

What's the Next Wave of Disruption in Science and Engineering?

Reference Models for Engineering

Johannes Brandstetter
ML in Poland 2025
October 17 2025

Content of this talk

- AI4Engineering (AI beyond LLMs)
- Reference models as the way forward
- I am giving my opinions which change over time and should not be taken too seriously!

Disclaimer: This talk is very much centered around data-driven simulations!
Might be transferable to other areas.

Status quo

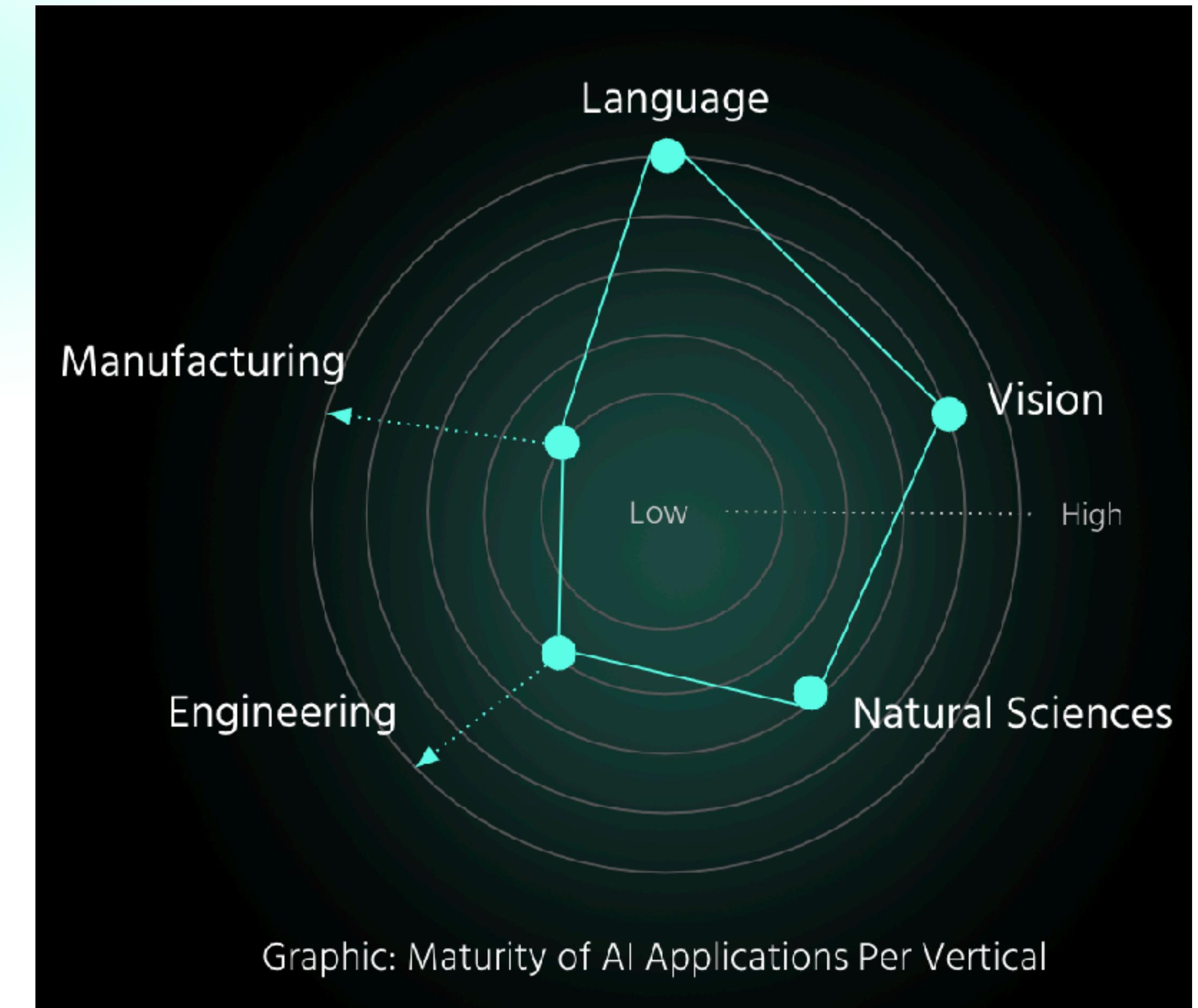
- In the next five years, AI breakthroughs will be built on **Transformers** which run on **Nvidia GPUs** built at **TSMC** with **ASML** machines (which uses **Zeiss technology** and **Trumpf lasers**):
 - AI race is an intertwined global affair
 - Focus is on text, images, videos, ...
 - “Let’s build things bigger and bigger and bigger until intelligence emerges”
- Big players built GPU centres with > 100k GPUs
 - Energy supply has silently become the next frontier
- OpenAI/Anthropic are valued 500/183 billion US dollars (and still far from profitable)
- Is there something else?



SIEMENS
ENERGY

The elevator pitch for investors

- AI systems have been scaled up for language and vision applications with tremendous success.
- Many verticals remain untouched.
 - Manufacturing, engineering, logistics,...
 - And my personal bet: **data-driven simulations, digital twins, ...**



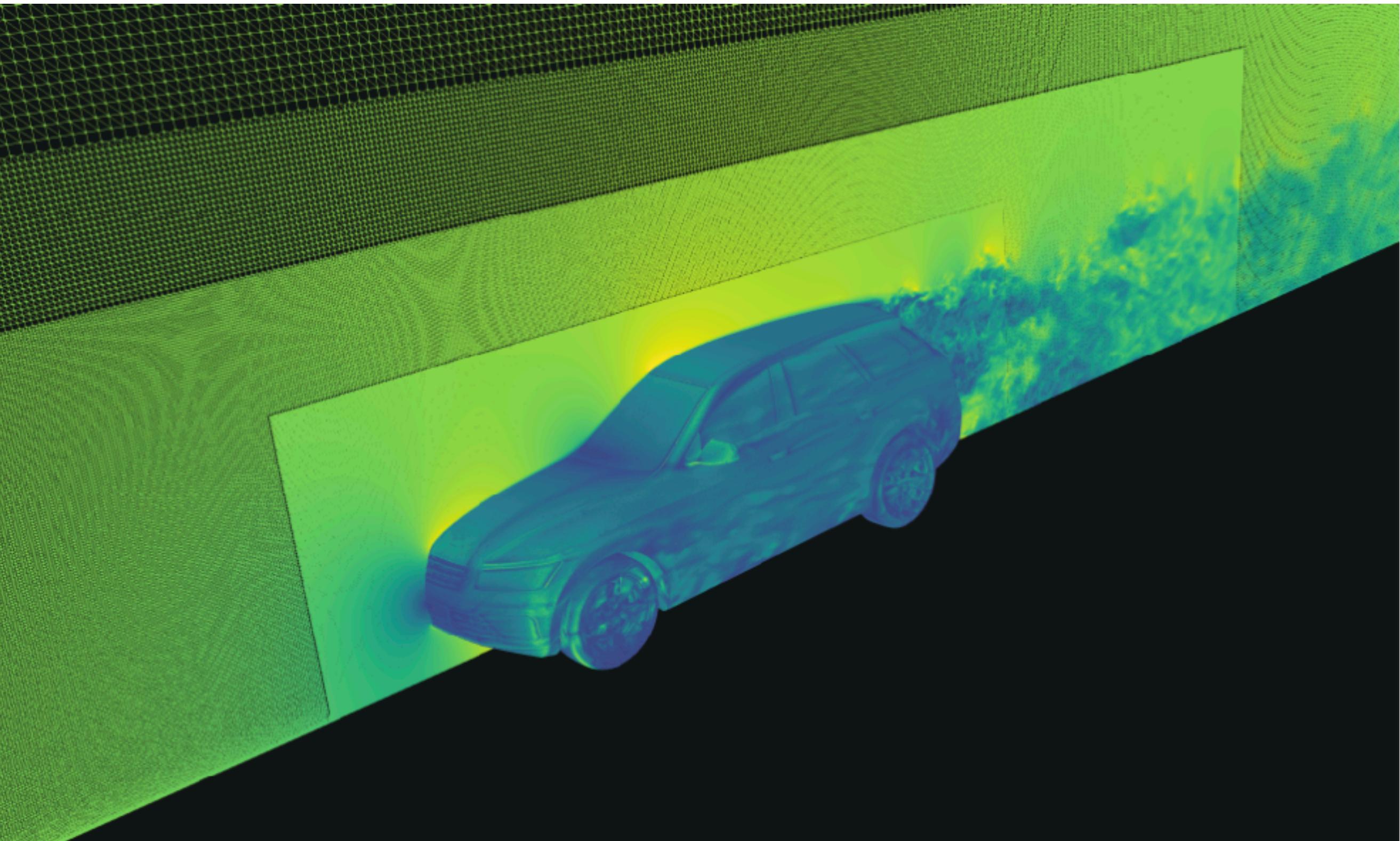
Reference models

- Foundation models are trained on internet-scale data:
 - GPT, Gemini, SAM
- Lots of claims on foundation models for physics:
 - One model for all physics?
- Reality is very different
- Reference models == foundation-like models for small, well-defined settings:
 - Weather, injection molding, external aerodynamics, semiconductors

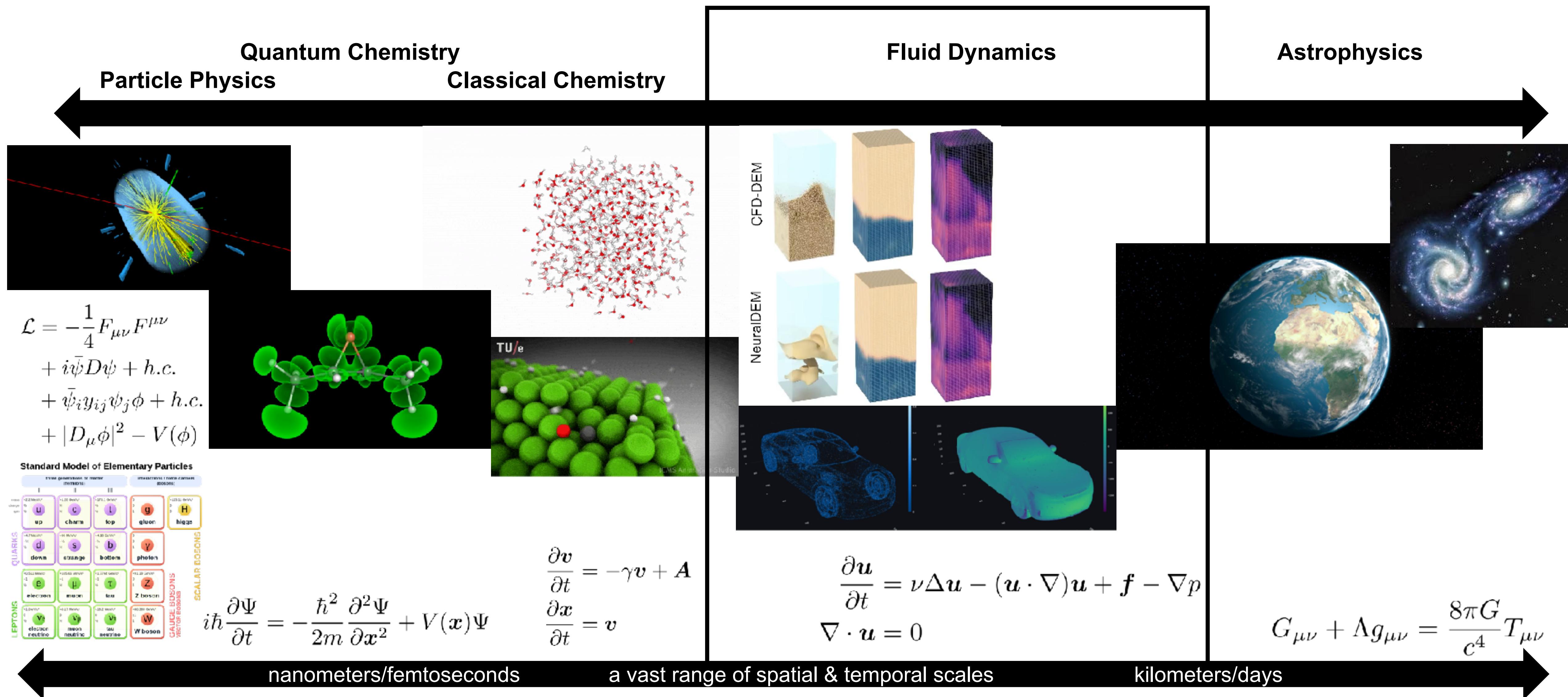
Reference models

Let's make it a bit more haptic

- Example CFD: turbulent / laminar flows, jets / wake flows / mixing layers, subsonic / transonic / supersonic, geometries, chemistry, multi-physics, ...
- We can write the input as something like
$$\mathbf{u}_{\text{CFD}} = [\mathbf{u}_{\text{geom}}, \mathbf{u}_{\text{preprocess}}, \mathbf{u}_{\text{boundary/initial}}, \mathbf{u}_{\text{mesh}}, \mathbf{u}_{\text{phys}}, \mathbf{u}_{\text{discr}}]$$
- To cover ALL CFD is nearly impossible
- We need to limit us to certain settings!



Nature is described by differential equations



Numerical solutions for PDEs

We discretize space and time

- Formulation of a time-dependent PDE:

$$\partial_t \mathbf{u} = F(t, \mathbf{x}, \mathbf{u}, \partial_{\mathbf{x}} \mathbf{u}, \partial_{\mathbf{xx}} \mathbf{u}, \dots) \quad (t, \mathbf{x}) \in [0, T] \times \mathbb{X}$$

$$\mathbf{u}(t, \mathbf{x}) = \mathbf{u}^0(\mathbf{x}) \quad \mathbf{x} \in \mathbb{X}$$

$$B[\mathbf{u}](t, \mathbf{x}) = 0 \quad (t, \mathbf{x}) \in [0, T] \times \partial \mathbb{X}$$

- We discretize domain into grid:

- Estimate the spatial derivatives, e.g. FDM
- Use spatial estimates for time update (e.g., Euler update)

$$f_x(x, y) \approx \frac{f(x + h, y) - f(x - h, y)}{2h}$$

$$f_y(x, y) \approx \frac{f(x, y + k) - f(x, y - k)}{2k}$$

$$f_{xx}(x, y) \approx \frac{f(x + h, y) - 2f(x, y) + f(x - h, y)}{h^2}$$

$$f_{yy}(x, y) \approx \frac{f(x, y + k) - 2f(x, y) + f(x, y - k)}{k^2}$$

$$f_{xy}(x, y) \approx \frac{f(x + h, y + k) - f(x + h, y - k) - f(x - h, y + k) + f(x - h, y - k)}{4hk}.$$

Weather as an example

The Sputnik Pangu weather moment

November 2022

Article | [Open access](#) | Published: 05 July 2023

Accurate medium-range global weather forecasting with 3D neural networks

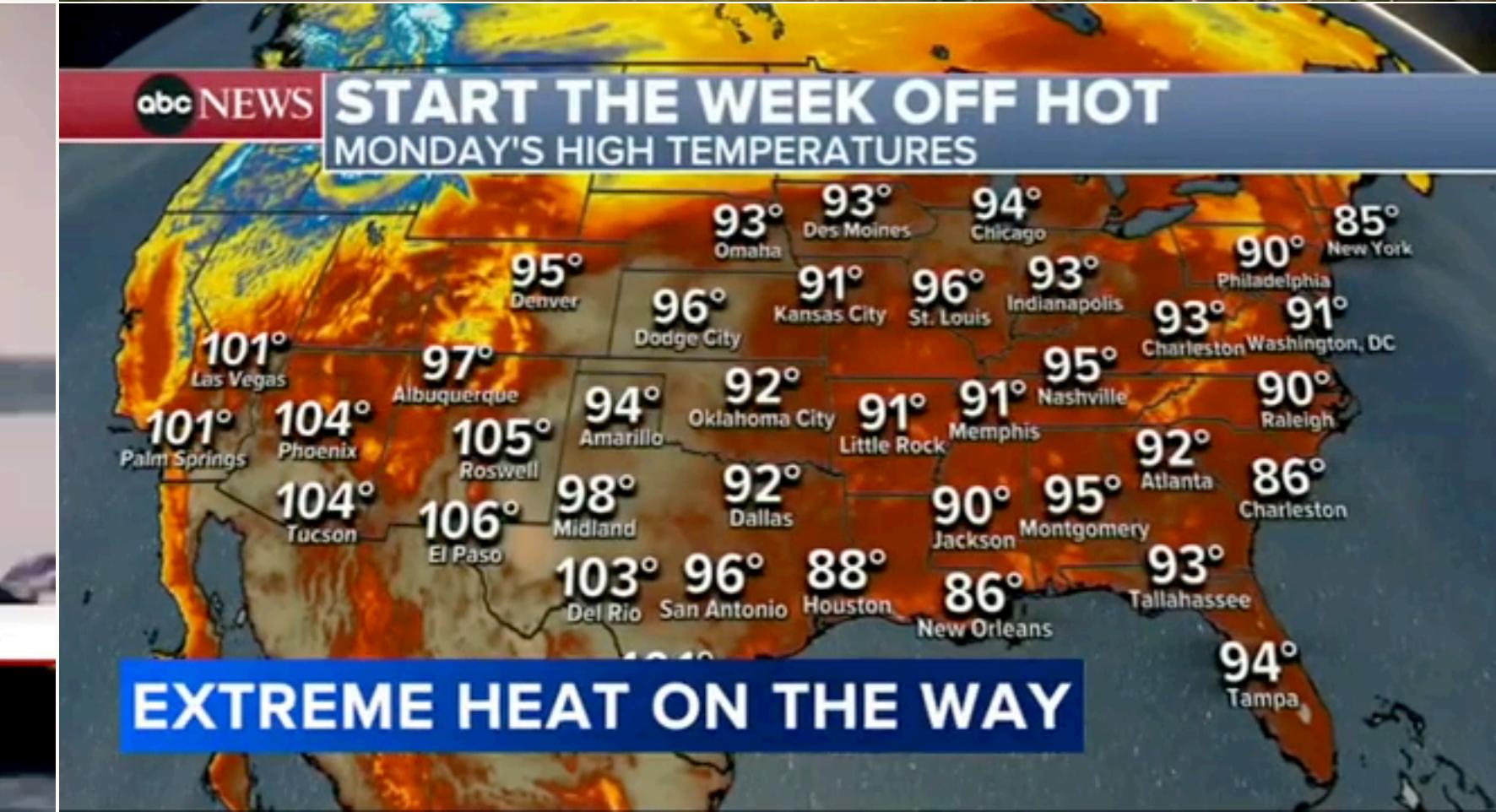
[Kaifeng Bi](#), [Lingxi Xie](#), [Hengheng Zhang](#), [Xin Chen](#), [Xiaotao Gu](#) & [Qi Tian](#) 

[Nature](#) **619**, 533–538 (2023) | [Cite this article](#)

286k Accesses | **1012** Citations | **1749** Altmetric | [Metrics](#)

Abstract

Weather forecasting is important for science and society. At present, the most accurate forecast system is the numerical weather prediction (NWP) method, which represents atmospheric states as discretized grids and numerically solves partial differential equations that describe the transition between those states¹. However, this procedure is computationally expensive. Recently, artificial-intelligence-based methods² have shown potential in accelerating weather forecasting by orders of magnitude, but the forecast accuracy is still significantly lower than that of NWP methods. Here we introduce an artificial-intelligence-based method for accurate, medium-range global weather forecasting. We show that three-dimensional deep networks equipped with Earth-specific priors are effective at dealing with complex patterns in weather data, and that a hierarchical temporal aggregation strategy reduces accumulation errors in medium-range forecasting. Trained on 39 years of global data, our program, Pangu-Weather, obtains stronger deterministic forecast results on reanalysis data in all tested variables when compared with the world's best NWP system, the operational integrated forecasting system of the European Centre for Medium-Range Weather Forecasts (ECMWF)³. Our method also works well with extreme weather forecasts and ensemble forecasts. When initialized with reanalysis data, the accuracy of tracking tropical cyclones is also higher than that of ECMWF-HRES.



The evolution of Earth system prediction

the
analytical
age

- 1820s:** Navier-Stokes equations formulated
- 1890s:** Cleveland Abbe establishes theoretical basis for weather prediction
- 1904:** Vilhelm Bjerknes outlines systematic approach to weather prediction
- 1917:** Lewis Fry Richardson develops first numerical weather prediction methods

the
numerical
age

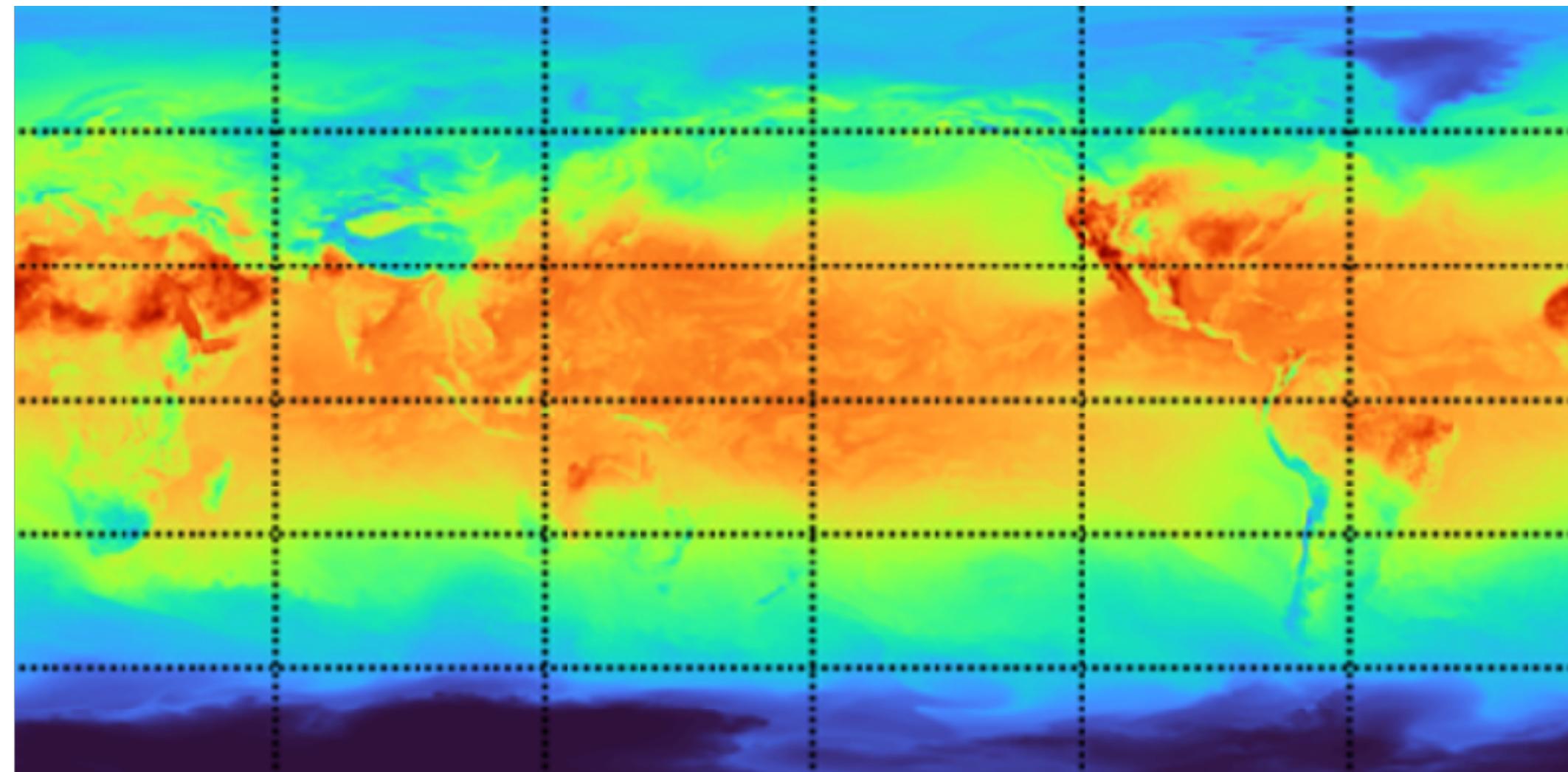
- 1922:** Richardson publishes "Weather Prediction by Numerical Process"
- 1946:** ENIAC computer developed, enabling first numerical calculations
- 1950:** First numerical weather prediction by Charney, Fjørtoft, and von Neumann
- 1955:** First operational numerical weather predictions
- 1960s:** Global circulation models emerge
- 1975:** ECMWF established, marking international collaboration
- 1979:** First coupled ocean-atmosphere models
- 1983:** European Centre's IFS model introduced
- 1990s:** Ensemble prediction systems developed

the AI age

- 2018:** First serious comparisons of AI vs physics models (Dueben and Bauer)
- 2019:** AI models skillful to multiple days (Weyn et al.)
- 2020:** WeatherBench starts to drive ML development (Rasp et al.)
- 2022:** GNNs outperform GFS at 1° (Keisler)
- 2023:** **Climax** demonstrates first foundation model principles
- 2023:** Pangu-Weather outperforms HRES at 0.25° (Bi et al.)
- 2024:** GenCast outperforms IFS ensemble (Price et al.)
- 2024:** ECMWF launches AIFS, Microsoft Research launches **Aurora**

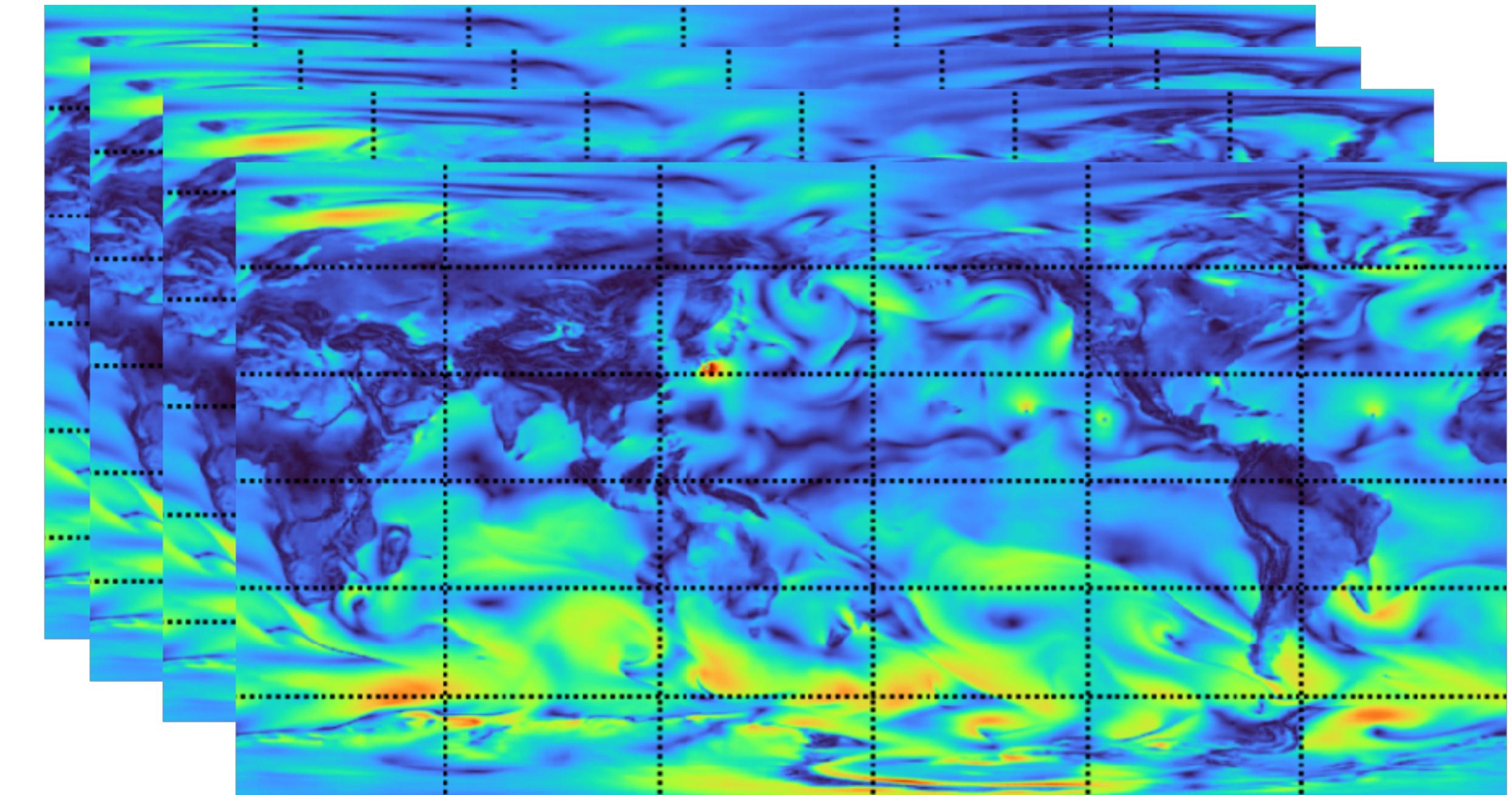
Weather data

Surface variable: temp 2m from the Earth's surface



Shape = H x W

Atmospheric variable: wind speed



Shape = L x H x W

Why Earth science needs Foundation Models

1. Exabytes of data

- Multiple scales and modalities
- Observations: satellites, weather stations
- NWP: forecasts, analysis, reanalysis

2. Transfer learning opportunities

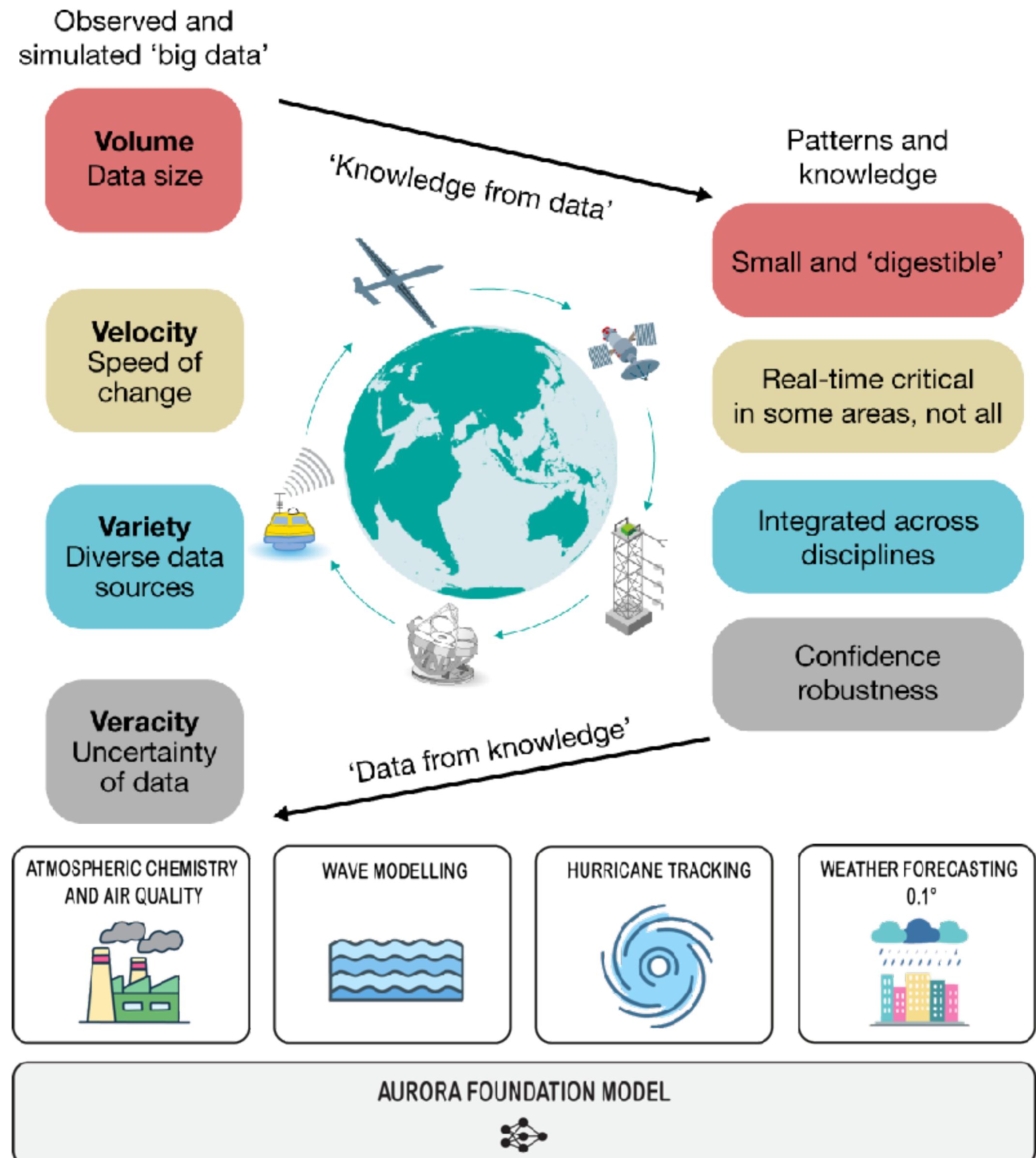
- Common physical principles
- Coupled interactions
- Effective in data scarce tasks

3. Compute & Infrastructure demands

- Current specialized AI models: limited scope
- Current NWP models: node-hours on supercomputers
- Foundation models:
 - **4–8 weeks development per fine-tuning task**
 - Inference takes minutes on 1 GPU
 - Improved flexibility, performance, robustness
 - **Unlocking new capabilities, sustainability!**

Enter Aurora:

- Pre-trained 1PB of diverse data
- Fine-tuned to tackle diverse tasks
- Strong performance on operational evals



Pre-training

- **Objective:** Predict global state of any variables at any resolution 6h ahead

- **Cost:**

- 150,000 steps
- 32 A100s
- 3 weeks

Variable	Units	Description
SURFACE-LEVEL METEOROLOGICAL VARIABLES		
2T	K	Temperature at 2 m above surface of land or sea
U10	m s^{-1}	Eastward component of wind at 10 m
V10	m s^{-1}	Southward component of wind at 10 m
WS	m s^{-1}	Wind speed at 10 m; equal to $(U10^2 + V10^2)^{1/2}$
MSL	Pa	Air pressure at mean sea level
ATMOSPHERIC METEOROLOGICAL VARIABLES		
U	m s^{-1}	Eastward component of wind
V	m s^{-1}	Southward component of wind
T	K	Temperature
Q	kg kg^{-1}	Specific humidity
Z	$\text{m}^2 \text{s}^{-2}$	Geopotential

Pretraining Datasets							
Name	Resolution	Timeframe	Surface Variables	Atmospheric Variables	Num levels	Size (TB)	Num frames
ERA5	$0.25^\circ \times 0.25^\circ$	1979-2020	2T, U10, V10, MSL	U, V, T, Q, Z	13	105.43	367,920
HRES-0.25	$0.25^\circ \times 0.25^\circ$	2016-2020	2T, U10, V10, MSL	U, V, T, Q, Z	13	42.88	149,650
IFS-ENS-0.25	$0.25^\circ \times 0.25^\circ$	2018-2020	2T, U10, V10, MSL	U, V, T, Q, Z	3	518.41	6,570,000
GFS Forecast	$0.25^\circ \times 0.25^\circ$	2015-2020	2T, U10, V10, MSL	U, V, T, Q, Z	13	130.39	560,640
GFS Analysis	$0.25^\circ \times 0.25^\circ$	2015-2020	2T, U10, V10, MSL	U, V, T, Q, Z	13	2.04	8,760
GEFS Reforecast	$0.25^\circ \times 0.25^\circ$	2000-2019	2T, MSL	U, V, T, Q, Z	3	194.02	2,920,000
CMCC-CM2-VHR4	$0.25^\circ \times 0.25^\circ$	1950-2014	2T, U10, V10, MSL	U, V, T, Q	7	12.6	94,900
ECMWF-IFS-HR	$0.45^\circ \times 0.45^\circ$	1950-2014	2T, U10, V10, MSL	U, V, T, Q	7	3.89	94,900
MERRA-2	$0.625^\circ \times 0.5^\circ$	1980-2020	2T, U10, V10, MSL	U, V, T, Q	13	5.85	125,560
IFS-ENS-Mean	$0.25^\circ \times 0.25^\circ$	2018-2020	2T, U10, V10, MSL	U, V, T, Q, Z	3	10.37	131,400
				Total	1,219.91	11,023,730	

Fine-tuning task #1: Air pollution forecasting

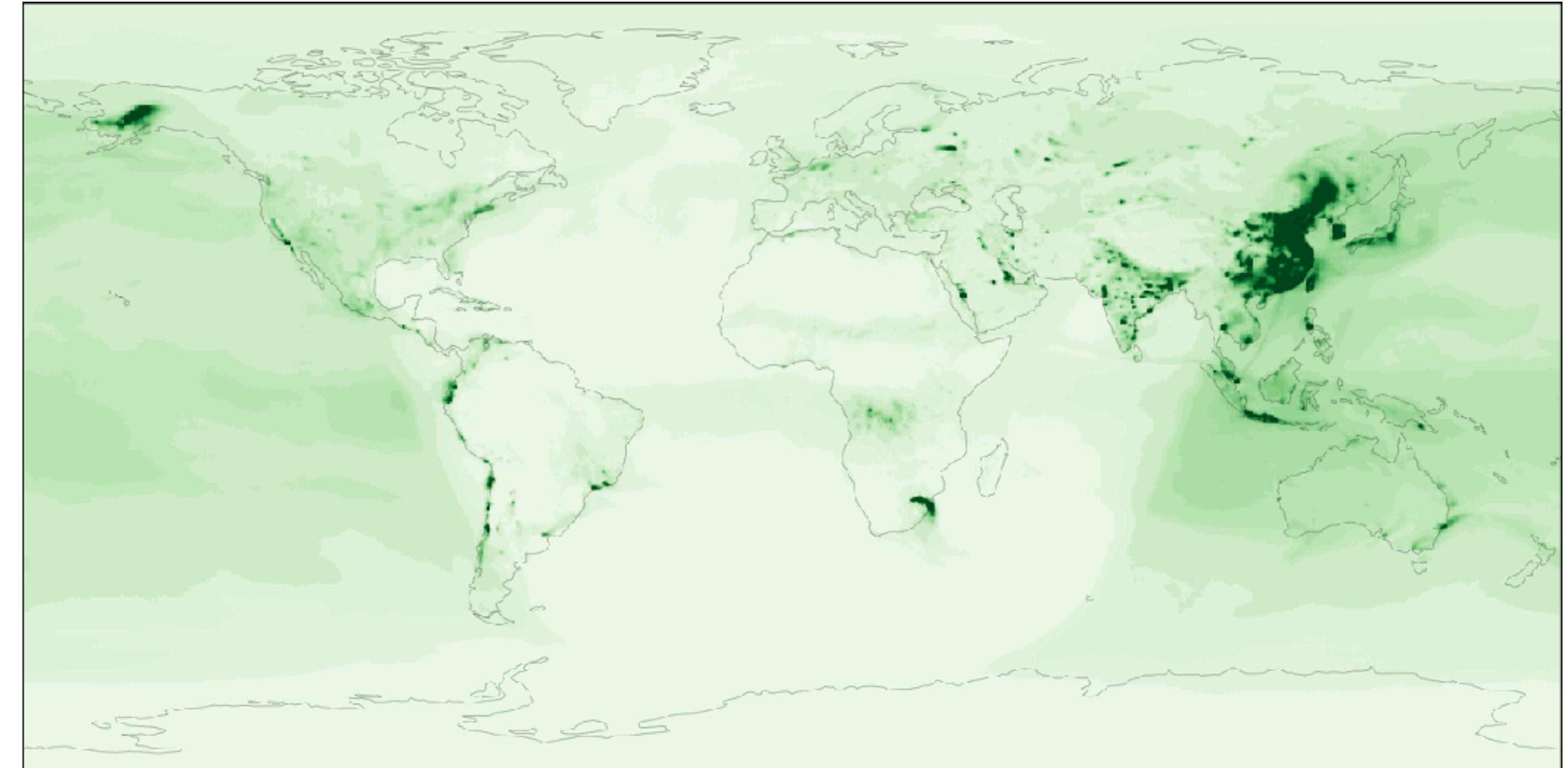
Setup: Model concentration of PM₁, PM_{2.5}, PM₁₀, CO, NO, NO₂, SO₂, O₃

Data: Copernicus Atmospheric Monitoring Service (CAMS) analysis, 0.4° resolution

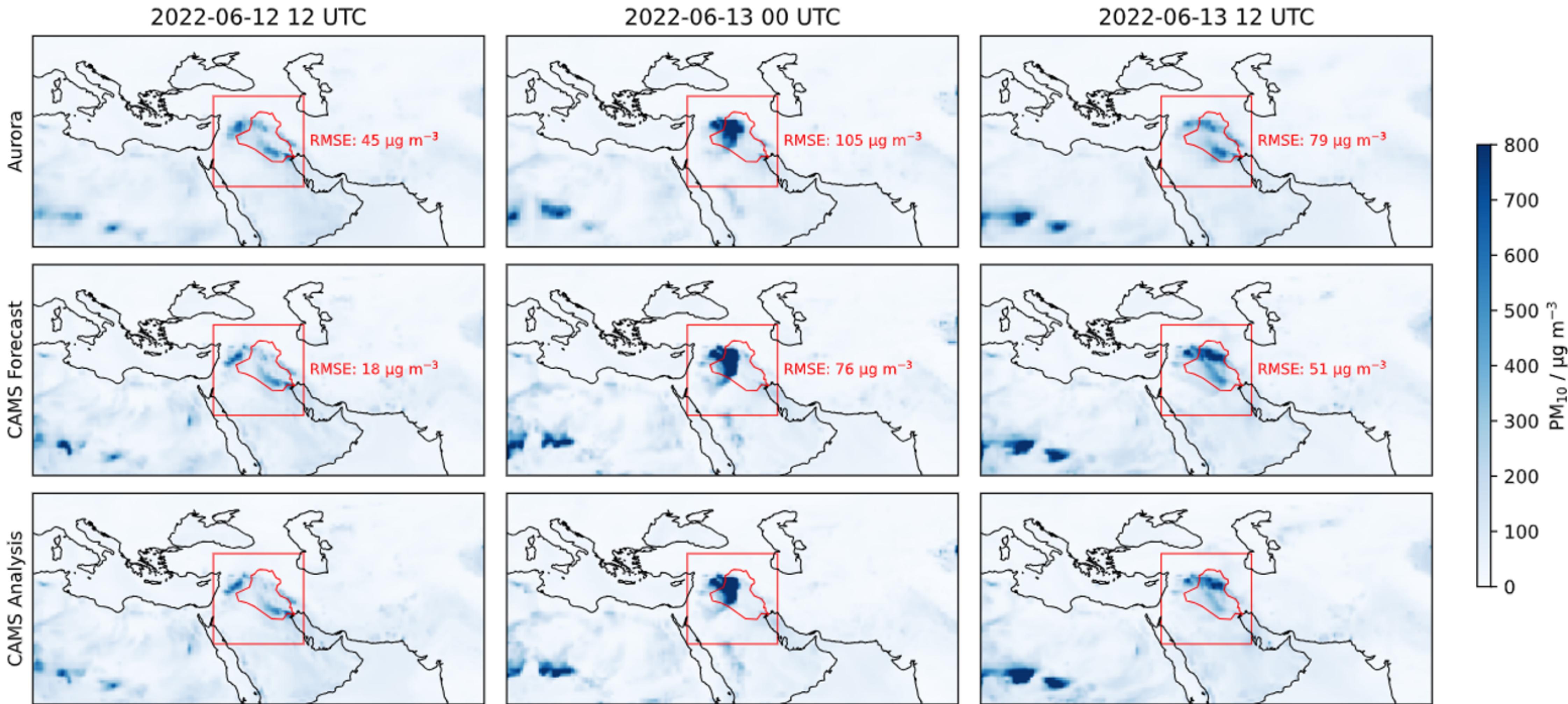
Baseline: Operational CAMS forecasts

Challenges:

- Adaptation to a new domain
- Data scarcity
- Non-stationary
- Lack of emission data



NO₂, like most variables in CAMS, is skewed towards high values in areas with high anthropogenic emissions. It also exhibits a strong diurnal cycle due to photolysis.



- Aurora accurately captures a severe sandstorm that hit Iraq on June 13, 2022.
- Initialization via CAMS analysis at 12 Jun 2022 00 UTC.

Fine-tuning task #2: Ocean wave forecasting

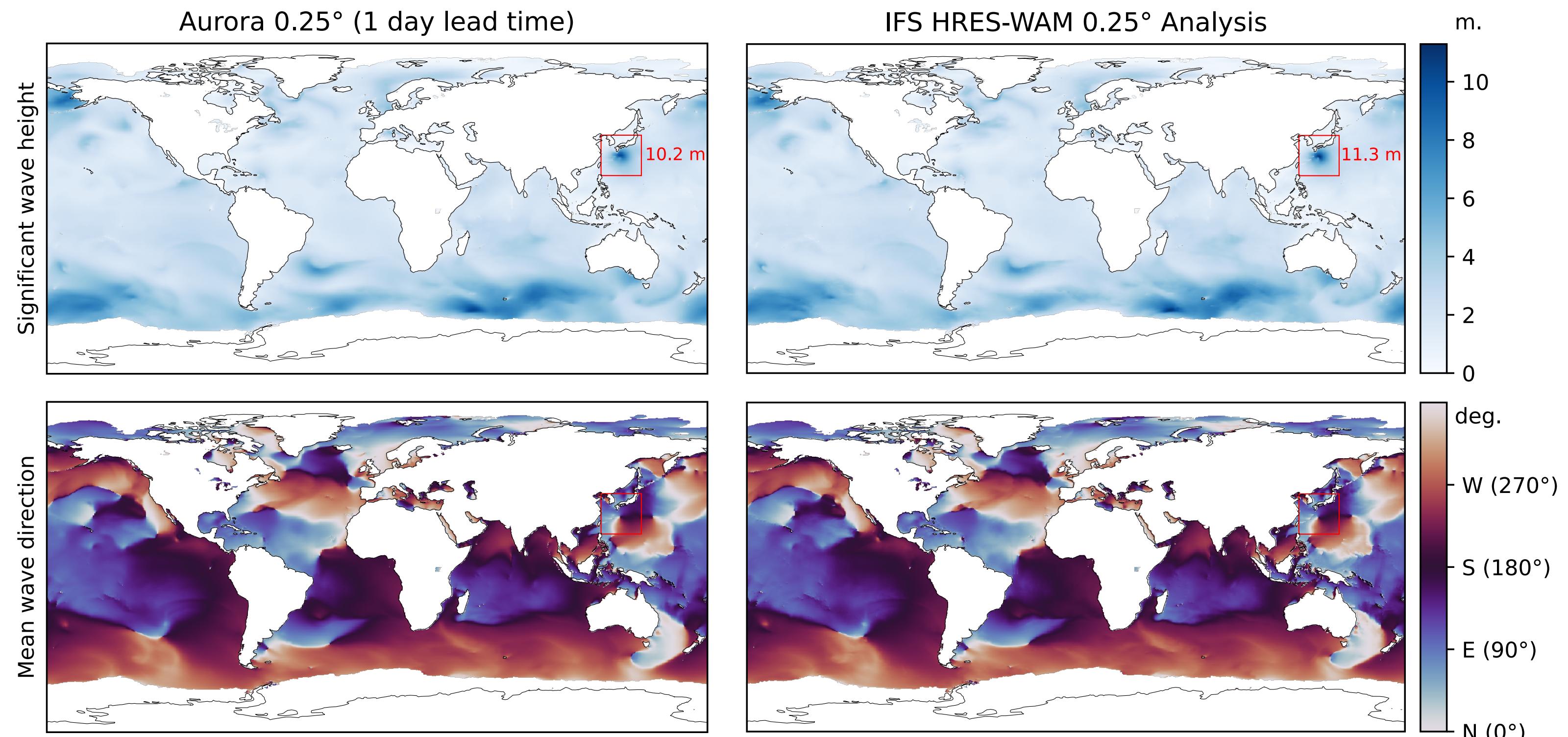
Setup: Model height, period, and direction of all wave components

Data: IFS HRES-WAM analysis, 0.25° resolution

Baseline: IFS HRES-WAM operational forecasts

Challenges:

- Adaptation to a new domain
- Data domain is not fixed (e.g., absence of swell, sea ice)
- How to model wave angles?



Aurora accurately predicts significant wave height and mean wave direction for Typhoon Nanmadol, the most intense tropical cyclone in 2022. The red box shows the location of the typhoon and the number is the peak significant wave height.

Fine-tuning task #3: High-resolution weather forecasting

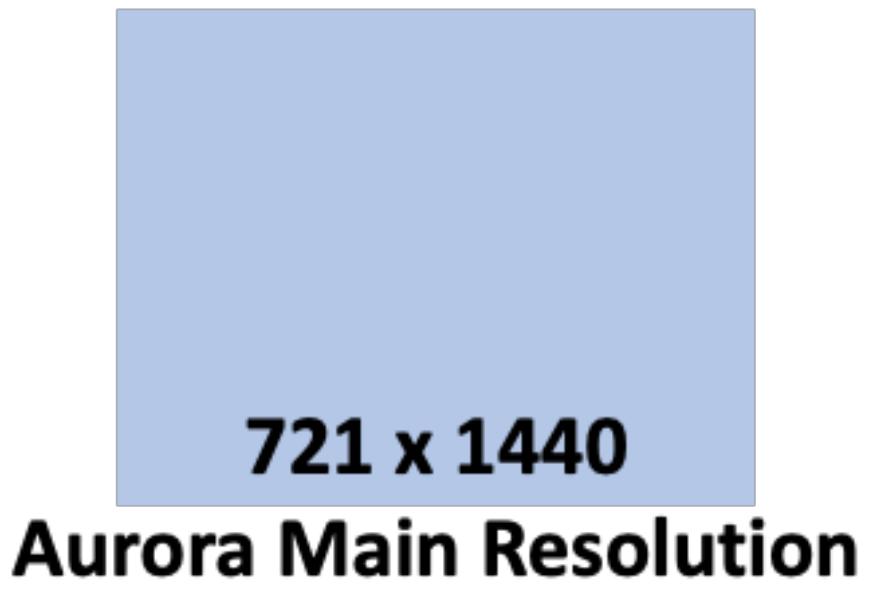
Setup: Model weather variables such as wind speed, temperature, specific humidity, etc.

Data: IFS HRES analysis, 0.1° resolution

Baseline: IFS HRES operational forecasts

Challenges:

- Adaptation to a new resolution
- Data scarcity
- Data complexity (~2GB per datapoint)



1801 x 3600
Aurora Highest Resolution

Storm Ciaran in Amsterdam



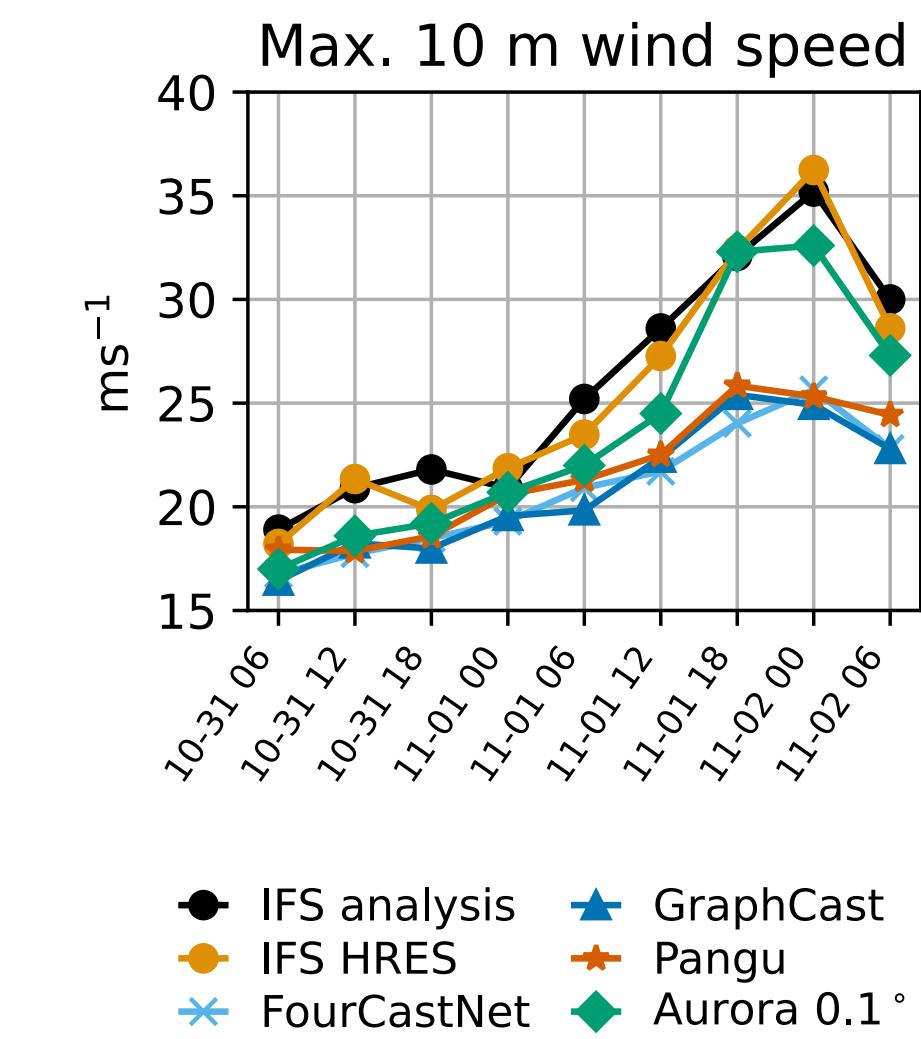
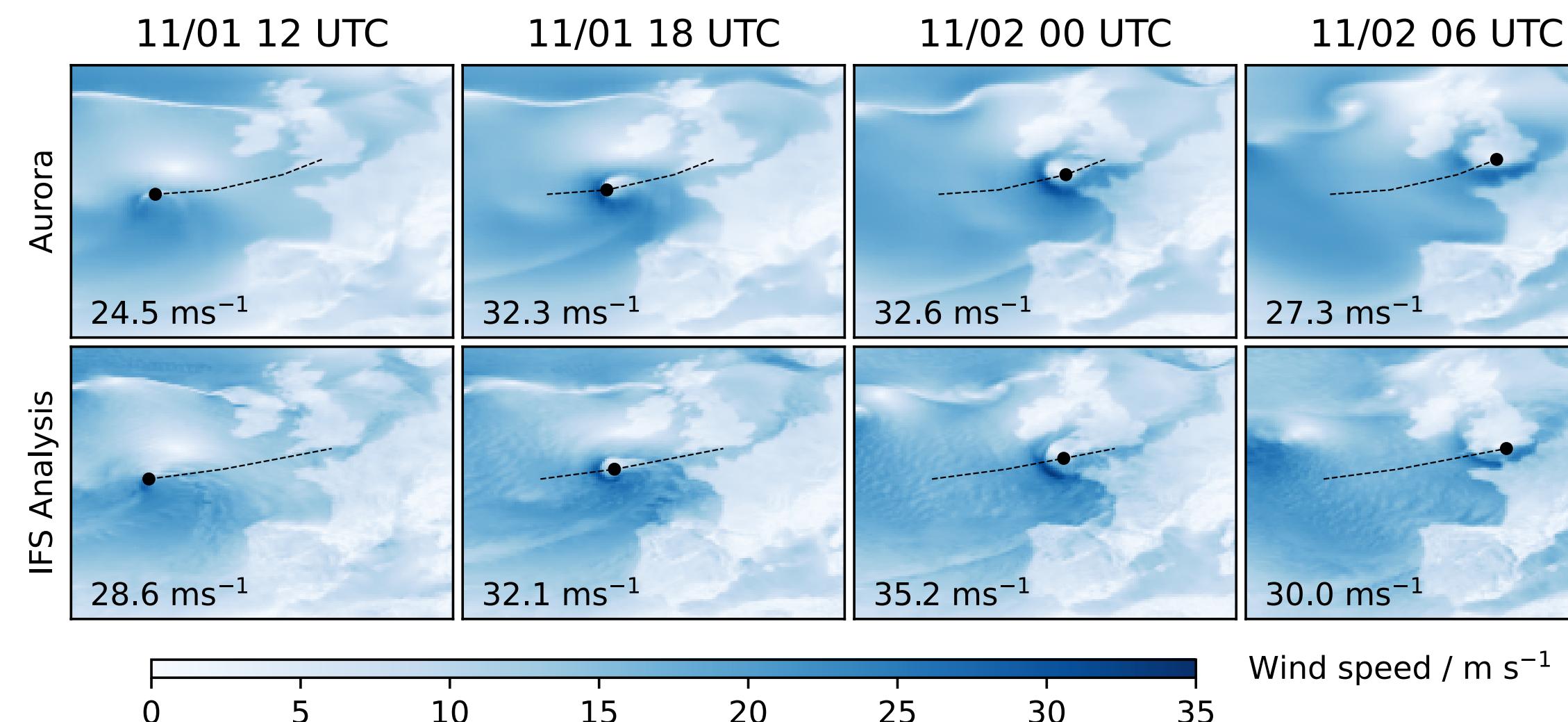
Setup: Model weather variables such as wind speed, temperature, specific humidity, etc.

Data: IFS HRES analysis, 0.1° resolution

Baseline: IFS HRES operational forecasts

Challenges:

- Adaptation to a new resolution
- Data scarcity
- Data complexity ($\sim 2\text{Gb}$ per datapoint)

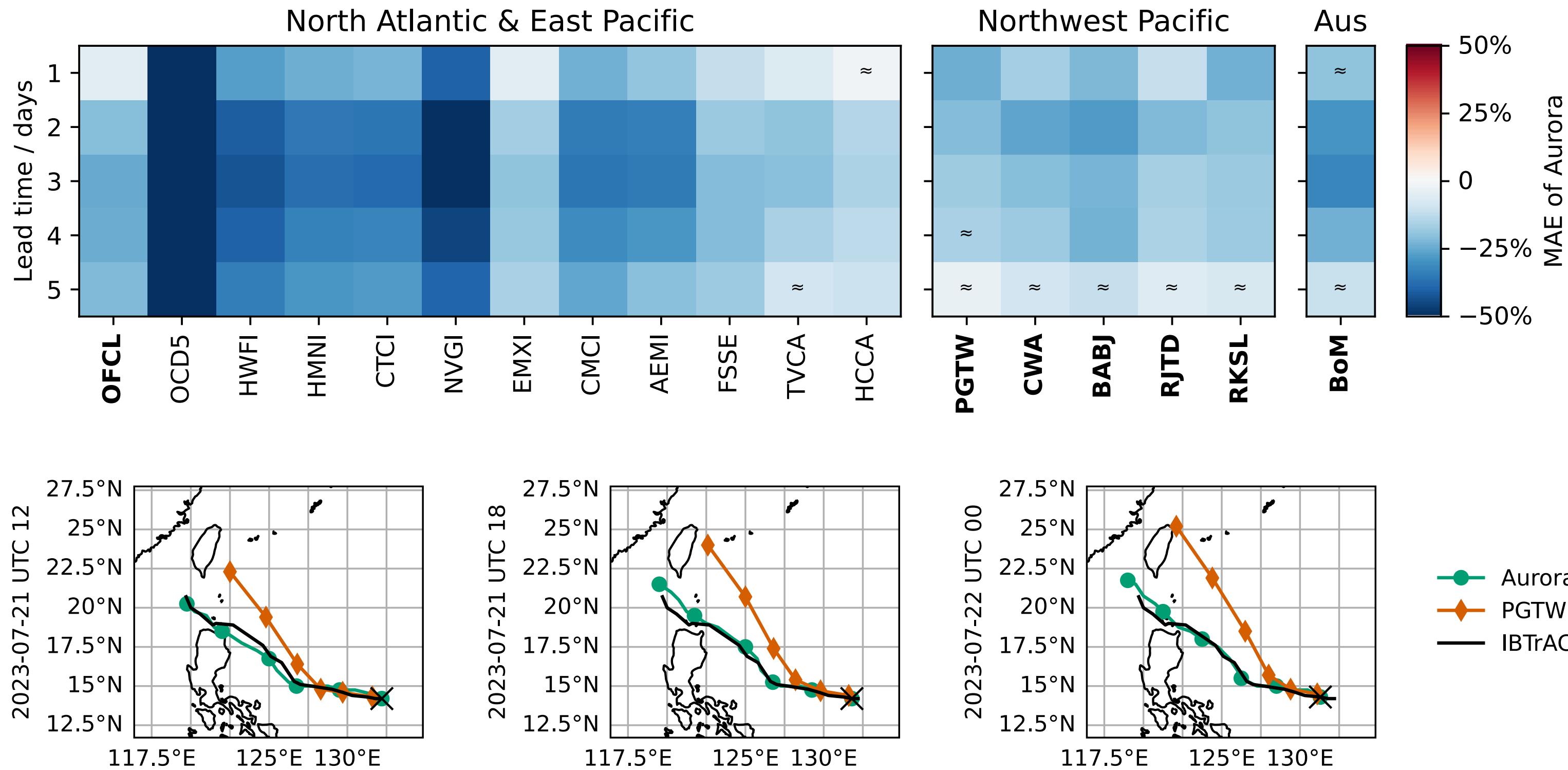


Charlton-Perez et al. (2024) showed that existing AIWP models were not able to capture the spike in maximum 10 m wind speed that occurs on 00 UTC 2 November 2023. Aurora is able to better match the IFS-HRES forecast of the sudden increase in 10 m wind speed.

Fine-tuning task #4: Tropical Hurricane Tracking

Data: Tropical hurricane tracks in 2023

Baselines: Operational forecasts issued by multiple centers worldwide



- First AI model to outperform multiple operational centers in various regions.

On 21 July, 2023 a tropical depression intensified into a tropical storm and was named Typhoon Doksuri. Doksuri would become the costliest Pacific typhoon to date, inflicting more than 28 billion USD in damage. Aurora correctly predicts that Doksuri will make landfall in the Northern Philippines, whereas PGTW predicts that it will pass over Taiwan.

Global medium-range weather forecasting

A breakdown of the success story

- Breakthroughs are obtained due to 3 reasons:
 1. **Model scale:** Vision Transformers / Swin Transformers have been proven as go-to method for 2D/3D vision applications): Pangu, Aurora, ...
 2. **Data scale:** ERA5 is a publicly available dataset which is easy-accessible and large enough. Aurora uses a plethora of similar datasets.
 3. **Tasks / Metrics:** MSE on next time step (6 hours ahead). We take a snapshot of the earth and predict 6 hours into the future. Tasks / metrics is the enabler route!

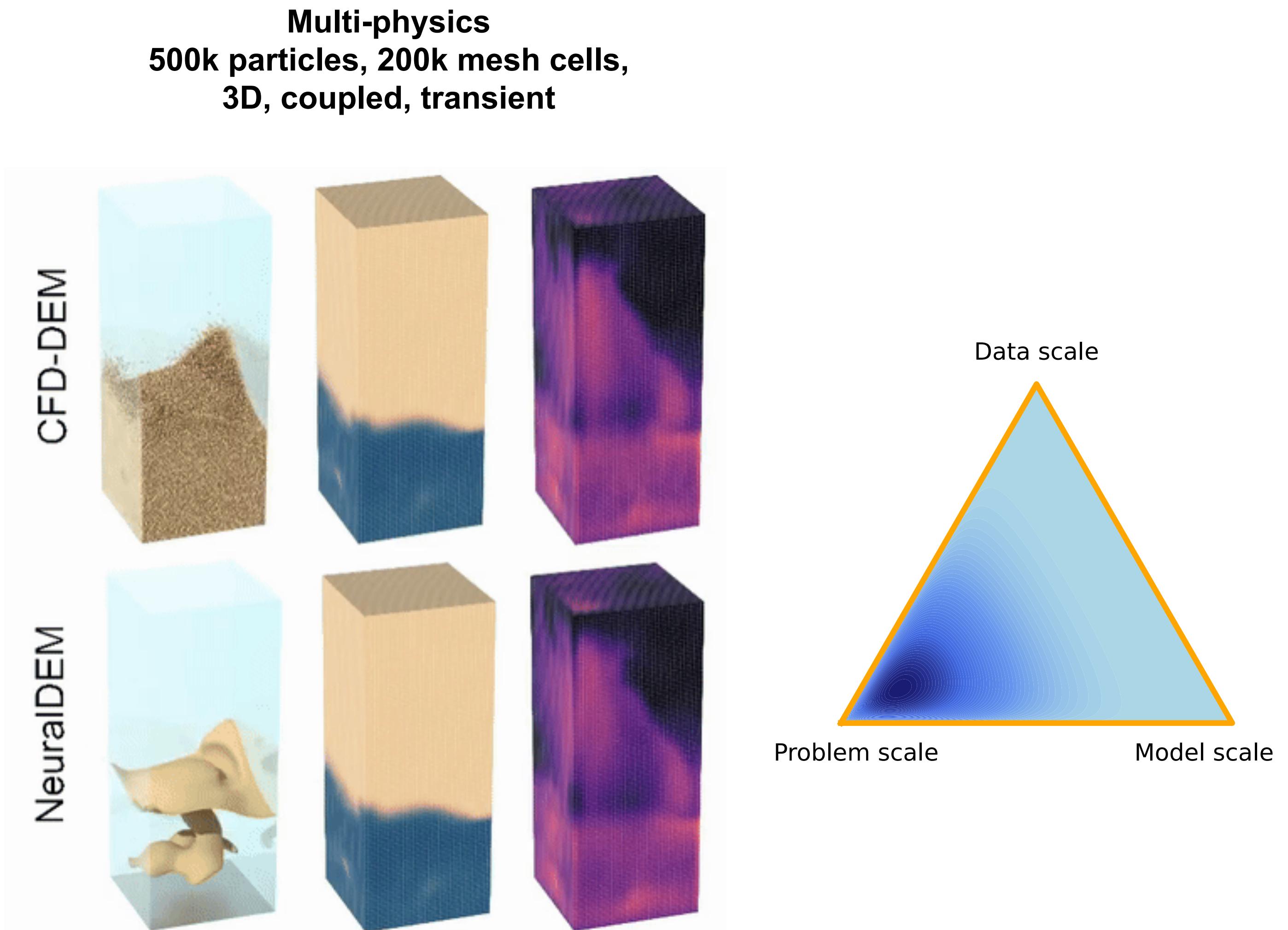
From weather to reference
models for engineering

The elevator pitch for scientists

We are facing new unknown ML challenges



CFD
1 billion mesh cells, 3D
Non-transient



Problem scale

We first need to build model frameworks that take the role of ViTs in weather modeling

Universal Physics Transformers: A Framework For Efficiently Scaling Neural Operators

Benedikt Alkin^{1,2} Andreas Fürst¹ Simon Schmid³ Lukas Gruber¹
Markus Holzleitner⁴ Johannes Brandstetter^{1,2}

¹ ELLIS Unit Linz, Institute for Machine Learning, JKU Linz, Austria

² Emmi AI GmbH, Linz, Austria

³ Software Competence Center Hagenberg GmbH, Hagenberg, Austria

⁴ MaLGa Center, Department of Mathematics, University of Genoa, Italy, Austria
{alkin, fuerst, brandstetter}@m1.jku.at

NeuralDEM – Real-time Simulation of Industrial Particulate Flows

Benedikt Alkin^{†,*,1,2} Tobias Kronlachner^{†,*,1,3} Samuele Papa^{†,‡,1,4,5}
Stefan Pirker³ Thomas Lichtenegger^{1,3} Johannes Brandstetter^{§,1,2}

¹Emmi AI GmbH, Linz, Austria

²ELLIS Unit Linz, Institute for Machine Learning, JKU Linz, Austria

³Department of Particulate Flow Modelling, JKU Linz, Austria

⁴University of Amsterdam, Amsterdam, Netherlands

⁵The Netherlands Cancer Institute, Amsterdam, Netherlands

AB-UPT: Scaling Neural CFD Surrogates for High-Fidelity Automotive Aerodynamics Simulations via Anchored-Branched Universal Physics Transformers

Benedikt Alkin^{*,1}, Maurits Bleeker^{*,1}, Richard Kurle^{*,1}, Tobias Kronlachner^{*,1},
Reinhard Sonnleitner¹, Matthias Dorfer¹, Johannes Brandstetter^{1,2}

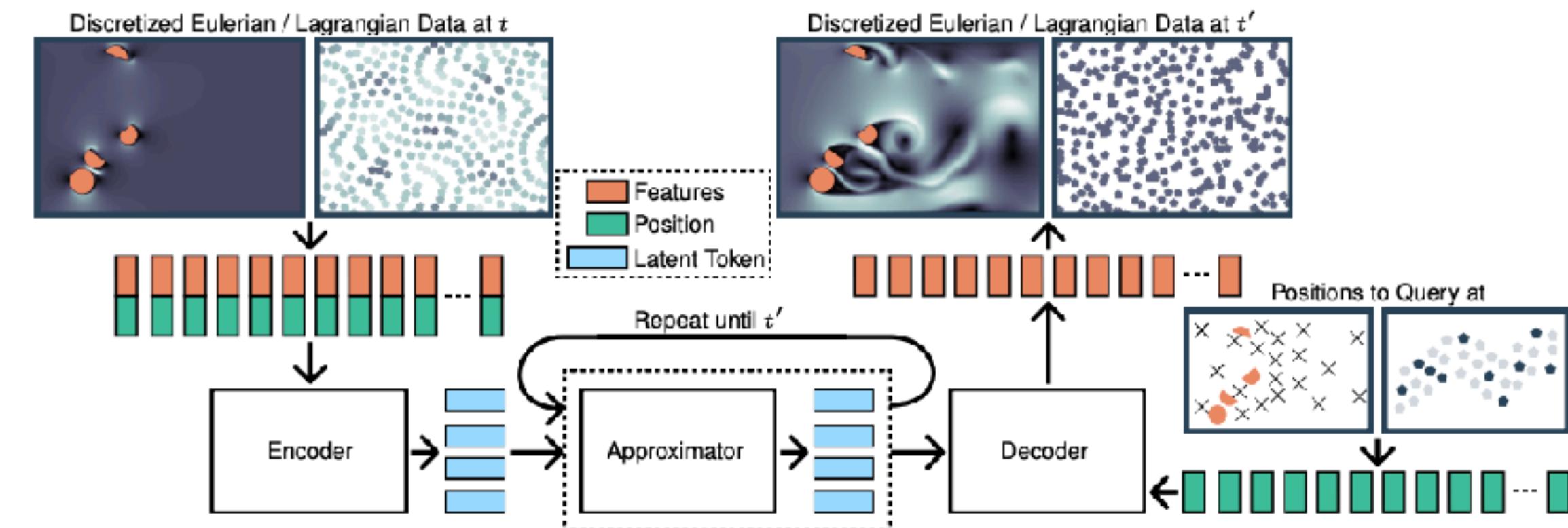
*Equal contribution

¹Emmi AI GmbH

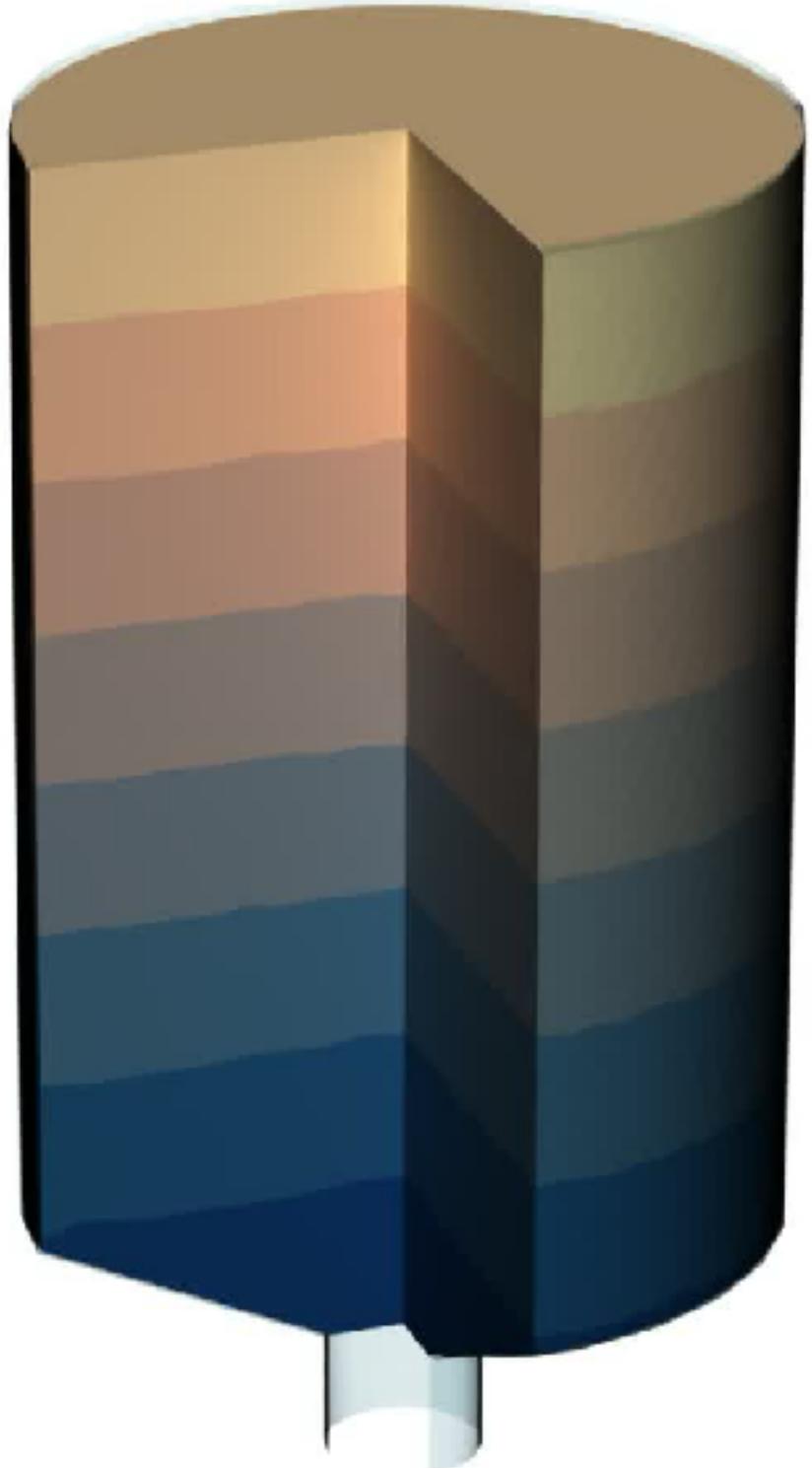
²Ellis Unit, LIT AI Lab, JKU Linz

Correspondence to benedikt@emmi.ai, johannes@emmi.ai

<https://github.com/Emmi-AI/AB-UPT>

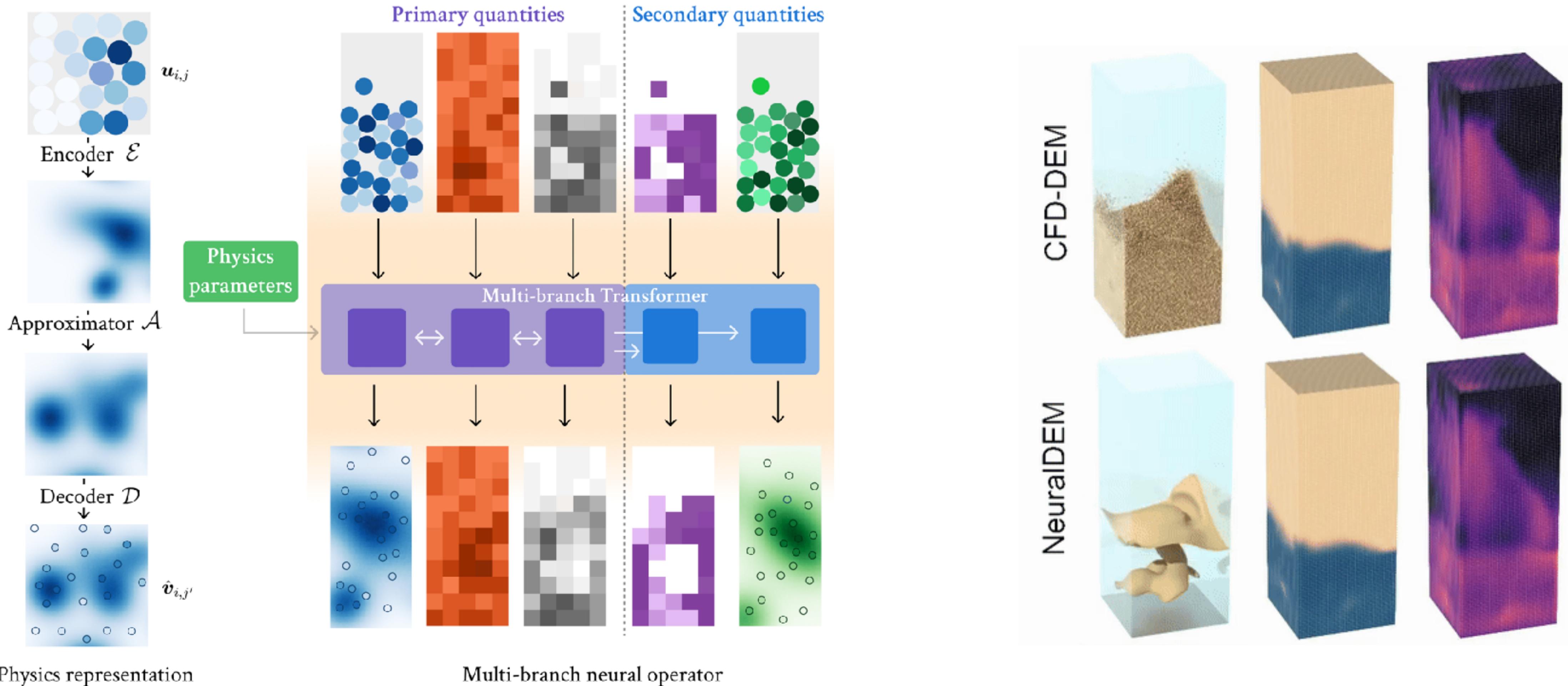


Field point of view



NeuralDEM
Alkin et al.

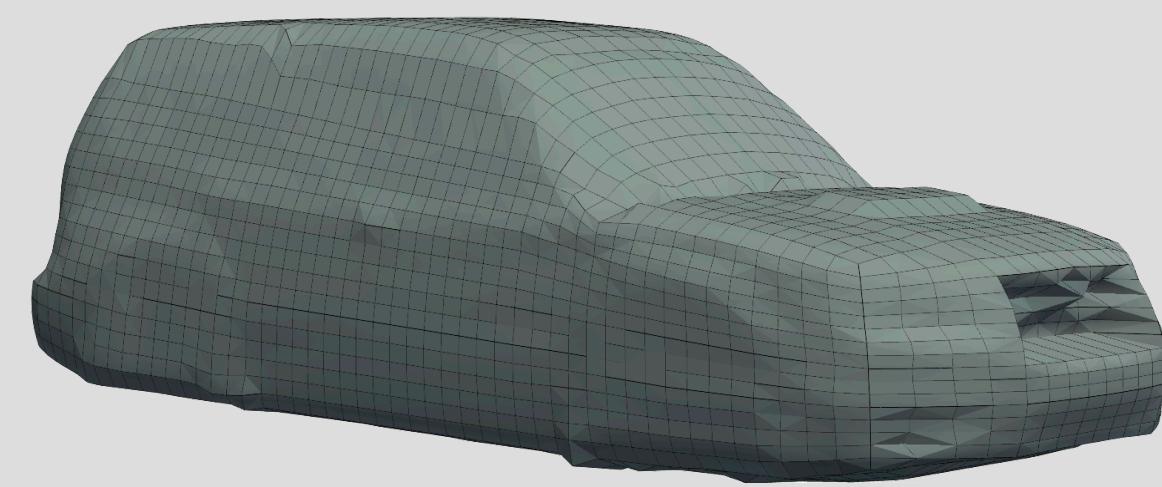
Multi-branch neural operators



Example:CFD

- Physics d.o.f. != simulation mesh
- The fine-grained simulation mesh is needed for numerics, not for ML.
- Quantities such as drag or lift coefficient need full surface resolution.
- Similarly, for many analyses full volume meshes need to be resolved.
- Data is scarce, yet lots of information is within one data sample (physics is the same!)

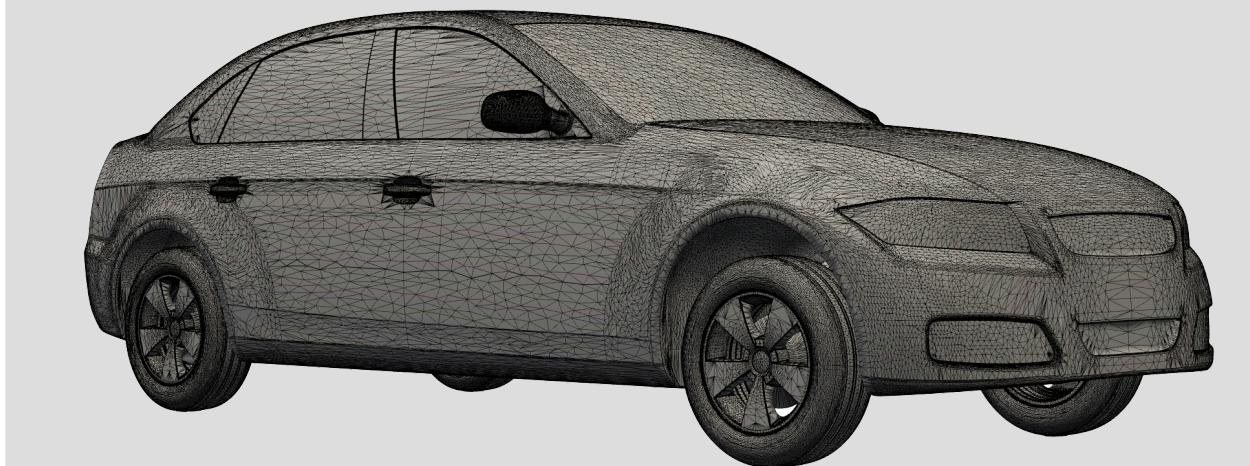
Small problem scale



Low number of simulation mesh cells

Simple physics phenomena

Large problem scale

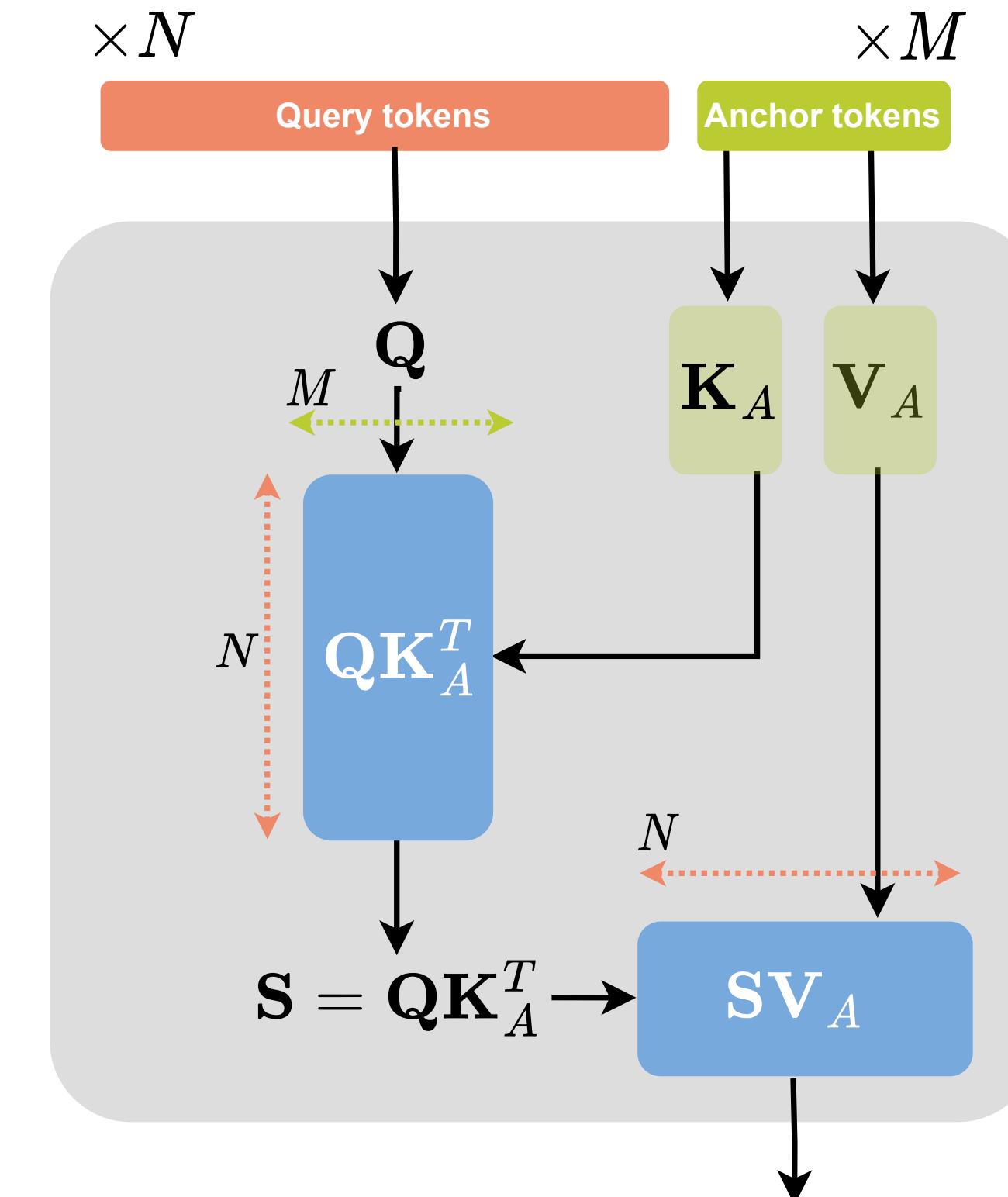
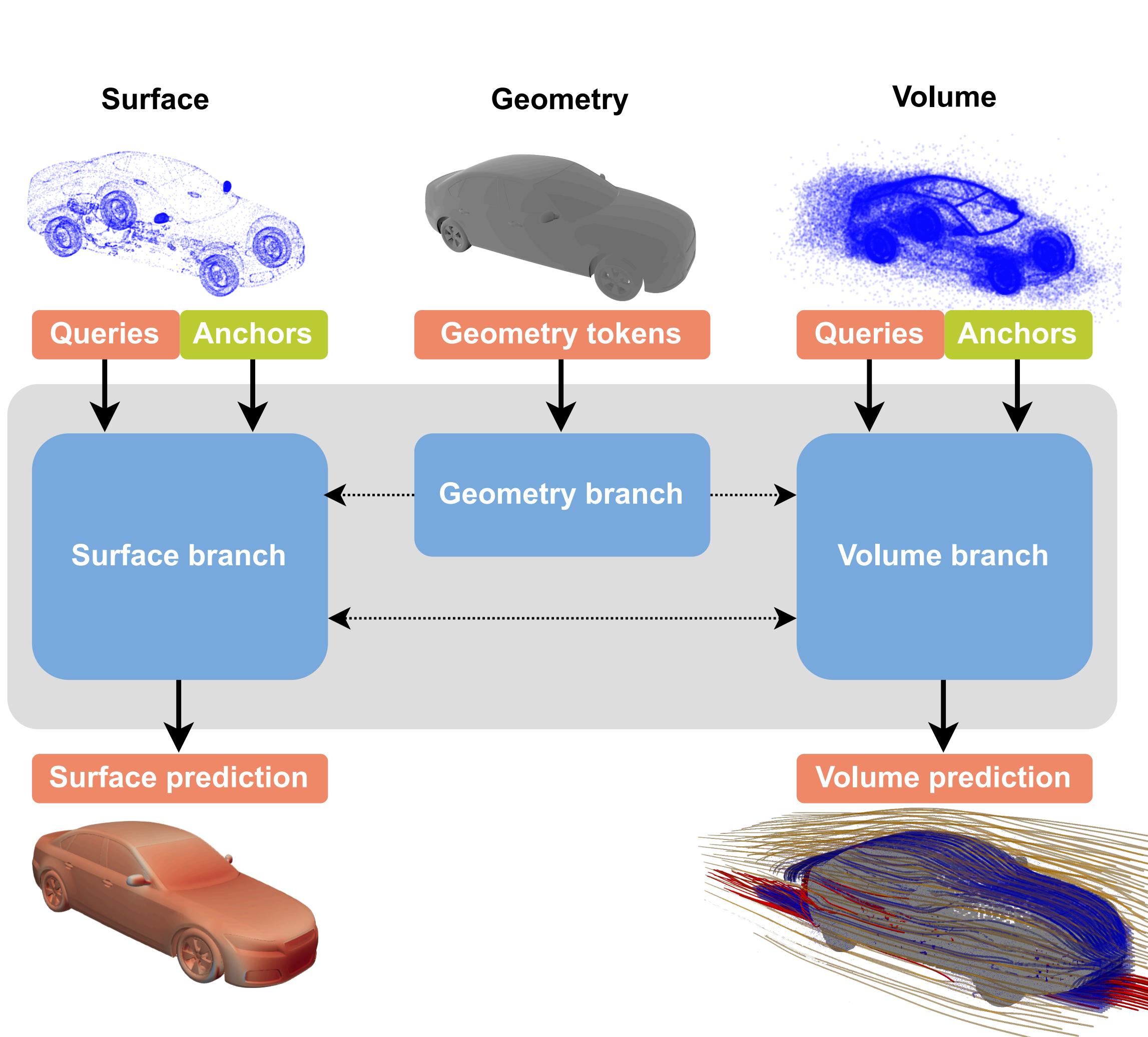


High number of simulation mesh cells

Highly complex physics phenomena

Multi-branch anchor attention

... opens lots of new modeling possibilities



AB-UPT: Scaling Neural CFD Surrogates for High-Fidelity Automotive Aerodynamics Simulations via Anchored-Branched Universal Physics Transformers

Benedikt Alkin^{*1}, Maurits Bleeker^{*1}, Richard Kurle^{*1}, Tobias Kronlachner^{*1},
Reinhard Sonnleitner¹, Matthias Dorfer¹, Johannes Brandstetter^{1,2}

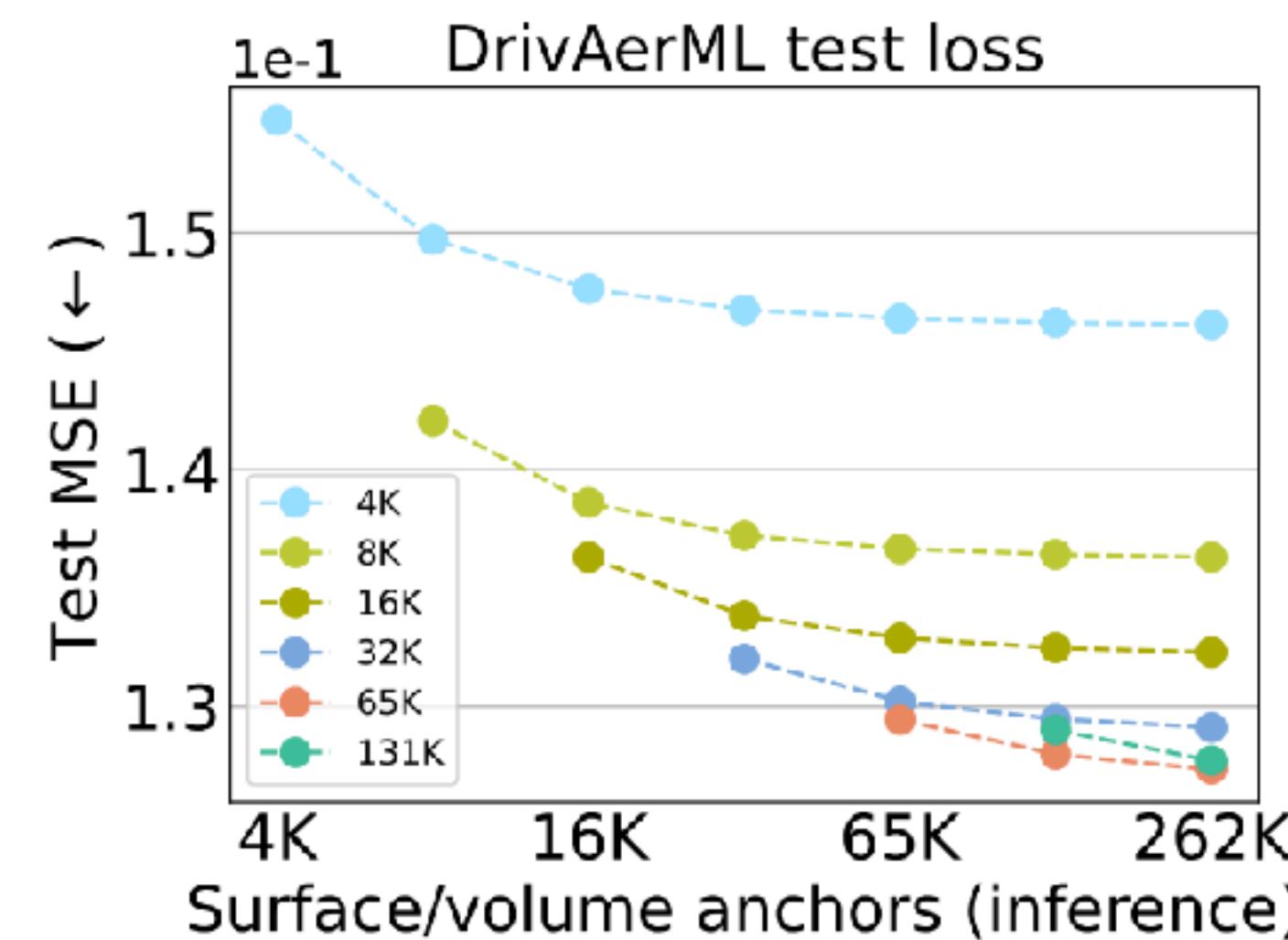
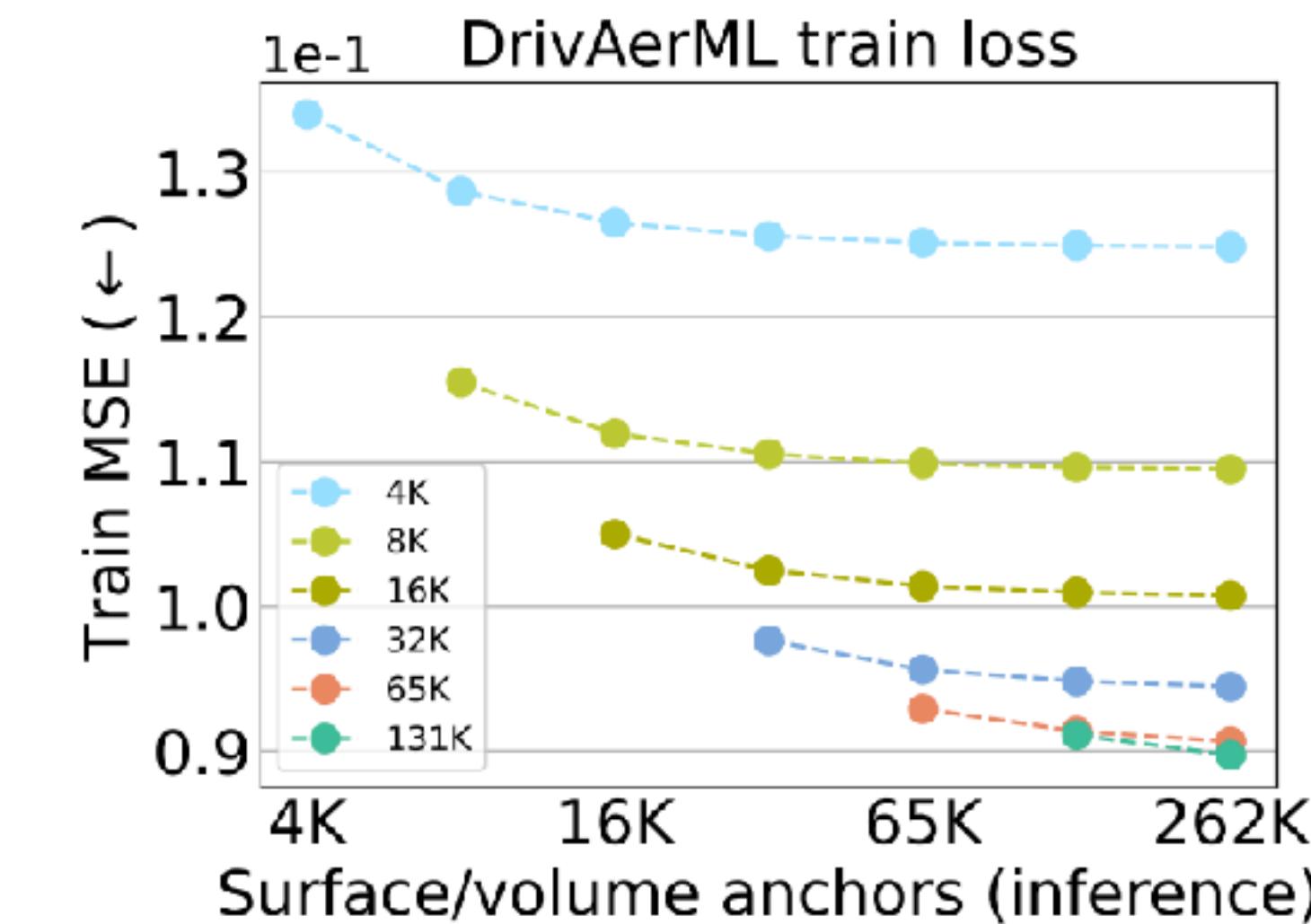
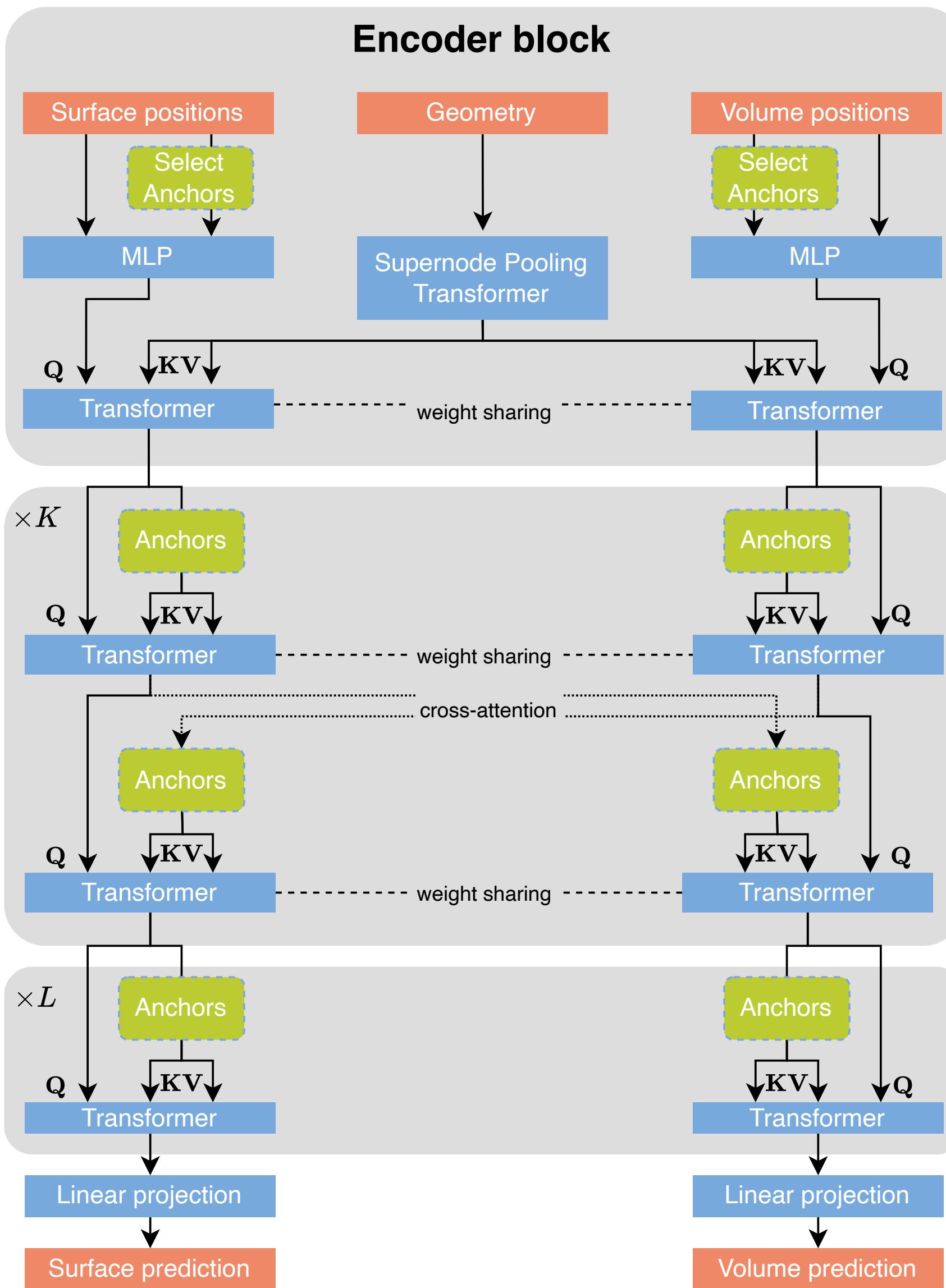
^{*}Equal contribution

¹Emmi AI GmbH

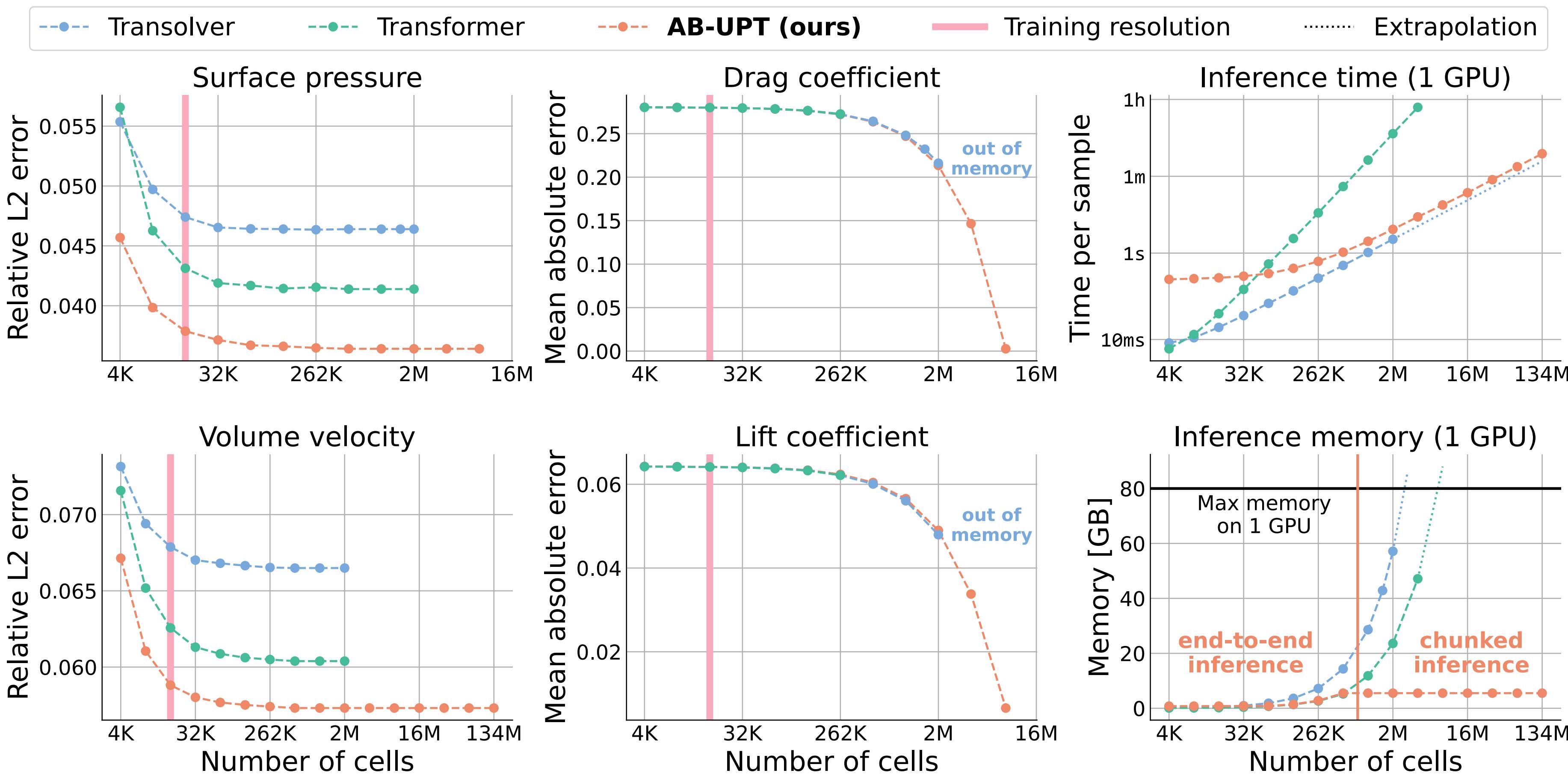
²Ellis Unit, LIT AI Lab, JKU Linz

Correspondence to johannes@emmi.ai

Anchor attention



Model properties



Data scale

Data is the oil in engineering

- There is no ERA5 in engineering.
- Engineering simulations are costly, compute-heavy, complicated, ...
- Data comes with different fidelities, which sometimes amount to orders of magnitude in compute requirements.
- Companies sit on their data.
- ML workflows are widely missing.
- Data is IP.

Example

Large-scale CFD

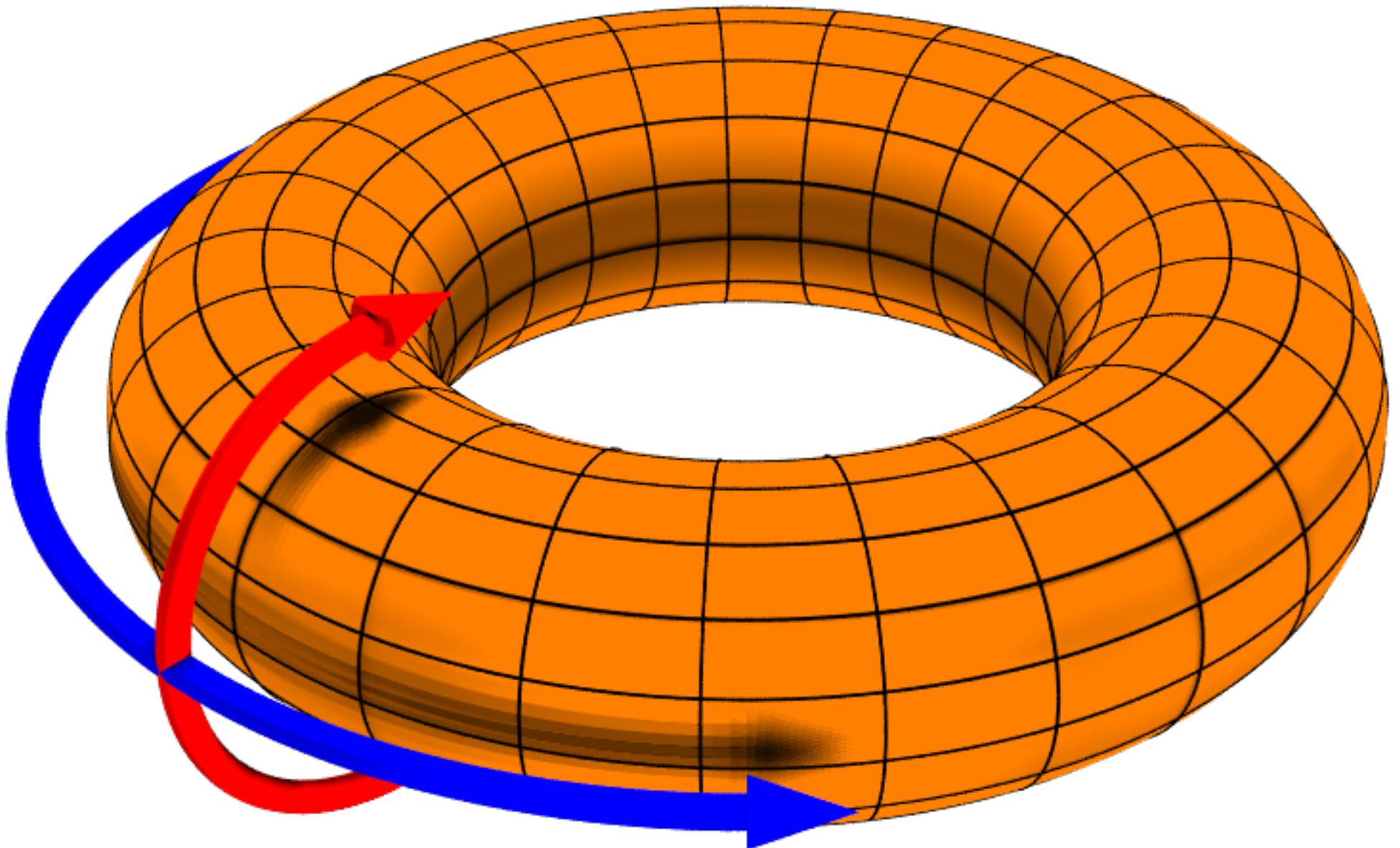


Nuclear fusion

5-dim gyrokinetic framework

**GyroSwin: 5D Surrogates for Gyrokinetic
Plasma Turbulence Simulations**

- Turbulence is a key driver of plasma confinement degradation, as it causes plasma to diffuse towards the reactor wall.
- Evolution of particles (ions and electrons) is described in terms of distribution function (3D space, 3D velocity space)
- Perpendicular fluctuations scale much smaller than the system size.
- Experimentally proved to be a good model for turbulence.



Fabian Paischer^{*1,3} Gianluca Galletti^{*1} William Hornsby² Paul Setinek¹

Lorenzo Zanisi² Naomi Carey² Stanislas Pamela² Johannes Brandstetter^{1,3}

¹ ELLIS Unit, Institute for Machine Learning, JKU Linz

² United Kingdom Atomic Energy Authority, Culham campus

³ EMMI AI, Linz

{paischer,galletti,brandstetter}@ml.jku.at

github.com/ml-jku/neural-gyrokinetics

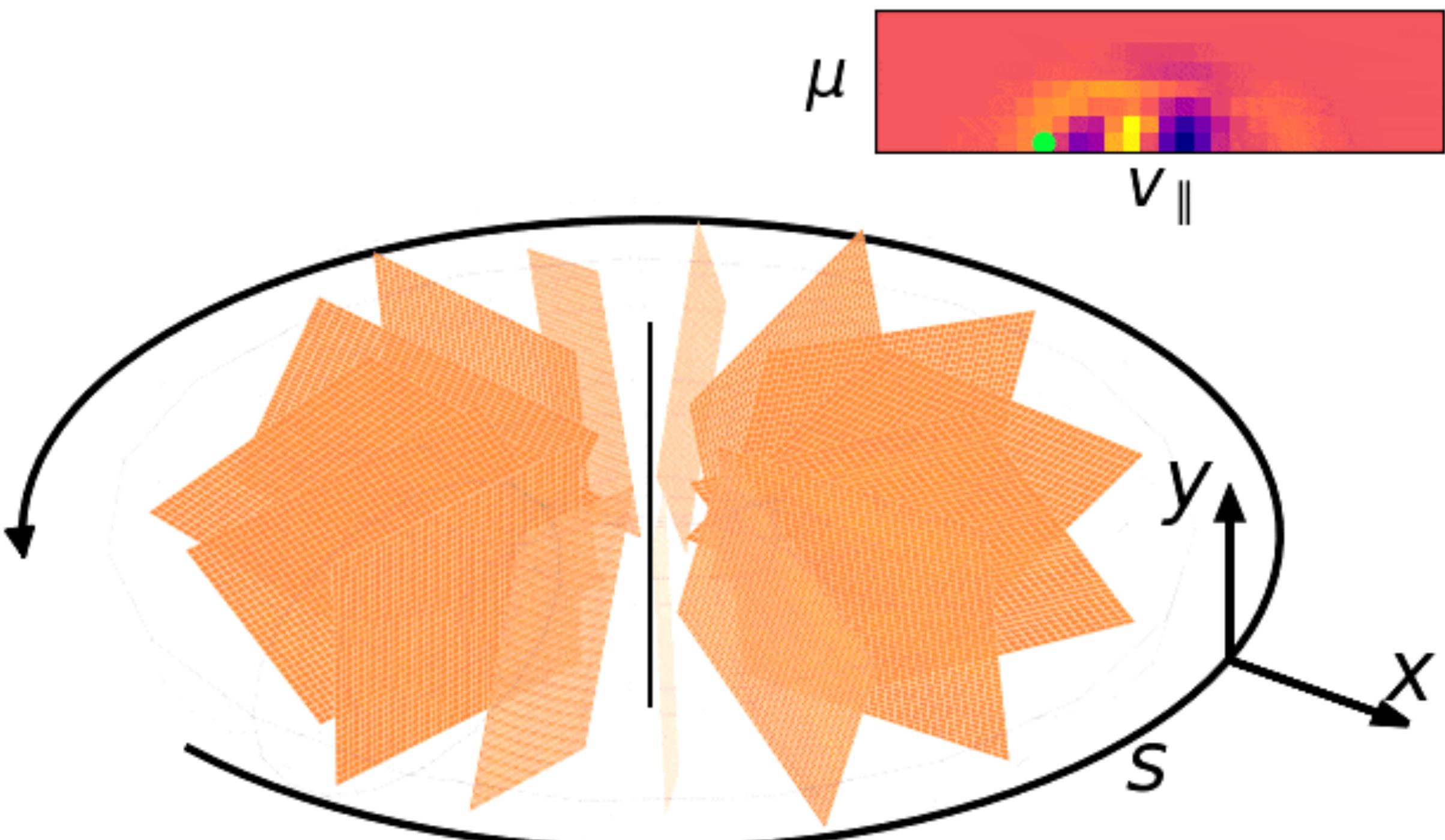
GyroSwin results

Table 1: Comparison of different surrogate approaches by capabilities.

Method	Average Flux	Diagnostics	Zonal Flows	Turbulence
Tabular Regressors, e.g., GPR, MLP	1D → 0D	X	X	X
SOTA Reduced Numerical modelling, e.g., QL	3D → 0D	3D → 1D	X	X
Neural Surrogates, e.g. GyroSwin (Ours)	5D → 0D	5D → 1D	5D → 1D	5D → 5D

Table 2: Evaluation for 5D turbulence modelling and nonlinear heat flux prediction. We evaluate all methods across six in-distribution (**ID**) and five out-of-distribution (**OOD**) simulations. For Q we report RMSE of time-averaged predictions after an autoregressive rollout. For f we report correlation time for autoregressive rollouts with threshold $\tau = 0.1$. Higher correlation time is better.

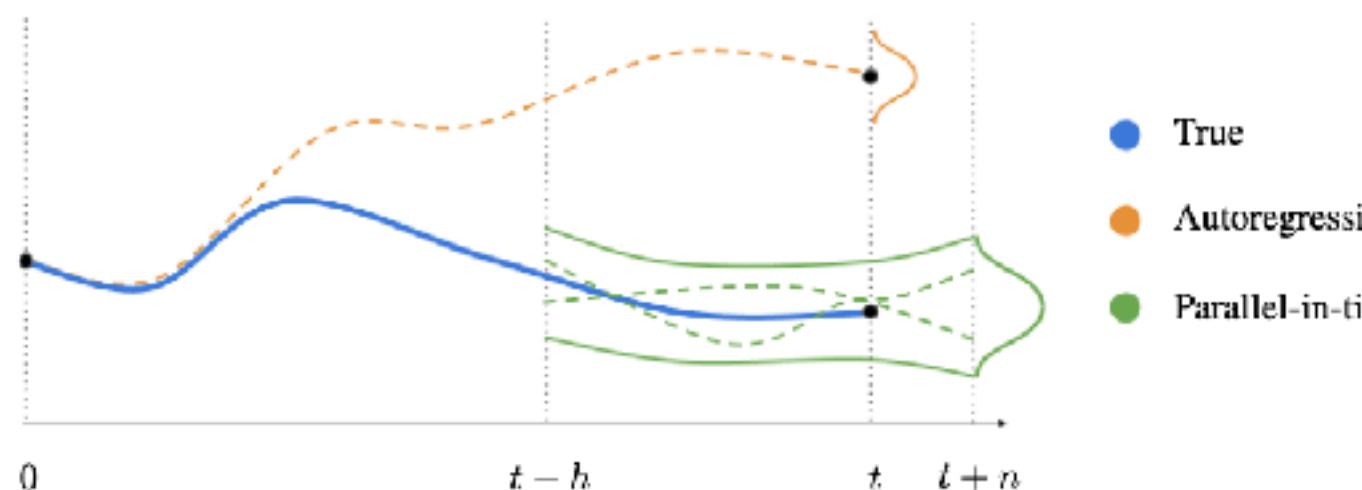
Method	Input	f		Q	
		ID (↑)	OOD (↑)	ID (↓)	OOD (↓)
<i>SOTA Reduced Numerical modelling</i>					
QL (Bourdelle et al., 2007)	3D	n/a	n/a	89.53 ± 11.76	95.22 ± 21.57
<i>Classical Regression Techniques</i>					
GPR (Hornshy et al., 2024)	0D	n/a	n/a	43.82 ± 10.84	59.28 ± 17.55
MLP	0D	n/a	n/a	50.50 ± 10.79	61.98 ± 18.41
<i>Neural Surrogate Models (48 simulations)</i>					
FNO (Li et al., 2021)	3D	9.33 ± 0.56	9.20 ± 0.58	119.88 ± 13.15	124.96 ± 23.27
PointNet (Qi et al., 2016)	5D	7.33 ± 0.21	7.40 ± 0.24	119.93 ± 13.15	125.05 ± 23.29
Transolver (Wu et al., 2024)	5D	9.83 ± 1.40	10.80 ± 1.46	119.93 ± 13.15	125.05 ± 23.28
ViT (Dosovitskiy et al., 2021)	5D	16.83 ± 1.49	19.20 ± 1.36	119.63 ± 13.13	125.13 ± 23.29
GyroSwin (Ours)	5D	26.50 ± 3.55	28.60 ± 8.82	67.68 ± 10.28	70.48 ± 17.21
<i>Scaling GyroSwin to 241 simulations</i>					
GyroSwin _{Small} (Ours)	5D	98.00 ± 27.53	76.40 ± 17.60	23.72 ± 4.05	53.54 ± 18.10
GyroSwin _{Medium} (Ours)	5D	94.17 ± 21.96	91.20 ± 18.61	37.24 ± 9.60	44.17 ± 17.68
GyroSwin (Ours)	5D	110.33 ± 19.74	111.80 ± 23.86	18.35 ± 1.56	26.43 ± 9.49



Open challenges

- Multi-fidelity datasets
- Real-world measurements (digital twins)
- Sim to real gap
- Transient simulations / data storage

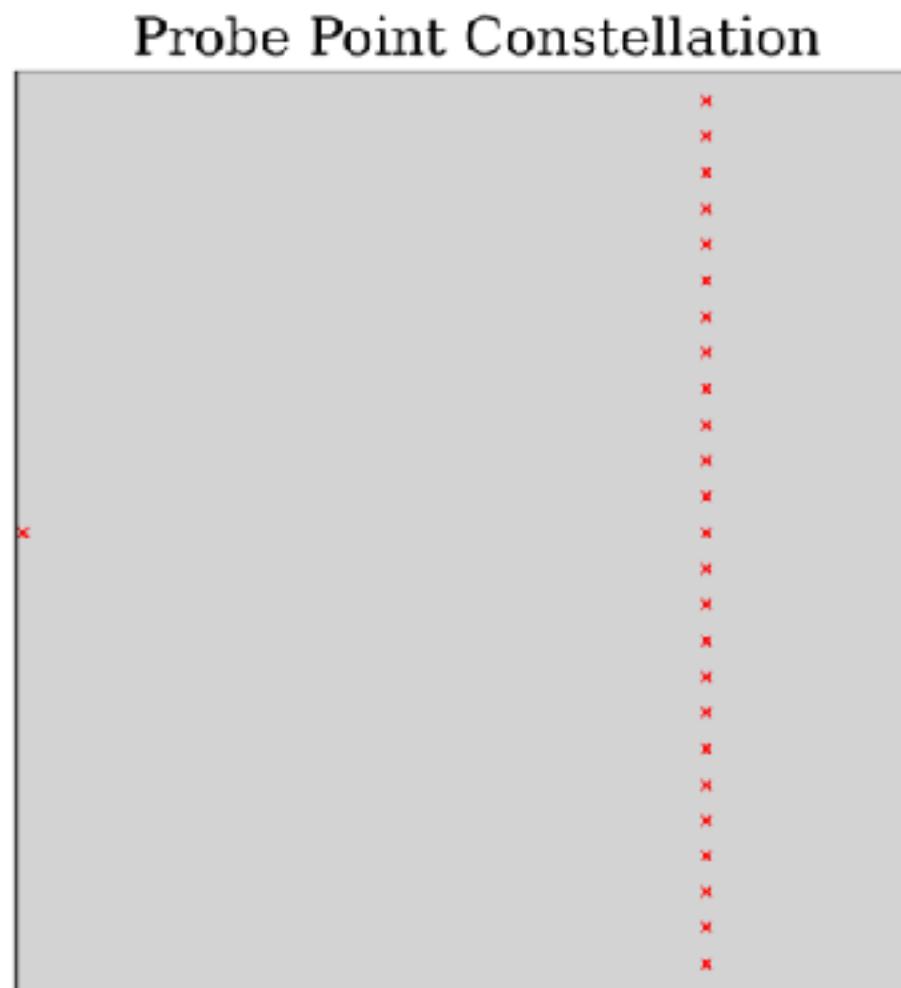
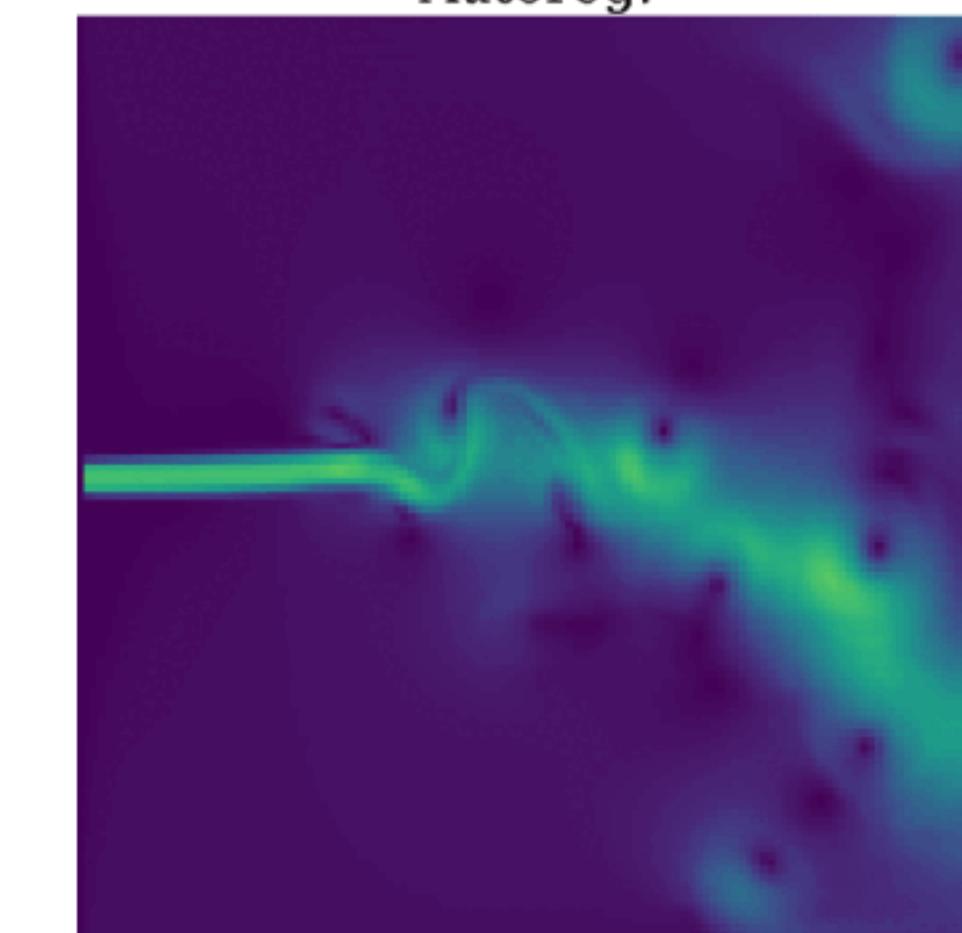
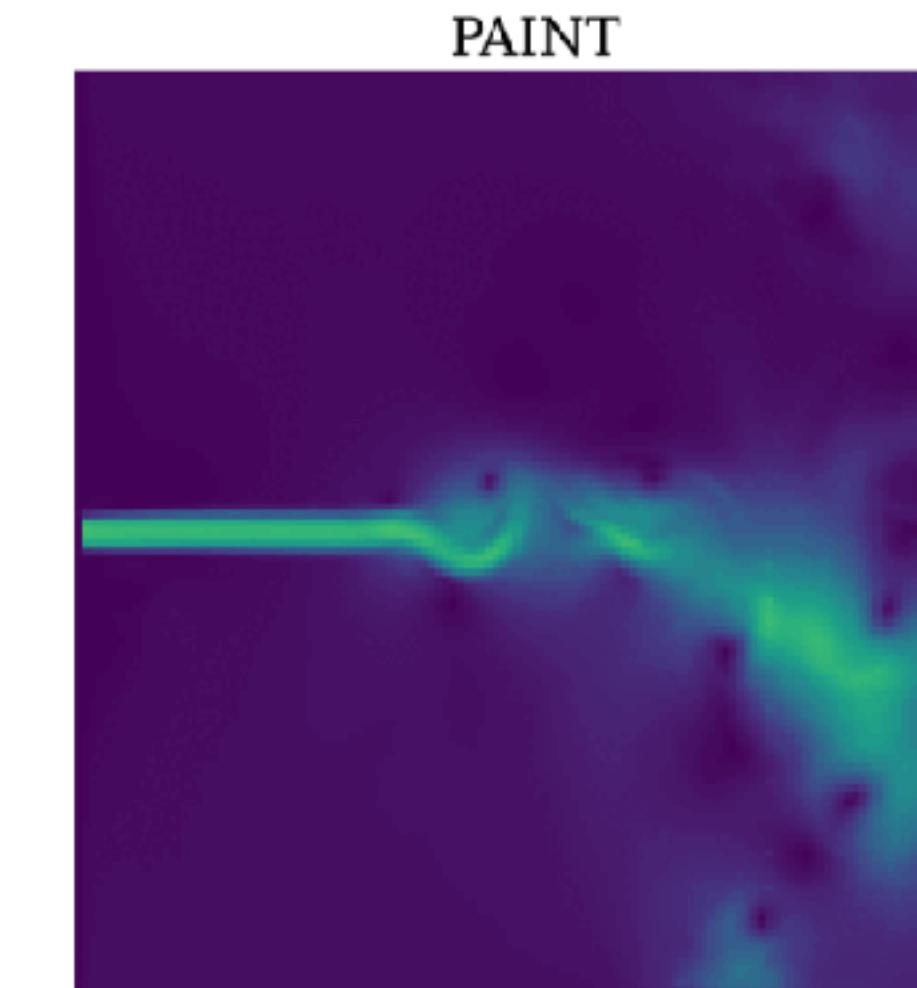
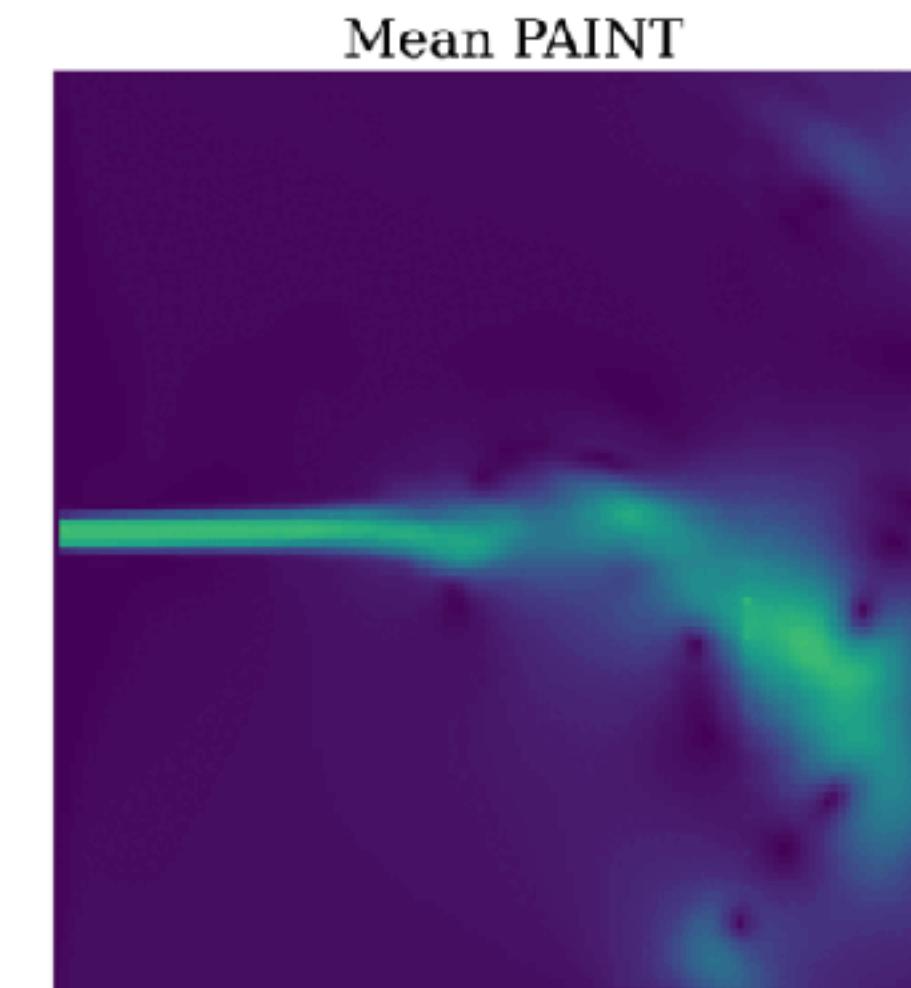
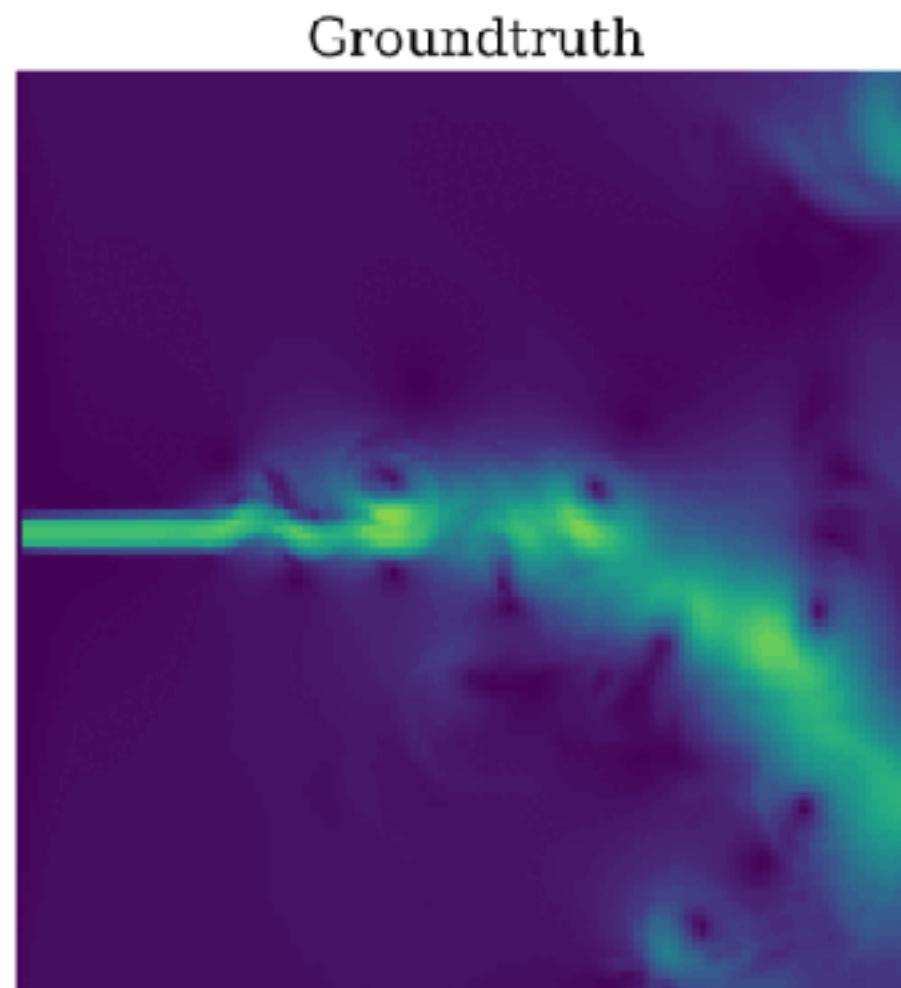
PAINT



PAINT: Parallel-in-time Neural Twins for Dynamical System Reconstruction

Andreas Radler^{* 1} Vincent Seyfried^{* 2,3}

Stefan Pirker² Johannes Brandstetter^{1,3} Thomas Lichtenegger^{2,3}



MSE

PAINT

Autoreg.

