NVIDIA

BUILDING EARTH DIGITAL TWINS FOR NVIDIA'S EARTH-2 INITIATIVE

Karthik Kashinath

Principal Engineer and Scientist, AI-HPC, NVIDIA

Engineering Lead, Earth-2, NVIDIA

AGENDA

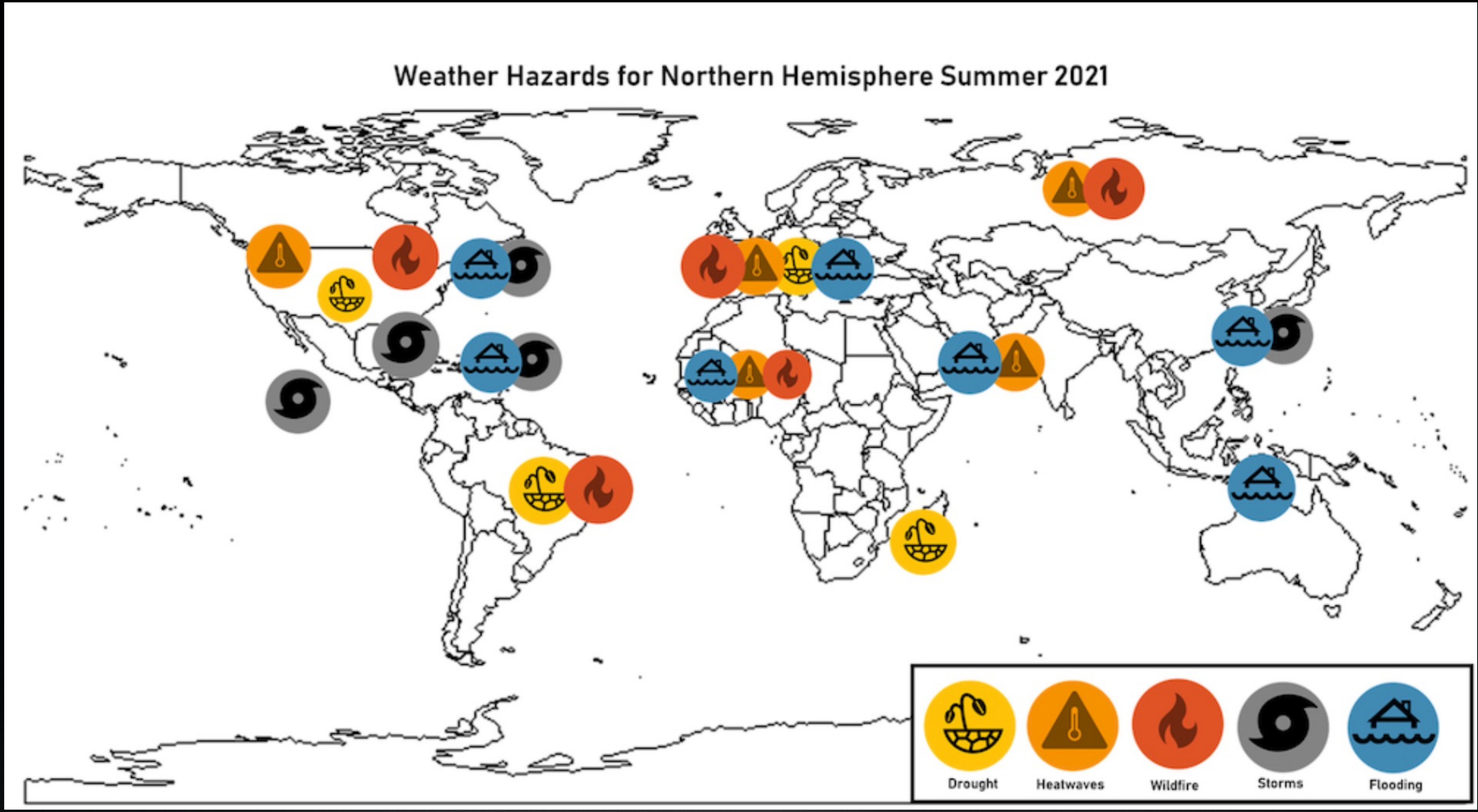
WHY CLIMATE SCIENCE NEEDS DIGITAL TWINS

CHALLENGES AND OPPORTUNITIES FOR PHYSICS-INFORMED ML

FOURCASTNET: A STEP TOWARDS DIGITAL TWIN EARTH

CHALLENGE: DRAMATIC RISE IN EXTREME WEATHER ACROSS THE GLOBE

EXTREME WEATHER EVERYWHERE



CLIMATE SCIENCE REQUIRES MILLION-X SPEEDUPS

Computational constraints limit model resolution

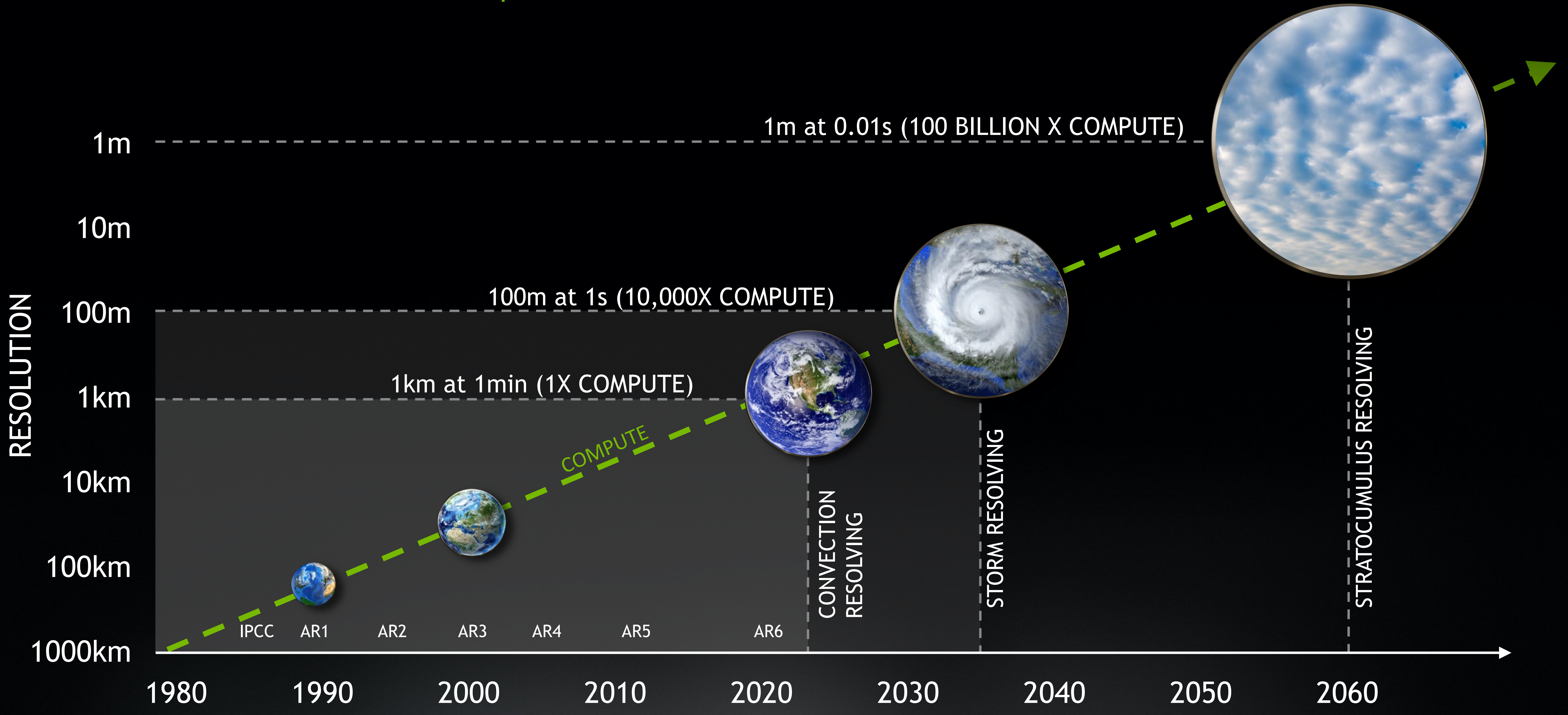
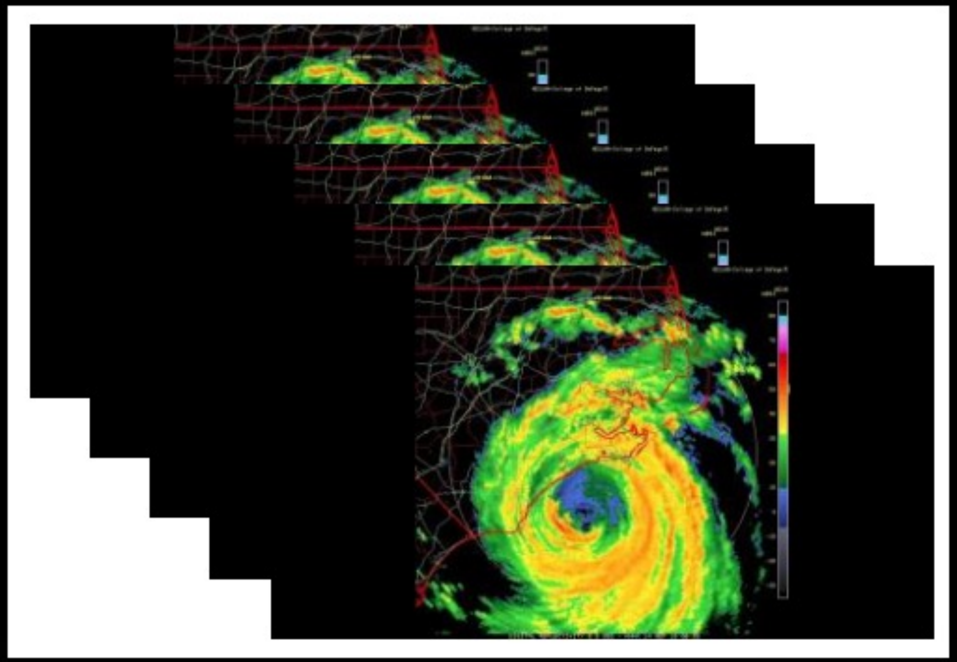


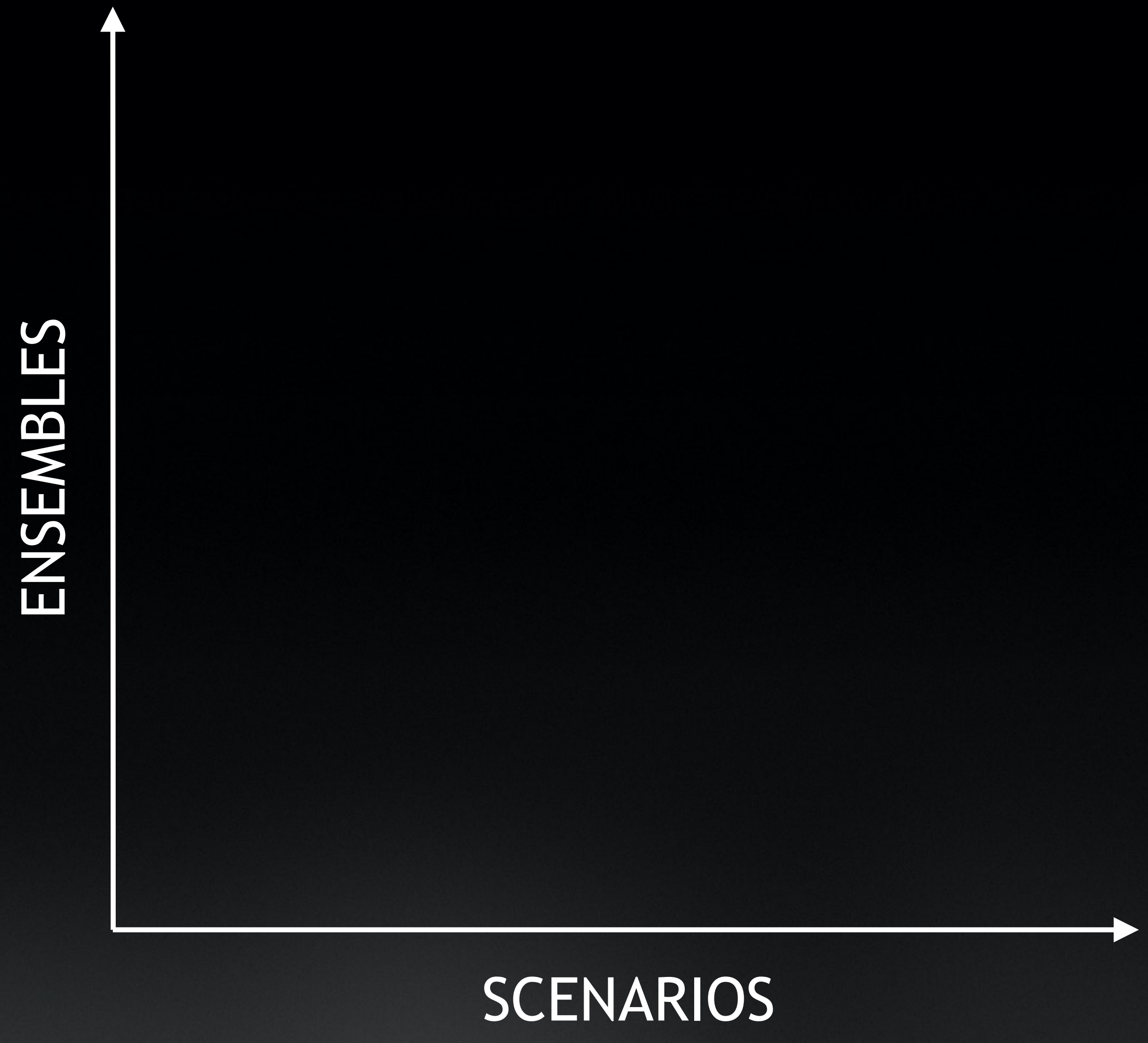
Figure adapted from: Schneider, T., Teixeira, J., Bretherton, C. *et al.* Climate goals and computing the future of clouds. *Nature Climate Change* **7,** 3–5 (2017). https://doi.org/10.1038/nclimate3190

CLIMATE SCIENCE REQUIRES MILLION-X SPEEDUPS

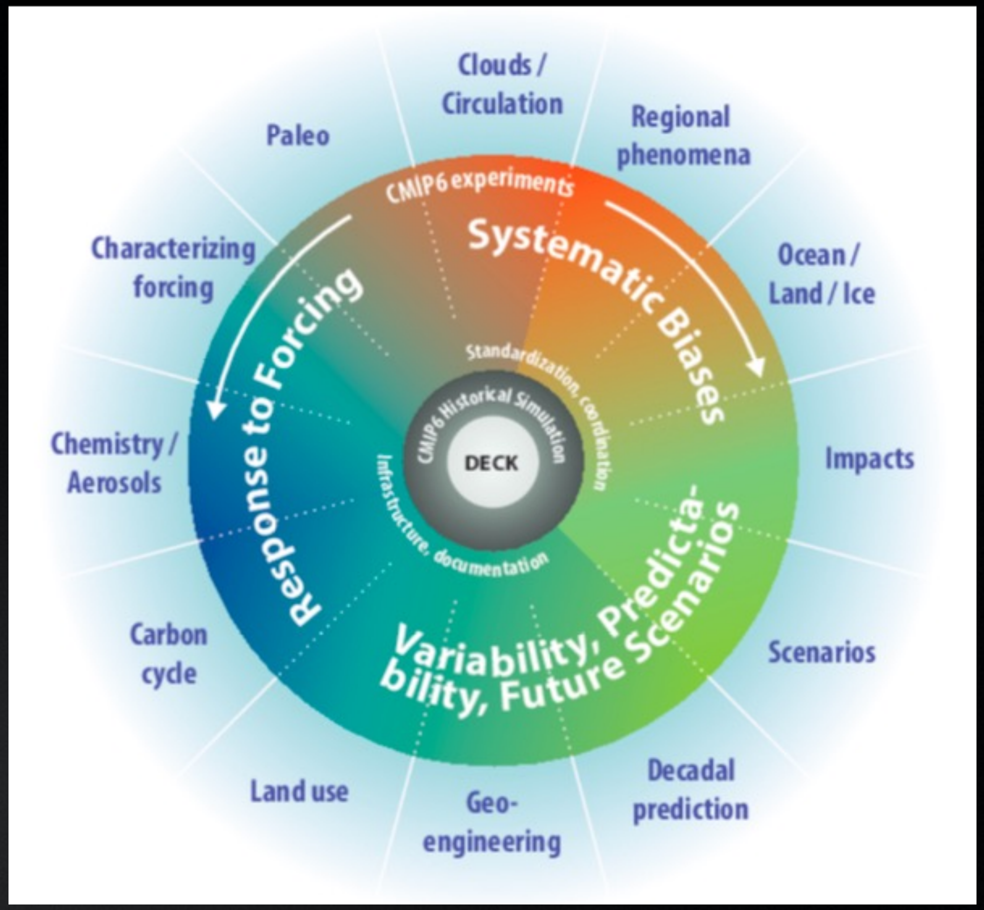
Computational constraints limit the size of ensembles and how many scenarios can be explored

10s —> 1000s OF MEMBERS



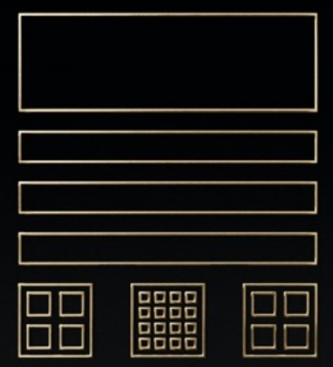


10s —> 1000s OF SCENARIOS

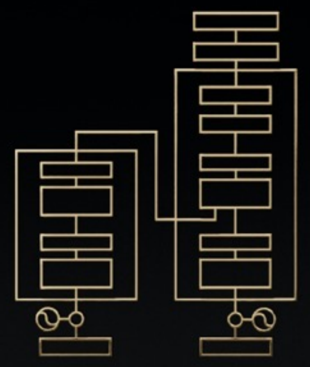


ADVANCES IN COMPUTING AND ML PROMISE MILLION-X SPEEDUPS

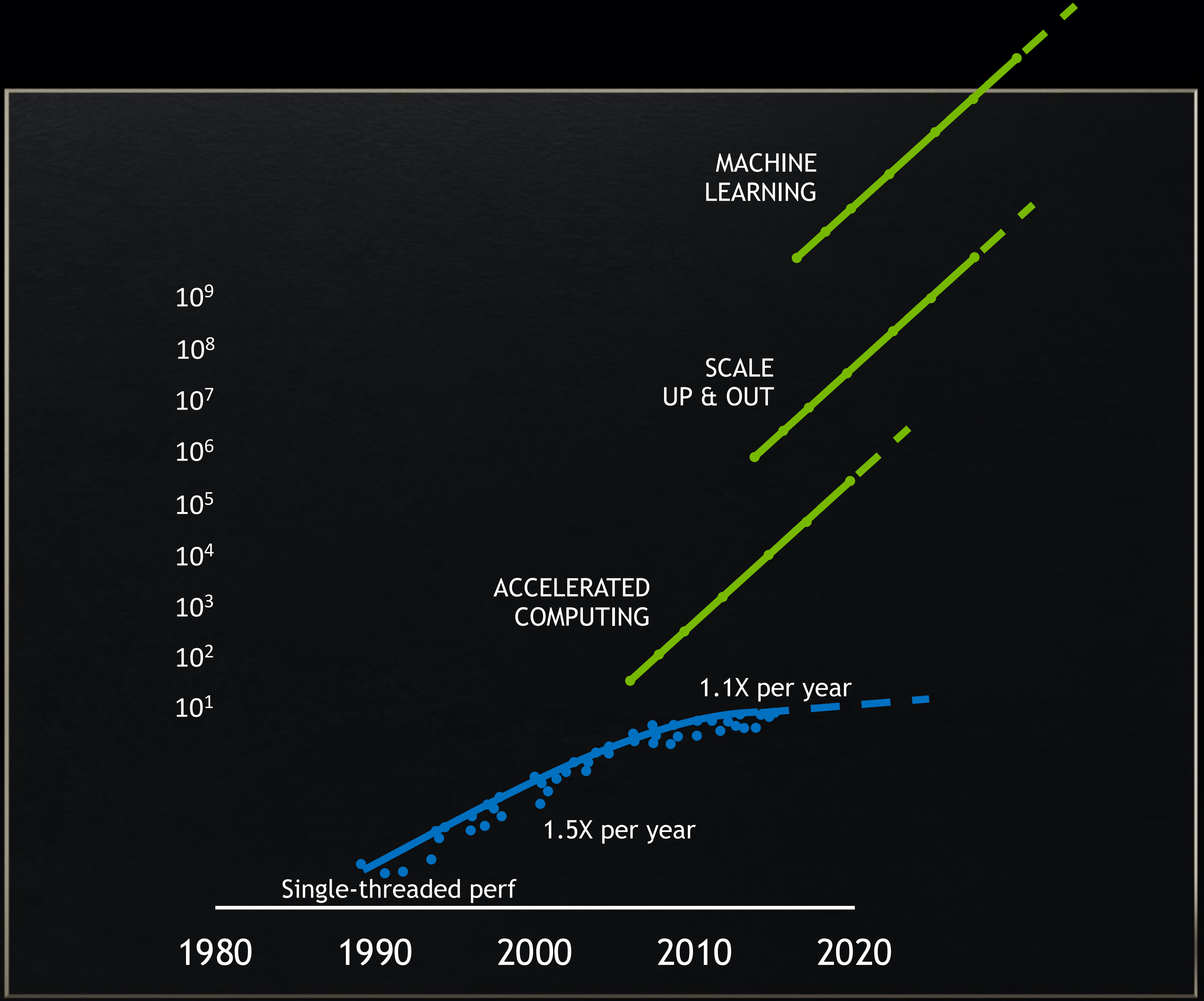
Accelerated Computing



AI



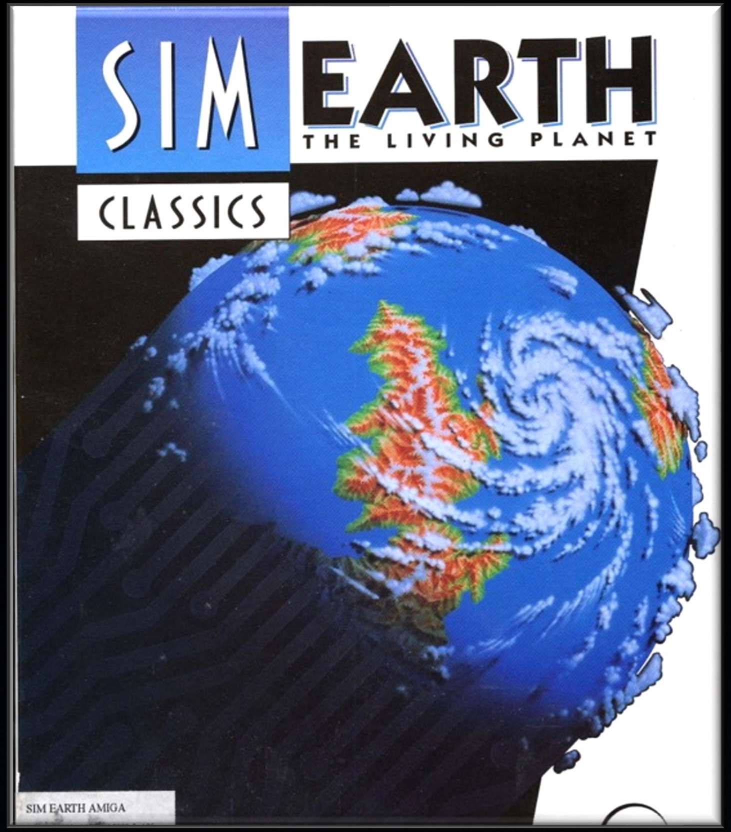
Data Center Scale



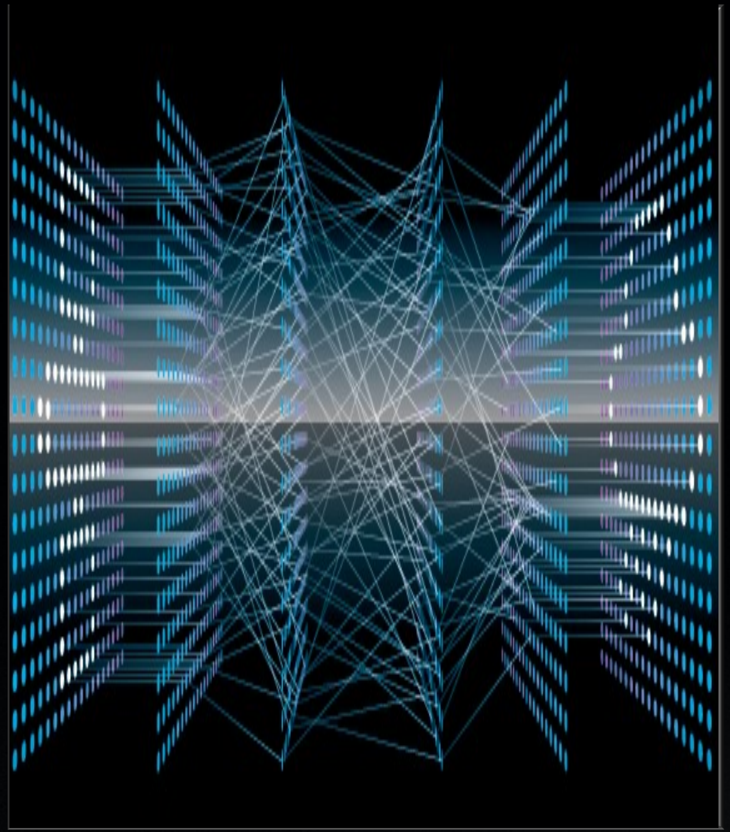
DESTINATION-EARTH: DIGITAL REPLICAS OF EARTH

Project DestinE envisions what Earth-system modeling could be

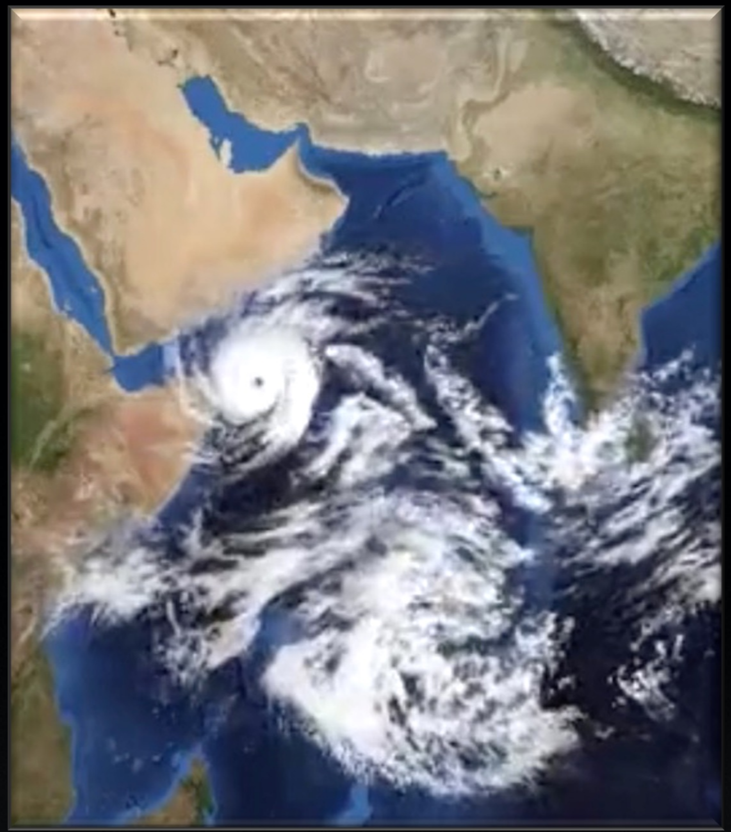
Intuitive User Interface: What-if Q & A



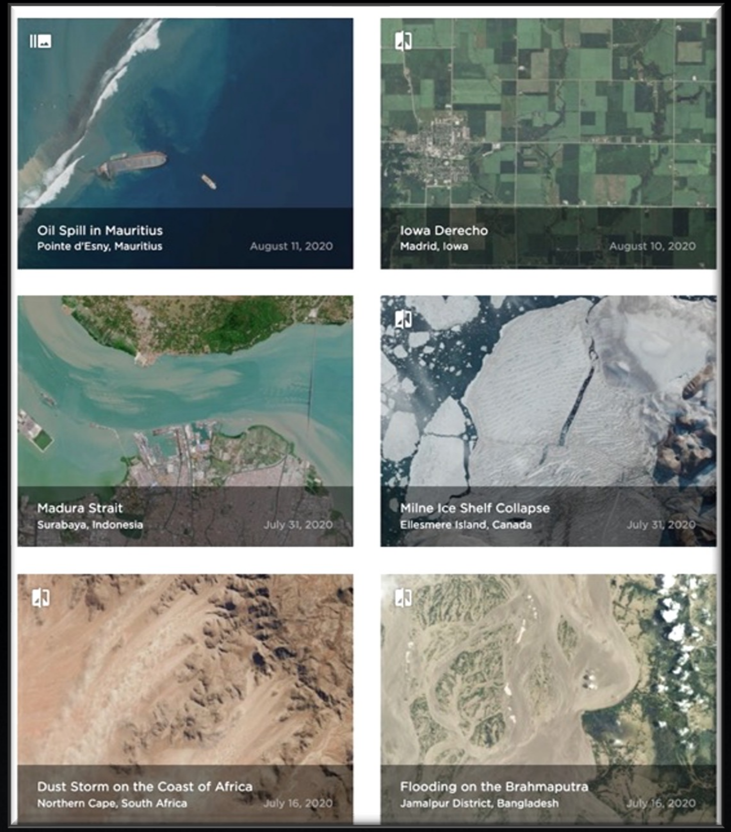
Data-driven Models



Storm-resolving Models



Unified Observations



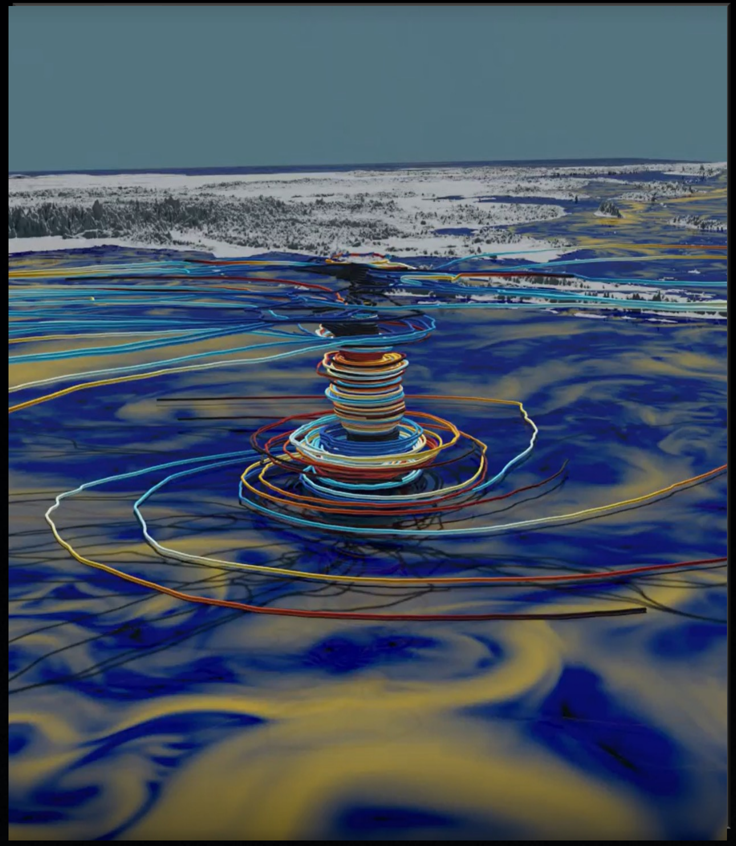
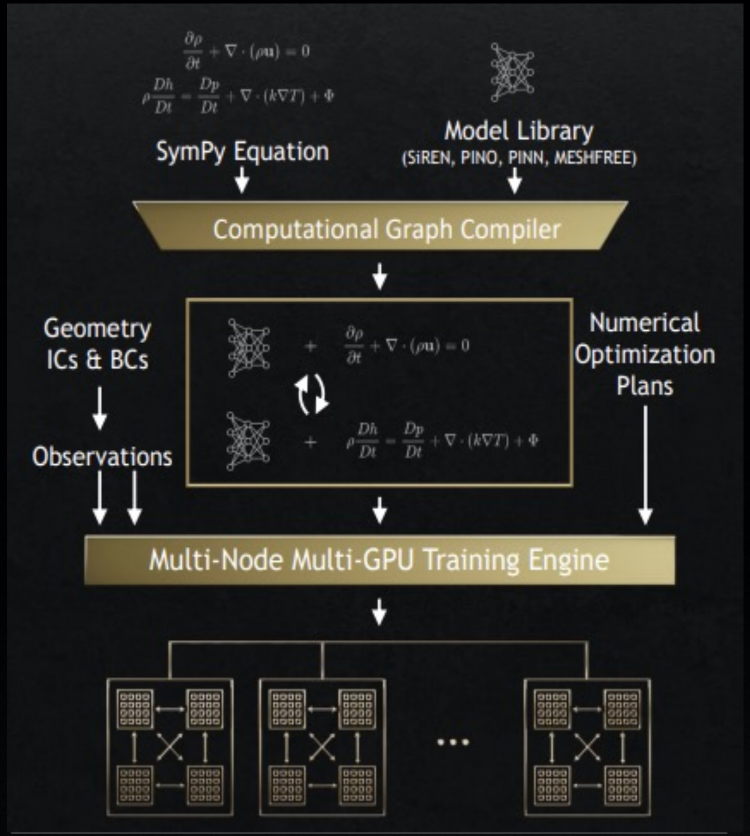
Exascale Compute

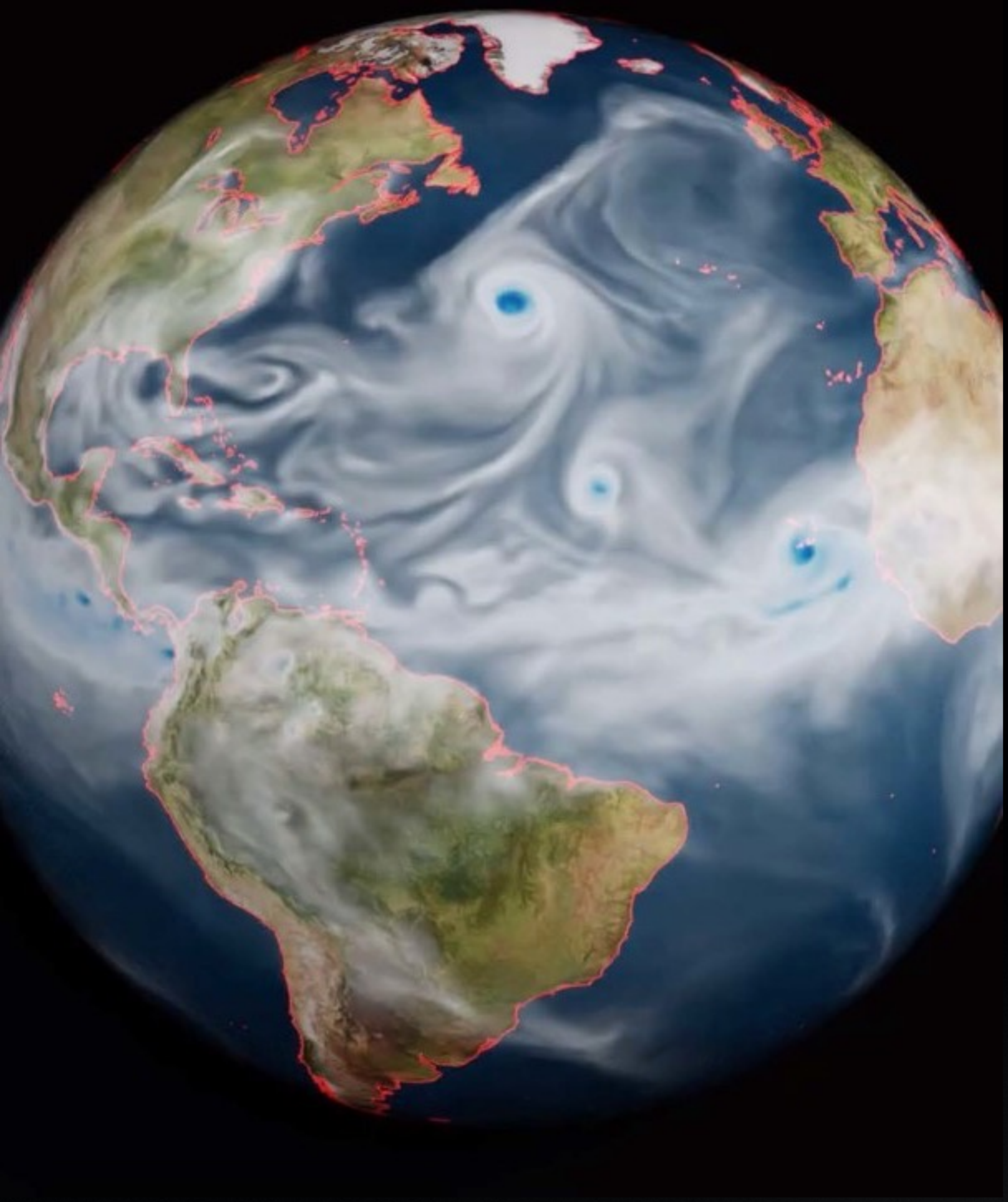
https://digital-strategy.ec.europa.eu/en/library/destination-earth

NVIDIA

DESTINATION-EARTH: DIGITAL REPLICAS OF EARTH

NVIDIA has technologies needed to make this vision a reality

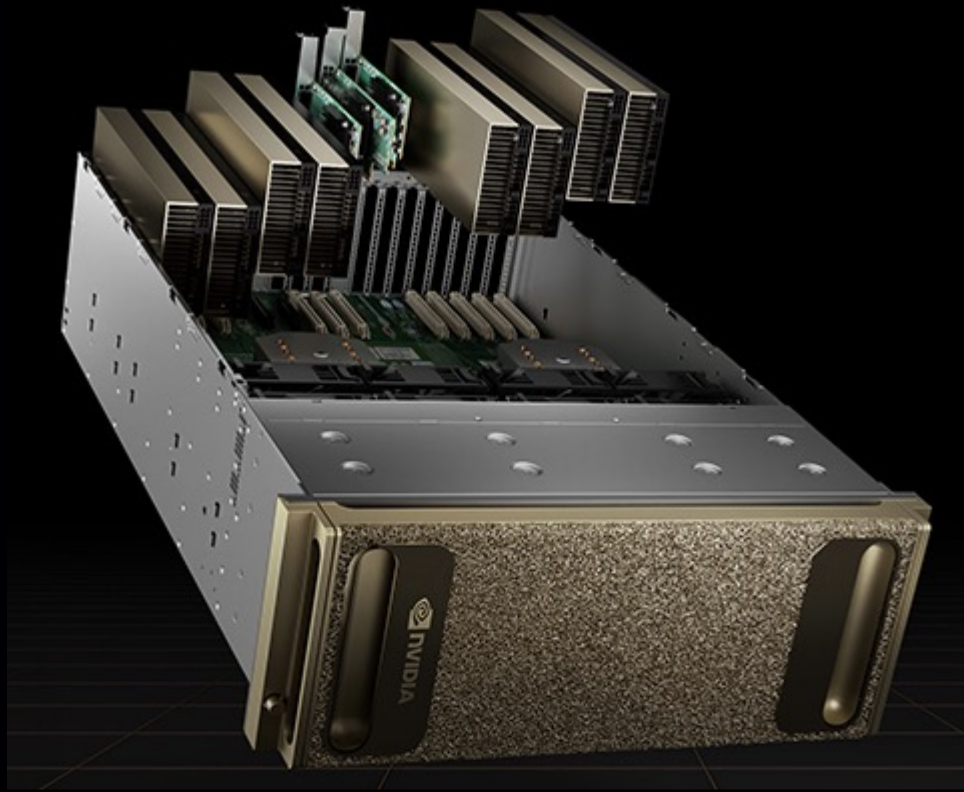
Intuitive User Interface: What-if Q & A	Data-driven Models	Storm-resolving Models	Unified Observations	Exascale Compute
				
OMNIVERSE	PHYSICS-ML / MODULUS	GPU-ACCELERATION	OMNIVERSE NUCLEUS	OVX SUPERPOD



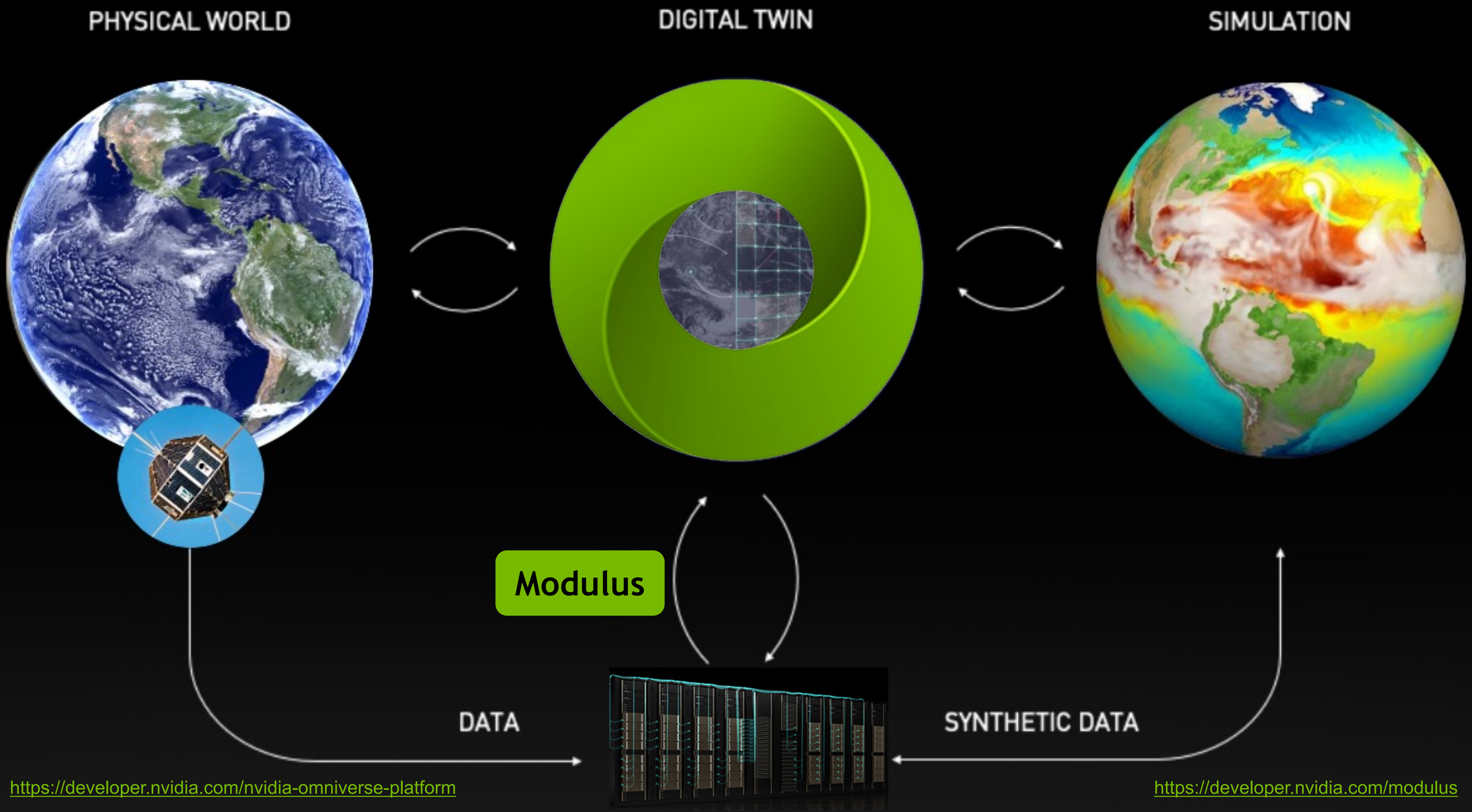
Earth-2

WHY?
INTERACTIVITY AT SCALE: UNFOLD AND EXTRACT INFORMATION

HOW?
DIGITAL TWINS TO MONITOR, PREDICT, MITIGATE, AND ADAPT



https://blogs.nvidia.com/blog/2021/11/12/earth-2-supercomputer/



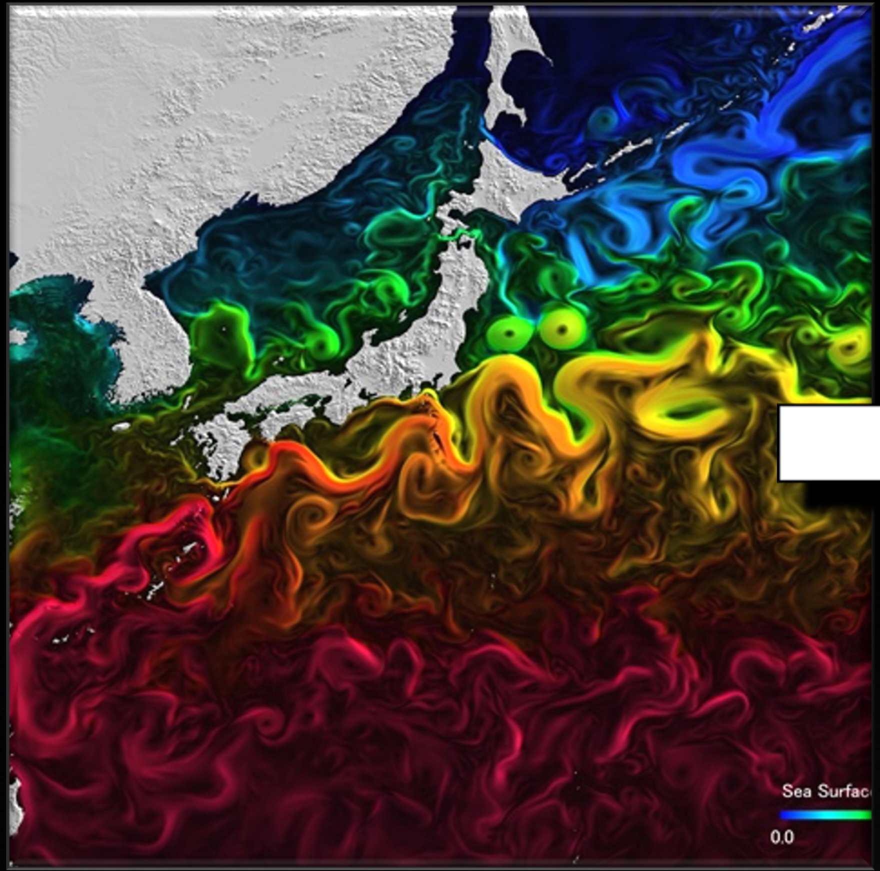
https://developer.nvidia.com/nvidia-omniverse-platform

https://developer.nvidia.com/modulus

DIGITAL TWIN USES
MODEL COUPLING TO ANSWER WHAT-IF QUESTIONS

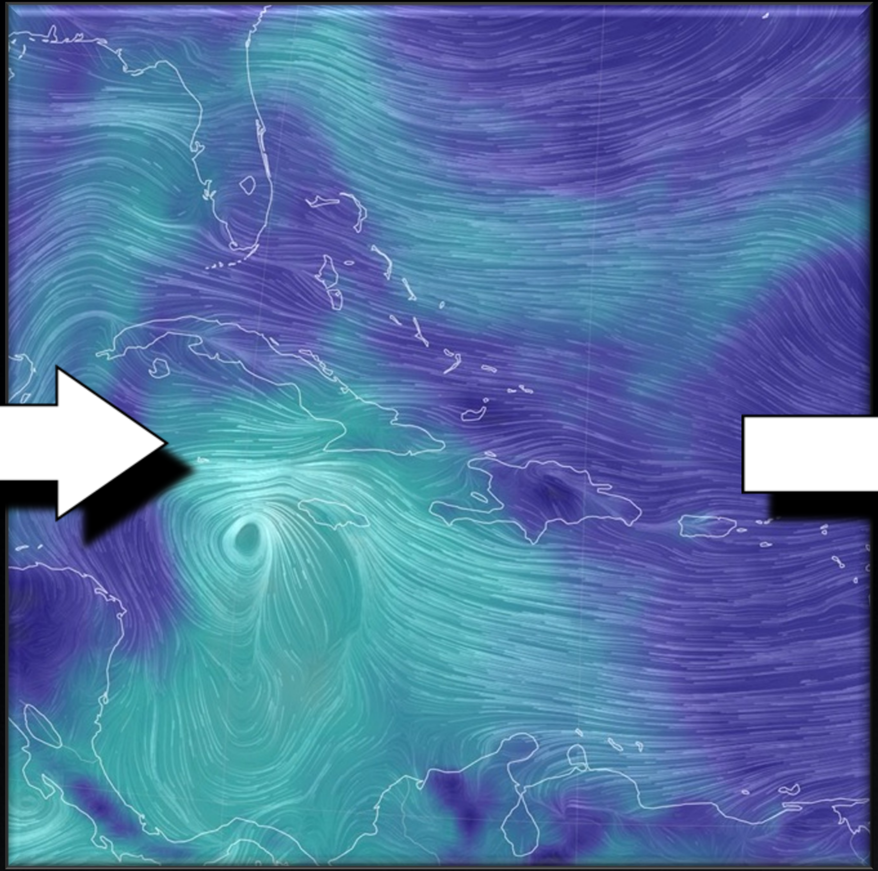
Coupling fast and accurate ML models together enables end-to-end analyses

Climate Twin



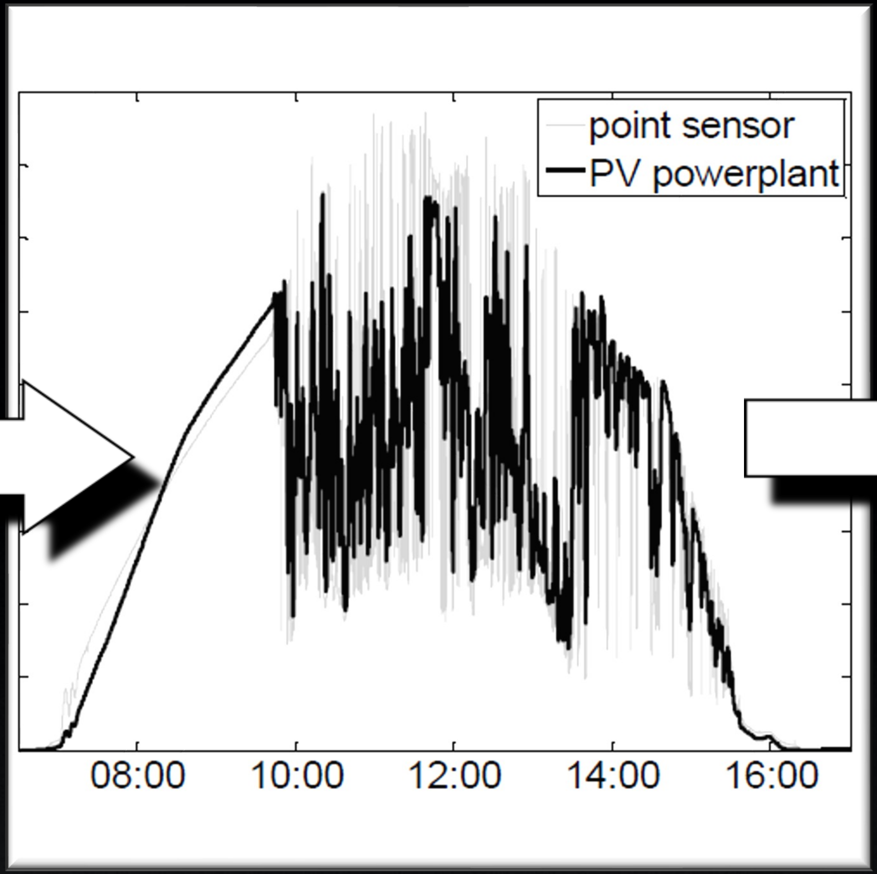
http://www.jamstec.go.jp/gallery/e/simulation/weather/002.html

Weather Twin



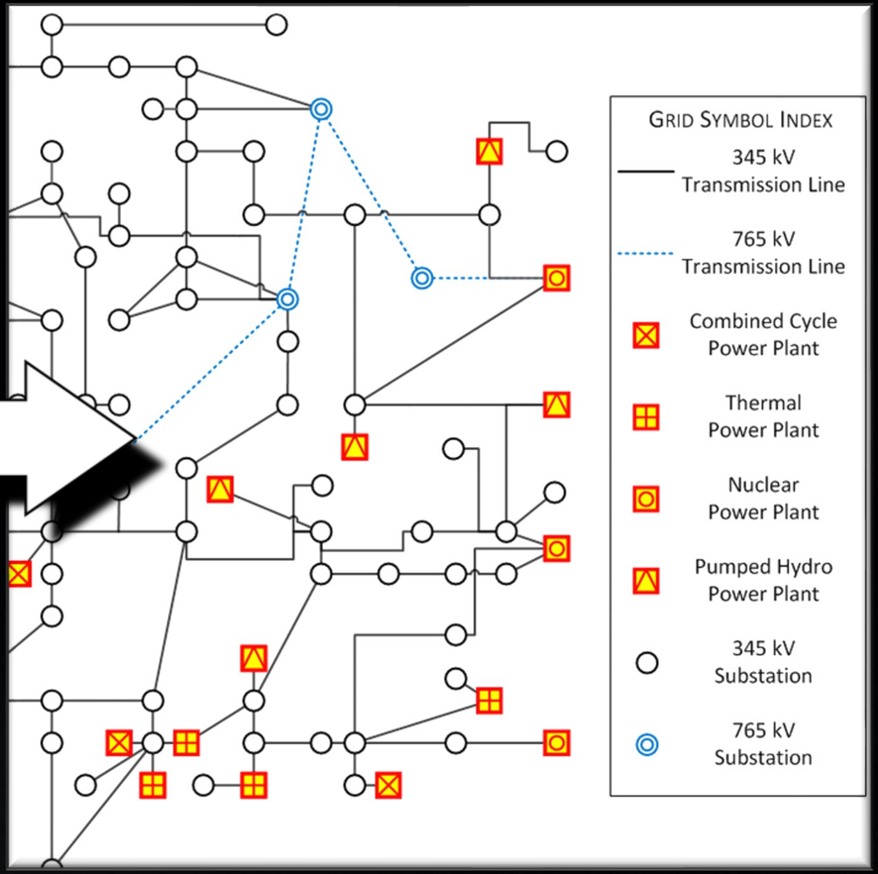
https://earth.nullschool.net/#current/wind/surface/level/overlay=total_precipitable_water/orthographic=-64.71,19.12,2322

Powerplant Twin



https://www.esig.energy/wiki-main-page/simulating-solar-power-plant-variability-a-review-of-current-methods/

Electric Grid Model



https://www.researchgate.net/figure/Overall-configuration-of-electric-power-grid-modelling_fig2_311217737

EARTH DIGITAL TWIN
MACHINE LEARNING CHALLENGES AND APPROACHES

CHALLENGES

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

APPROACHES

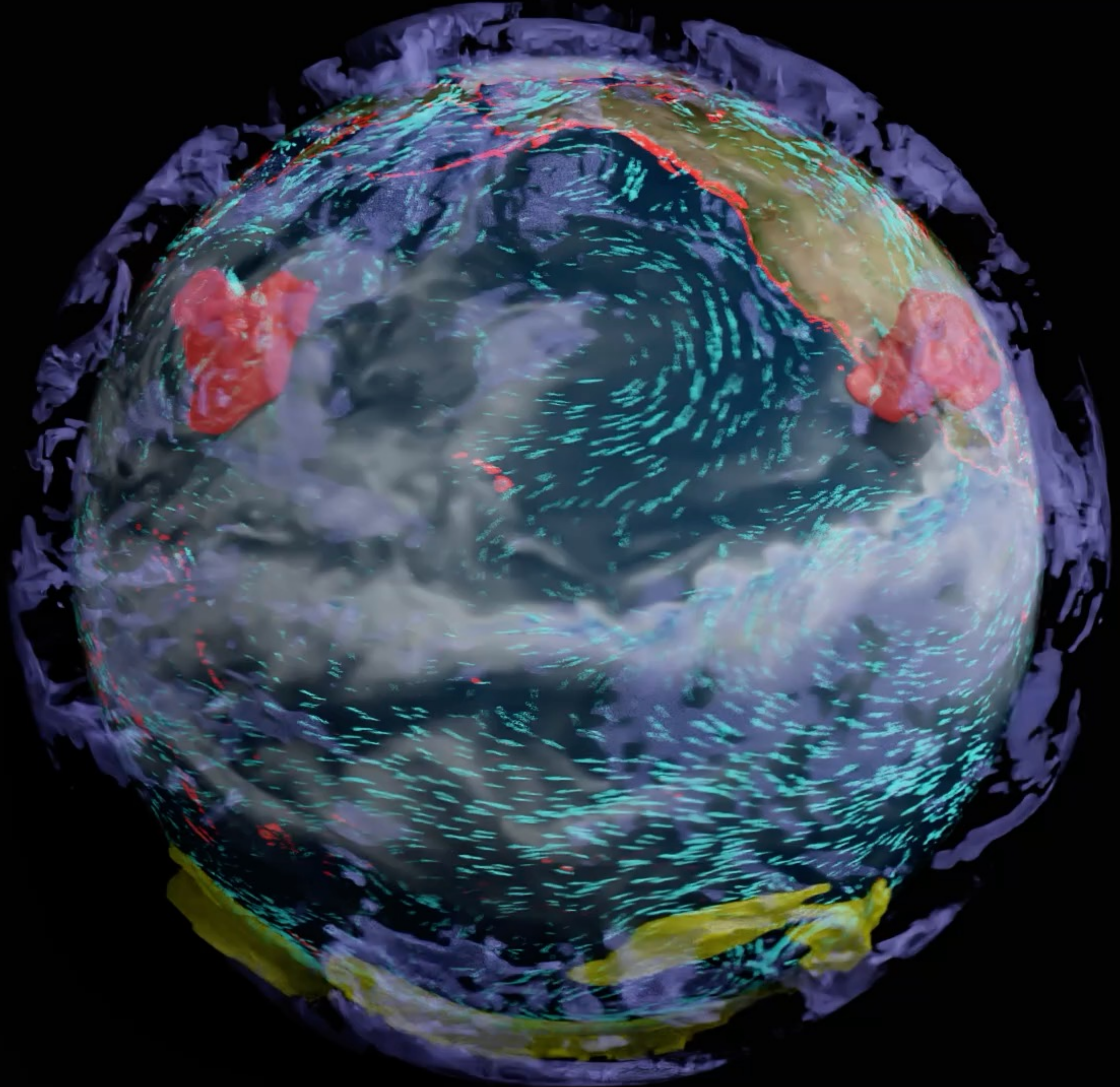
- Emulation
- Super-resolution
- Segmentation
- Online learning
- Reinforcement Learning
- …

YEAR 2100

+3 C Global Temperature

+60% Extreme Tropical Cyclones

+400% Extreme Atmospheric Rivers



2100, SEP 14

- TROPICAL CYCLONES
- ATMOSPHERIC RIVERS
- CLOUDS
- WINDS

FourCastNet

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


J. Pathak
NVIDIA


S. Subramanian
LBL


P. Harrington
LBL

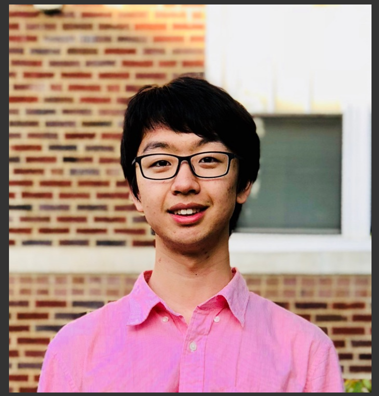

S. Raja
U. Michigan


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

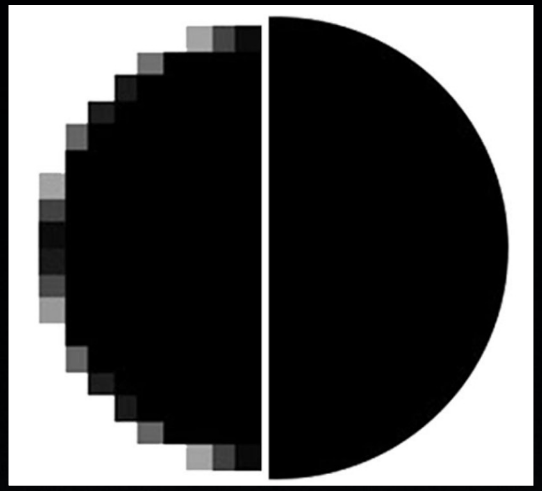

A. Anandkumar
Caltech/NVIDIA

MULTI-SCALE MODELING: RESOLUTION-FREE LEARNING

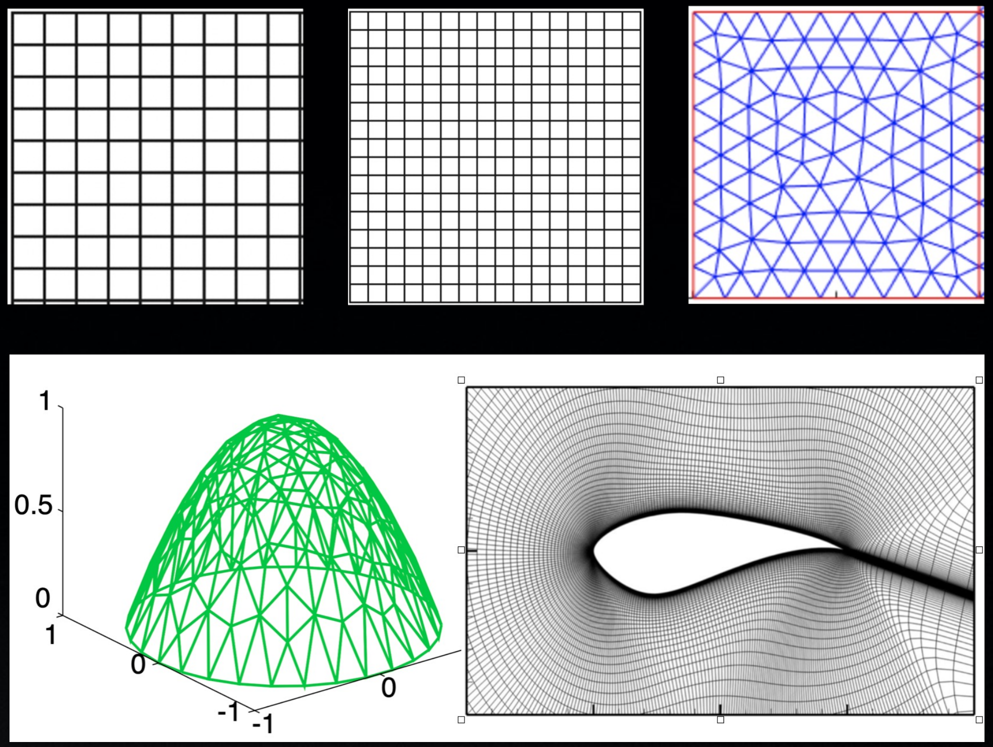
Learning Operators Between Continuous Inputs and Outputs

NEURAL NETWORK
Function: Mapping in Finite Dimensions

Discretized Input



NEURAL OPERATOR
Operator: Mapping in Infinite Dimensions (Function Space)



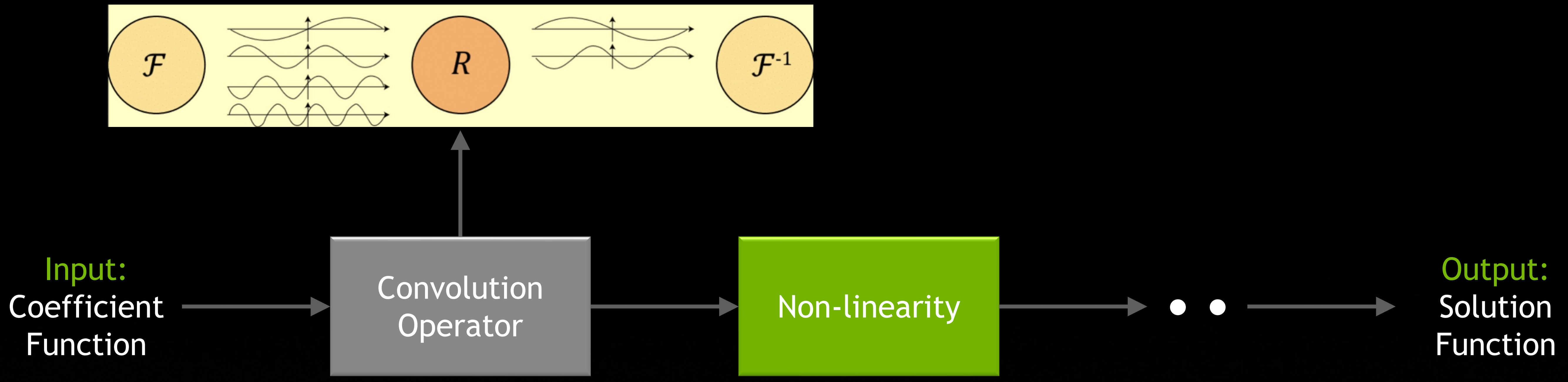
Continuous Function

Can be applied to simulations, experiments / observations, or both, spanning multiple orders of magnitude in spatio-temporal scales, thus allowing scientists to estimate dynamics better by pooling existing data

Li, Kovachki, et al. (2021), Fourier neural operator for parametric partial differential equations, ICLR 2021

FOURIER NEURAL OPERATOR

New Framework for Learning PDEs

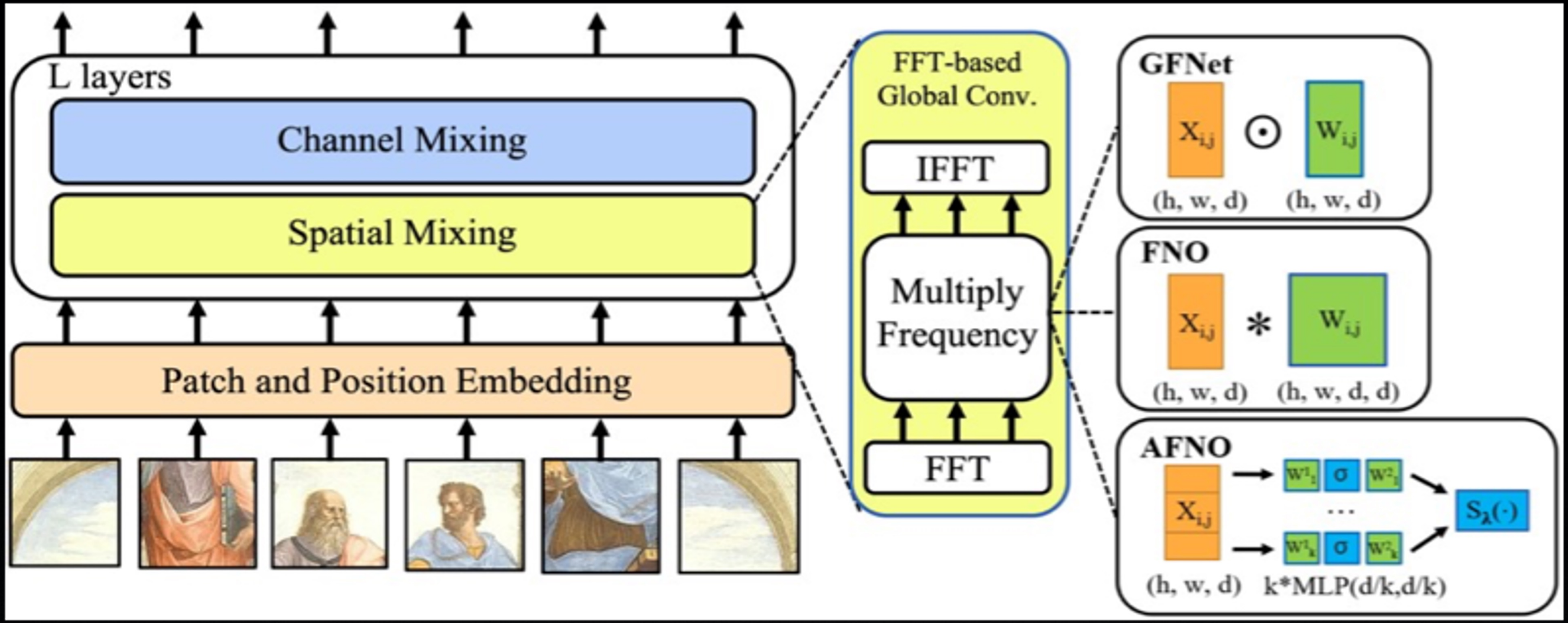


- Fourier transform for global convolution
- Efficient (few modes)
- Learn solution operator
- Continuous, mesh-independent, resolution-invariant

Li, Kovachki, et al. (2021), Fourier neural operator for parametric partial differential equations, ICLR 2021

ADAPTIVE FNO: CONNECTION BETWEEN FNO AND TRANSFORMERS

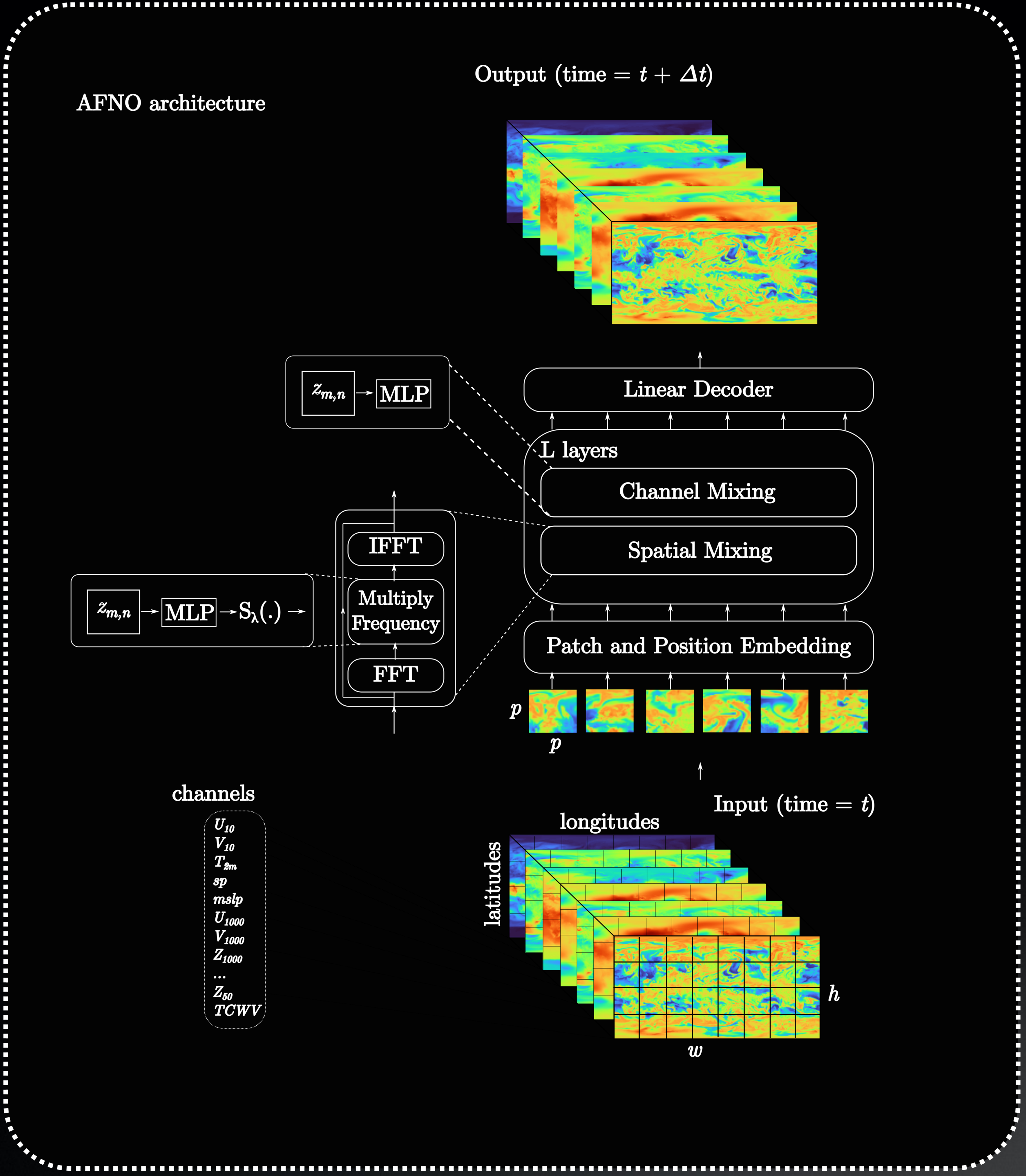
Efficient self-attention through FNO



- Transformers are a special case of neural operators
- Self-attention can be generalized to continuous integration
- FNO can be used in self-attention layers
- AFNO: Efficient token-mixing; Block-diagonal channel-mixing; Adaptive weight-sharing

Guibas, Mardani, et al. (2022), Efficient token mixing for transformers via Adaptive FNO, ICLR 2022

FOURCASTNET (FOURIER FORECASTING NETWORK)



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

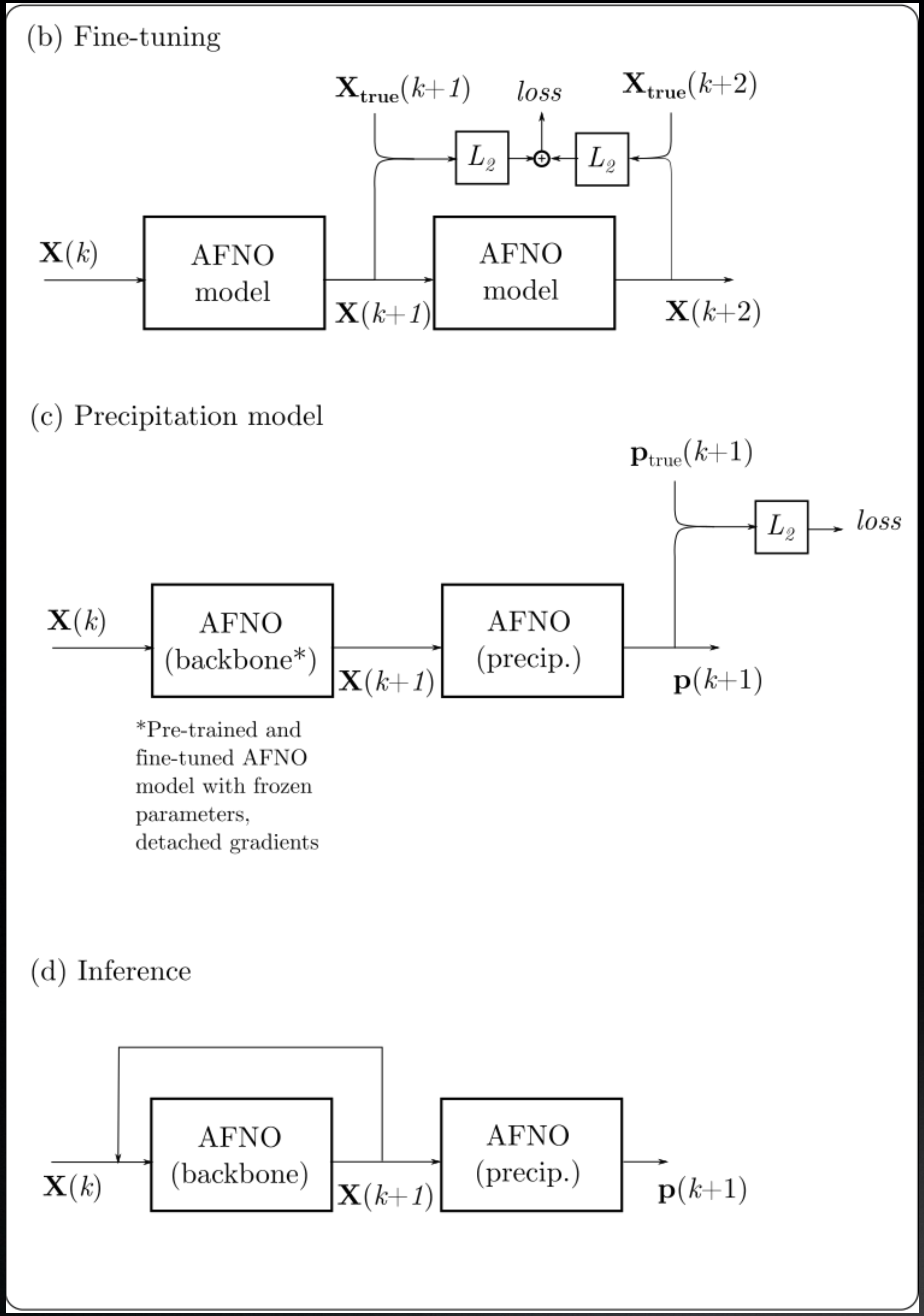
Vertical Level	Variables
Surface	$U_{10}, V_{10}, T_{2m}, sp, mslp$
$1000hPa$	U, V, Z
$850hPa$	T, U, V, Z, RH
$500hPa$	T, U, V, Z, RH
$50hPa$	Z
Integrated	$TCWV$

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

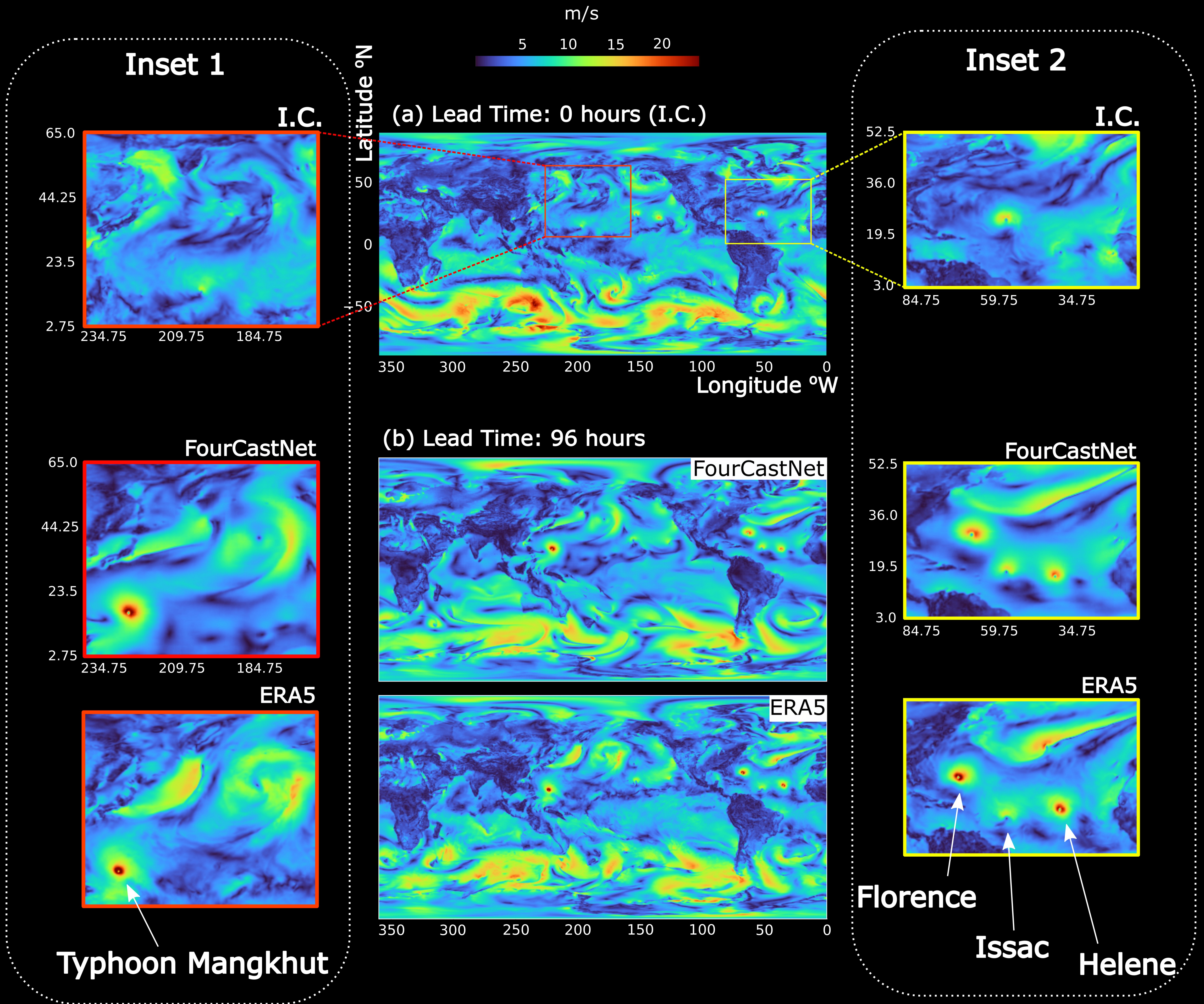
Vertical Level	Variables
Surface	$U_{10}, V_{10}, T_{2m}, sp, mslp$
$1000hPa$	U, V, Z
$850hPa$	T, U, V, Z, RH
$500hPa$	T, U, V, Z, RH
$50hPa$	Z
Integrated	$TCWV$

Extending to include radiation processes, vapor transport, clouds

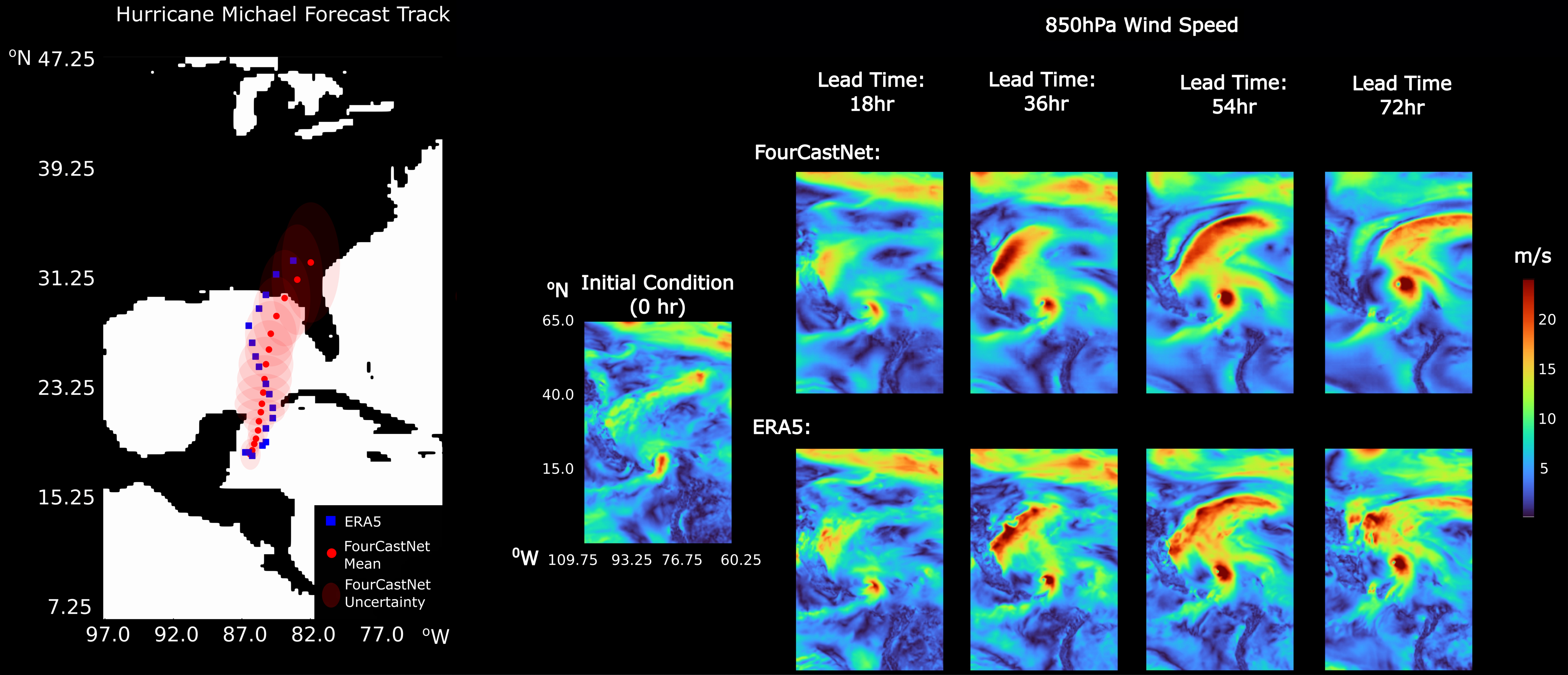
Training set: 1979 to 2015
Validation set: 2016, 2017
Held out: 2018 onwards

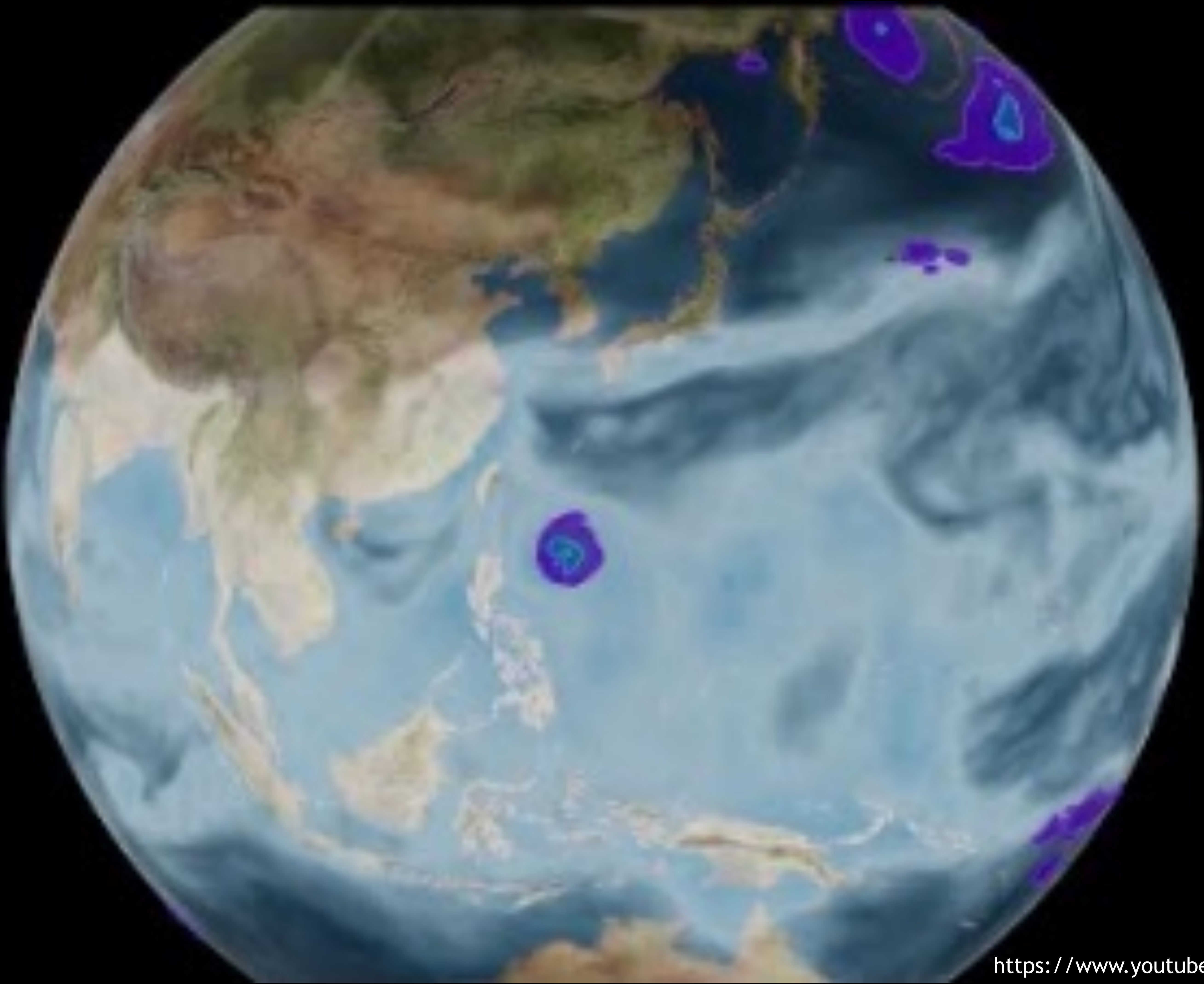
Pathak et al. (2022), FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators, arXiv:2202.11214

EXCELLENT SKILL ON FORECASTING SURFACE WINDS



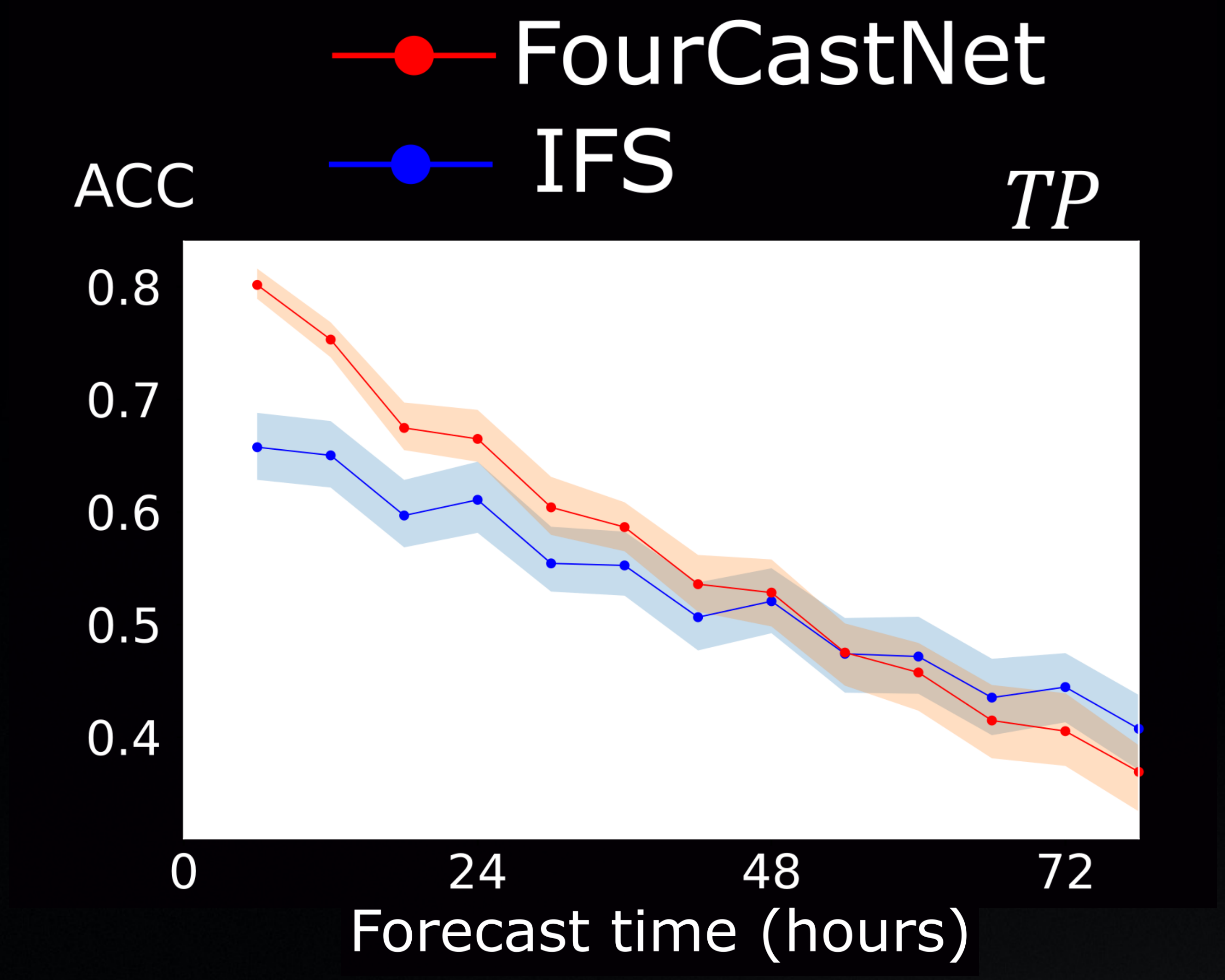
FOURCASTNET PREDICTS HURRICANE PATHS AND INTENSITIES



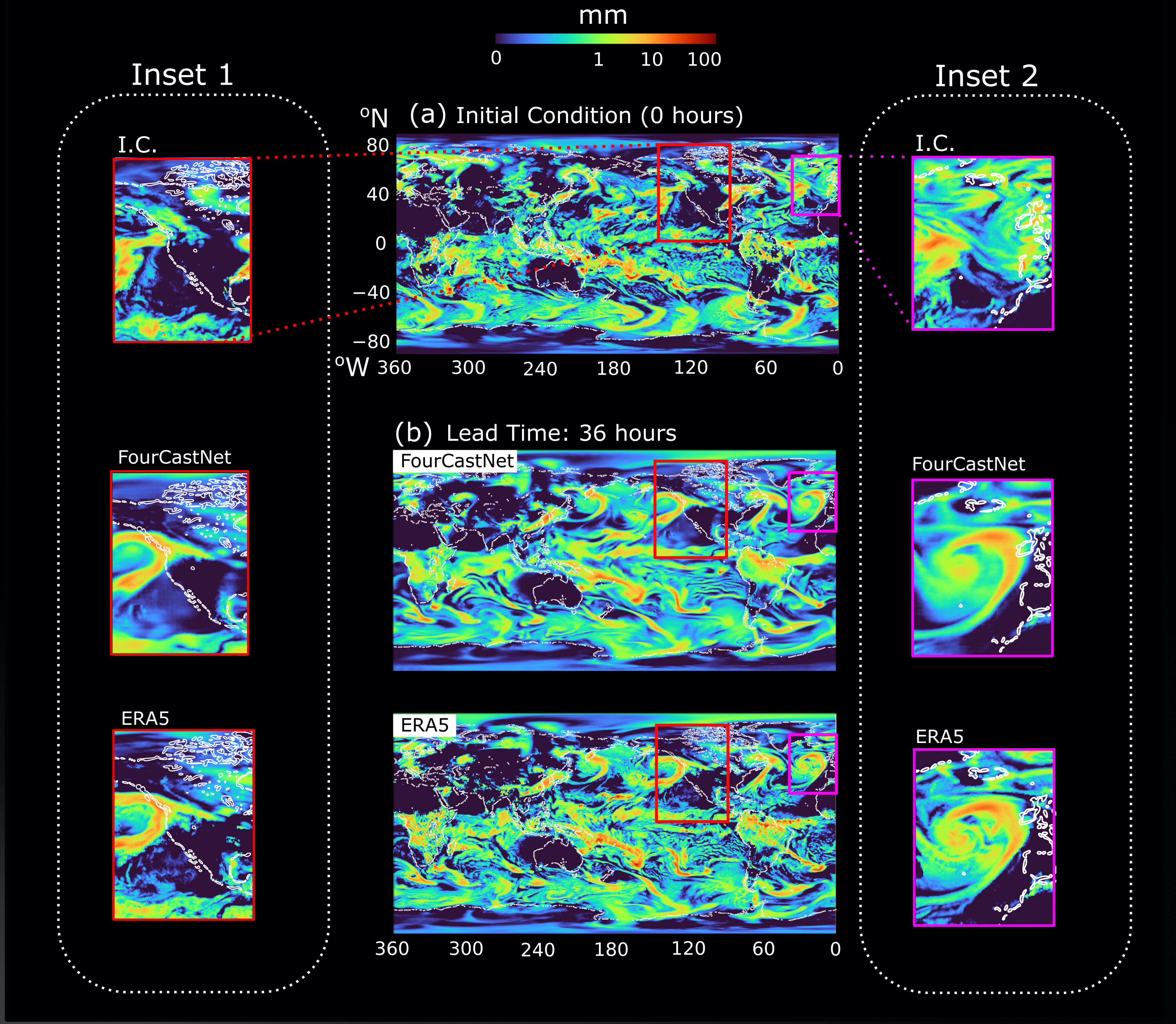


https://www.youtube.com/watch?v=JnGPxZ9glVk

UNPRECEDENTED SKILL ON PRECIPITATION FORECASTS

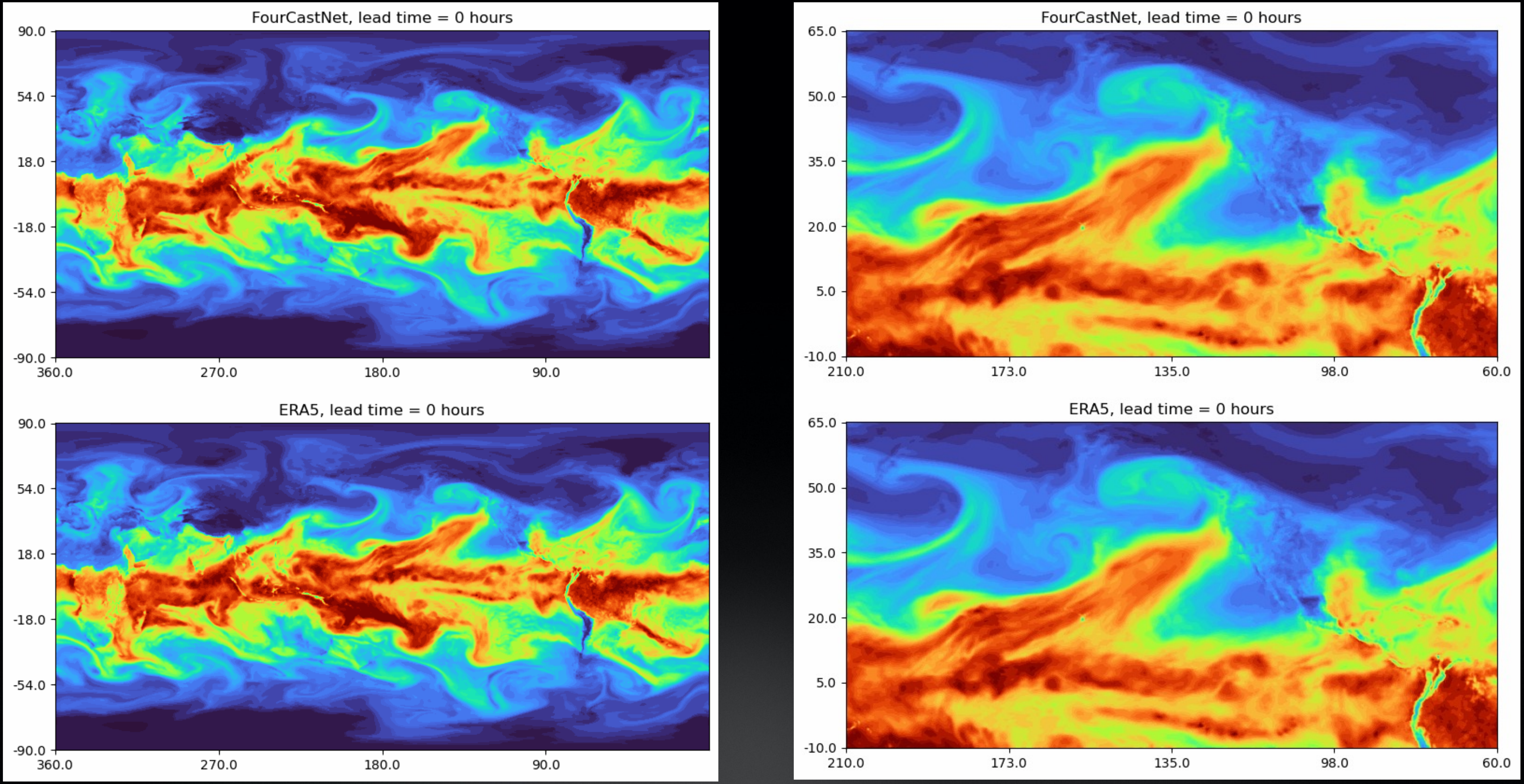


Note: Ground truth is ERA5, NOT observations

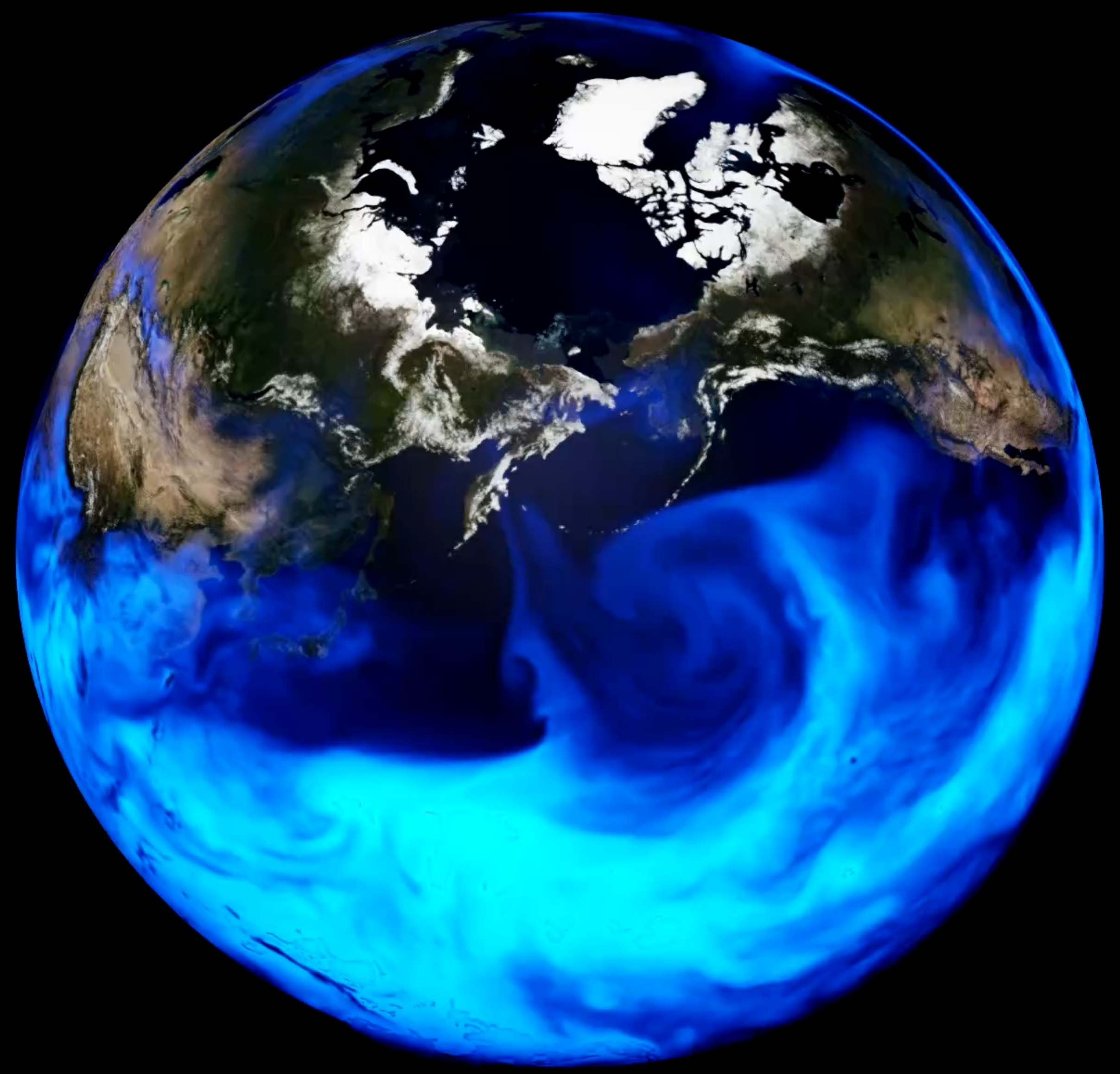


Pathak et al. (2022), FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators, arXiv:2202.11214

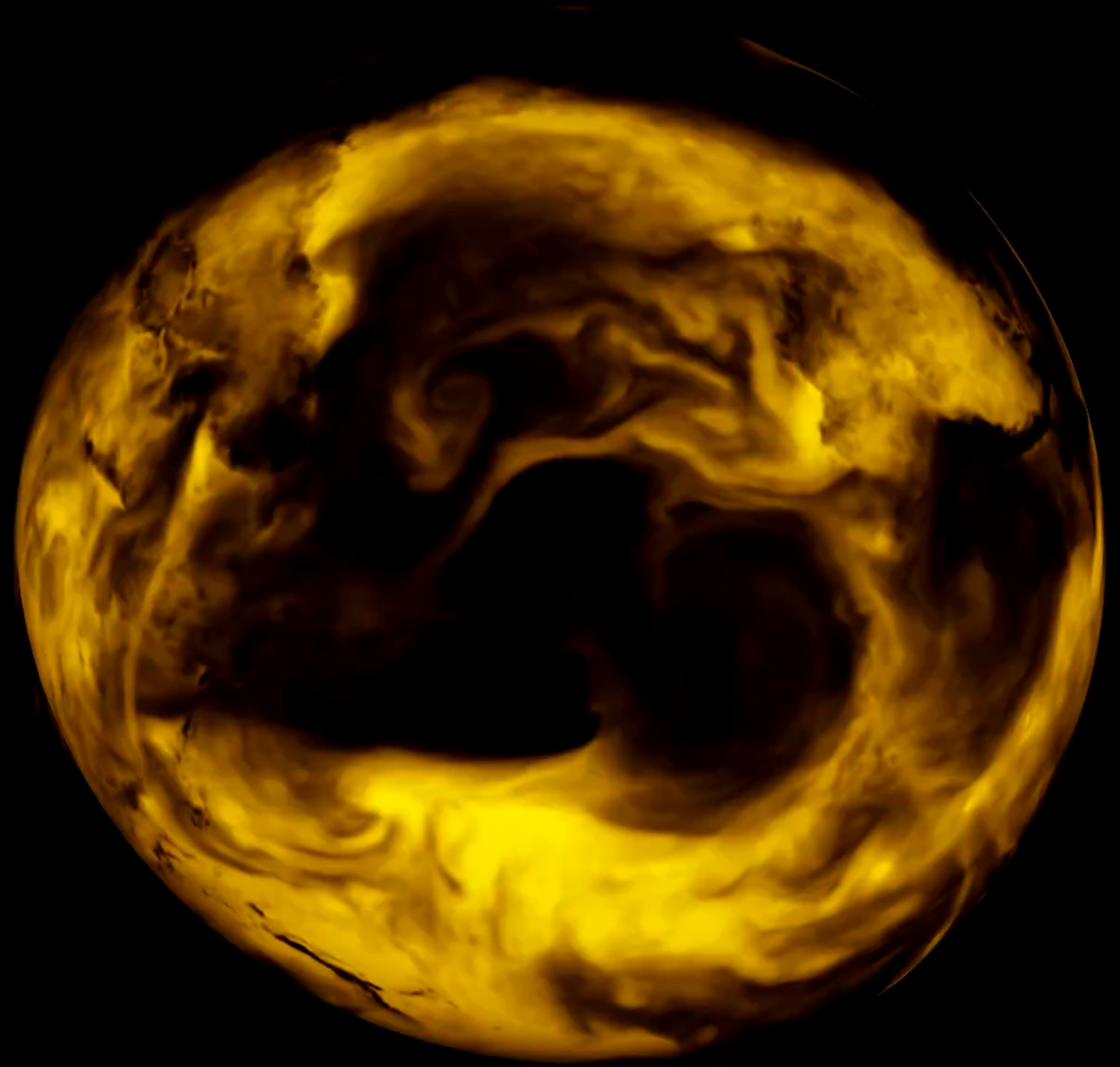
ACCURATE TOTAL COLUMN WATER VAPOR DYNAMICS



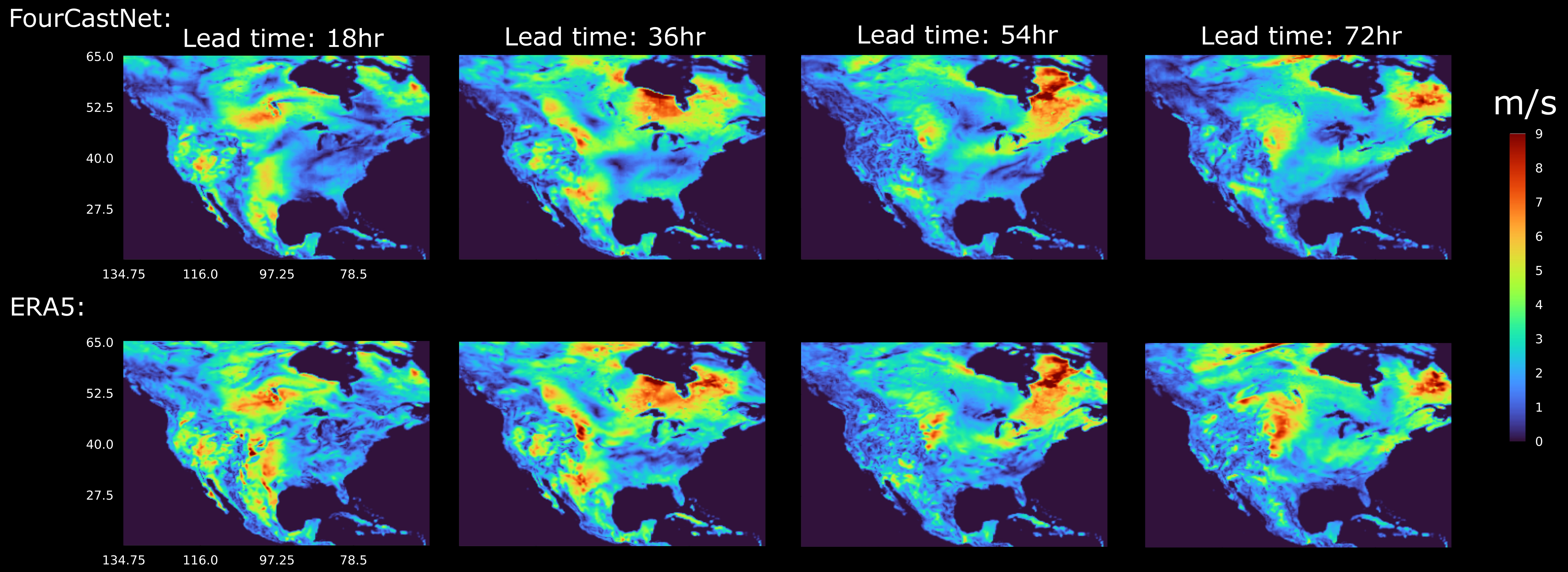
Ground Truth



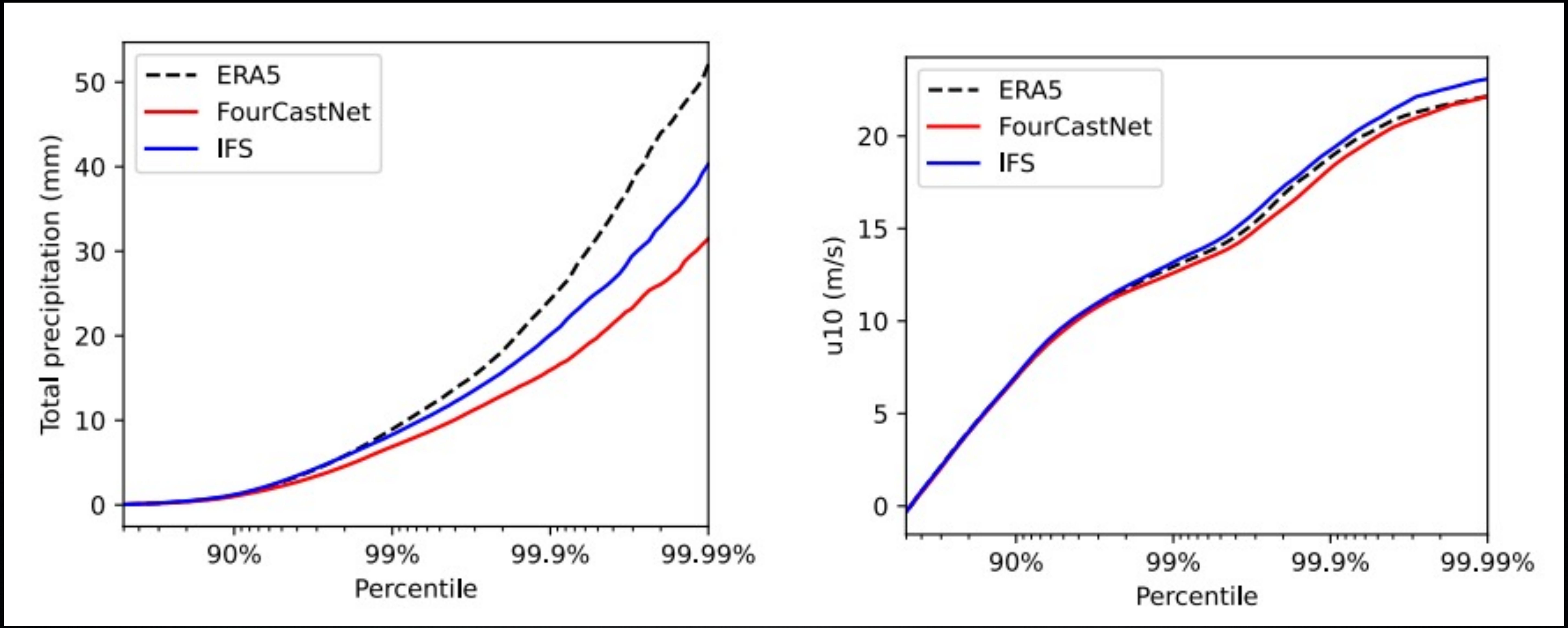
FourCastNet



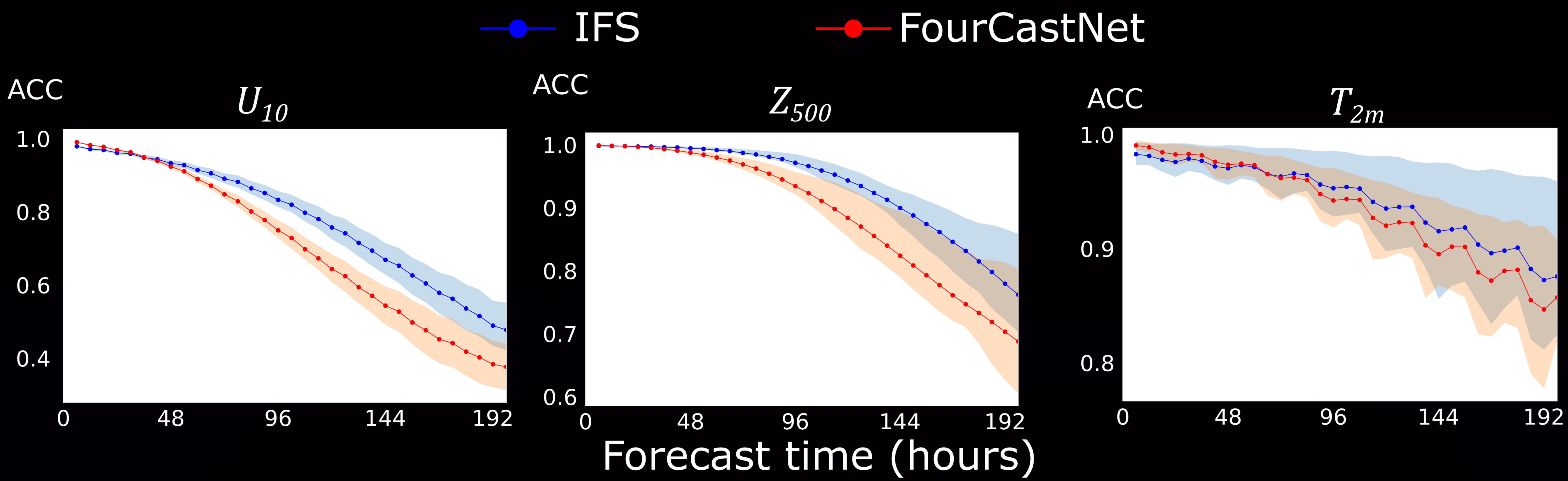
FOURCASTNET PREDICTS NEAR-SURFACE WIND FIELDS ACCURATELY: IMPORTANT IMPLICATIONS FOR WIND ENERGY PLANNING



PERFORMANCE ON EXTREME PRECIPITATION AND WIND EVENTS



SHORT-TERM FORECAST SKILL APPROACHING IFS

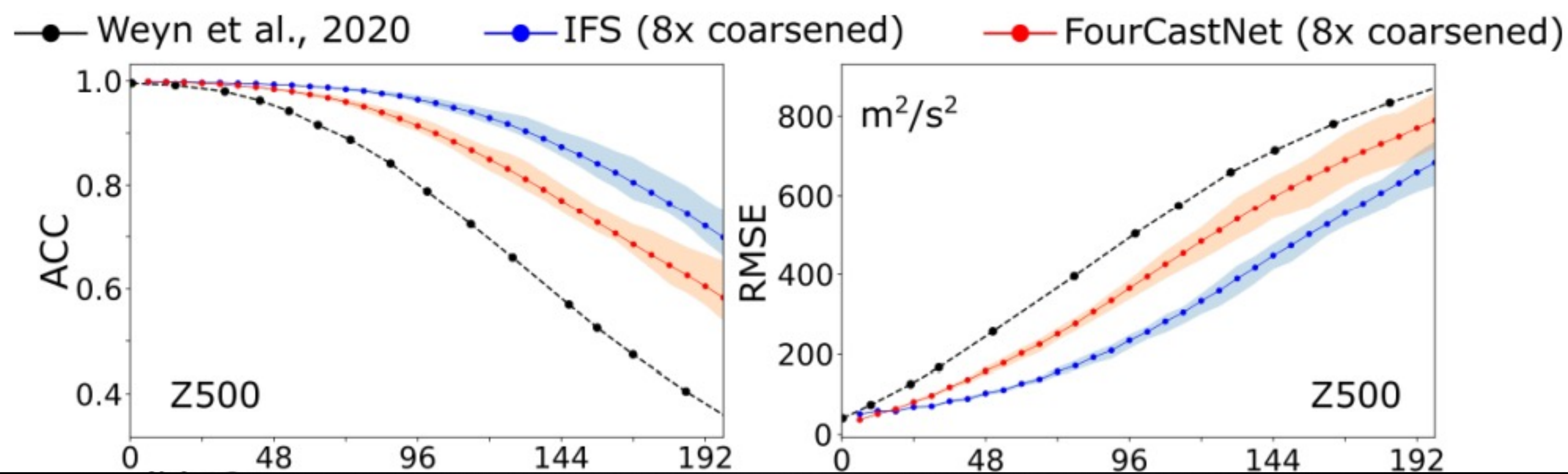


Note: Ground truth is ERA5, NOT observations

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**

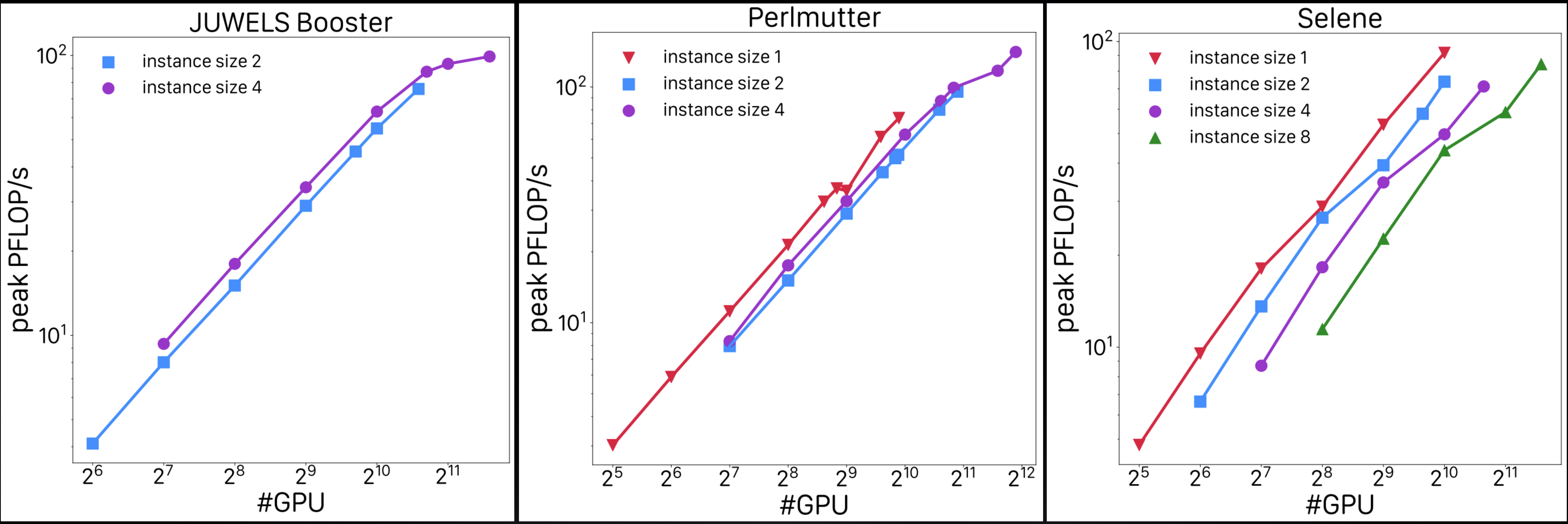
- 100-member ensemble forecast in 7 seconds
- 100-member ensemble forecast consumes 7 kJ
- 4 to 5 orders-of-magnitude speedup over NWP
- 4 orders-of-magnitude smaller energy footprint

Caveats

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

SCALING RESULTS (1 OF 2)

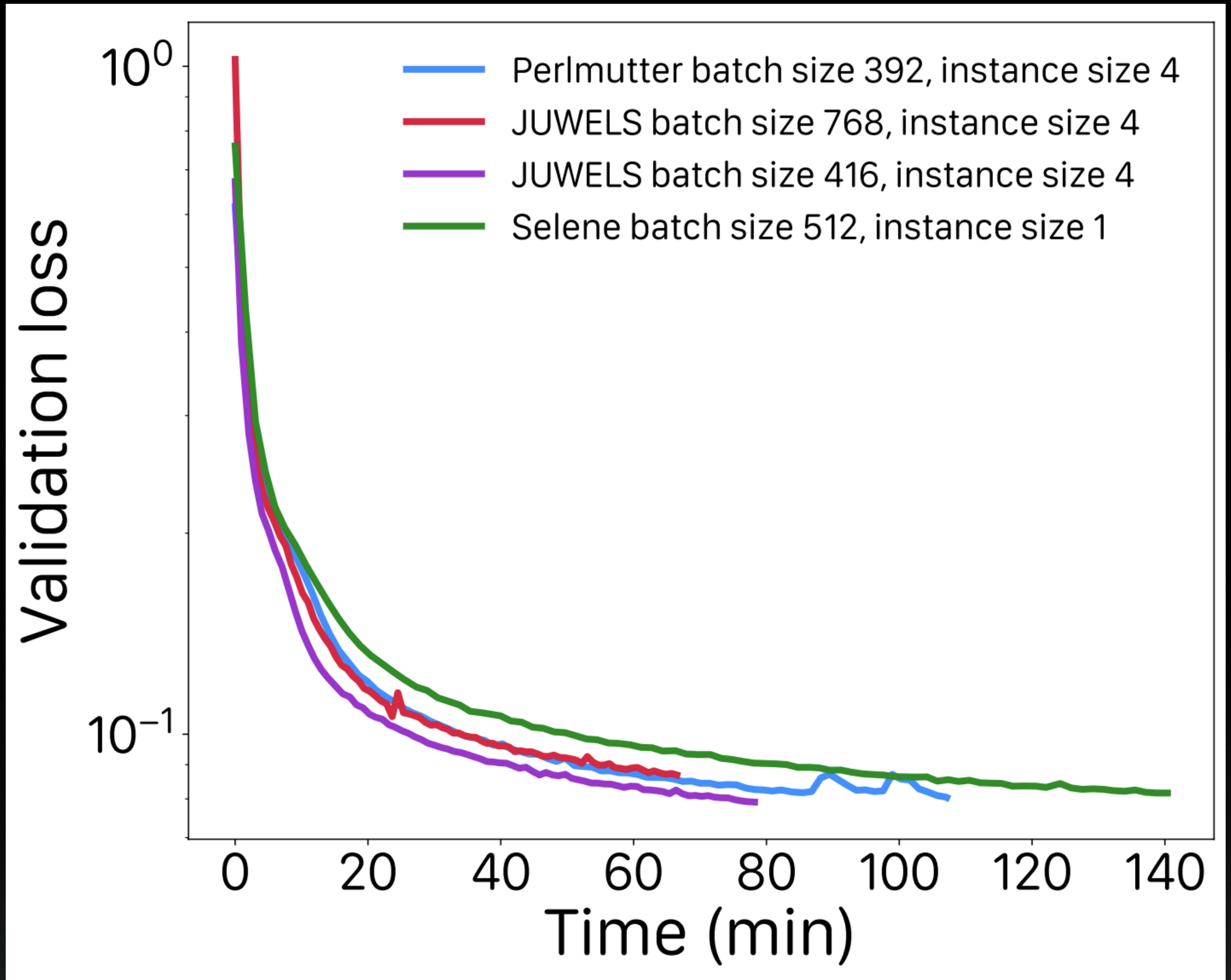
FourCastNet scaled efficiently upto ~ 4000 GPUs on three supercomputing systems:
JUWELS Booster, Perlmutter, and Selene



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

SCALING RESULTS (2 OF 2)

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



SCALING CHALLENGES

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 WEATHER AND CLIMATE EXTREMES

MONITOR | FUSE DATA | ASSIMILATE DATA

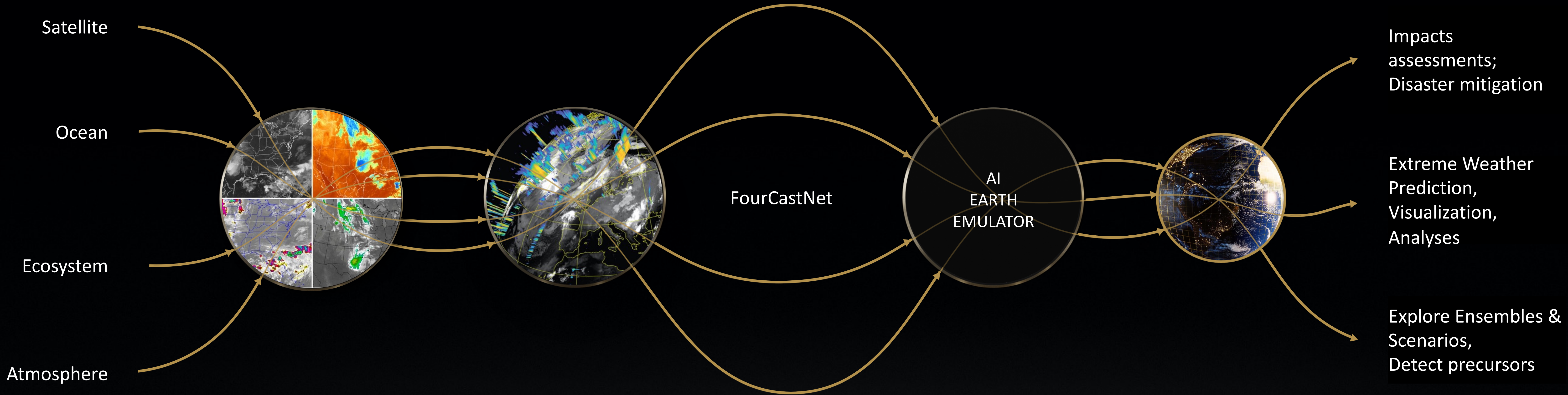
PHYSICS-ML EMULATION

PREDICT | ANALYZE | VISUALIZE

ERA5 Reanalysis Data (ECMWF) / Hi-res simulations

Modulus Model Training and Inference

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)

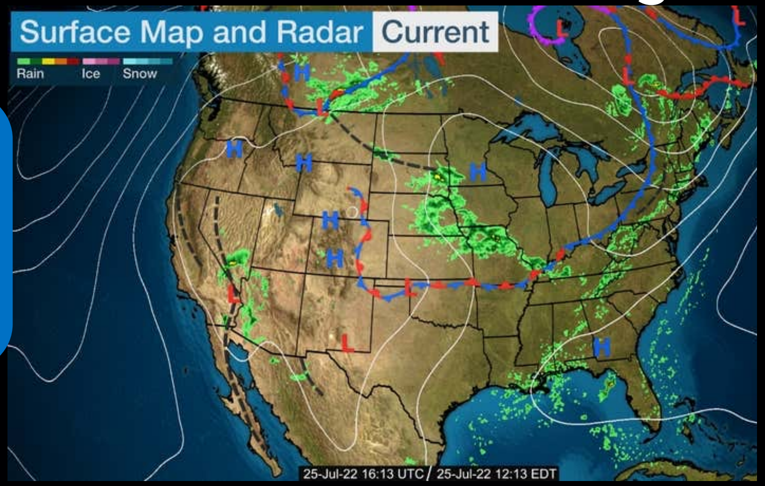
1. Predict, visualize, detect and track extremes
2. Compare skill to traditional NWP models
3. Checkpoint / restart ensembles around events
4. Assess extreme weather impacts, mitigate disasters
5. Interactively investigate impact of changing climate scenarios on behavior of extremes
6. Detect precursors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

TODAY

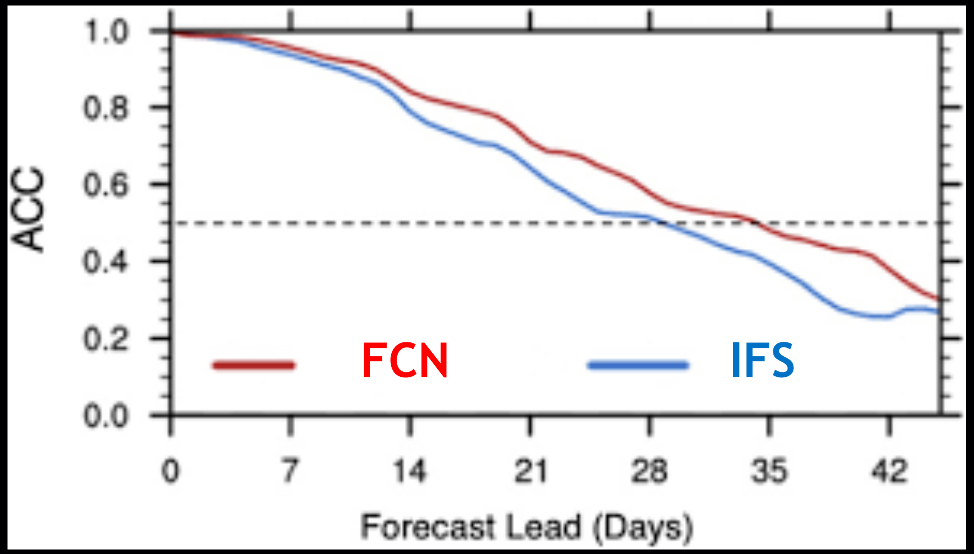
- Unprecedented skill
- 1000-member ensemble in seconds
- 4 to 5 orders-of-magnitude speedup over NWP
- 4 orders-of-magnitude smaller energy footprint

FCN-Real-time: Next-gen Weather Forecasting

The FourCast Channel

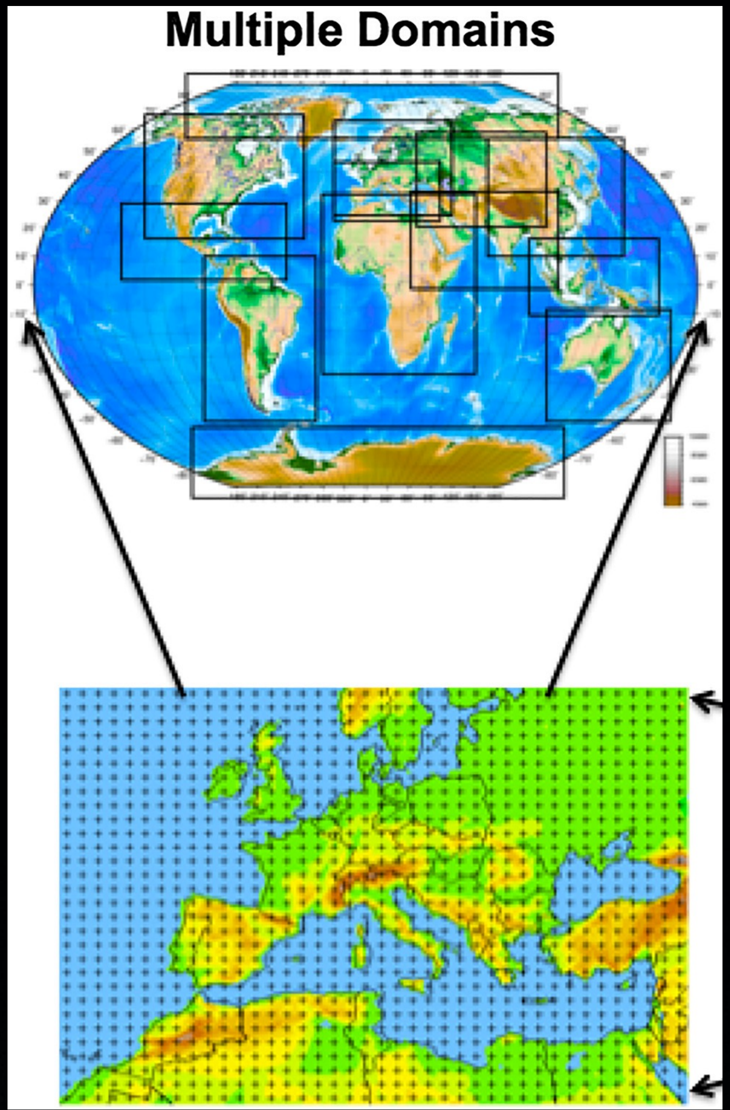


FCN-Big: Scale-up and scale-out to beat IFS (Gold Standard - ECMWF)



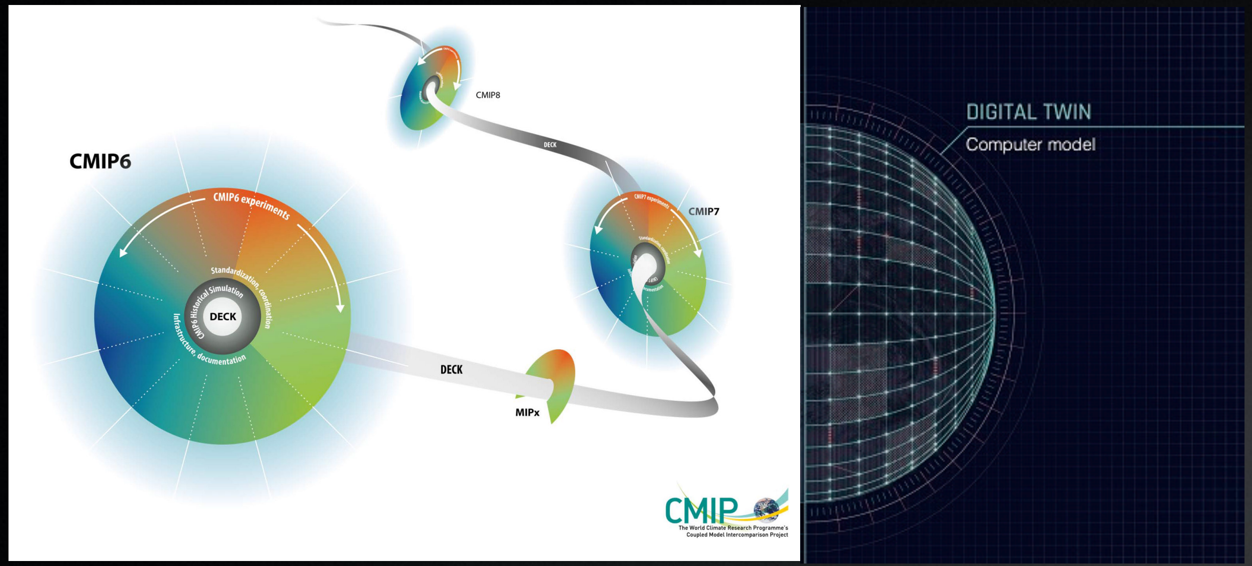
SAMPLE ONLY!
NOT REAL DATA!

FCN-Regional: Nested global-local FCN to democratize regional weather forecasting

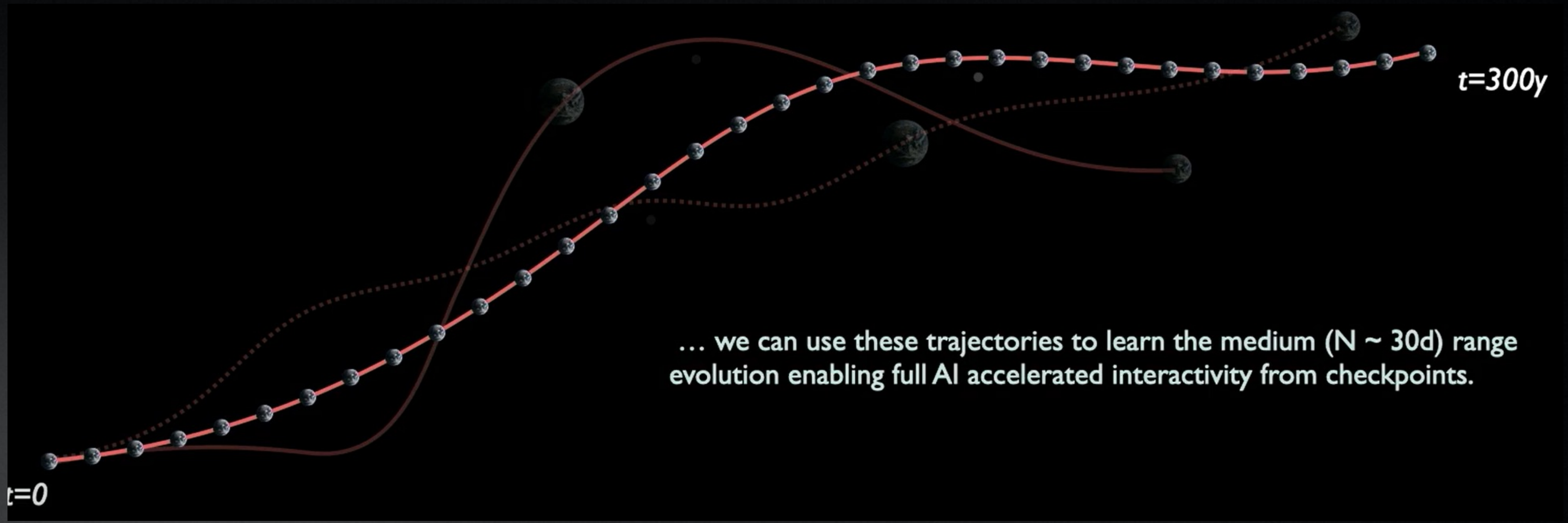


FCN-Probabilistic: Calibrated Ensembles

FCN-Climate: CMIP6-initialized extremes in 2030-2050

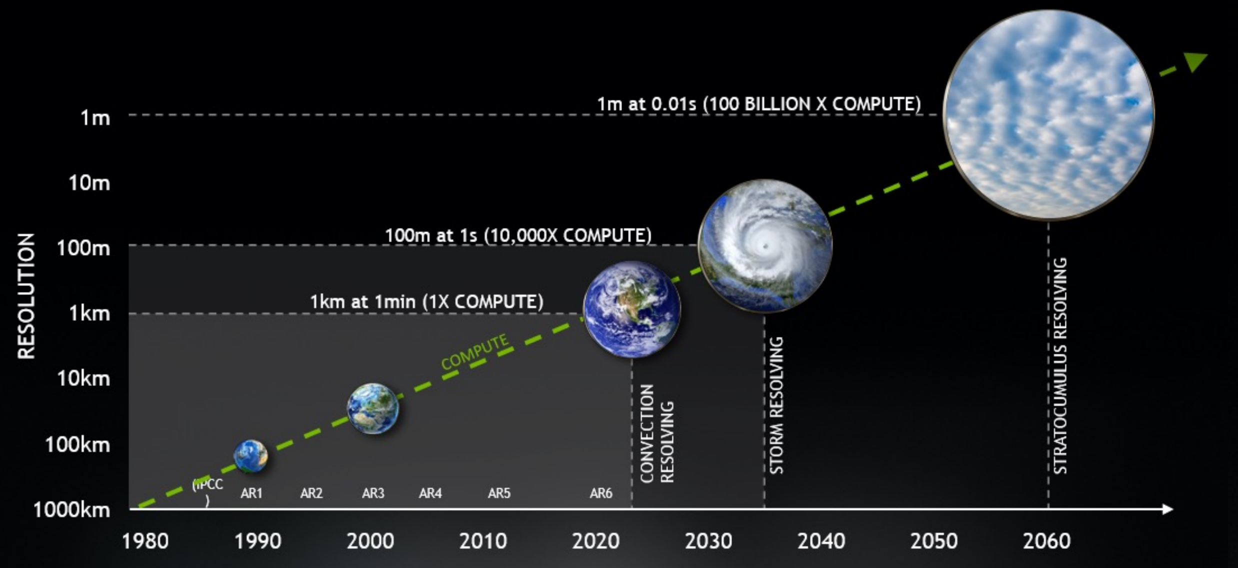


FCN-Climate+: Climate tethering at 1-10 km scale



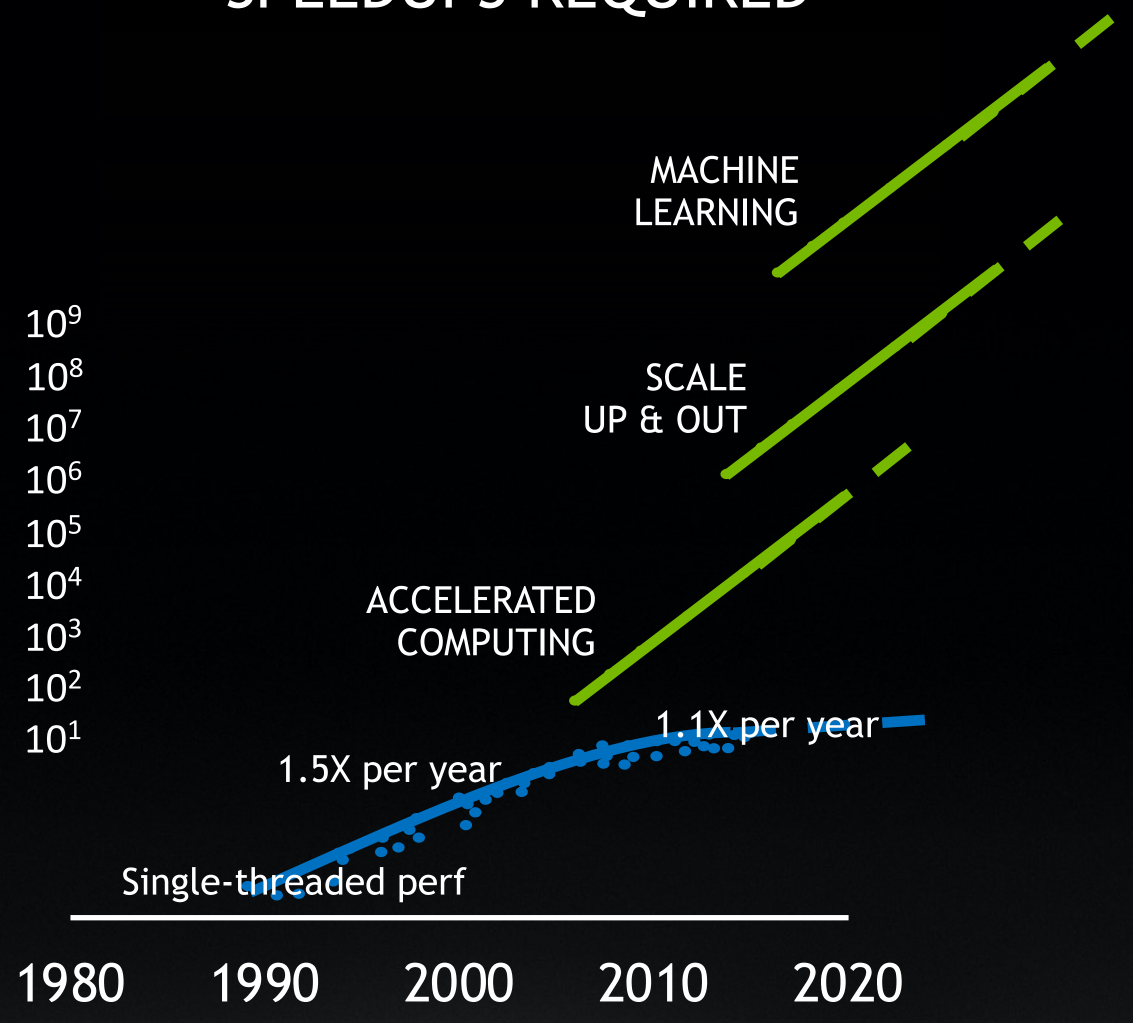
SYNTHESIS

CLIMATE MODELING DEMANDS EXTREME COMPUTE



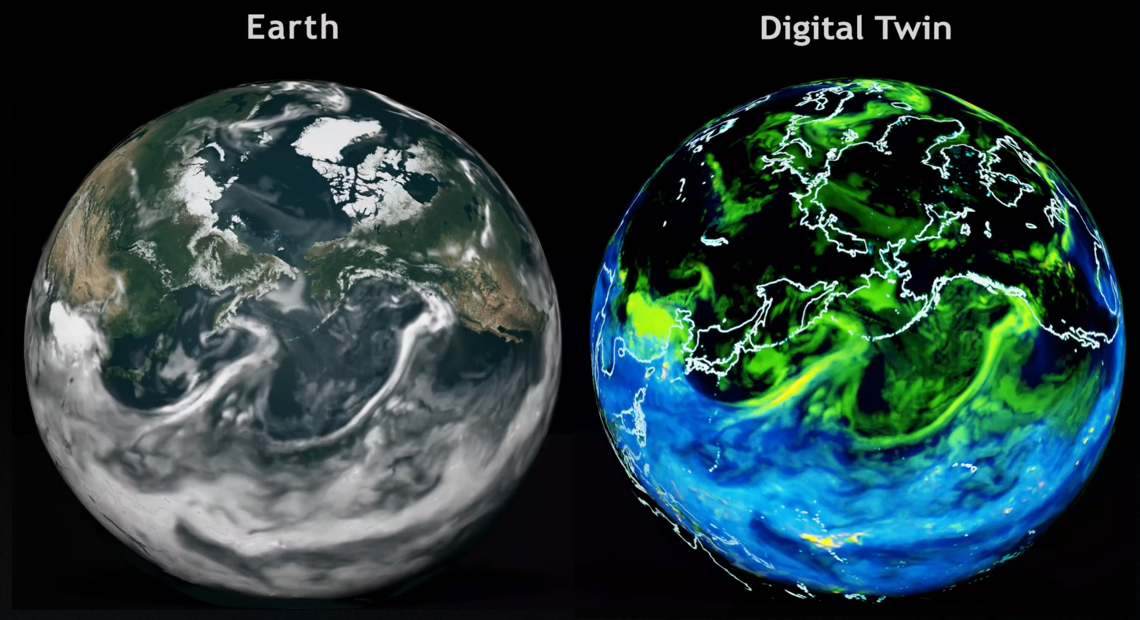
Resolution, Ensembles, Scenarios

MILLION X SPEEDUPS REQUIRED



Computing, Physics-ML

AI-POWERED DIGITAL TWINS PROMISE ACTIONABLE RESULTS IN ACTIONABLE TIME



Earth-2

THANK YOU!