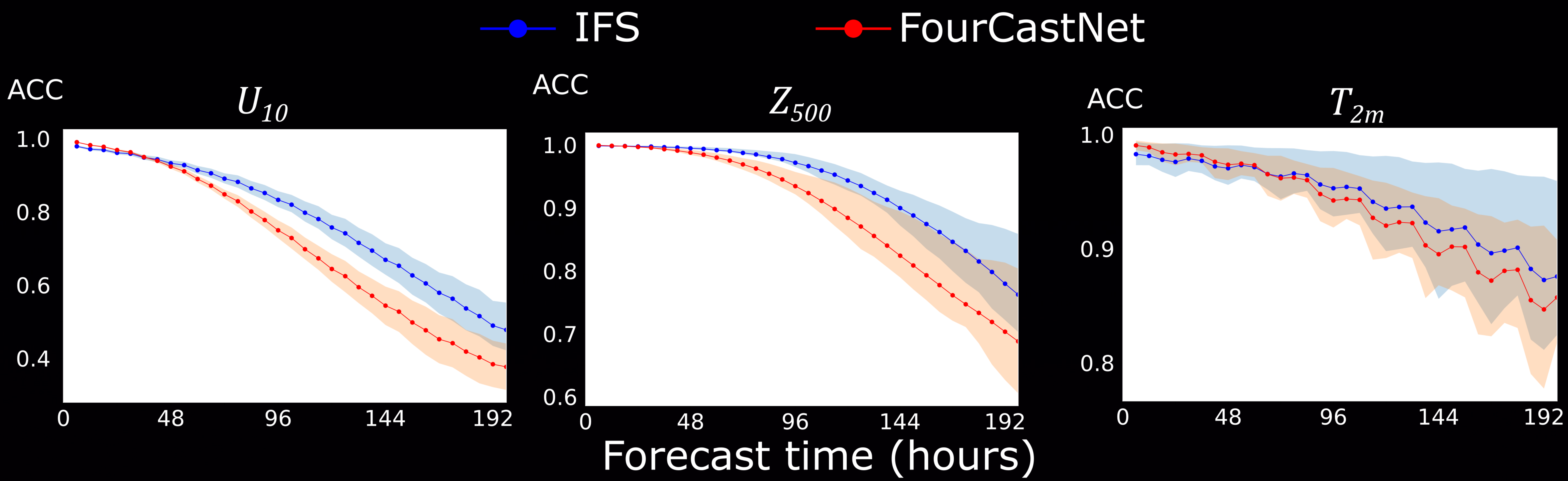
FOURCASTNET PREDICTS NEAR-SURFACE WIND FIELDS ACCURATELY: IMPORTANT IMPLICATIONS FOR WIND ENERGY PLANNING



PERFORMANCE ON EXTREME PRECIPITATION AND WIND EVENTS



SHORT-TERM FORECAST SKILL APPROACHING IFS

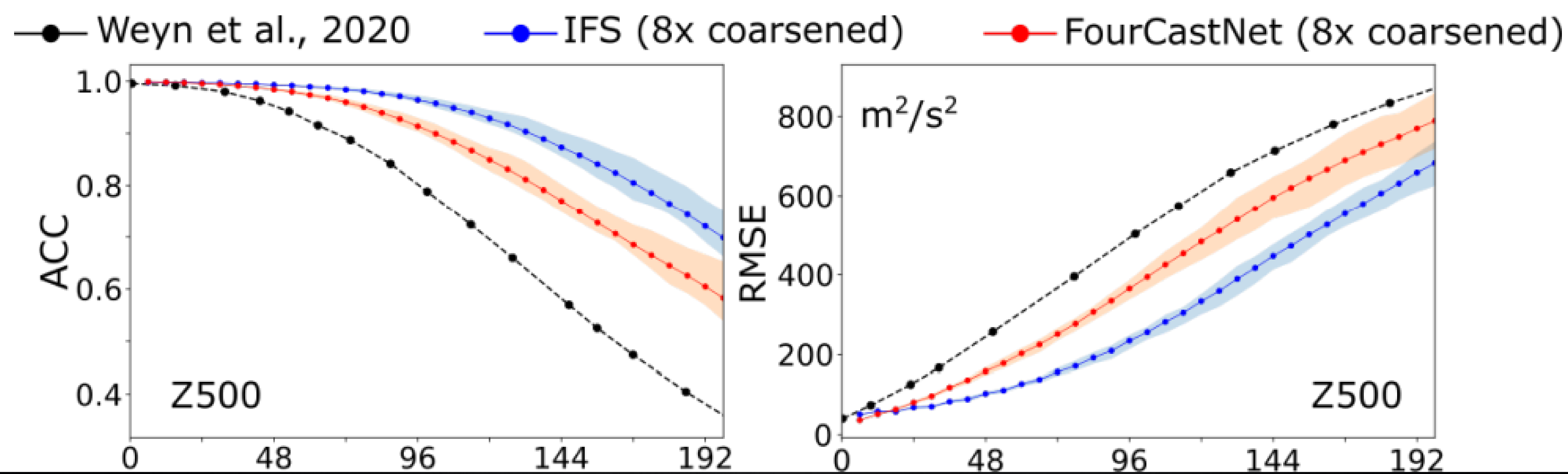


Note: Ground truth is ERA5, NOT observations

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

COMPARISON AGAINST STATE-OF-ART (DLWP, WEYN ET AL.)

8X higher resolution, significantly higher skill at weather timescales



Note: DLWP can predict reliably at S2S timescales

COMPUTATIONAL PERFORMANCE

Latency and Energy consumption for a 24-hour 100-member ensemble forecast				
	IFS	FCN - 30km (actual)	FCN - 18km (extrapolated)	IFS / FCN(18km) Ratio
Nodes required	3060	1	2	**1530**
Latency (Node-seconds)	984000	7	22	**44727**
Energy Consumed (kJ)	271000	7	22	**12318**

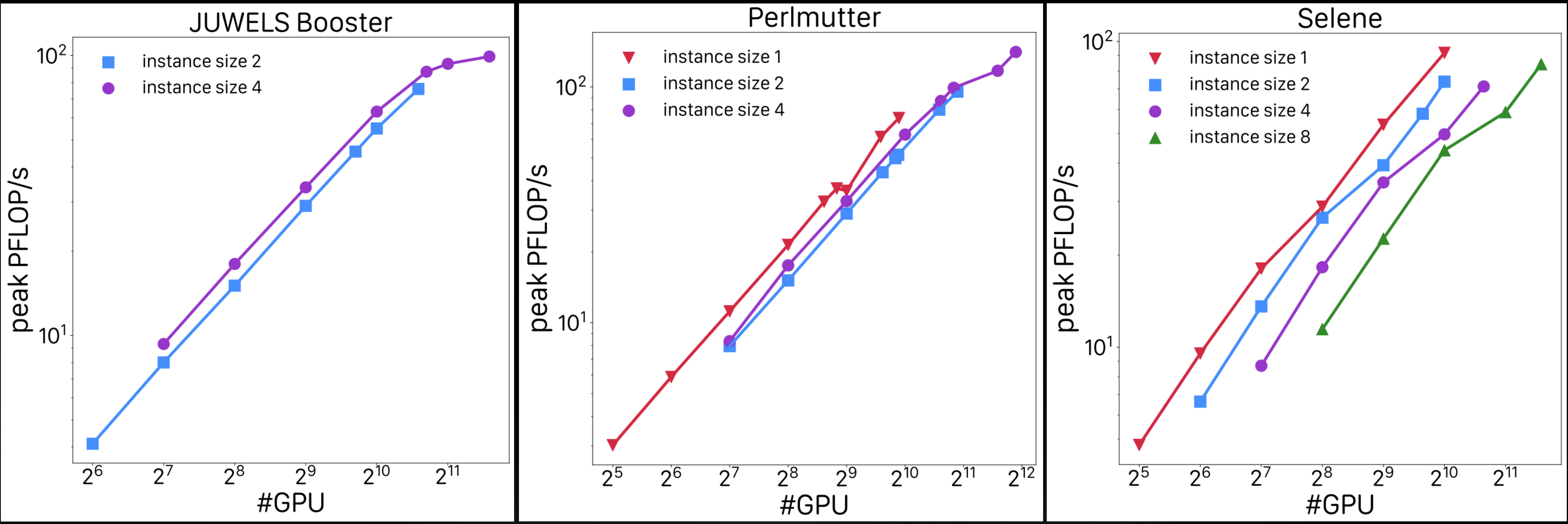
- 100-member ensemble forecast in 7 seconds
- 100-member ensemble forecast consumes 7 kJ
- 4 to 5 orders-of-magnitude speedup over NWP
- 4 orders-of-magnitude smaller energy footprint

Caveats

- FourCastNet is *not* physics constrained
- Orders-of-magnitude *fewer* variables and levels

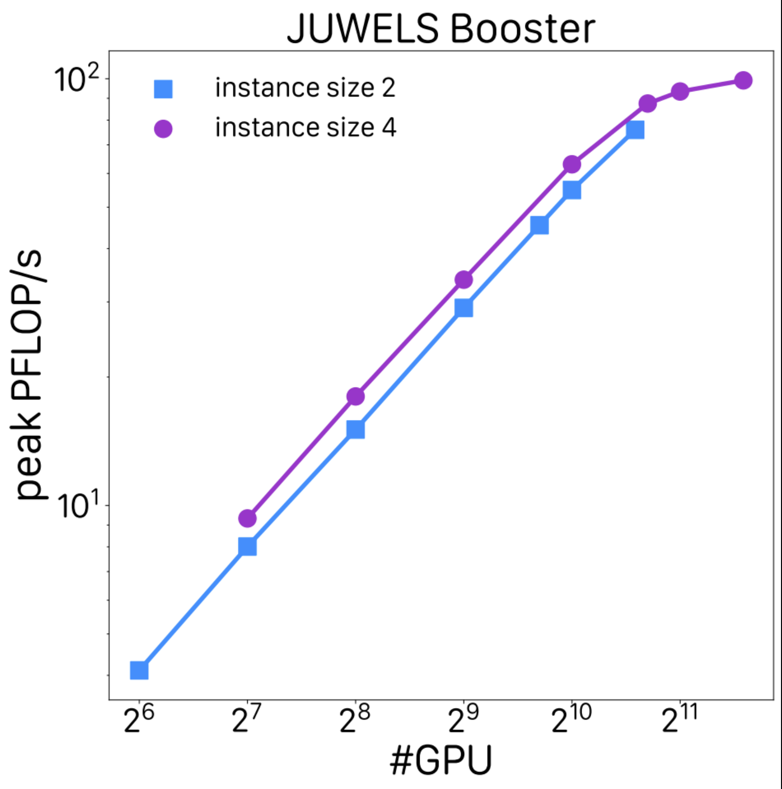
SCALING RESULTS (1 OF 3)

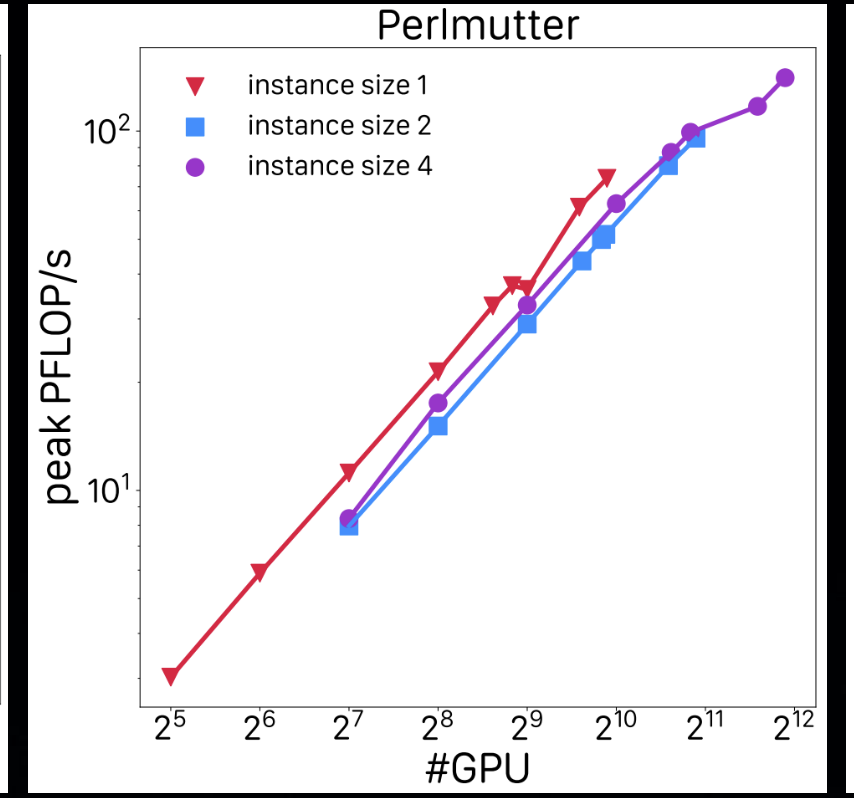
FourCastNet scaled efficiently upto ~ 4000 GPUs on three supercomputing systems:
JUWELS Booster, Perlmutter, and Selene

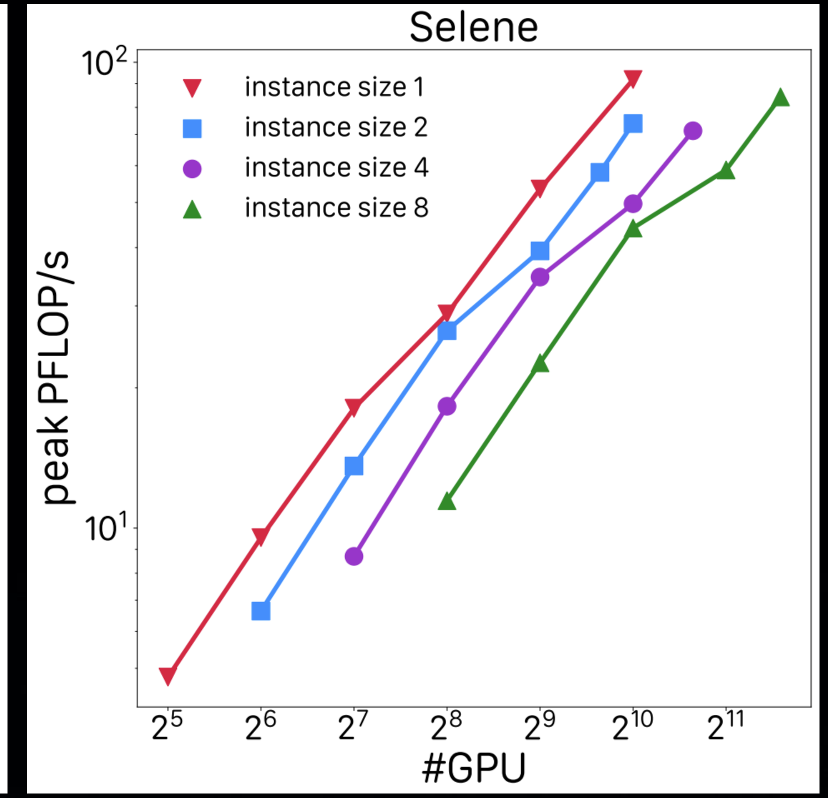


SCALING RESULTS (2 OF 3)

FourCastNet scaled efficiently upto ~ 4000 GPUs on three supercomputing systems: JUWELS Booster, Perlmutter, and Selene







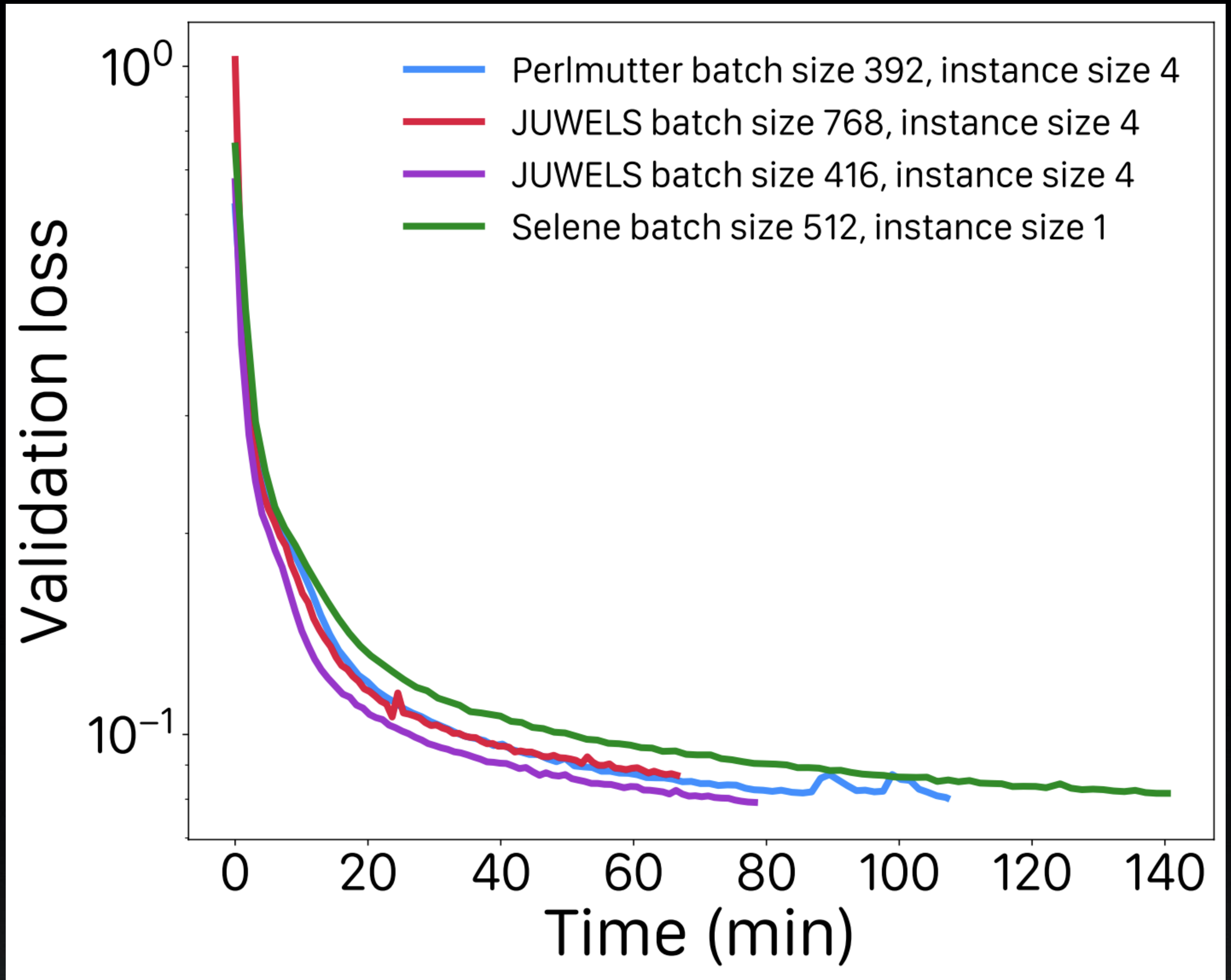
Peak performance is 140.8 petaFLOPS in mixed precision (averaged over a full epoch)

Model instance size	Phase	FP32	FP16	Tensor
1	training	28	6	8298
	validation	13	4	2840

FLOPS count per iteration per GPU for different precisions in units of 10^9 .

SCALING RESULTS (3 OF 3)

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



SCALING CHALLENGES HEADING TOWARDS 1 KM RESOLUTION

Model	Complexity (FLOPs)	Parameter Count	Interpretation
FNO	$Nd^2 + Nd \log N$	Nd^2	Global Conv.
AFNO	$Nd^2/k + Nd \log N$	$(1 + 4/k)d^2 + 4d$	Adaptive Global Conv.

p = patch size,
N = sequence size = ***dim_x * dim_y*** / ***p***^2;
d = embedding dimension, ***k*** = block count

	Current (25 km)	Intermediate (5 km)	Large (1 km)
N (p = 1)	1M	25M	625M
FFTs	720 x 1440 (***d*** of them)	3600 x 7200 (***d*** of them)	18k x 36k (***d*** of them)
Matrix Multiplies	[4***d*** x ***d***] * [***d***] (***N*** of them)	[4***d*** x ***d***] * [***d***] (***N*** of them)	[4***d*** x ***d***] * [***d***] (***N*** of them)

EARTH DIGITAL TWIN FOR EXTREME WEATHER

MONITOR | FUSE DATA | ASSIMILATE DATA

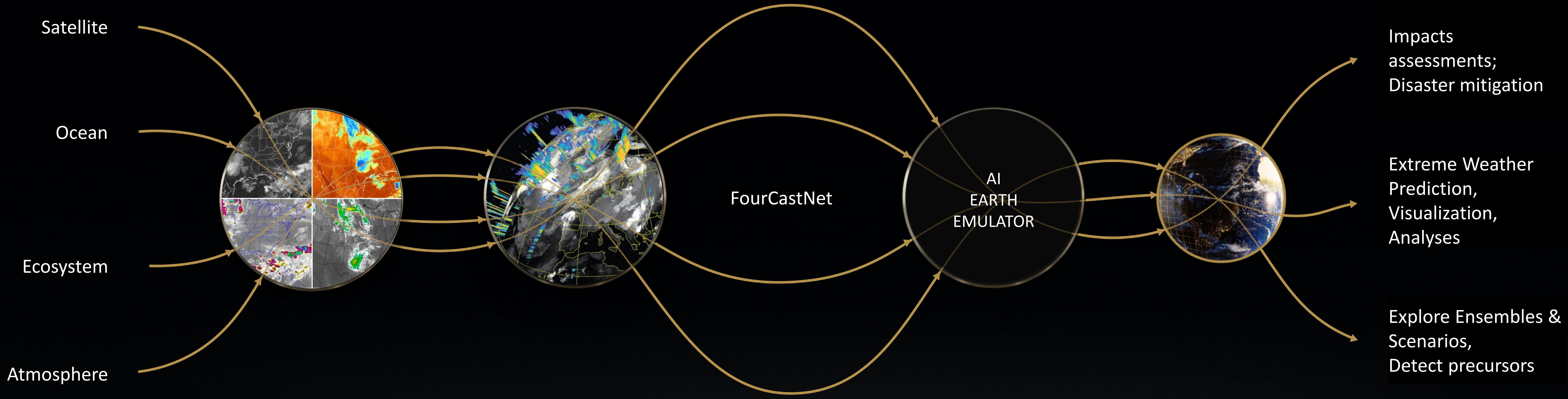
ERA5 Reanalysis Data (ECMWF) / Hi-res simulations

PHYSICS-ML EMULATION

Modulus Model Training and Inference

PREDICT | ANALYZE | VISUALIZE

Omniverse



Current Data Input:
Atmospheric winds and geopotential heights ~ 20 channels
10 TB, 30-km spatial resolution, 5 vertical pressure levels

Future Data Input:
Atmosphere, Ocean, Land, Ecosystems, Ice
10 PB, 5-km resolution, 20 vertical levels, 300 variables

Current Training and Inference:
16 hours on 128 GPUs, 0.25 seconds for 7-day forecast

Future Training and inference:
200 hours on 16384 GPUs, 4 seconds for 7-day forecast
(projected estimates – W.I.P)

1. Predict, visualize, detect and track extremes
2. Compare skill to traditional NWP models
3. Checkpoint / restart ensembles around events
4. Assess extreme weather impacts, mitigate disasters
5. Interactively investigate impact of changing climate scenarios on behavior of extremes
6. Detect precursors of extremes

NVIDIA

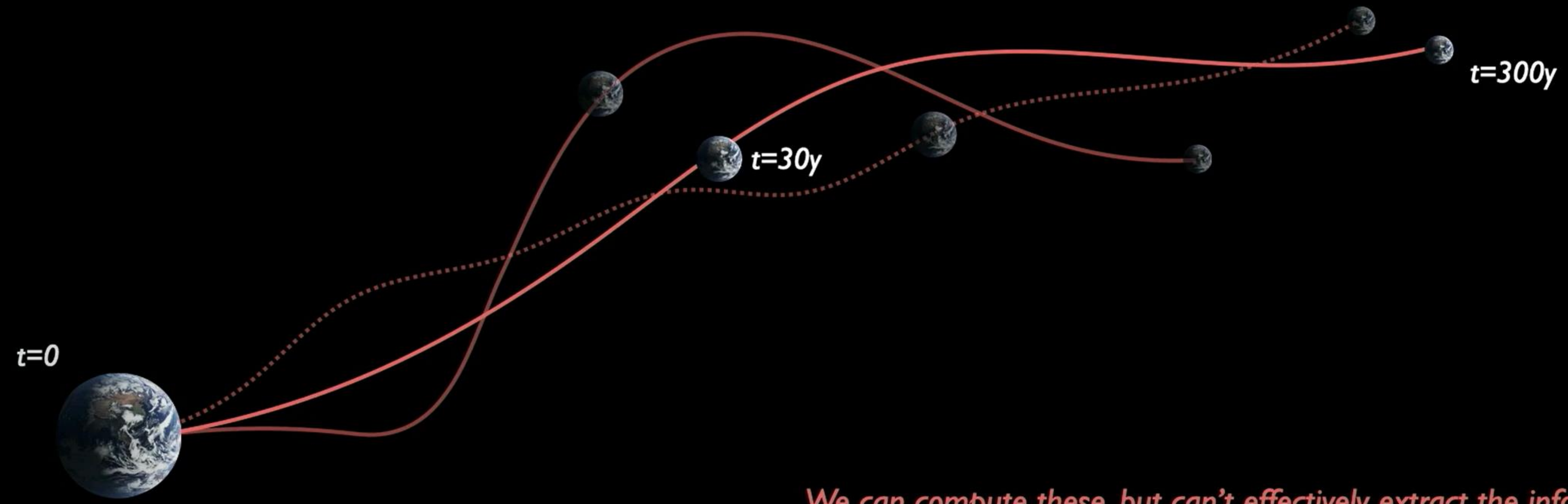
TODAY

- Unprecedented skill
- 1000-member ensemble in seconds
- 4 to 5 orders-of-magnitude speedup over NWP
- 4 orders-of-magnitude smaller energy footprint

LOOKING AHEAD

- Physics constraints
- Full state vector (ocean, land, sea ice)
- Higher resolutions
- Generative models for fine-scales
- Uncertainty Calibration
- Observational ground truth / diagnostics
- Weather ⟶ Climate
- ML Data Assimilation

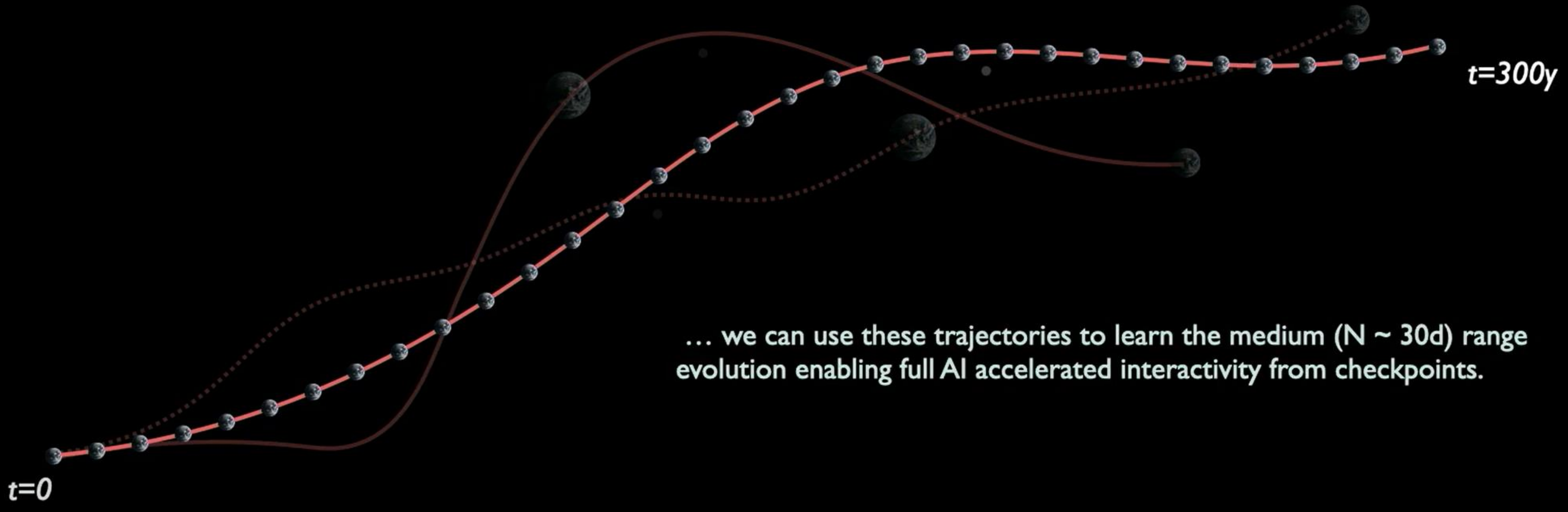
Climate as a trajectory in a Tera-dimensional (10^{12}) trajectory phase space



We can compute these, but can't effectively extract the information content from an XByte trajectory, let alone interact with it.

Source: Bjorn Stevens' keynote on "Digital Twins for Understanding and Adapting to Climate Change" at GTC Spring 2022

If we can compute these trajectories

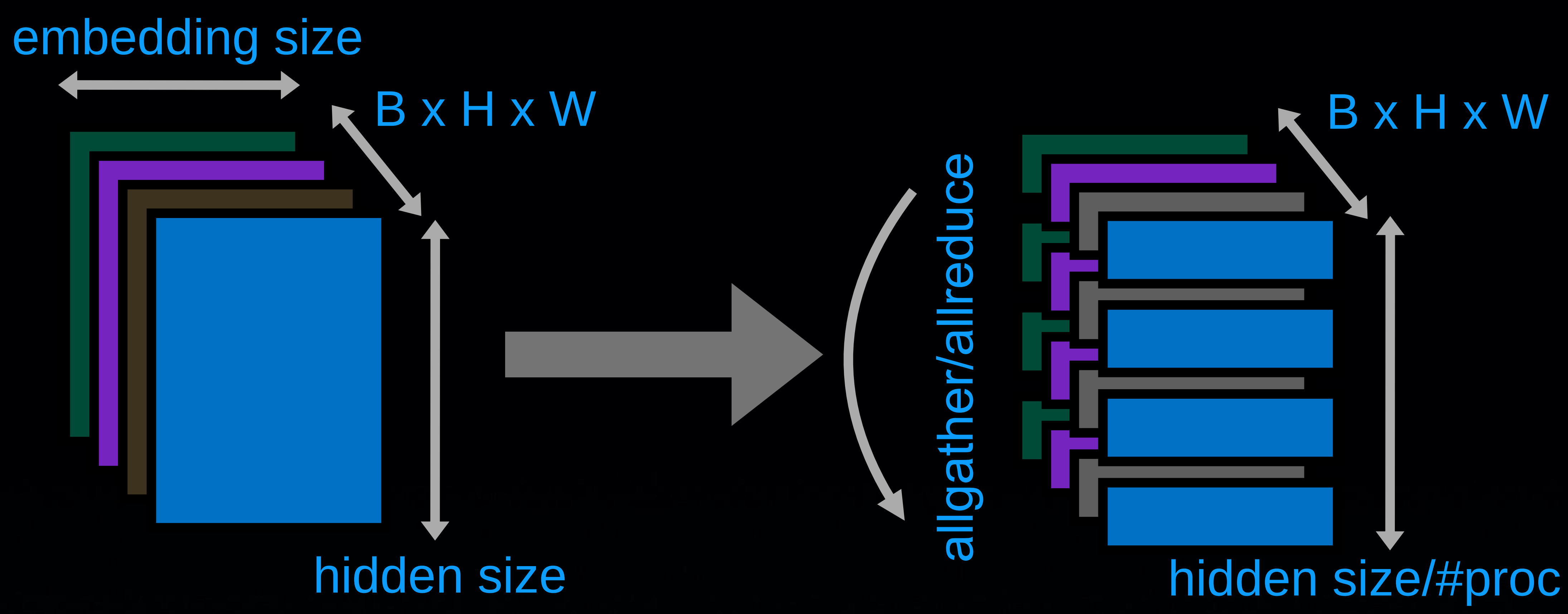


… we can use these trajectories to learn the medium (N ~ 30d) range evolution enabling full AI accelerated interactivity from checkpoints.

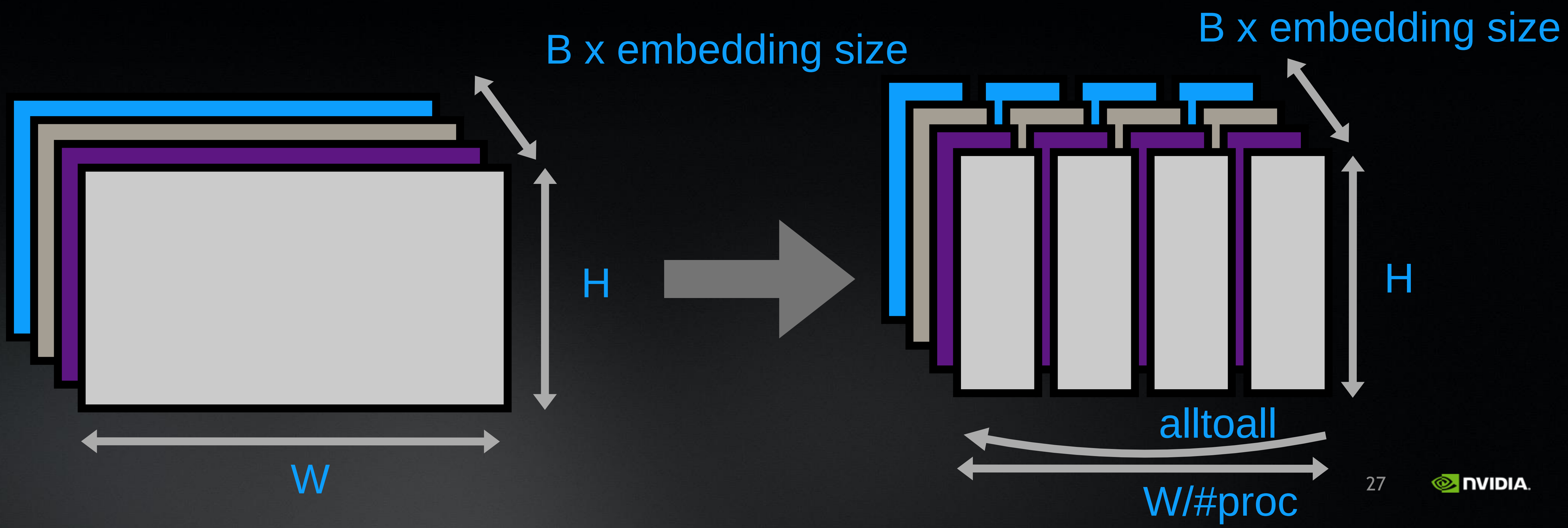
If only as a first step for learning the entire system $N \rightarrow \infty$

PARALLELIZATION METHODS I

Feature parallelization
splitting channel dim, dense layers become distributed matrix multiplications



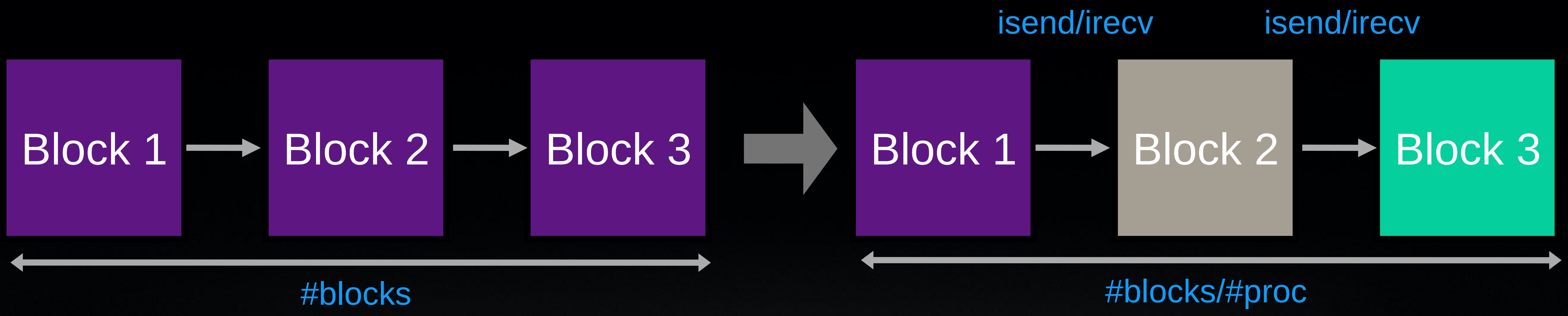
Domain decomposition
splitting height and width, FFTs become distributed, LayerNorm needs to exchange stats



NVIDIA

PARALLELIZATION METHODS II

Layer pipelining
place only a few transformer blocks
on each GPU



Data parallelism
split the global batch across GPU

Hybrid parallelism
mix and match different types