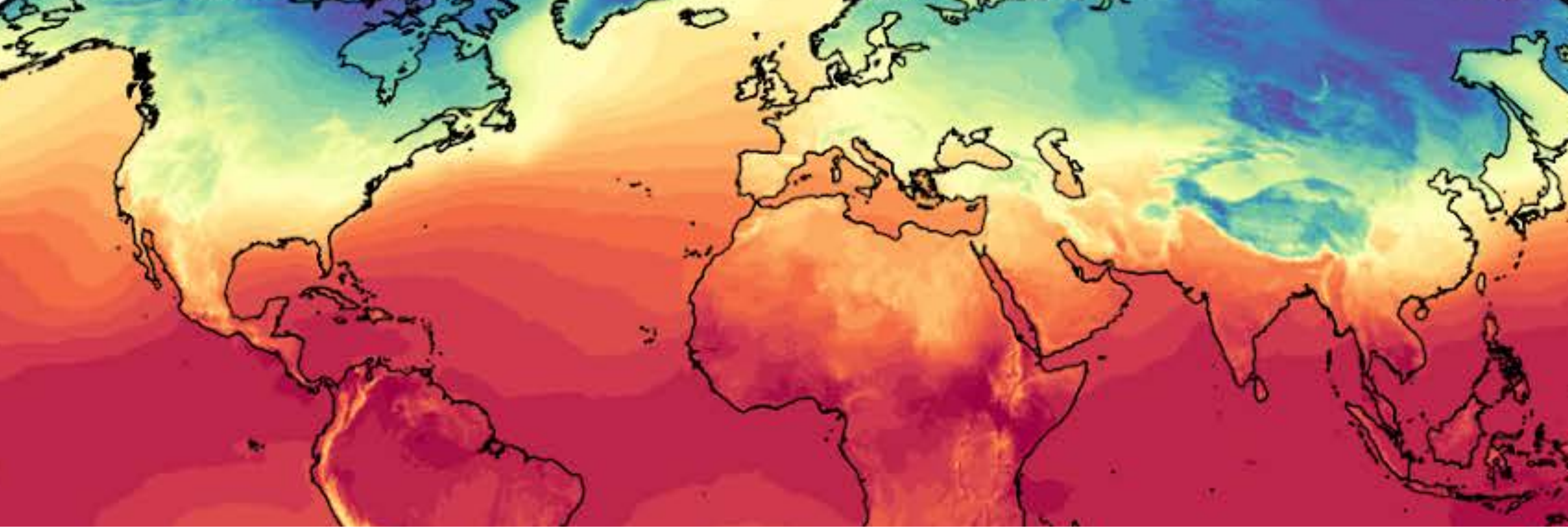


AtmoRep:
A machine learning based multi-purpose model for atmospheric dynamics

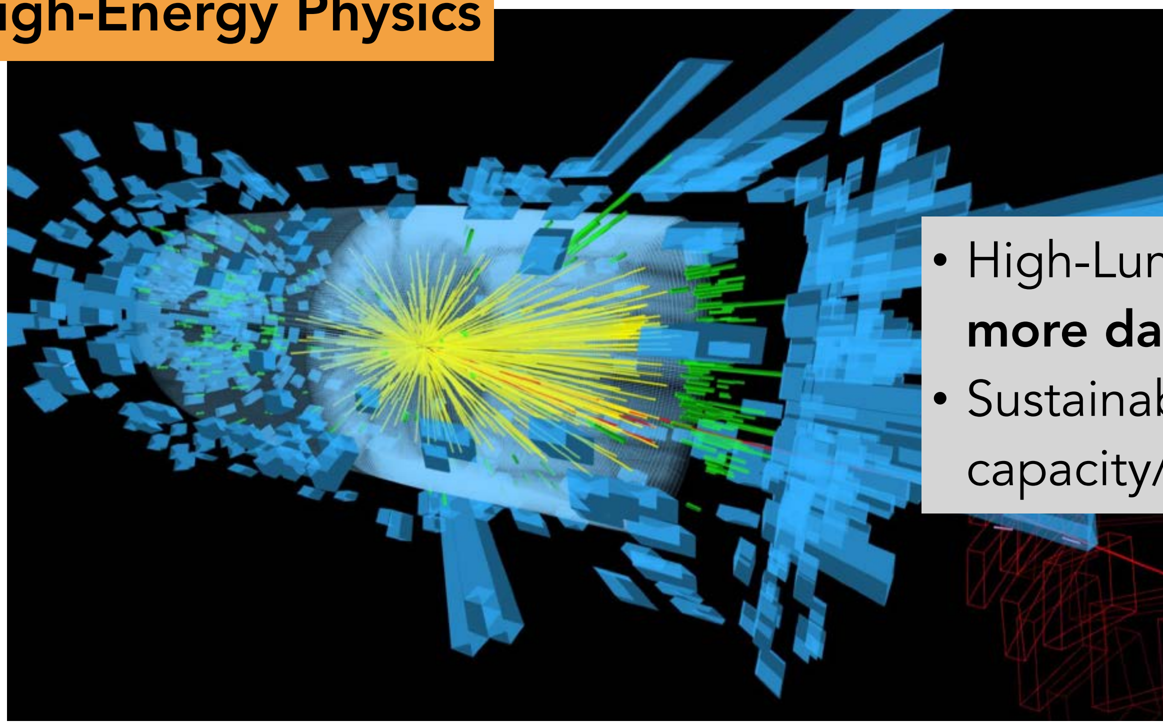
Christian Lessig, Ilaria Luise, Martin Schultz,
Michael Langguth, Alberto di Meglio et al.

ELLIIT Symposium, Linköping | 10th October 2024

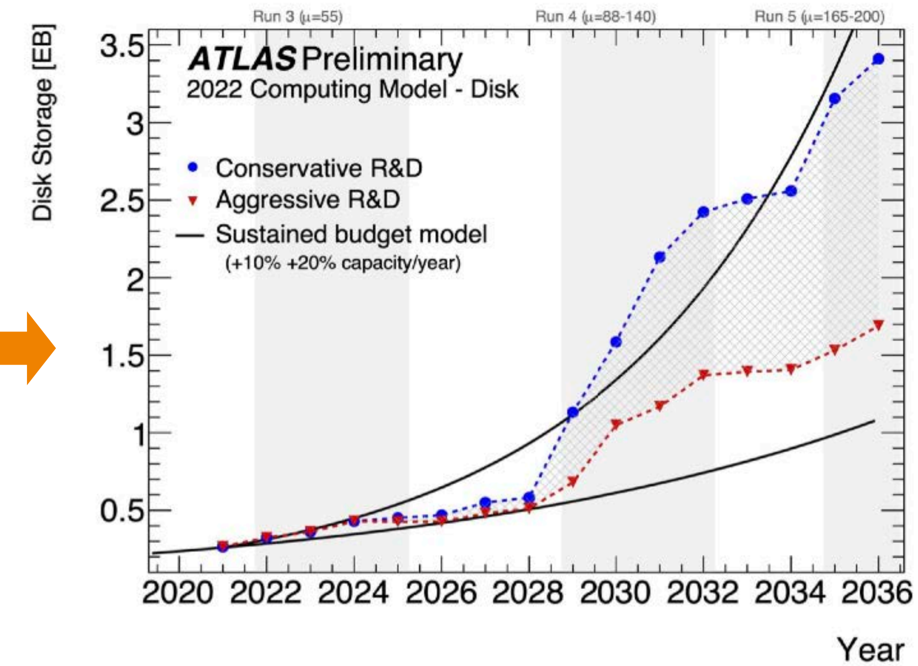
CERN | CIPEA CERN Innovation Programme on Environmental Applications
ECMWF
JÜLICH Forschungszentrum

The future of observational data

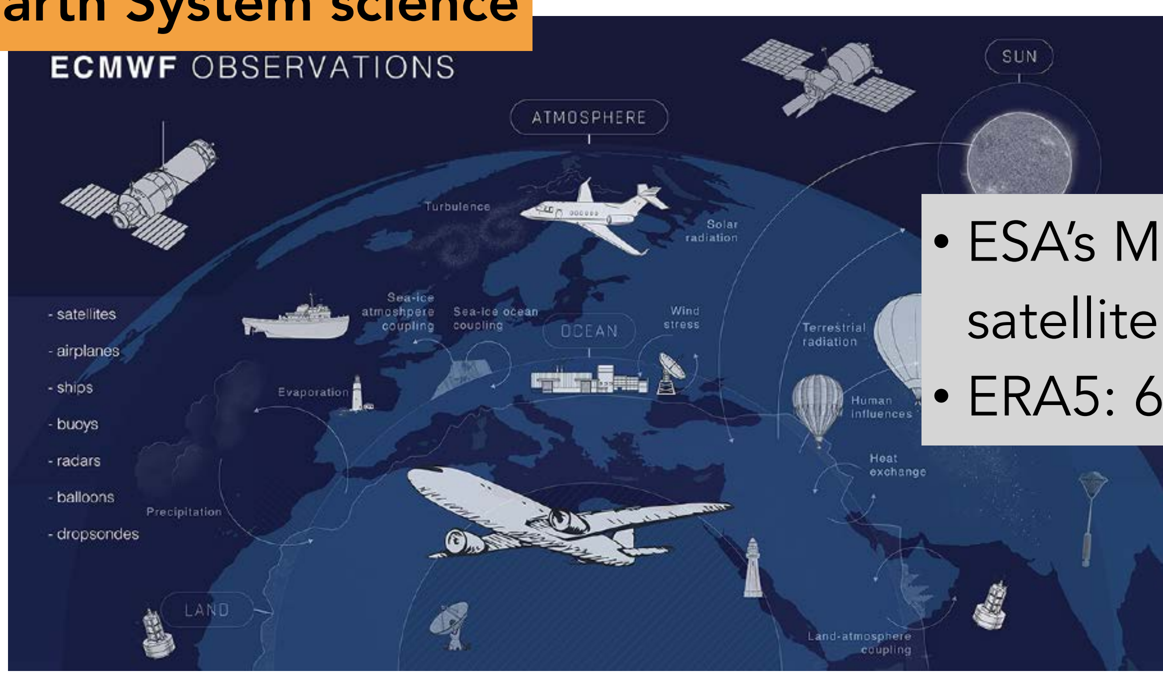
High-Energy Physics



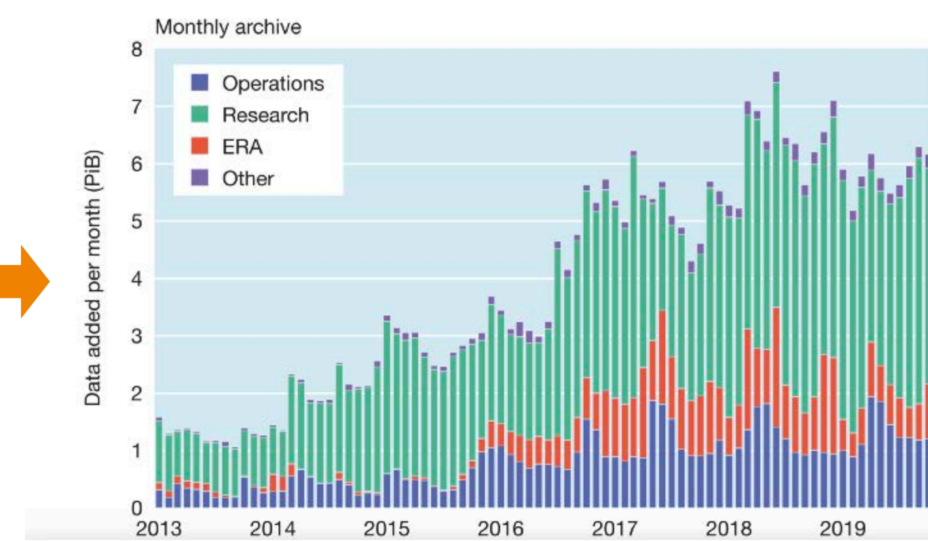
- High-Lumi LHC: **6 times more data by 2030**
- Sustainable model +20% capacity/year



Earth System science



- ESA's MetOp-SG satellite: **864 GB/day**
- ERA5: 6+ PB



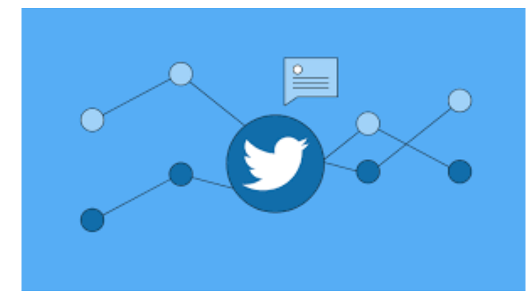
Need to find sustainable ways to store all these data

Ilaria Luise, CERN - ilaria.luise@cern.ch

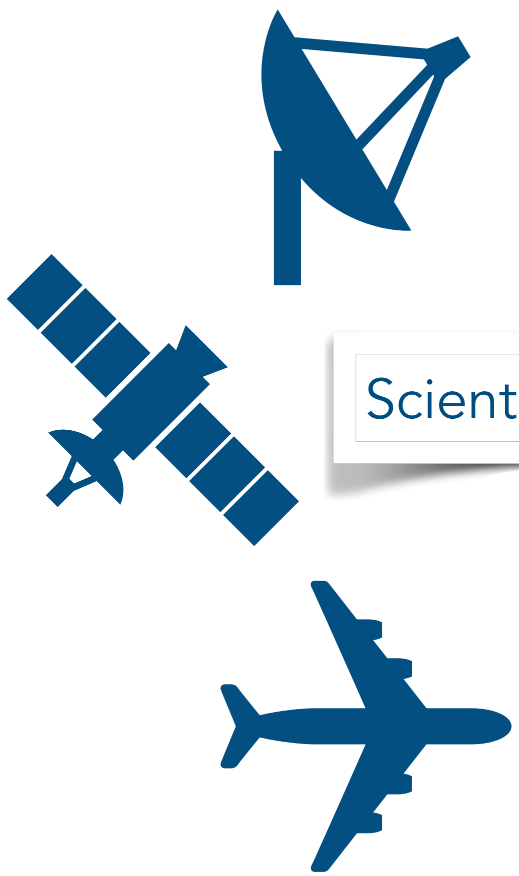
The future of observational data: multi-source

Data are getting **more and more multi-modal and the relationship between them is very complex to model** (and requires all kinds of approximations)

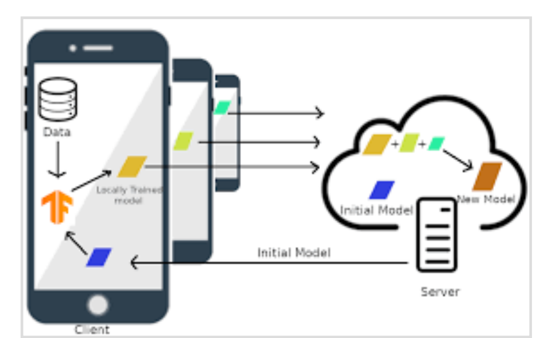
Social media



Economic growth
GDPs, birth rates…



Scientific Data → **Policy-oriented scientific models** ← New data types



Data from distributed devices

Conventional approaches for analysing and processing the data come to their limits

Ilaria Luise, CERN - ilaria.luise@cern.ch

The role of machine learning

Machine learning has been proven a very good tool to:

- Extract information from (very large) datasets
- Efficiently analyse very large amounts of data
- Easily handle data from different sources

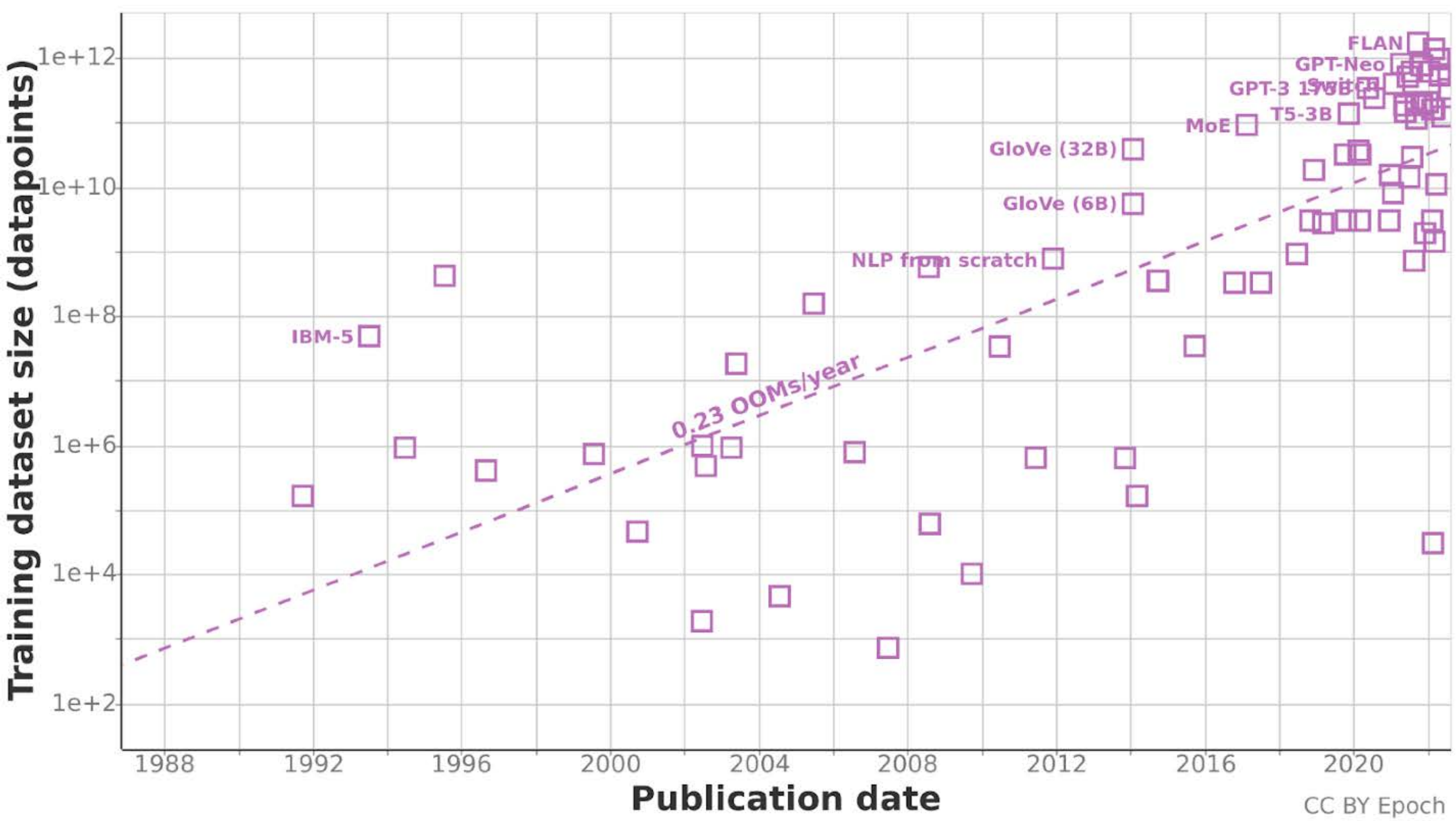
Ilaria Luise, CERN - ilaria.luise@cern.ch

The role of machine learning

Machine learning has been proven a very good tool to:

- Extract information from (very large) datasets
- Efficiently analyse very large amounts of data
- Easily handle data from different sources
- Scalability to HPC environments

Observation based datasets in physics are comparable or larger than these!

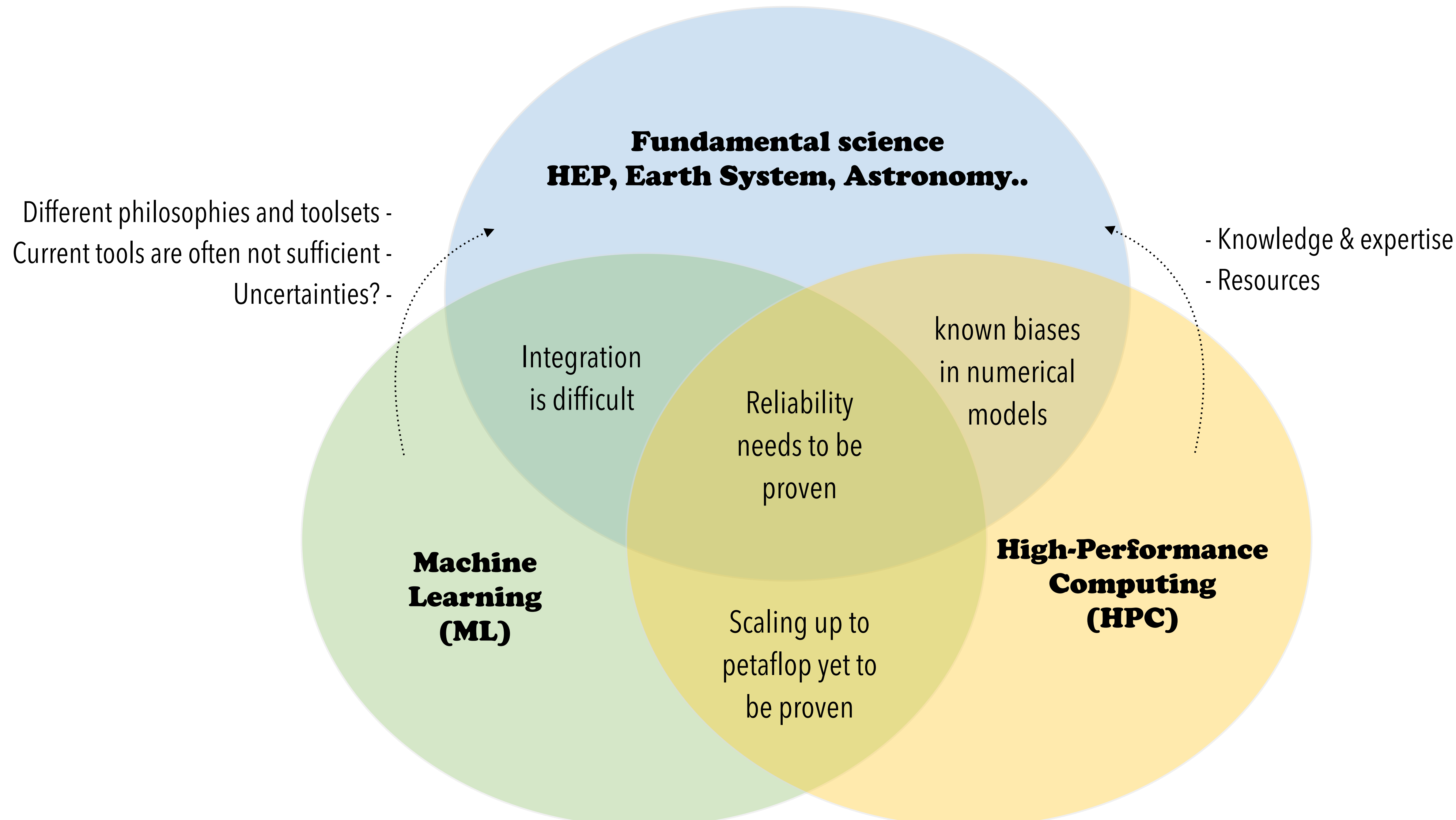


Can we use these tools for fully data-driven science?

Ilaria Luise, CERN - ilaria.luise@cern.ch

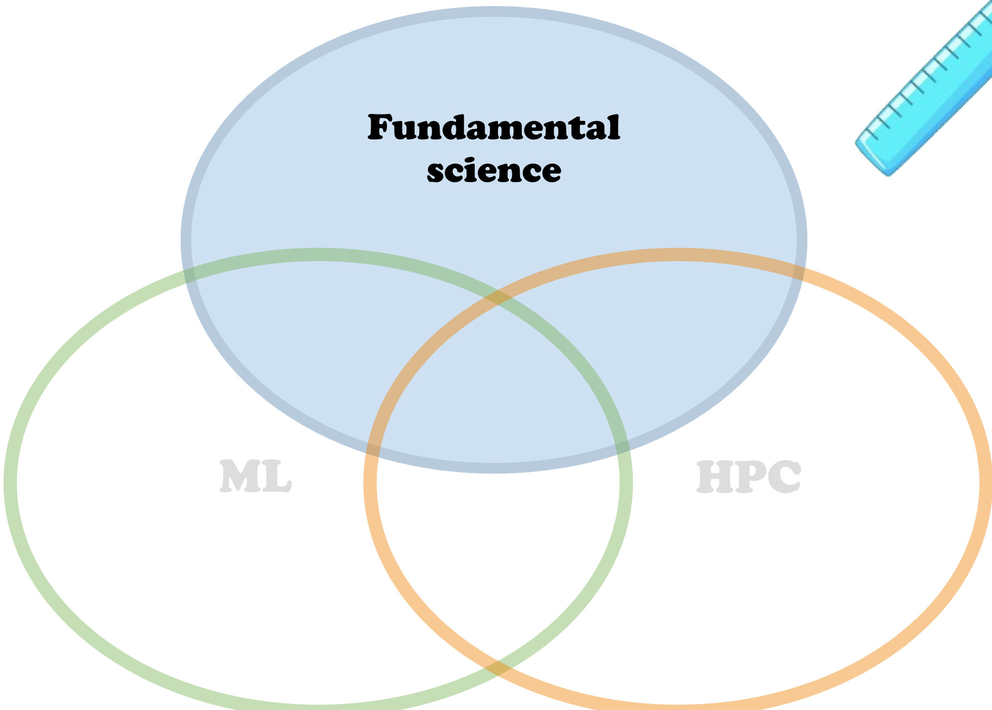
Where do we stand?

In short: we don't know how to extract scientific knowledge from large scale ML models



Slide based on material from Peter Dueben, ECMWF - Magdeburg September 2022

Scientific opportunities



Multi-scale dependencies:

- **Model complex higher-order, statistical relationships between observations, fields, …**
- improve current simulations

Compact representations:

- **Condense dataset information in a compact representation**
- eg. condense the info in a few GB rather than TB

Multi-source models:

- **Enable multimodal and multi-source learning**
- eg. build models based on scientific data, GDP, birth rate etc..

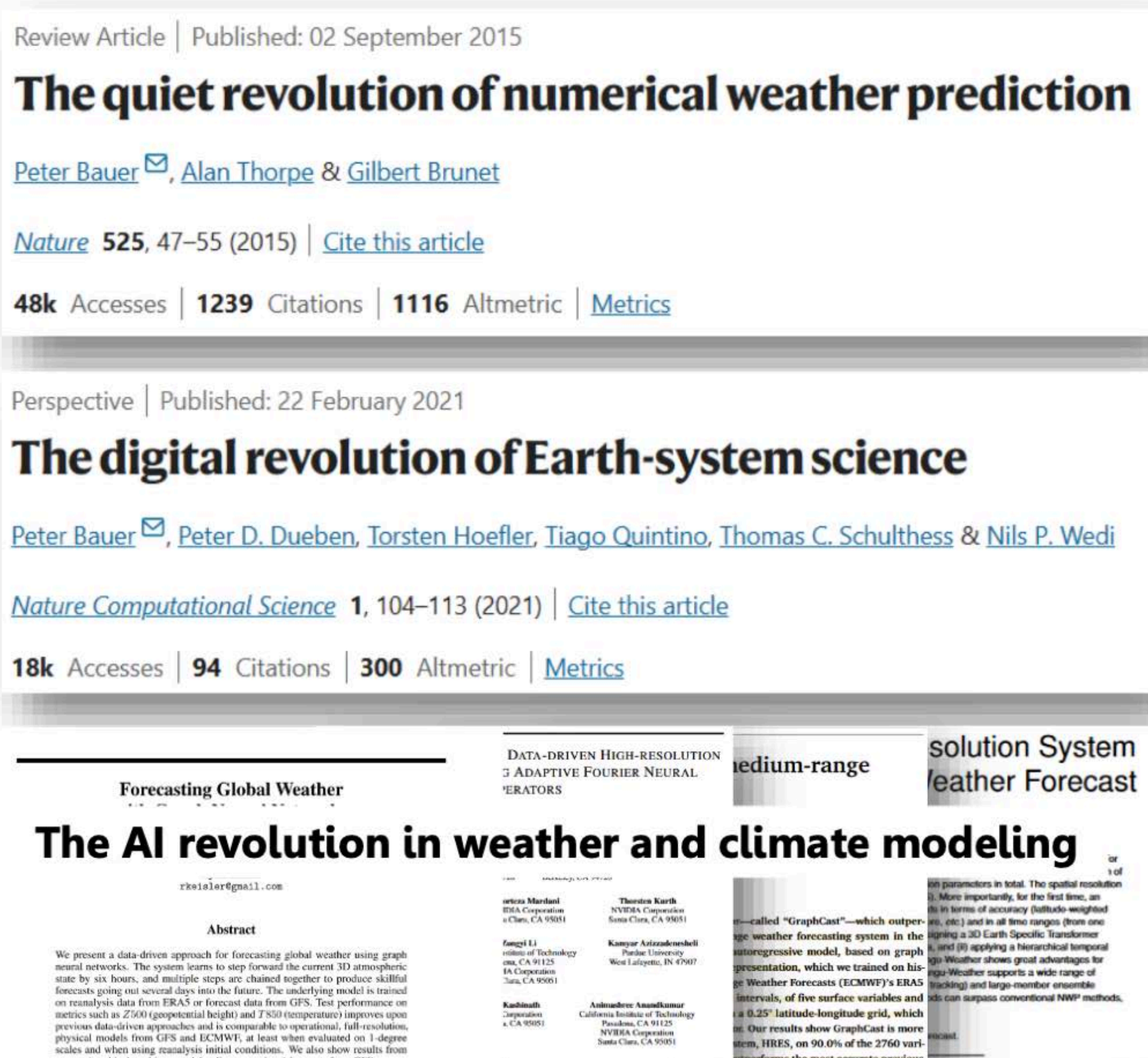
New discoveries:

- **Explore the potential of unsupervised learning to extract new information directly from data**
- Learn unknown correlation patterns

Ilaria Luise, CERN - ilaria.luise@cern.ch

The first breakthrough: weather & climate

First time that an AI-model trained on TBs of pre-processed observations outperforms the numerical models for a 10 day forecasts


Review Article | Published: 02 September 2015

The quiet revolution of numerical weather prediction

Peter Bauer, Alan Thorpe & Gilbert Brunet

Nature 525, 47–55 (2015) | Cite this article

48k Accesses | 1239 Citations | 1116 Altmetric | Metrics

Perspective | Published: 22 February 2021

The digital revolution of Earth-system science

Peter Bauer, Peter D. Dueben, Torsten Hoefler, Tiago Quintino, Thomas C. Schulthess & Nils P. Wedi

Nature Computational Science 1, 104–113 (2021) | Cite this article

18k Accesses | 94 Citations | 300 Altmetric | Metrics

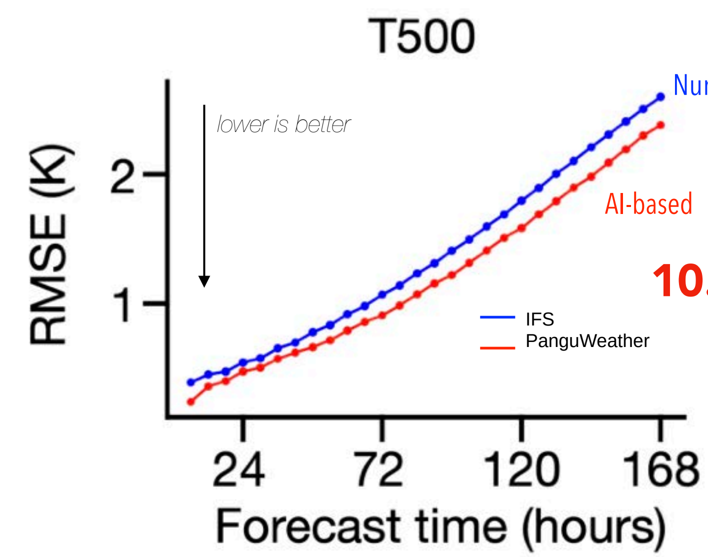
Forecasting Global Weather

The AI revolution in weather and climate modeling

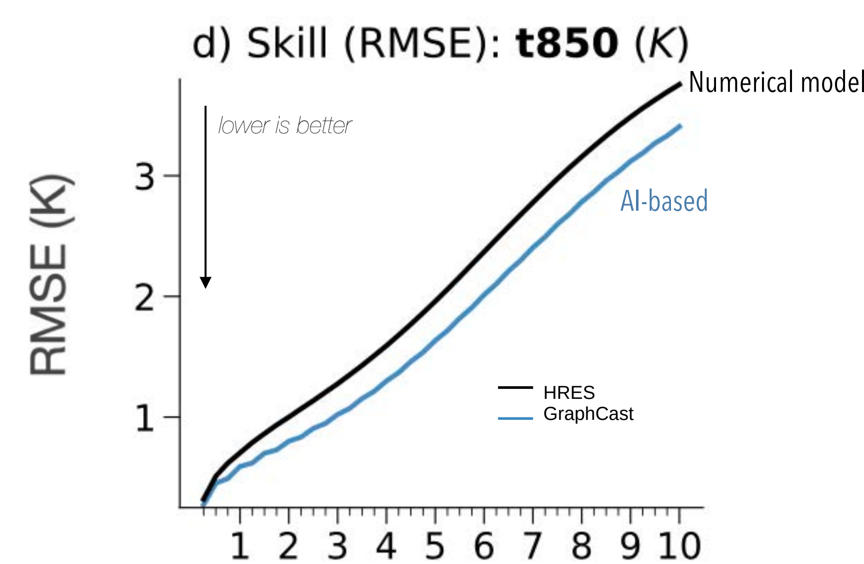
1960-2010

2005-2025

2022-



10.000 faster



From Martin Schultz - ESA-ECMWF workshop 2024

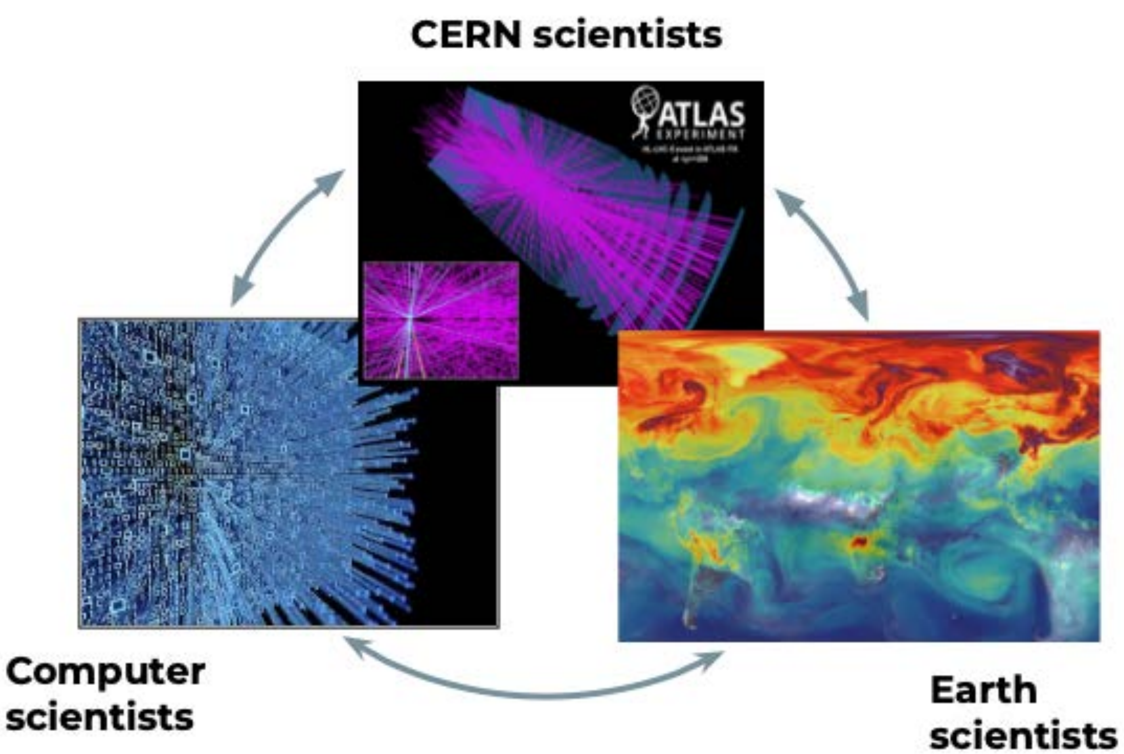
Can we go beyond?

AtmoRep: a probabilistic multi-purpose model for atmospheric dynamics

AtmoRep: Introduction

Atmosphere:

- Set of **complex non-linearly coupled phenomena** involving a **wide range of scales**
- Very large amounts of **observational data available in a format suitable for large scale ML**



<u>Common challenges:</u>

Model complex, nonlinear phenomena and improve current simulations

Earth science: eg. better understand convection phenomena
CERN: eg. particle-jet showers reconstruction

Explore potential of unsupervised learning for scientific applications

Earth science: eg. early detection of extreme events
CERN: eg. anomaly detection

Condense dataset information in a compact representation

eg. condense the info in a few GB rather than TB

Common Goal:

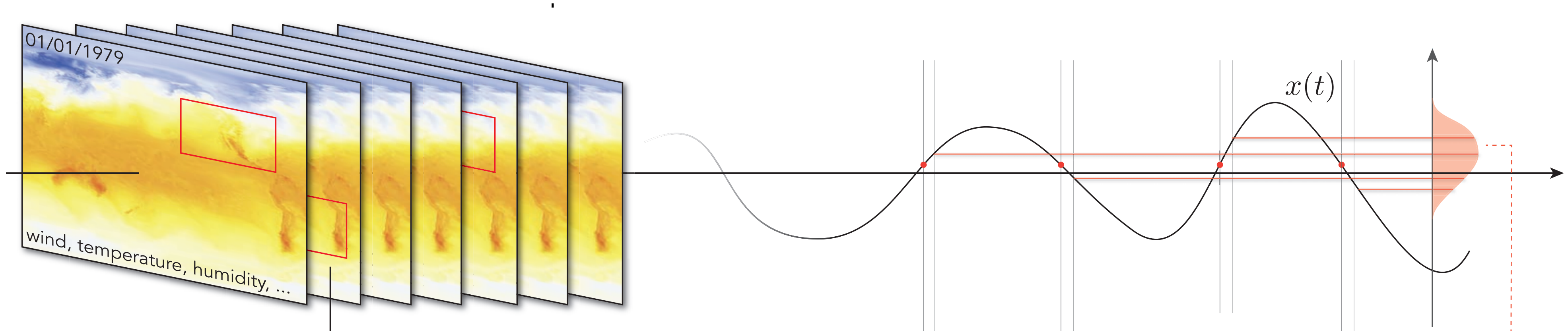
Use unsupervised learning to build a <u>task-independent</u> data-driven model to <u>encapsulate complex physics phenomena</u>

Ilaria Luise, CERN - ilaria.luise@cern.ch

What is a task-independent model for us?

Encapsulate the spatio-temporal evolution of a dynamical system

Probability of getting the state y given the initial state x and the auxiliary info α $\rightarrow$ $p(y\,|\,x, \alpha)$ $\leftarrow$ *Auxiliary info: position, absolute time etc..*



Training

$$p(y\,|\,x, \alpha) \approx p_\theta(y\,|\,x, \alpha)$$

foundation model:
neural network that models data distribution for a specific domain

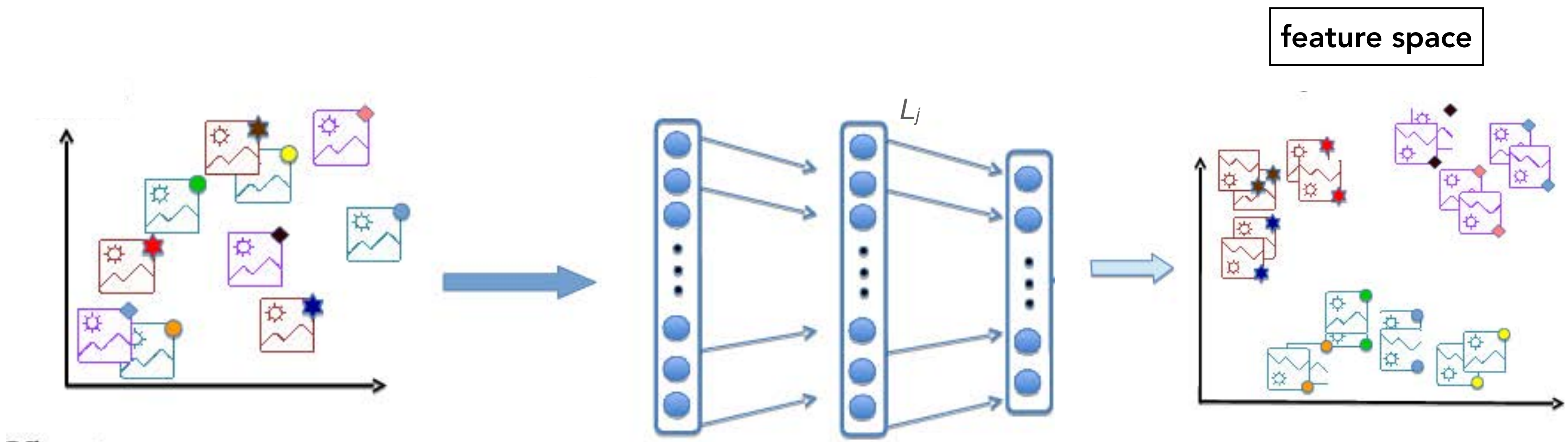
The distribution can be approximated by a large neural network

Ilaria Luise, CERN - ilaria.luise@cern.ch

Key Ingredient: Representation learning

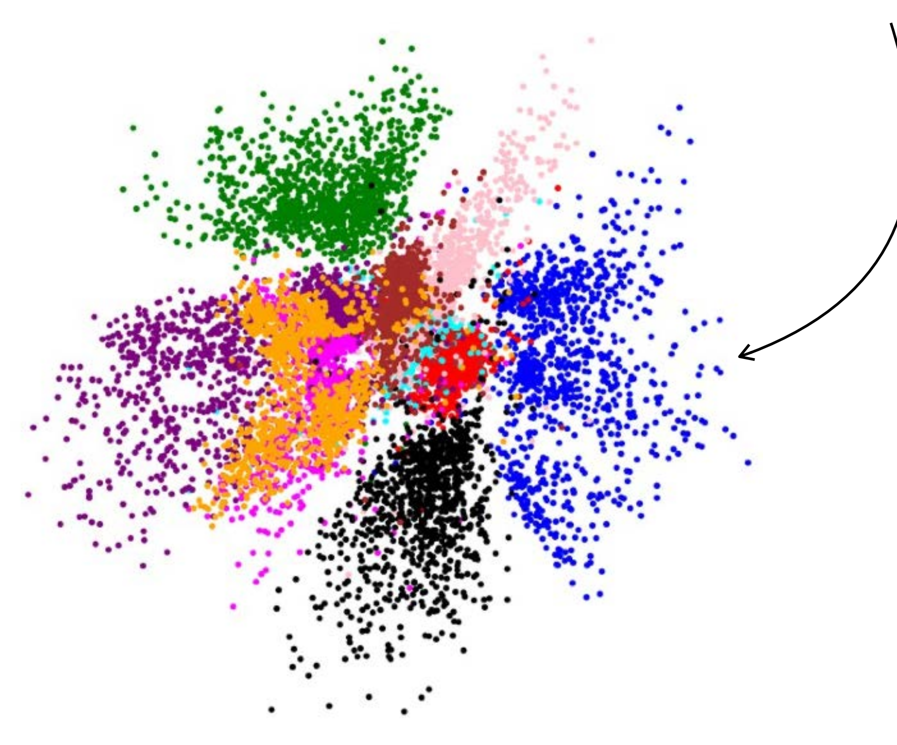
Representation learning:

- Learn a **task-independent representation** of the data in the **feature space** of the neural network



$$L_j : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

Effective "encoding" of the data useful for many applications

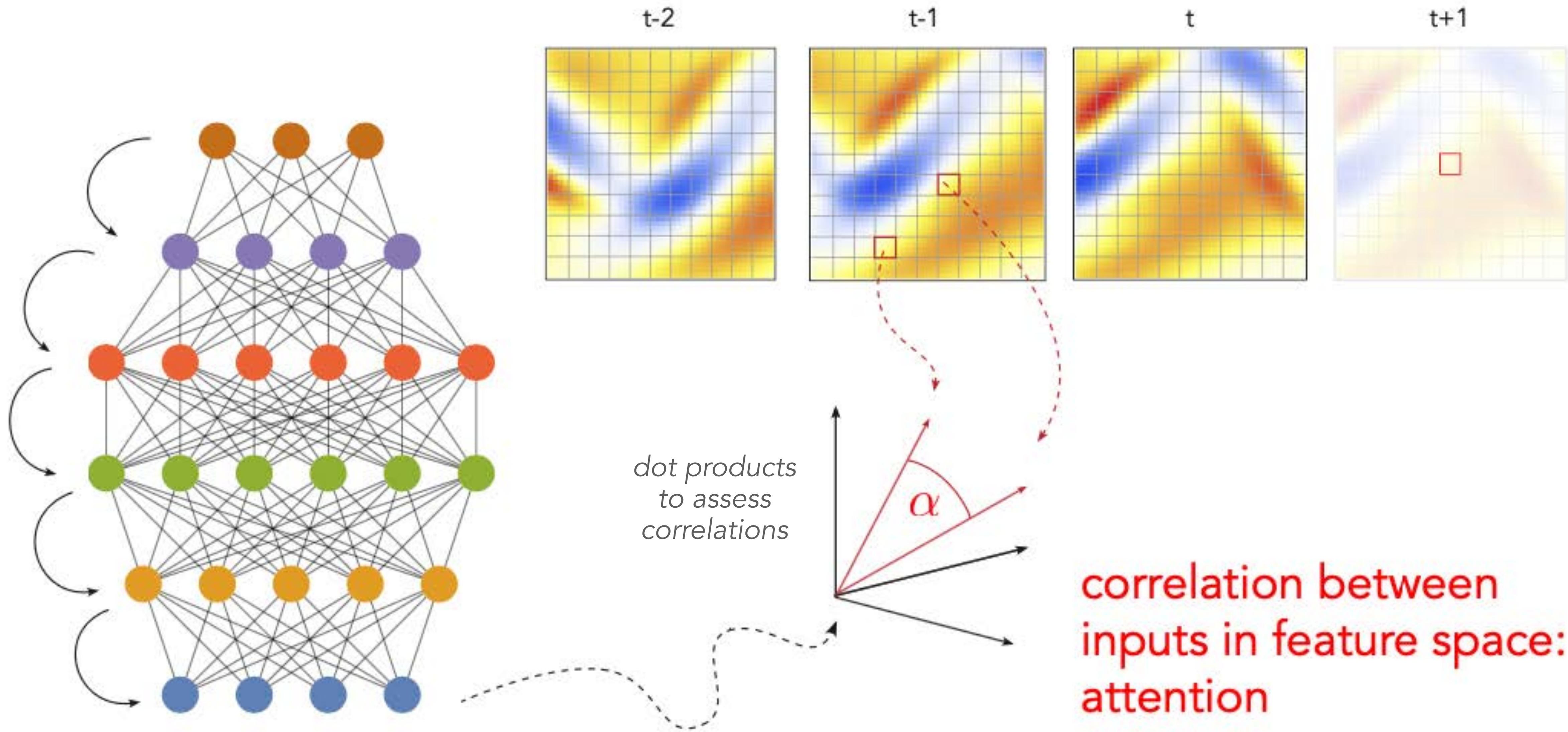


Data are clustered according to common properties in the "feature space"

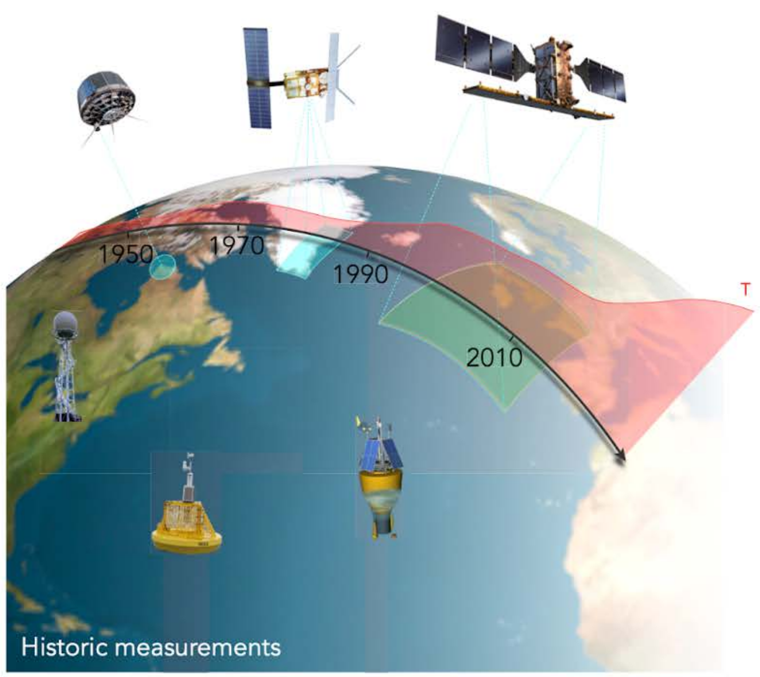
https://arxiv.org/pdf/1206.5538v3.pdf

Attention in transformers for flow models

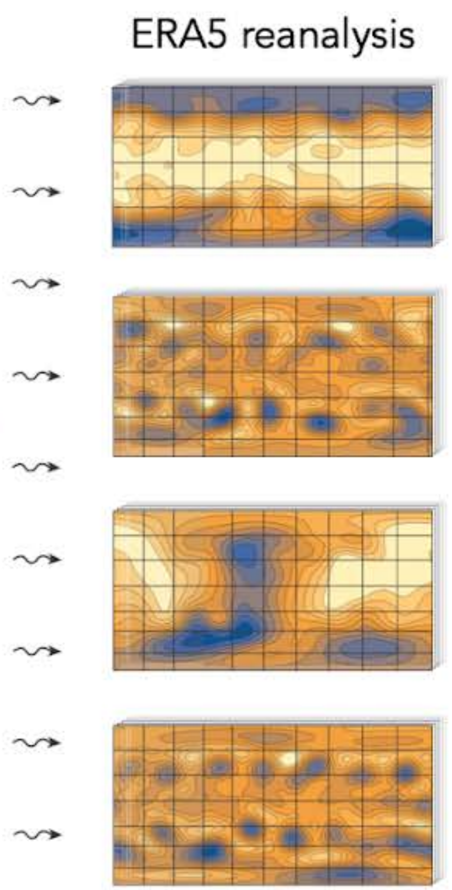
- Vision transformers: Transformers applied to imaging
- Split the image into “tokens” (group of points/cells) of finite size, eg. 8x8
- Project each token in the feature space through an embedding network
- Compute the “attention” across tokens



The project in a nutshell

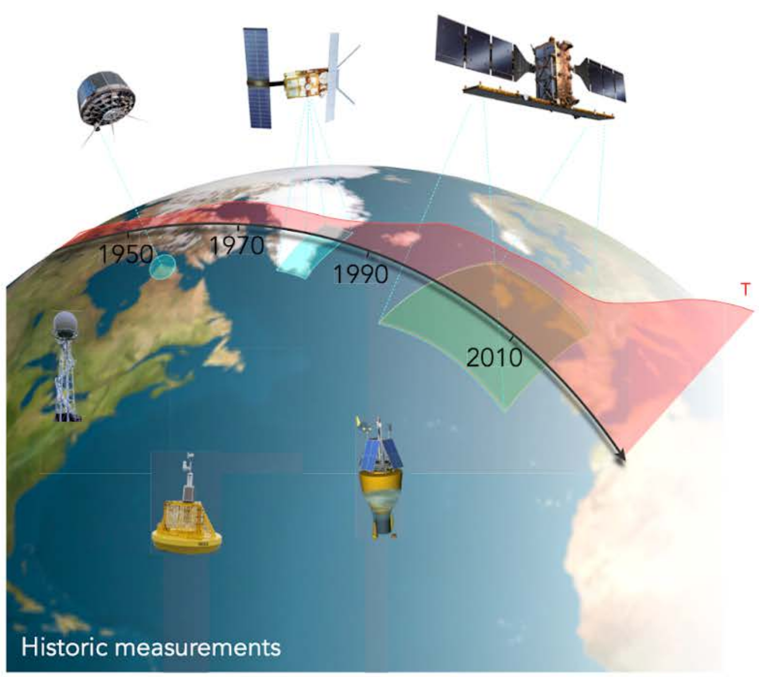


Observations



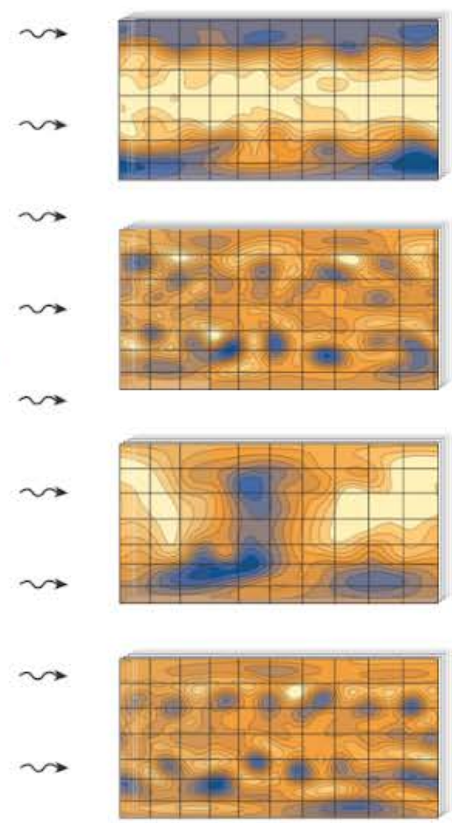
50 TB

The project in a nutshell

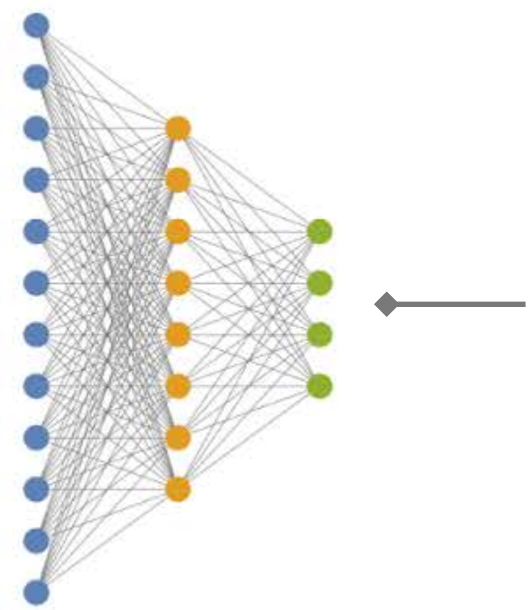


Observations

ERA5 reanalysis



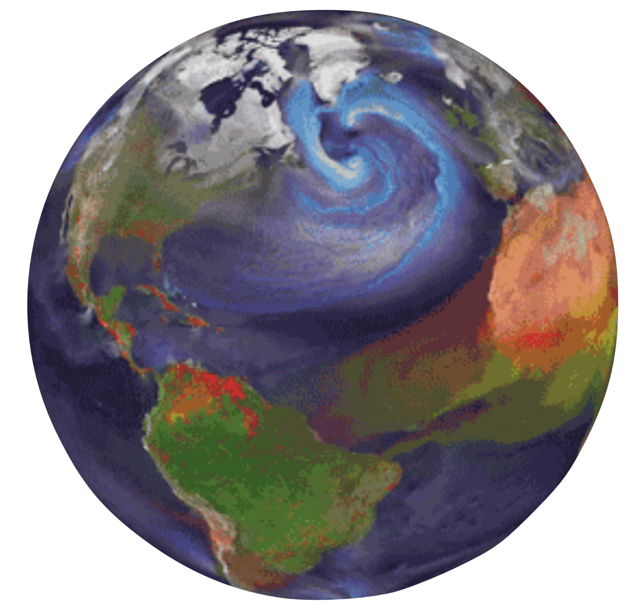
50 TB



large scale
machine learning

Transformers architecture

Spatio-temporal representation of atmospheric dynamics



Model
given by the trained neural network

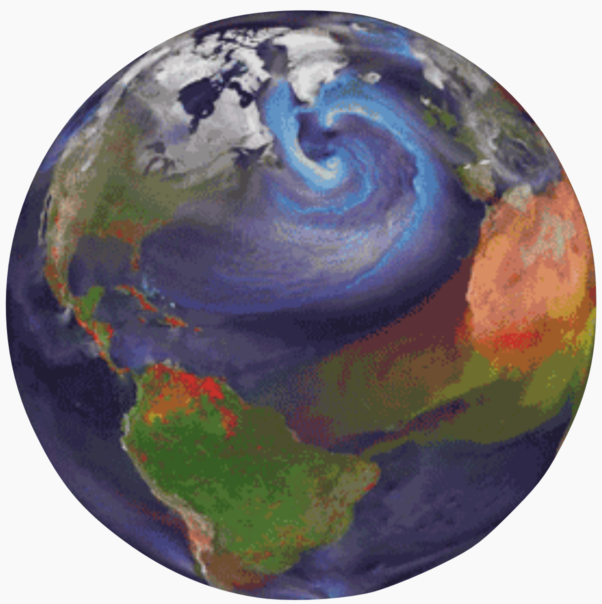
R&D at Juelich SSC:
$4x10^6$ GPU hours granted in 2023

JÜLICH Forschungszentrum

Ilaria Luise, CERN - ilaria.luise@cern.ch

Applications: one model for multiple purposes

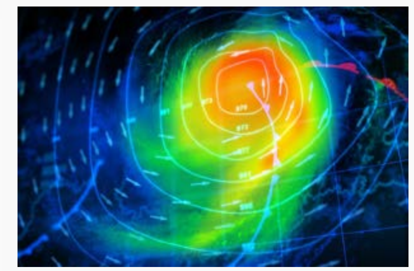
Spatio-temporal representation of atmospheric dynamics



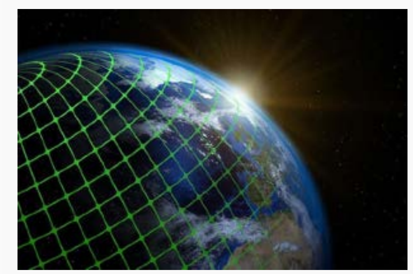
Model
given by the trained neural network

Task-dependent adaptable smaller networks

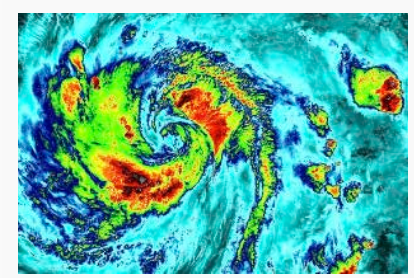
Adaptation



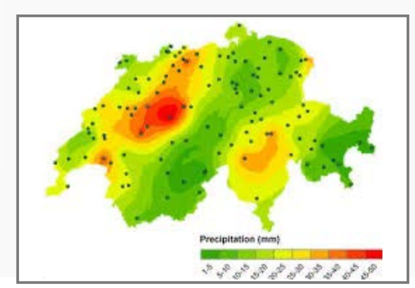
Weather predictions



Downscaling



Bias corrections



Spatio-temporal Interpolations (WIP)

Physics-related applications = uncertainties
Need for a ***stochastic approach***

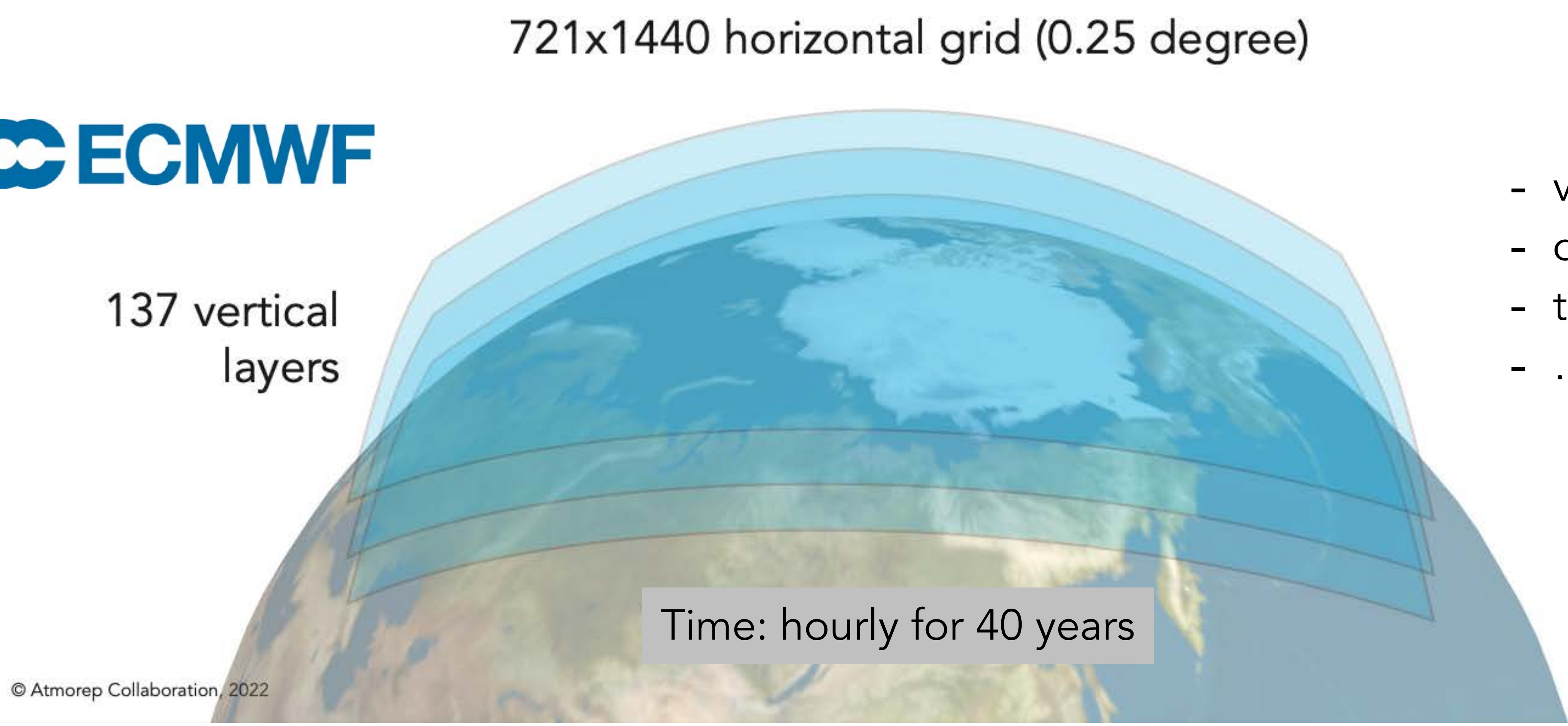
Ilaria Luise, CERN - ilaria.luise@cern.ch

Key Ingredient: The dataset

Subset of ERA5 reanalysis used at the moment for training:

- Physical fields: vorticity, divergence (or wind velocity), vertical velocity, temperature, specific humidity, total precipitation
- Space: 721 x 1440 x 5 vertical layers
- Time: **randomly sample** over 24 time steps per day for 365 days for 40 years

ECMWF



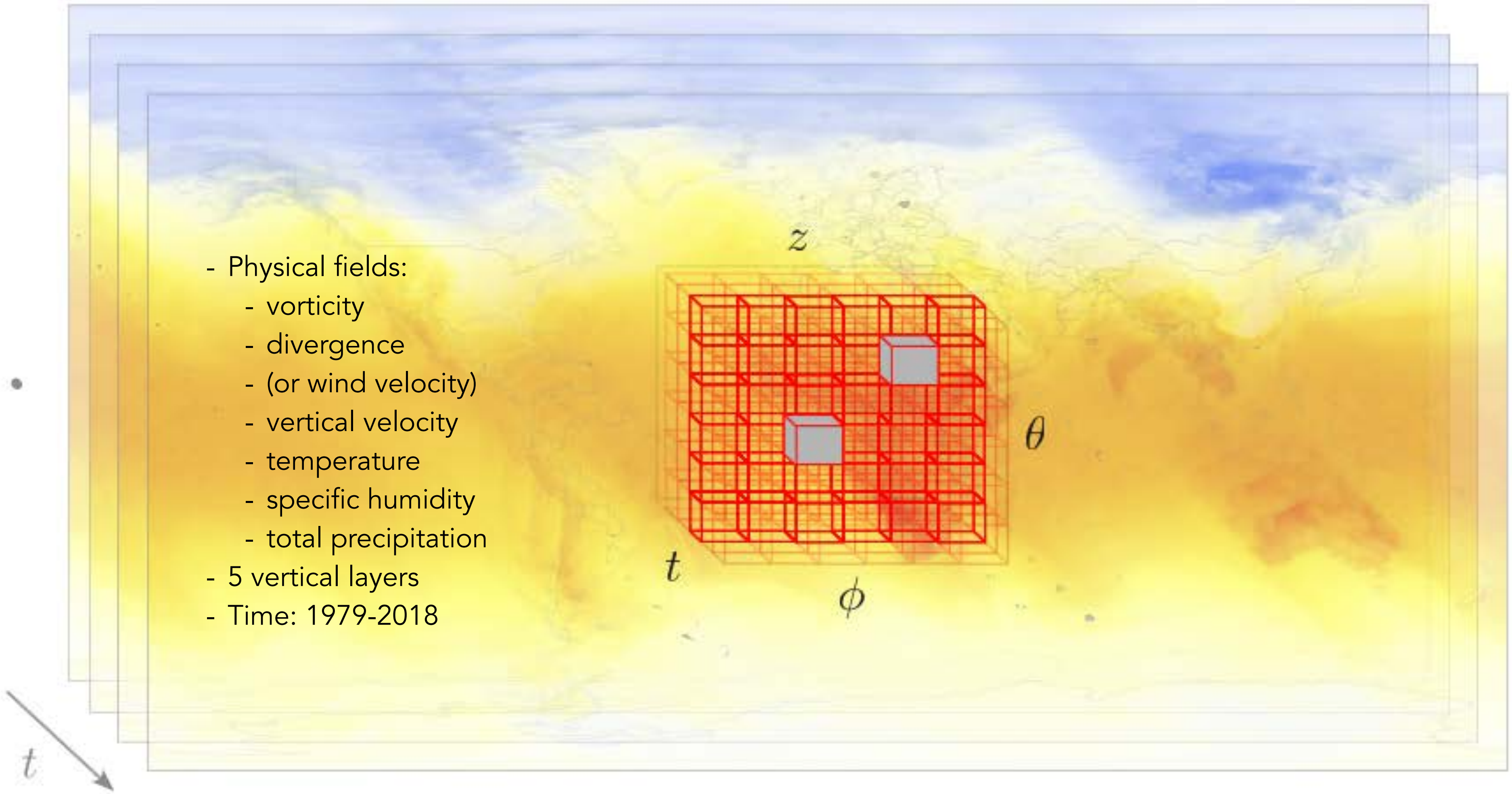
- vorticity
- divergence
- temperature
- …

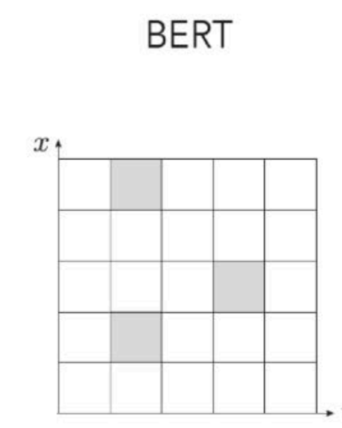
Ilaria Luise, CERN - ilaria.luise@cern.ch

Key Ingredient: The training protocol

Use an extension of BERT masked language modelling from self-supervised trainings in NLP

Random sampling of neighbourhoods for training





Split cube in small space-time regions (3D cubes) → tokens

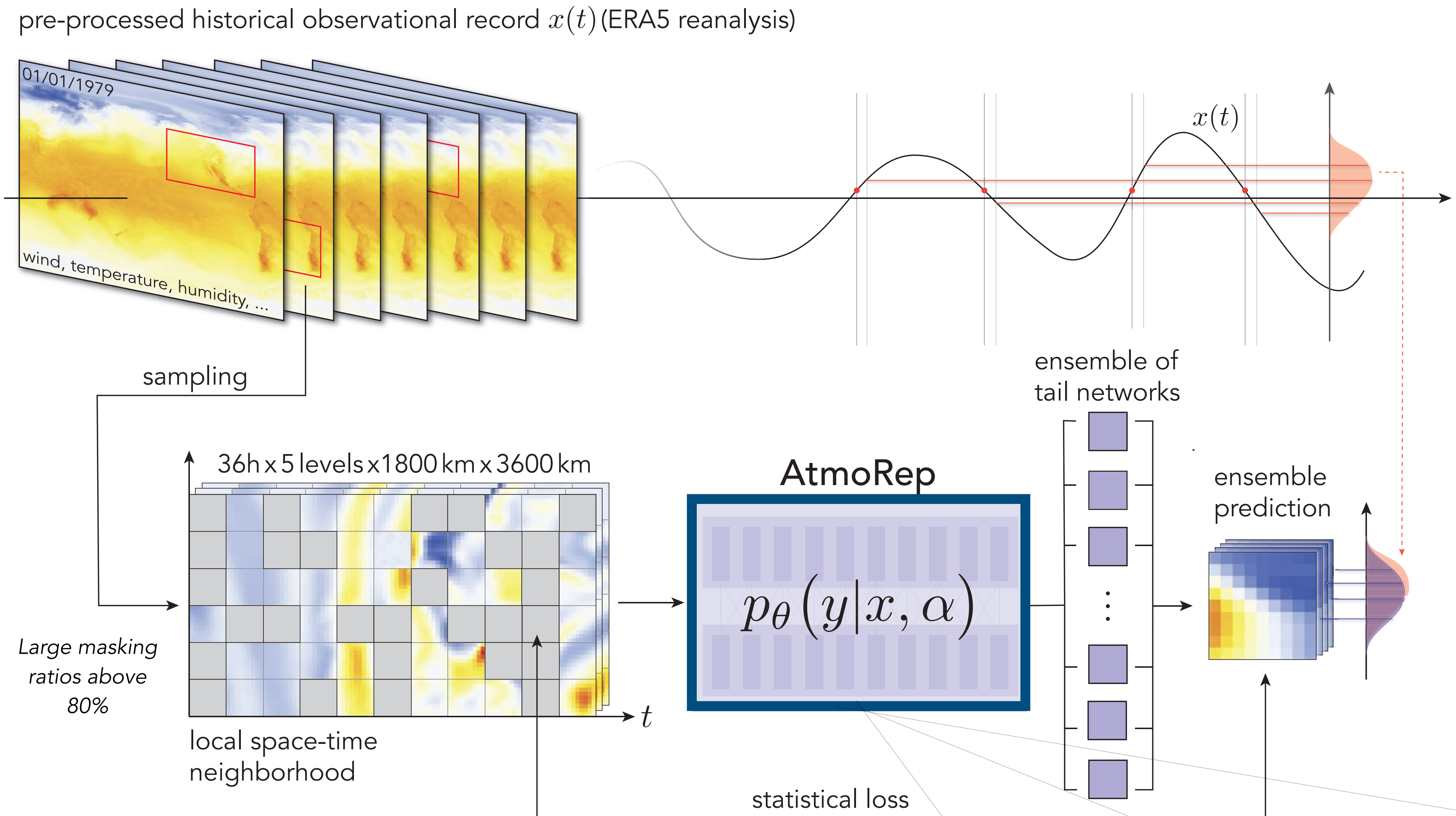
Mask random tokens within the hyper-cube and predict them

Large masking ratios above 80%

Default: 12 x 6 x 12 tokens with 3 x 9 x 9 grid points

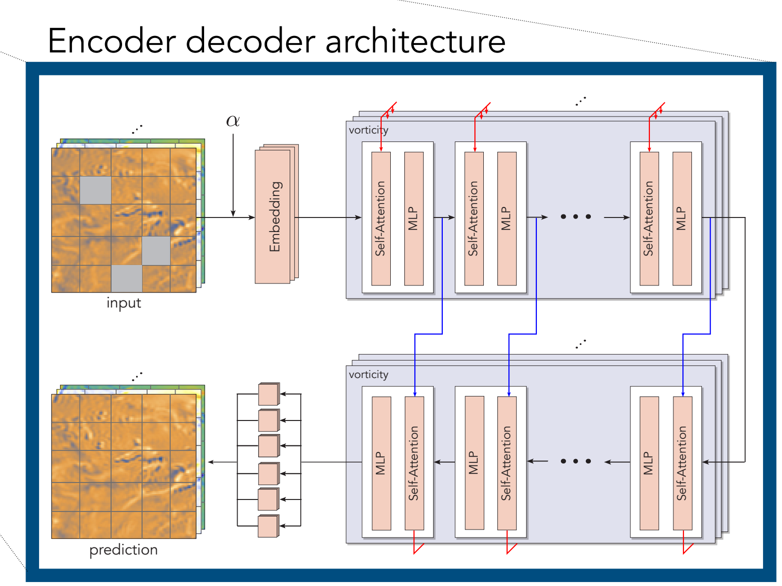
Ilaria Luise, CERN - ilaria.luise@cern.ch

The AtmoRep workflow



New stochastic approach

ensemble predictions with 16 members



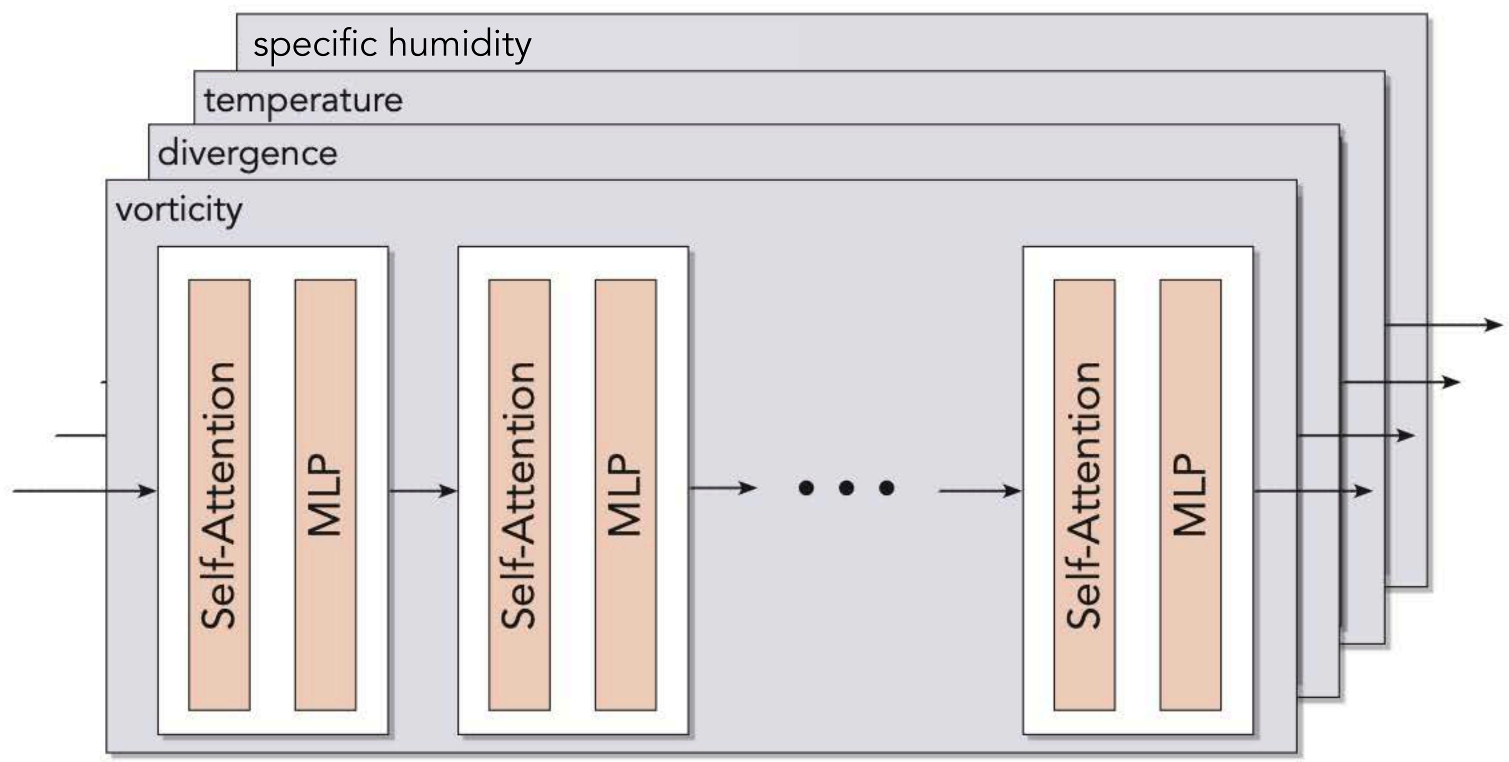
Approximate the 4-Dim PDF of the process using a Transformers-based network with 3.5 billion parameters

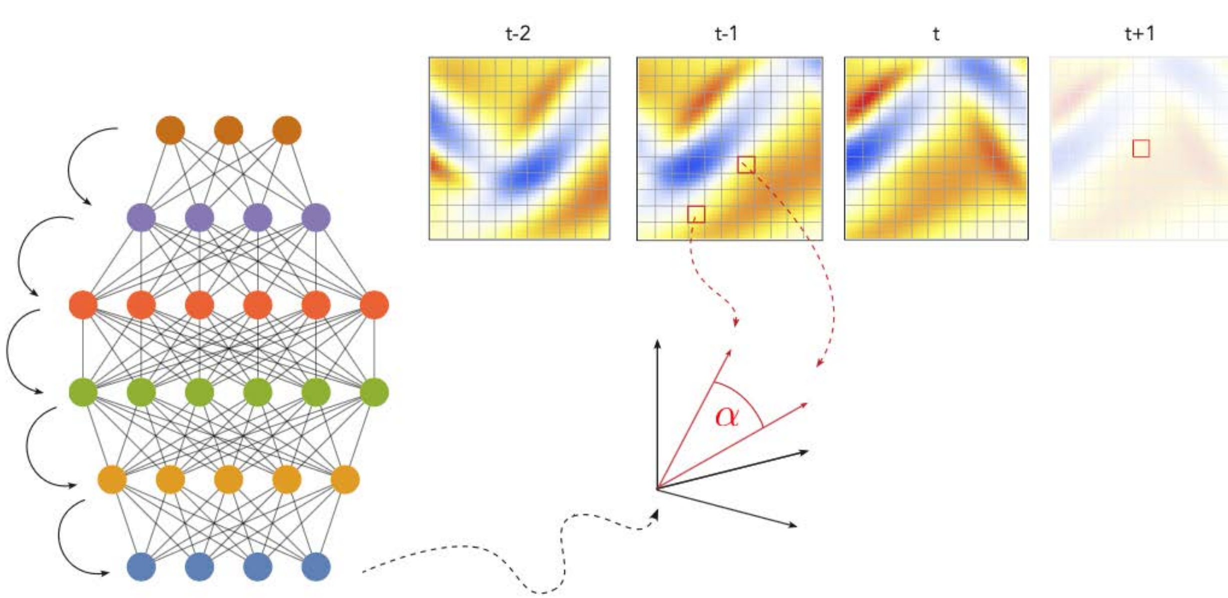
Network Architecture - Part 1

Single-field pre-training to encapsulate the spatio-temporal evolution of each field independently

Self attention

$$\sigma(Q K^T) V$$





Why?

- *Flexibility (add new fields w/o training from scratch)*
- *Computationally cheaper*
- *More memory efficient*

Network Architecture - Part 2

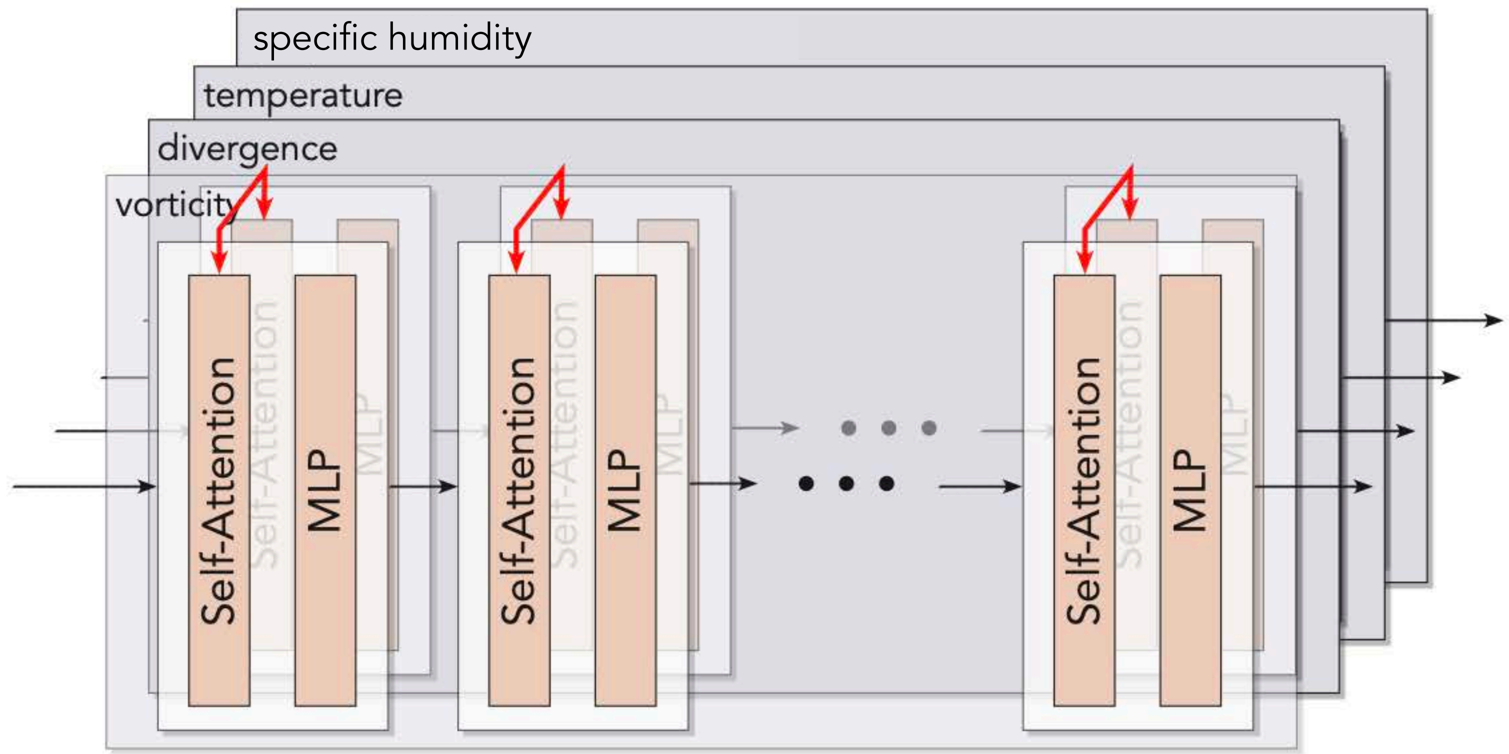
Couple the fields to encapsulate the interdependencies across variables and the correlation patterns

Self attention

$$\sigma\left(Q\,K^T\right)V$$

Cross attention

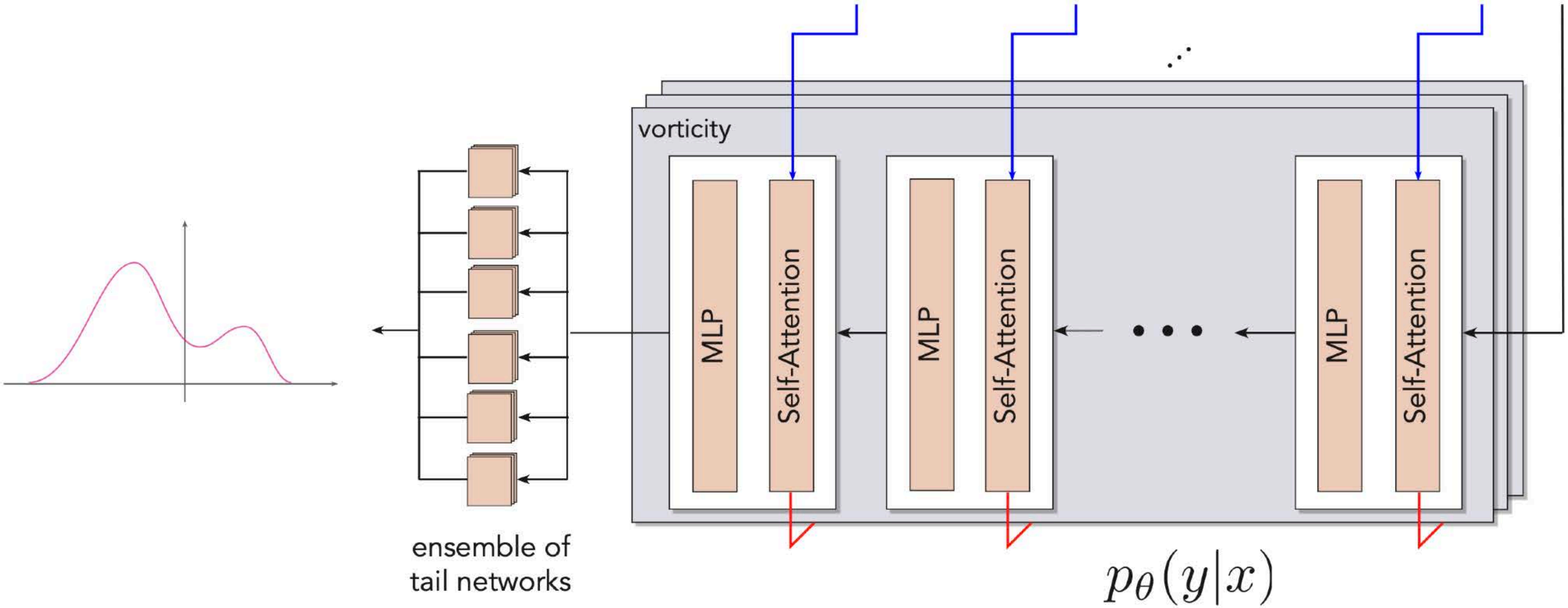
$$\sigma\left(Q_\varsigma\,K_\mu^T\right)V_\mu$$



Ilaria Luise, CERN - ilaria.luise@cern.ch

Key ingredient: statistical loss

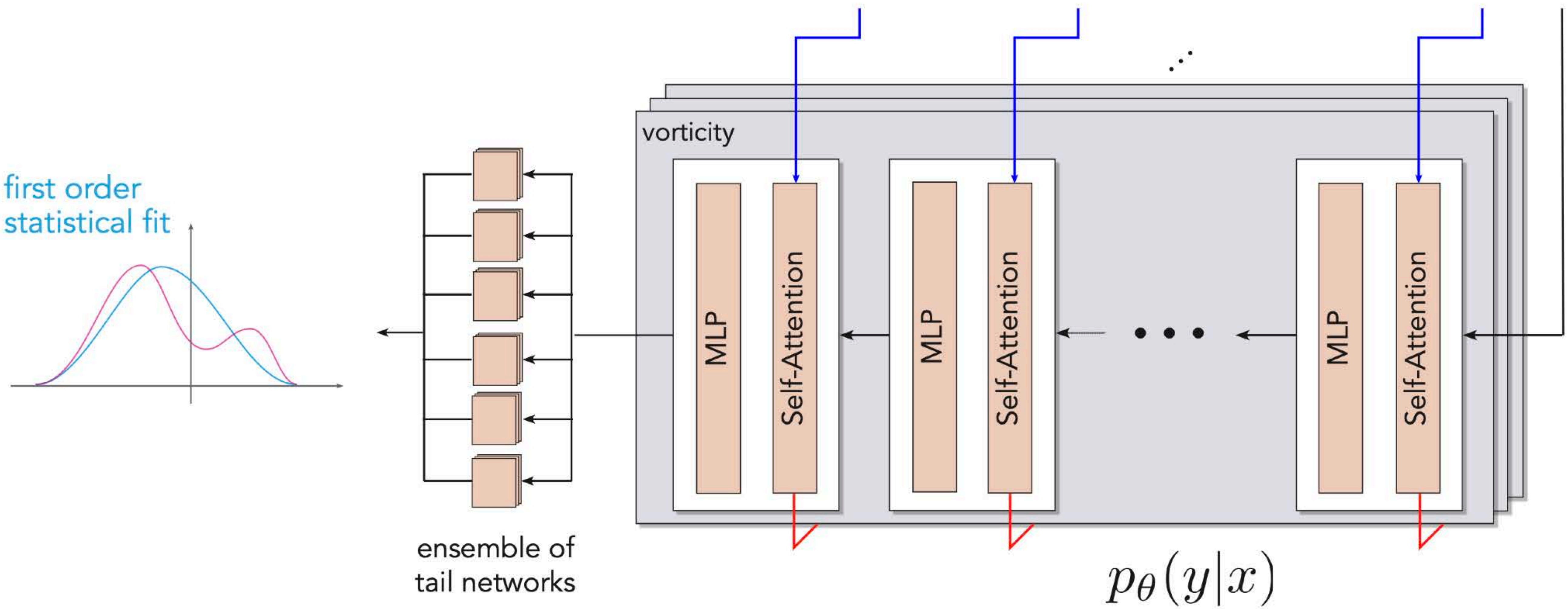
Loss: RMSE Loss + custom statistical loss



Ilaria Luise, CERN - ilaria.luise@cern.ch

Key ingredient: statistical loss

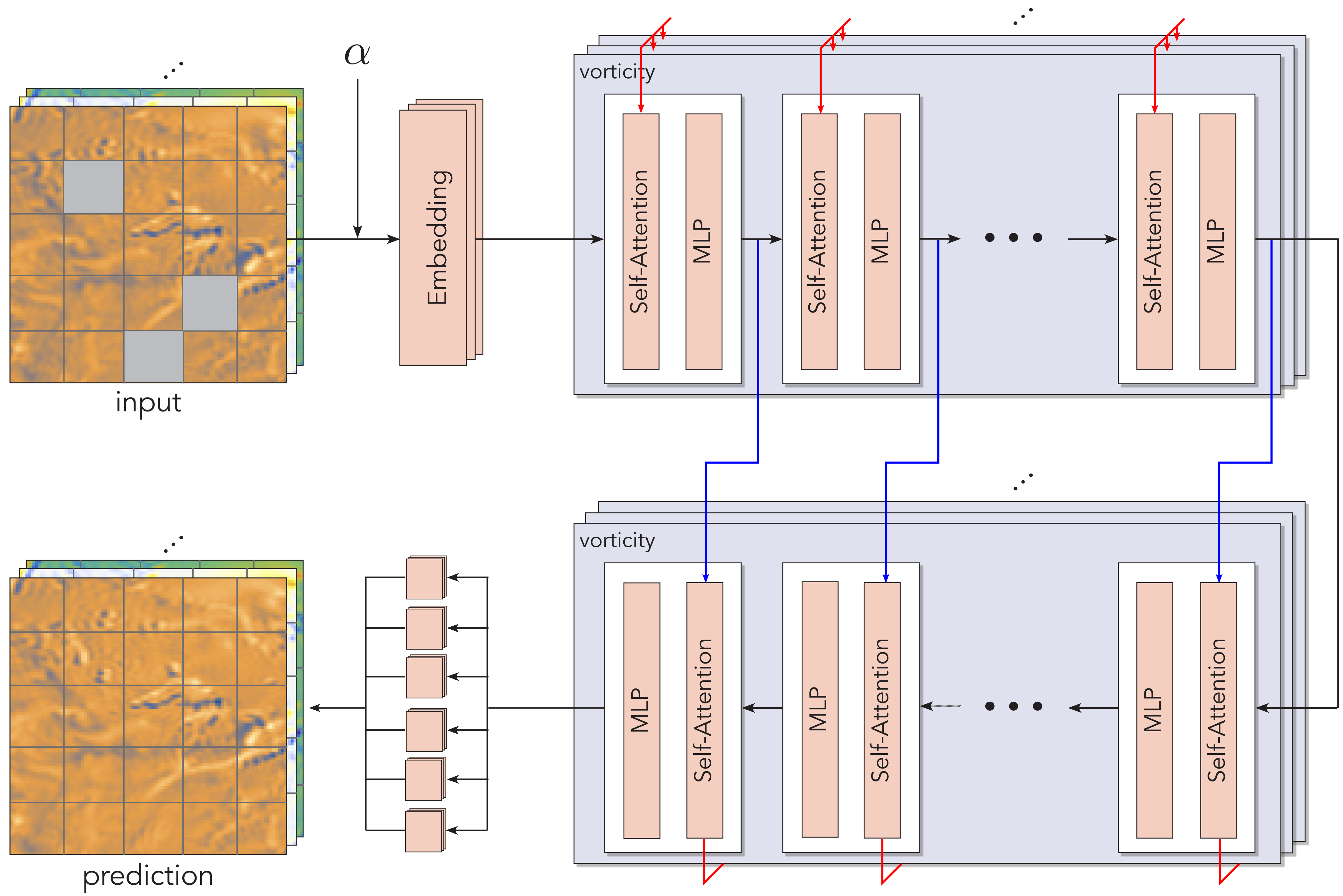
Loss: RMSE Loss + custom statistical loss



Ilaria Luise, CERN - ilaria.luise@cern.ch

Overall Network Architecture

- Num Heads 16
- Num Layers 10
- Num MLP Layers 2
- Trained on 32 nodes



Ilaria Luise, CERN - ilaria.luise@cern.ch

Task-specific fine-tuning

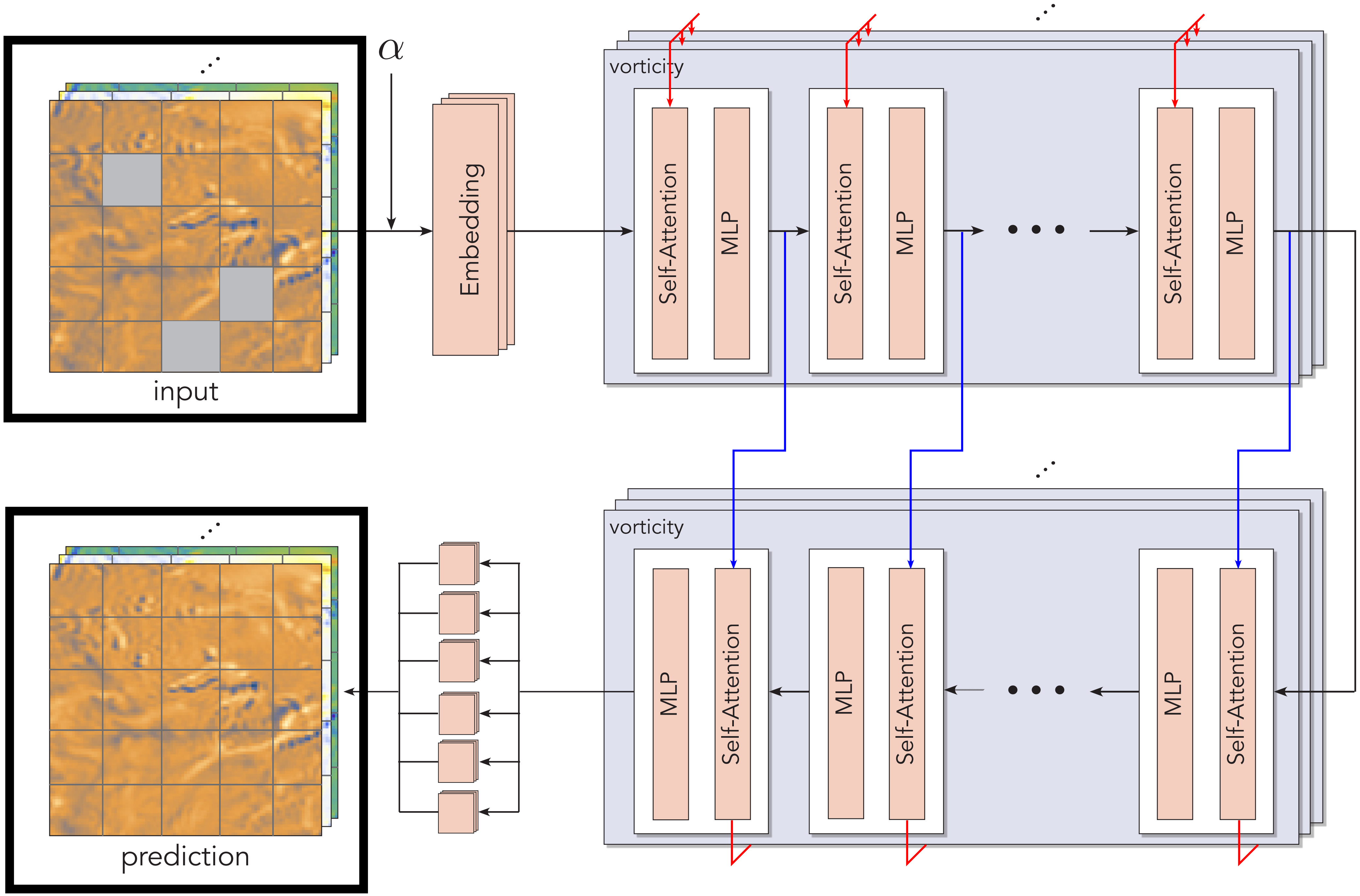
Goal: improve model performance for a specific task
e.g. forecasting, downscaling…

Examples:

e.g. fix masking scheme

OR

Change target dataset



Ilaria Luise, CERN - ilaria.luise@cern.ch

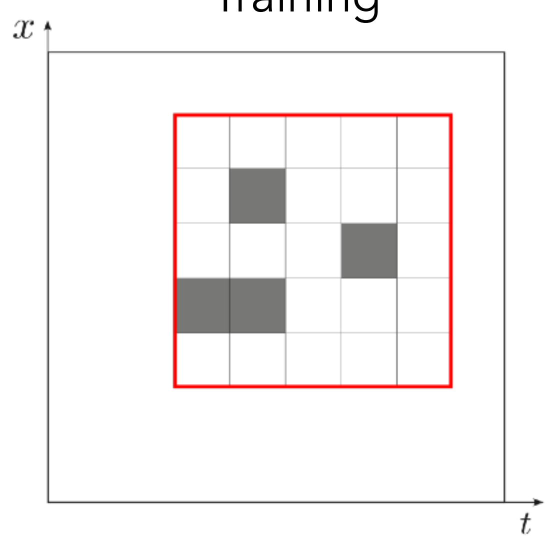
Task-specific fine-tuning

Goal: improve model performance for a specific task
e.g. forecasting, downscaling…

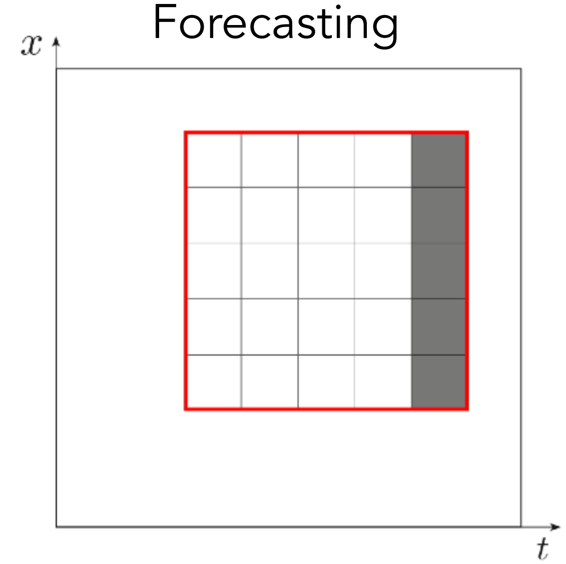
Examples:

e.g. fix masking scheme

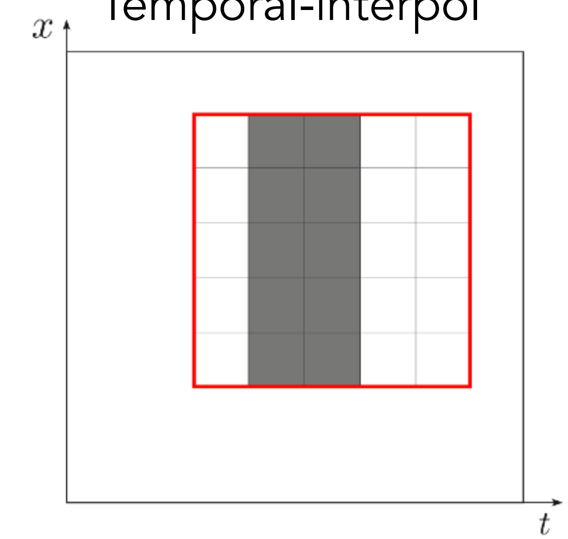
Training



Forecasting

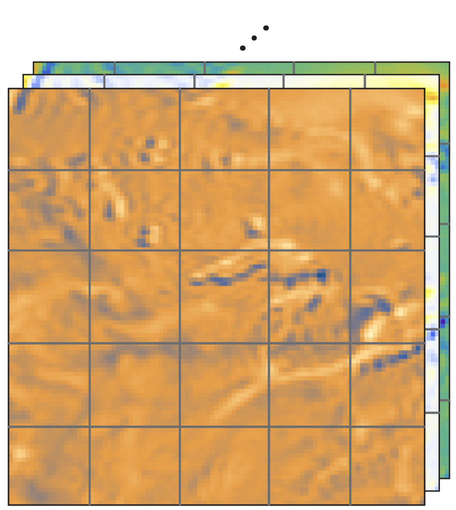


Temporal-interpol



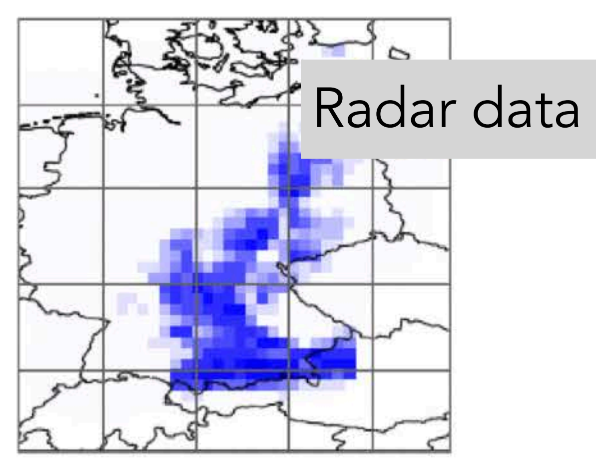
OR

Change target dataset



ERA5

Radklim data



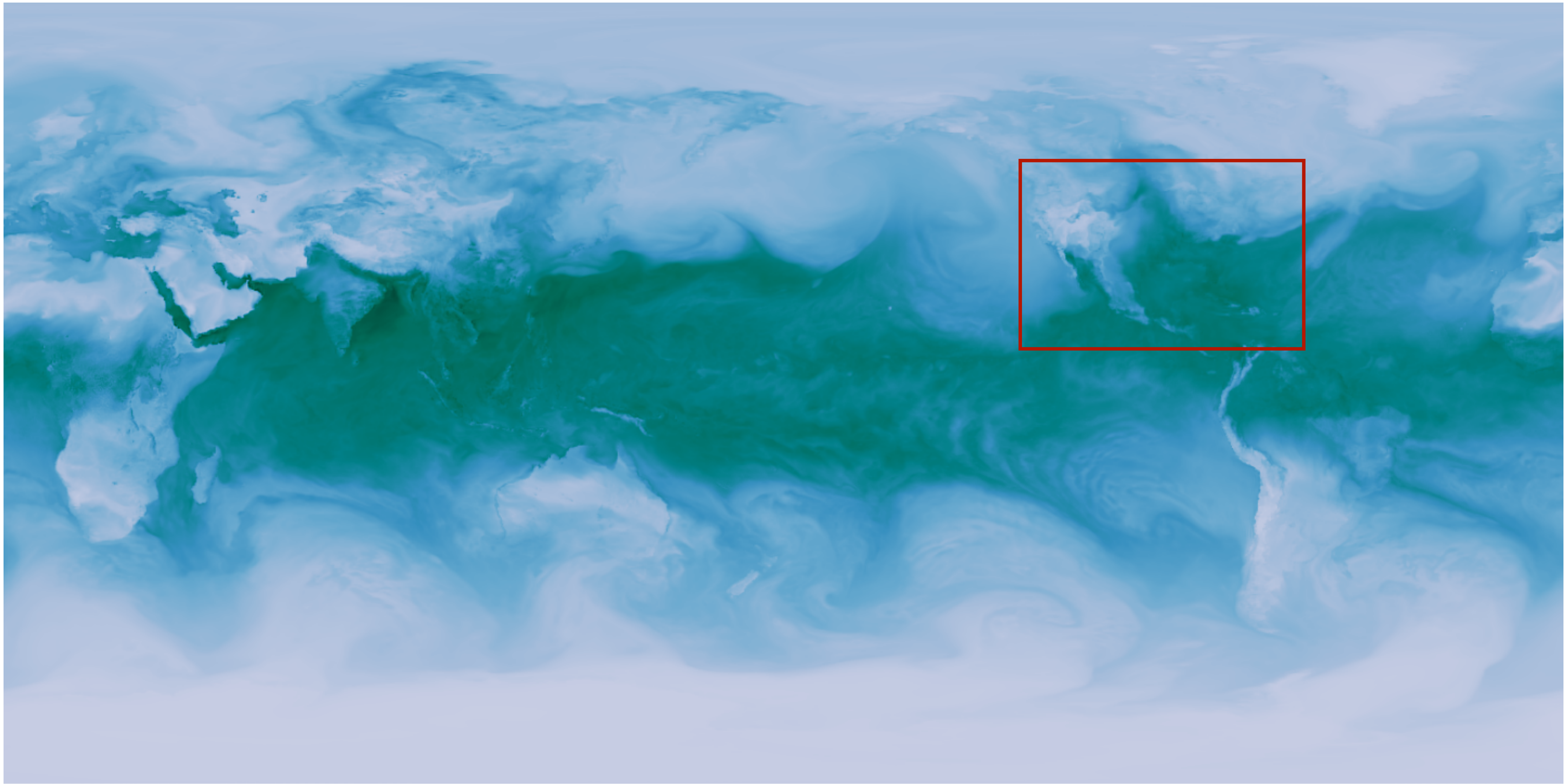
COSMO-REA6

Ilaria Luise, CERN - ilaria.luise@cern.ch

Short term weather forecasting

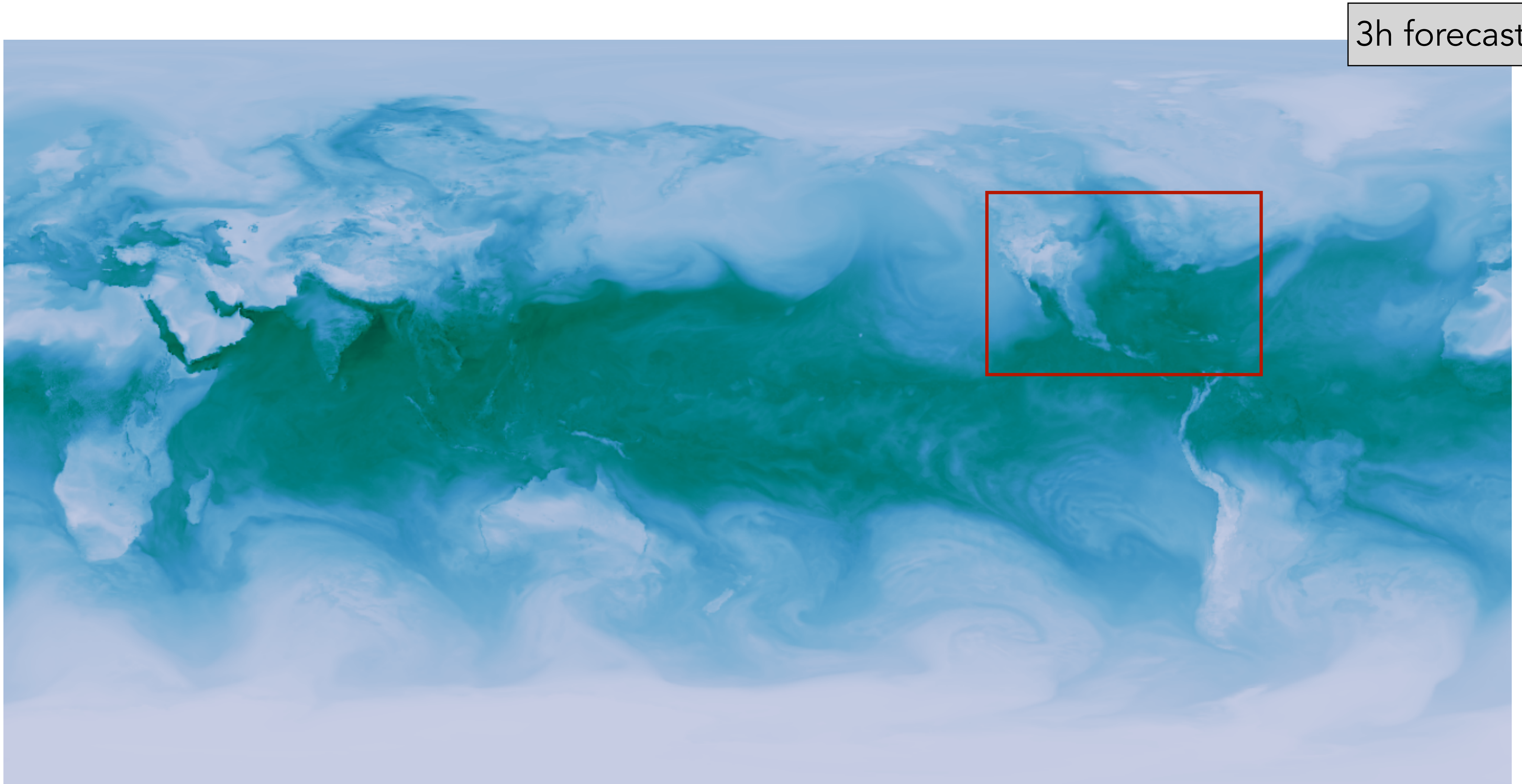
Results: Target - ERA5

specific humidity, June 15th 2018 13:00 UTC



Results: Prediction - AtmoRep

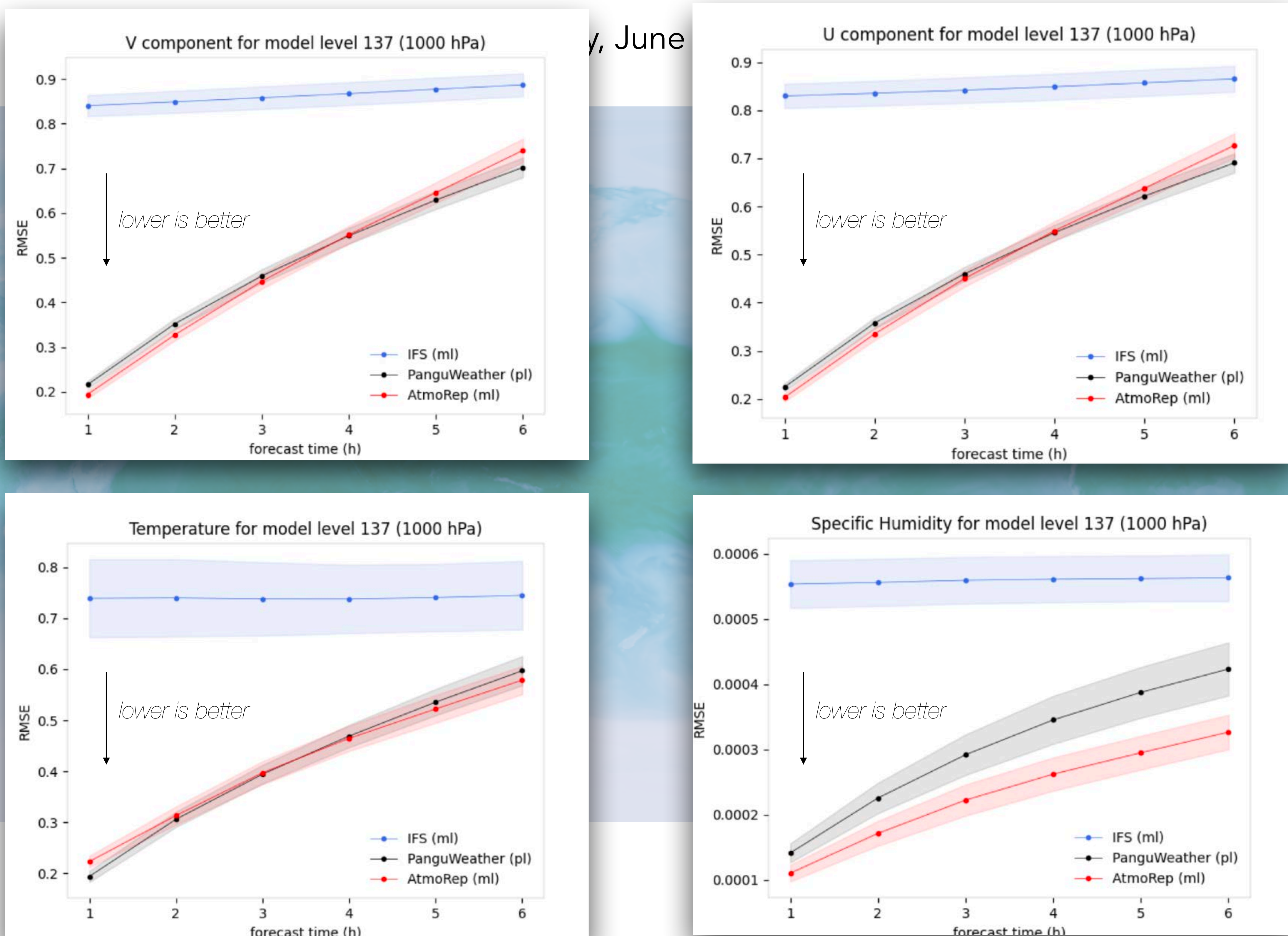
specific humidity, June 15th 2018 13:00 UTC



Ilaria Luise, CERN - ilaria.luise@cern.ch

Results: Prediction - AtmoRep

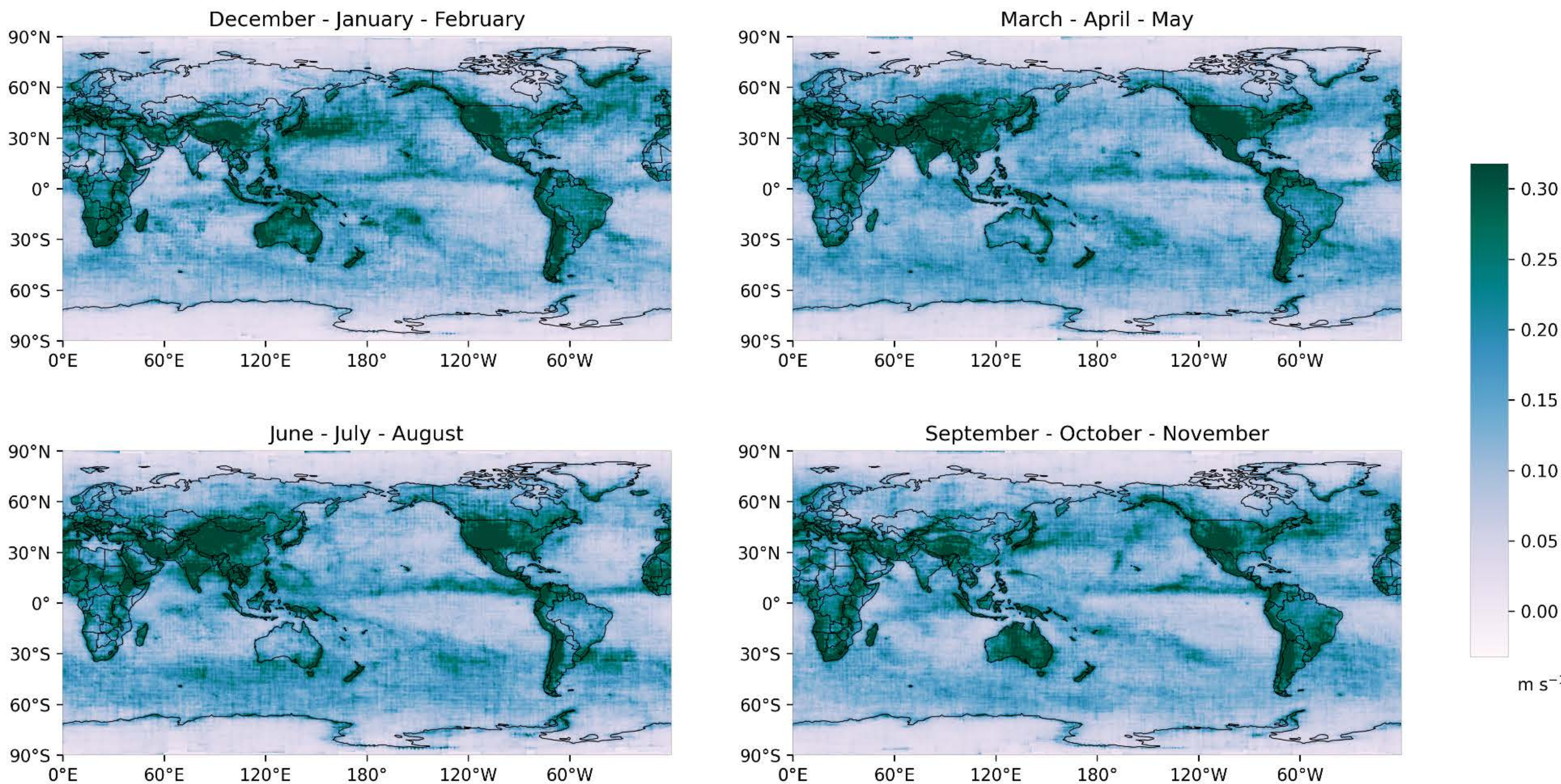
y, June



Ilaria Luise, CERN - ilaria.luise@cern.ch

Results: model errors

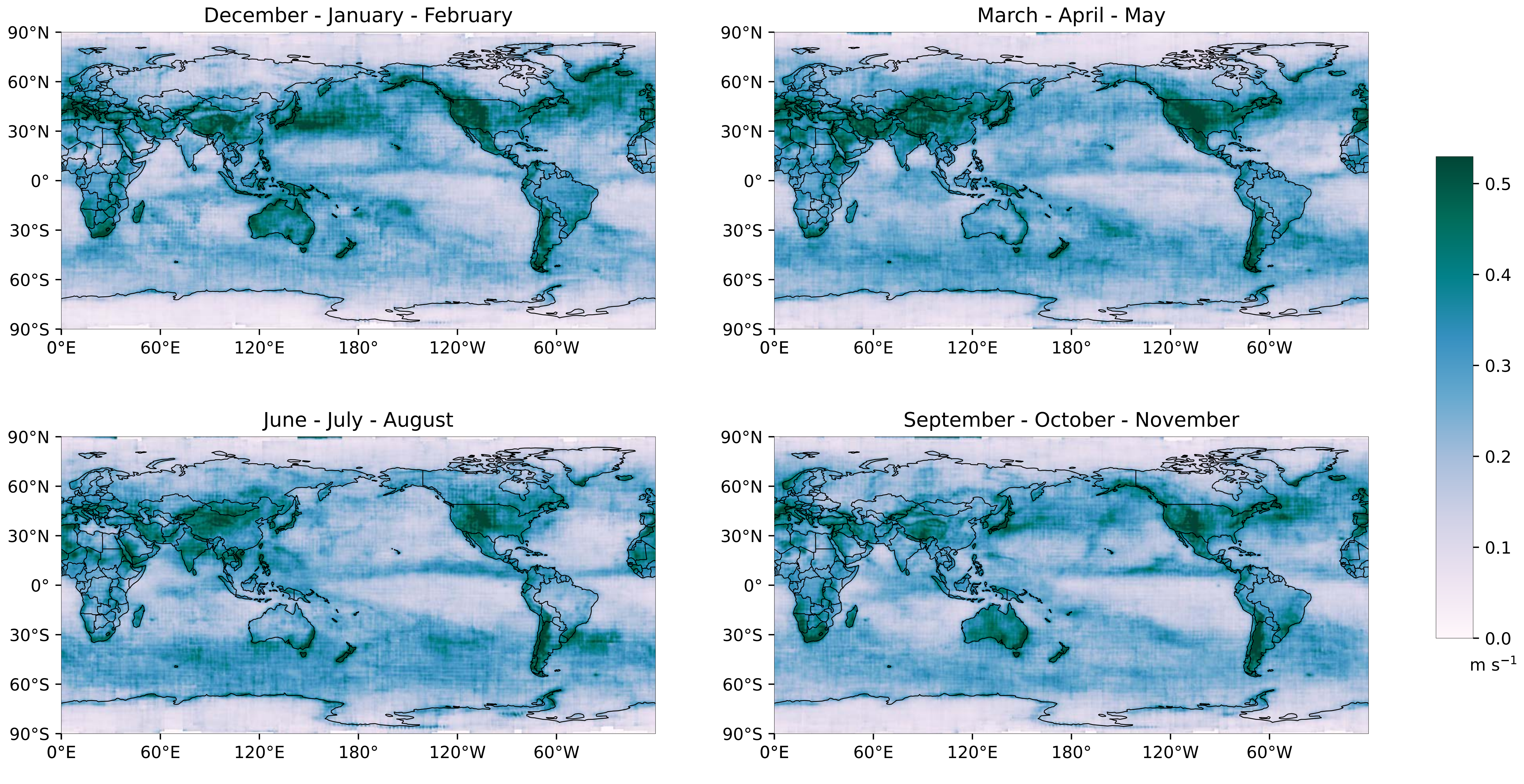
Wind velocity U-component: Average error w.r.t the ground truth



Ilaria Luise, CERN - ilaria.luise@cern.ch

Results: ensemble variability

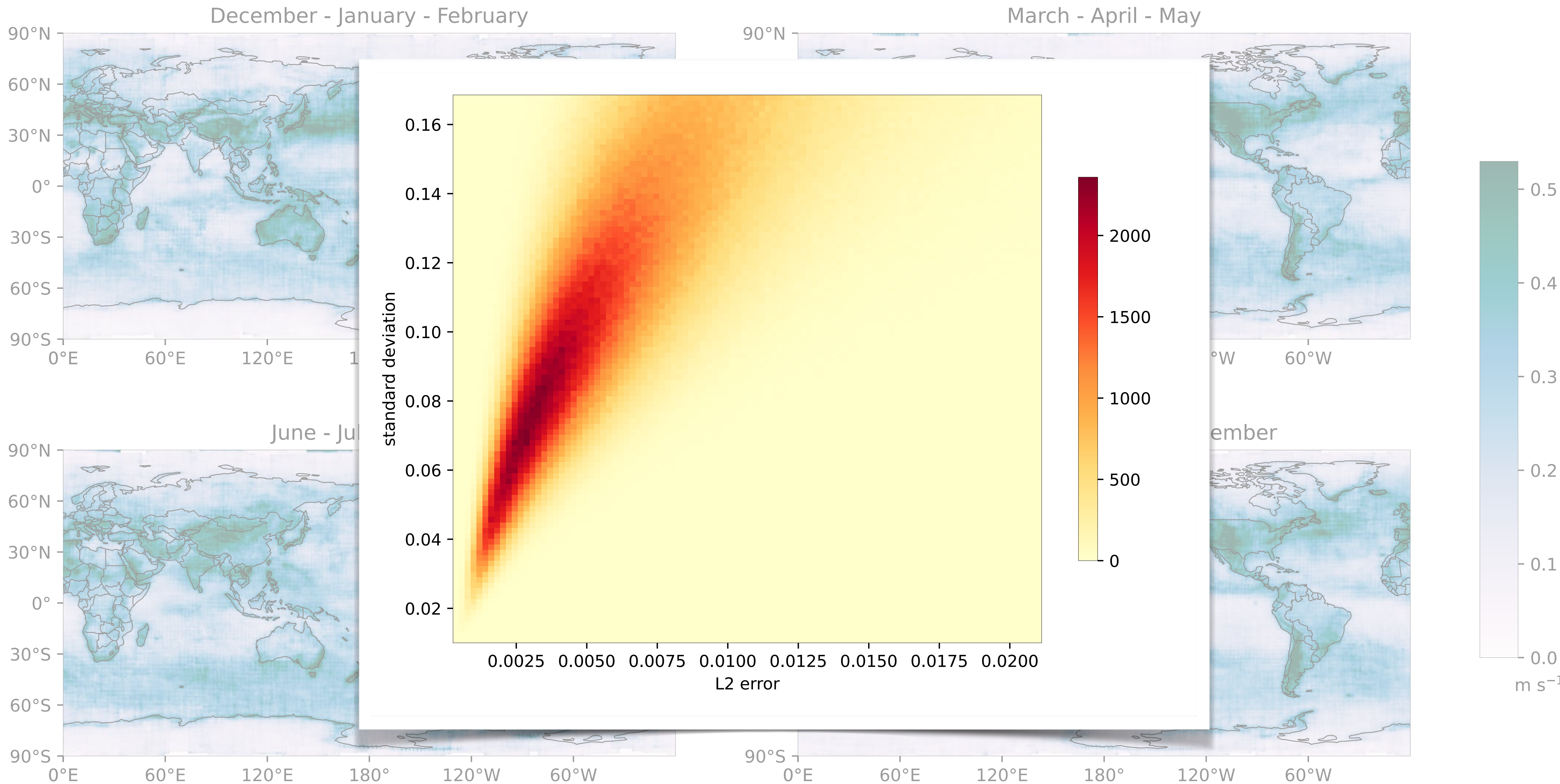
Wind velocity U-component: standard deviation of the ensemble



Ilaria Luise, CERN - ilaria.luise@cern.ch

Results: ensemble variability

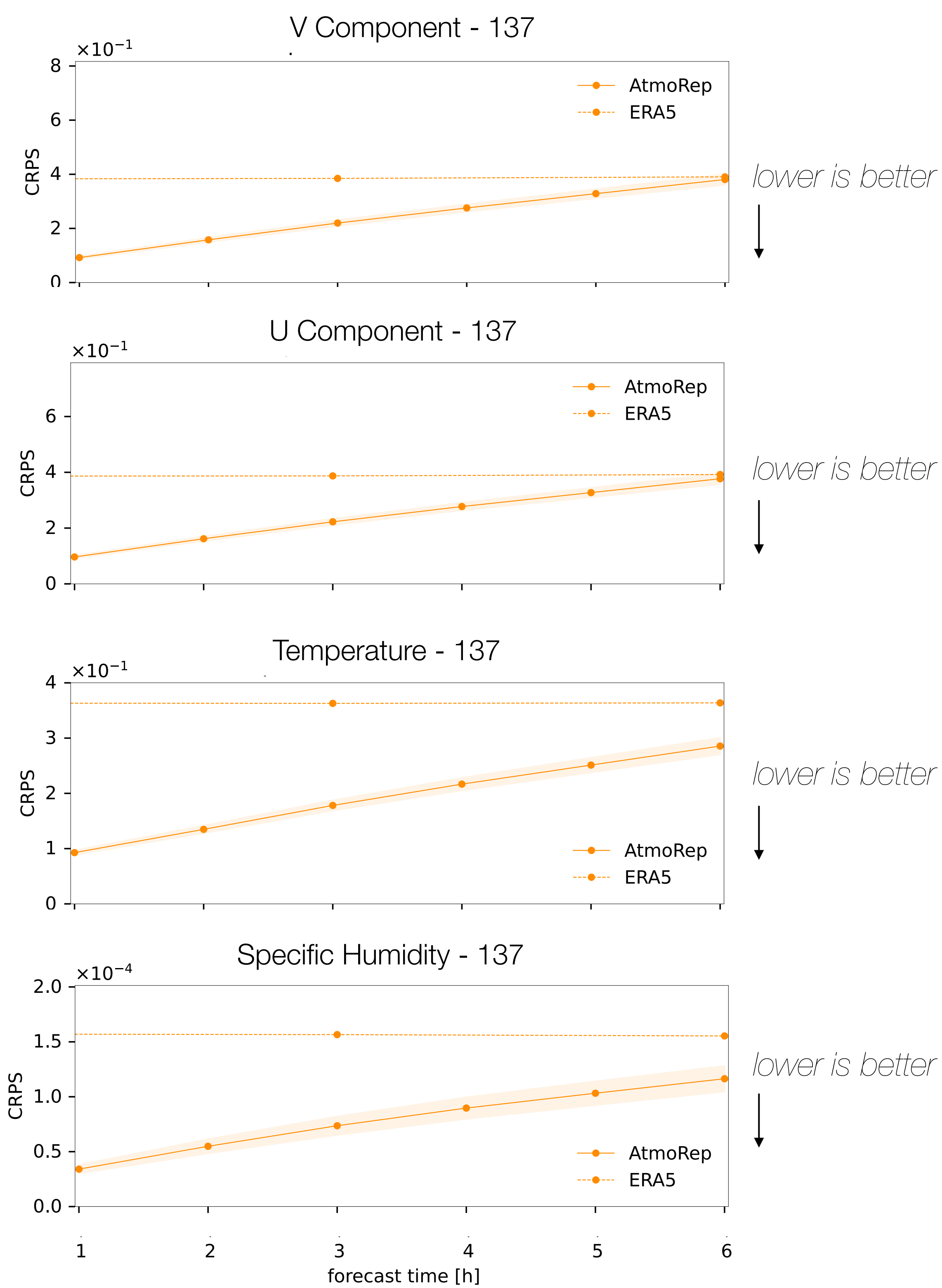
Wind velocity U-component: standard deviation of the ensemble



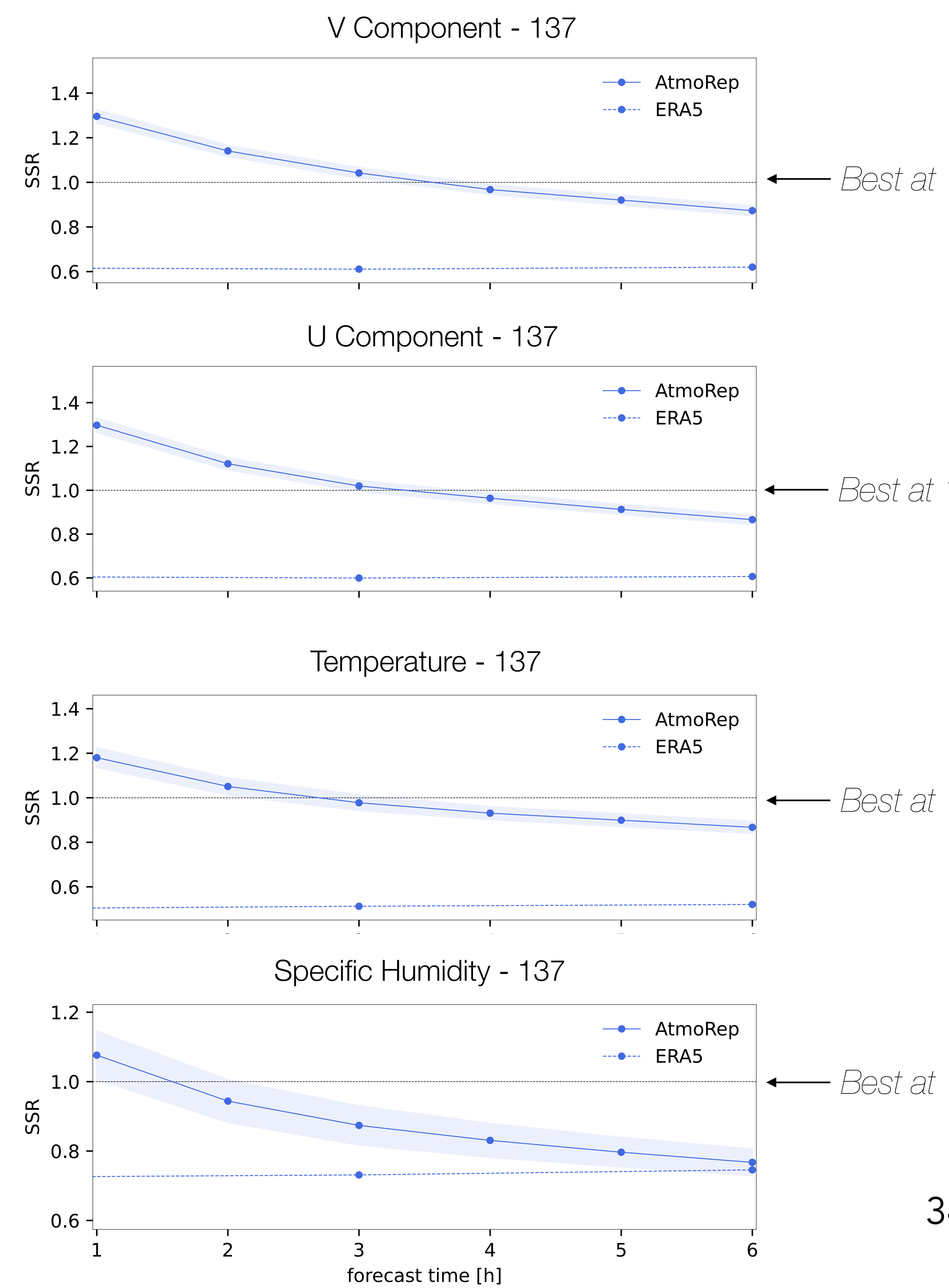
Ilaria Luise, CERN - ilaria.luise@cern.ch

Results: Ensemble - metrics

CRPS



Spread-Skill Ratio (SSR)

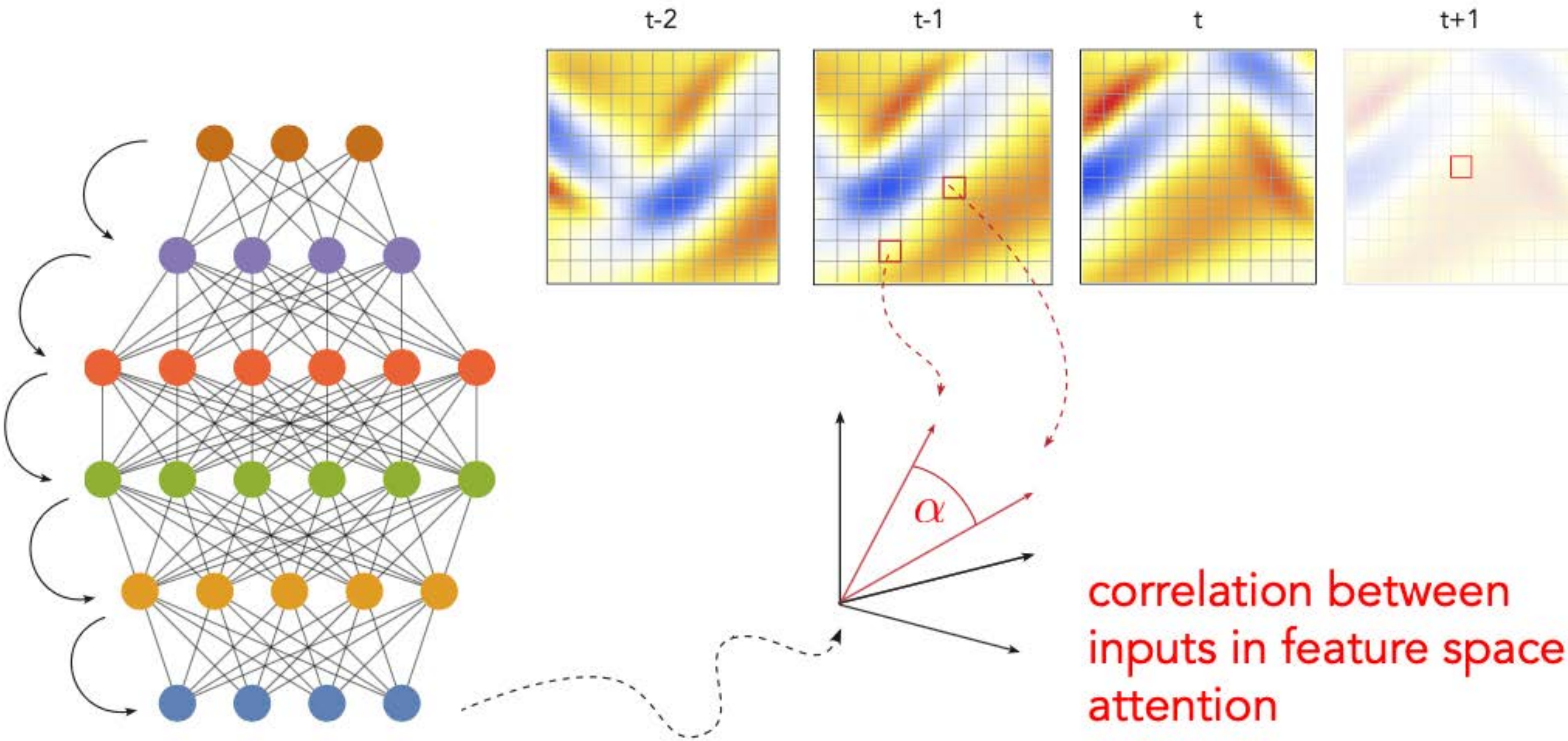


Ilaria Luise, CERN - ilaria.luise@cern.ch

Attention maps and interpretability

Inspect the self-attention mechanism:

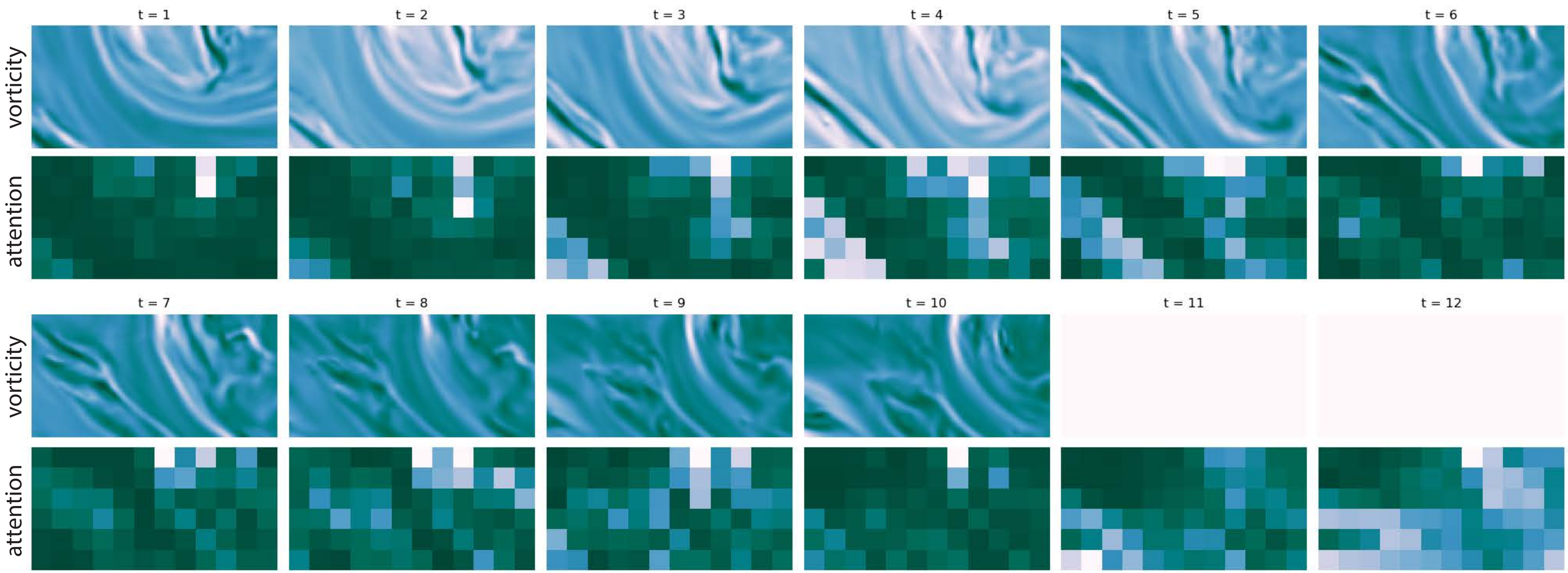
can we identify physics phenomena (e.g. hurricane formation) before they are even created?



Ilaria Luise, CERN - ilaria.luise@cern.ch

Attention maps and interpretability

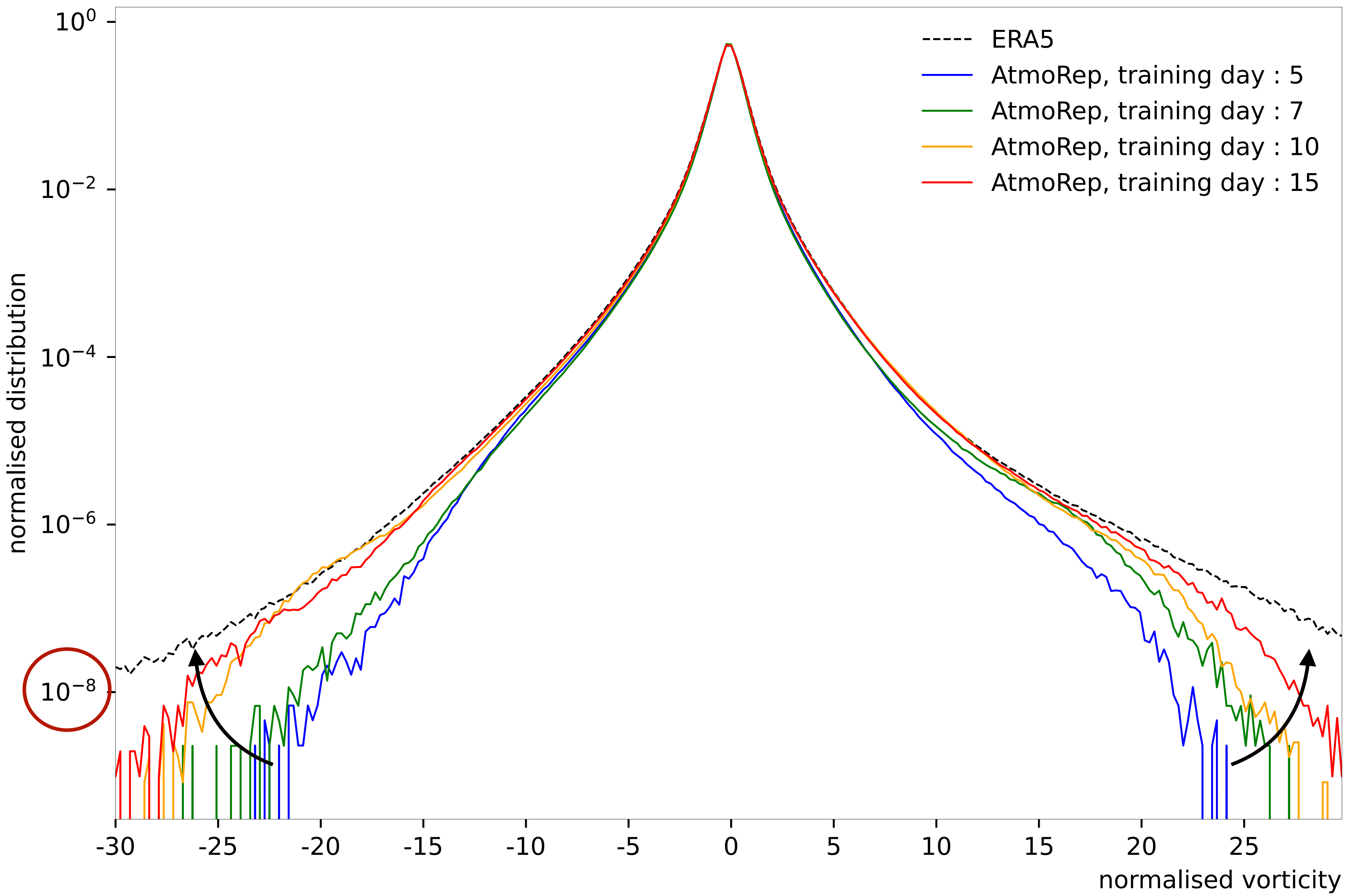
Inspect the self-attention mechanism:
can we identify physics phenomena (e.g. hurricane formation) before they are even created?



Ilaria Luise, CERN - ilaria.luise@cern.ch

Limitations: extreme events

Normalised vorticity, ~1M of random samples per epoch, 2018

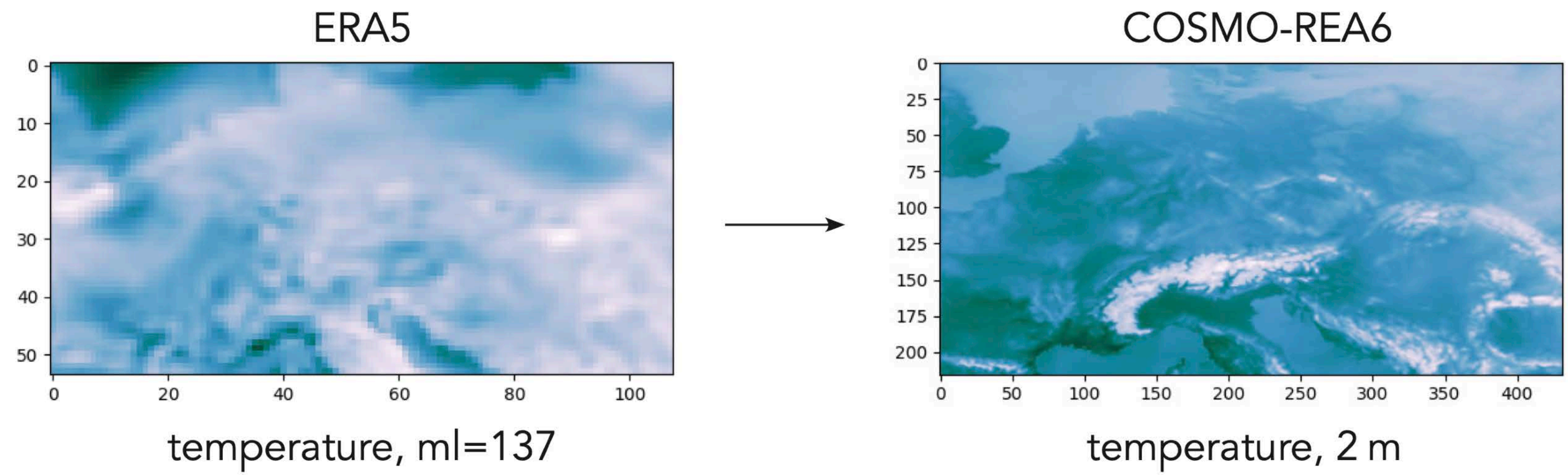


Ilaria Luise, CERN - ilaria.luise@cern.ch

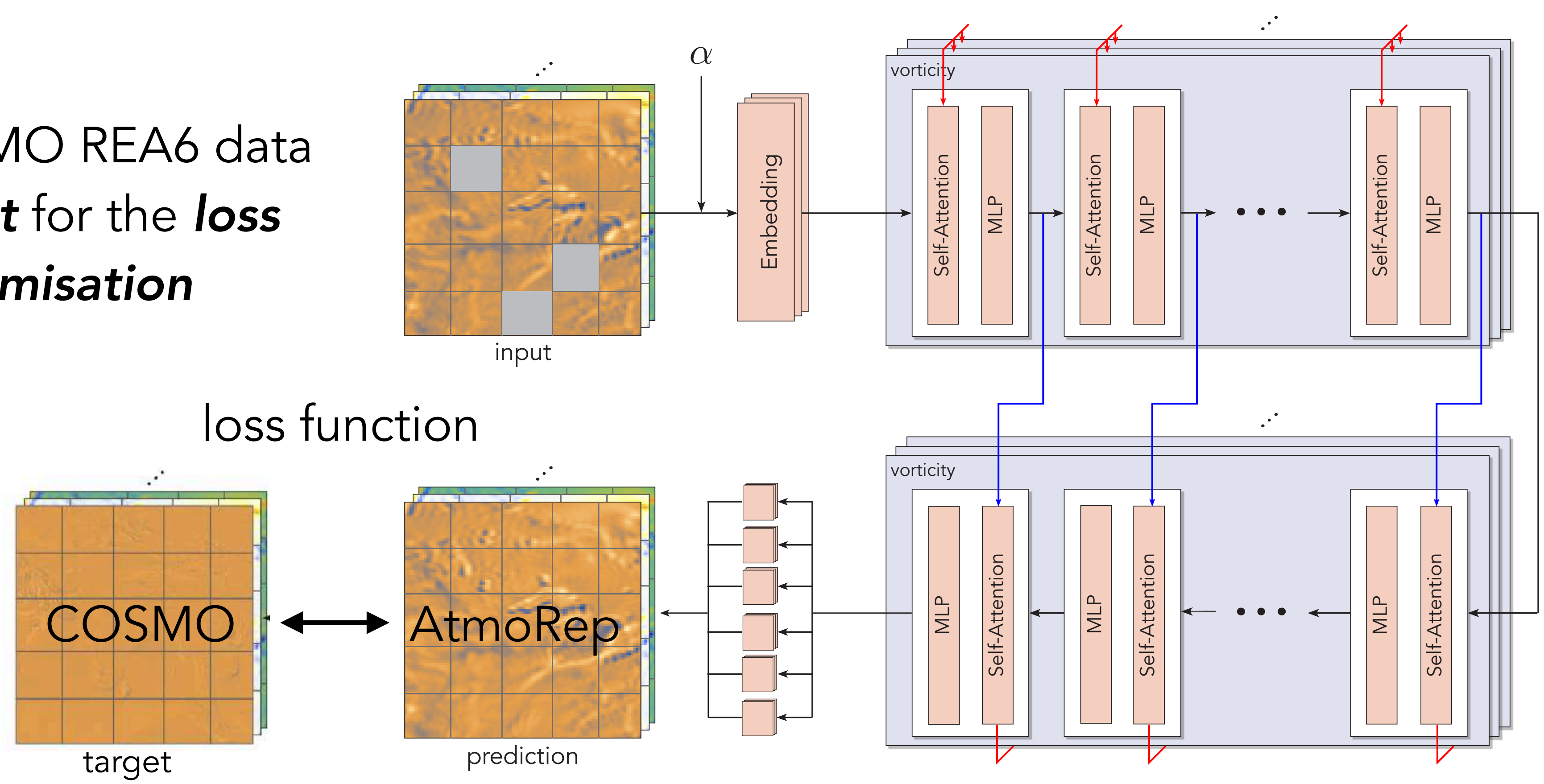
Towards a multi-modal approach

Fine tuning on other datasets: data driven precipitation corrections & downscaling

Downscaling



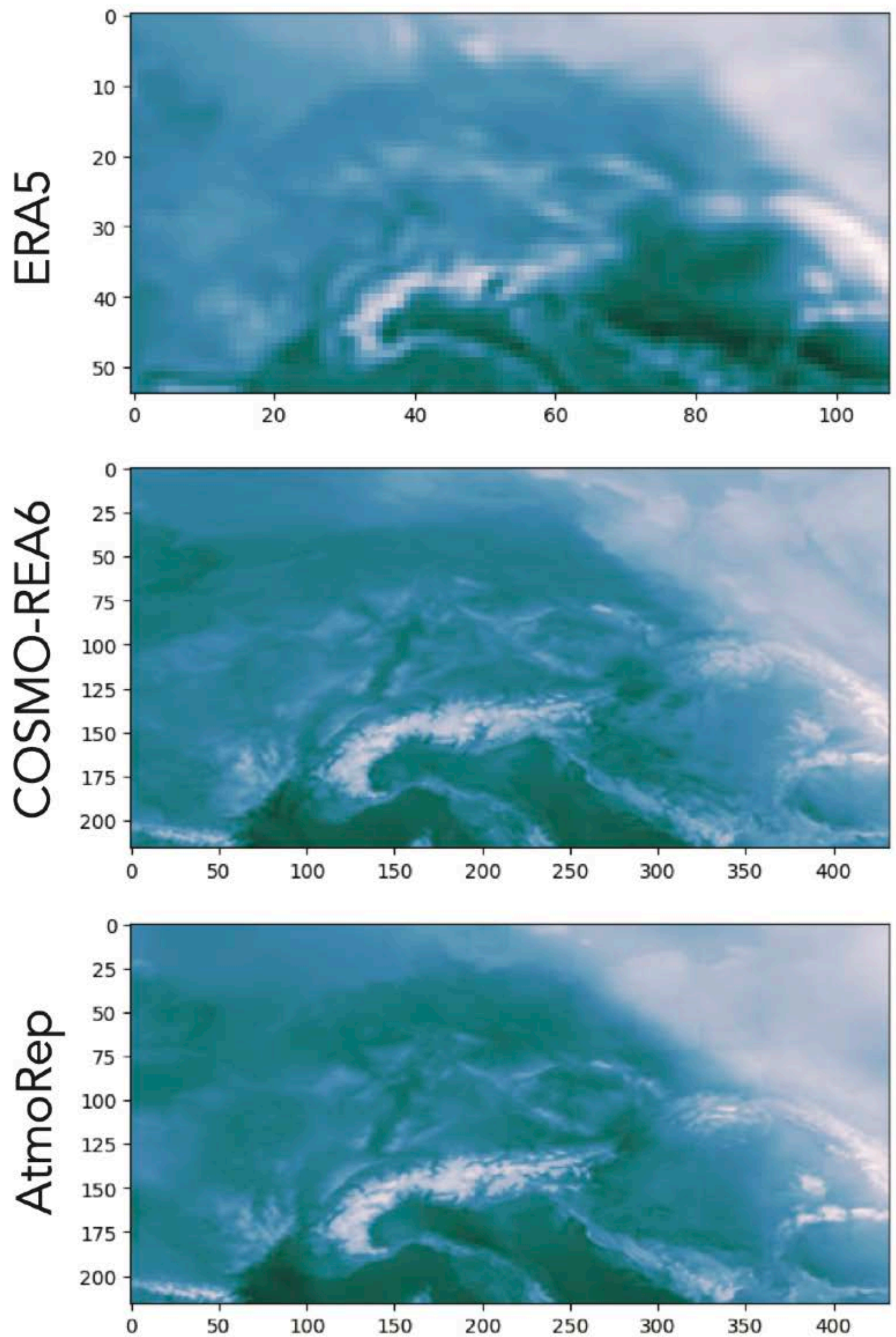
Use COSMO REA6 data as ***target*** for the ***loss minimisation***



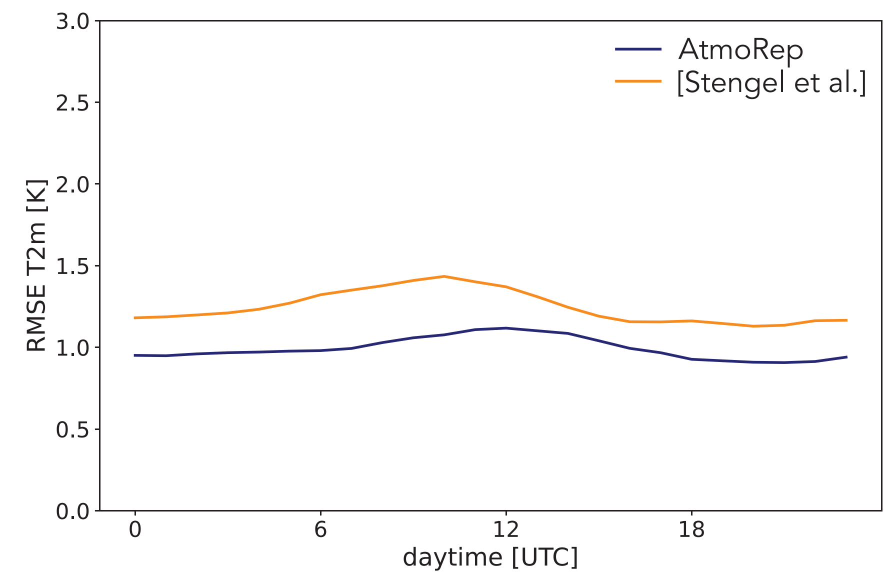
Ilaria Luise, CERN - ilaria.luise@cern.ch

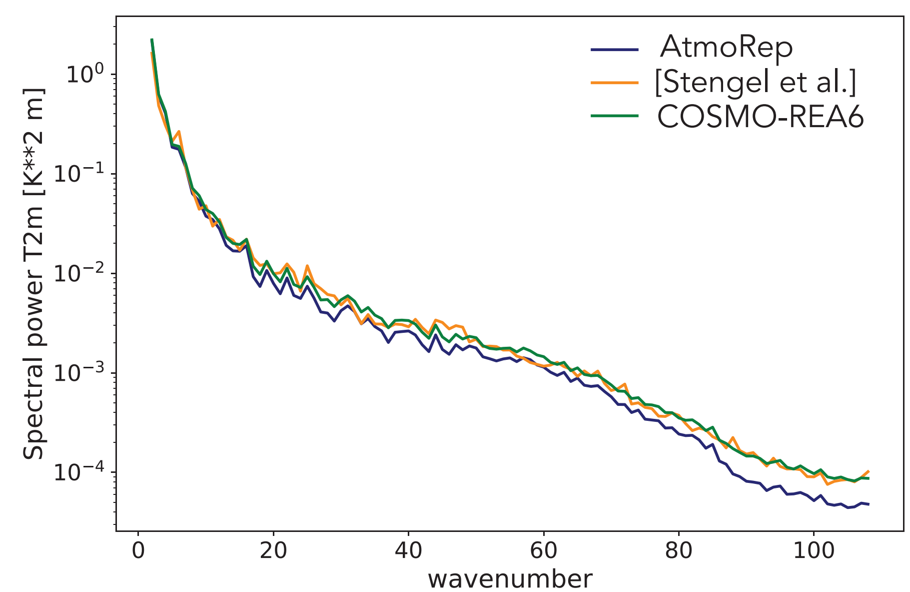
Downscaling

Use the COSMO REA6 dataset (6 km resolution vs ~32 km in ERA5) to create a downscaled version of ERA5



Comparison with a competing
AI-based model for downscaling:



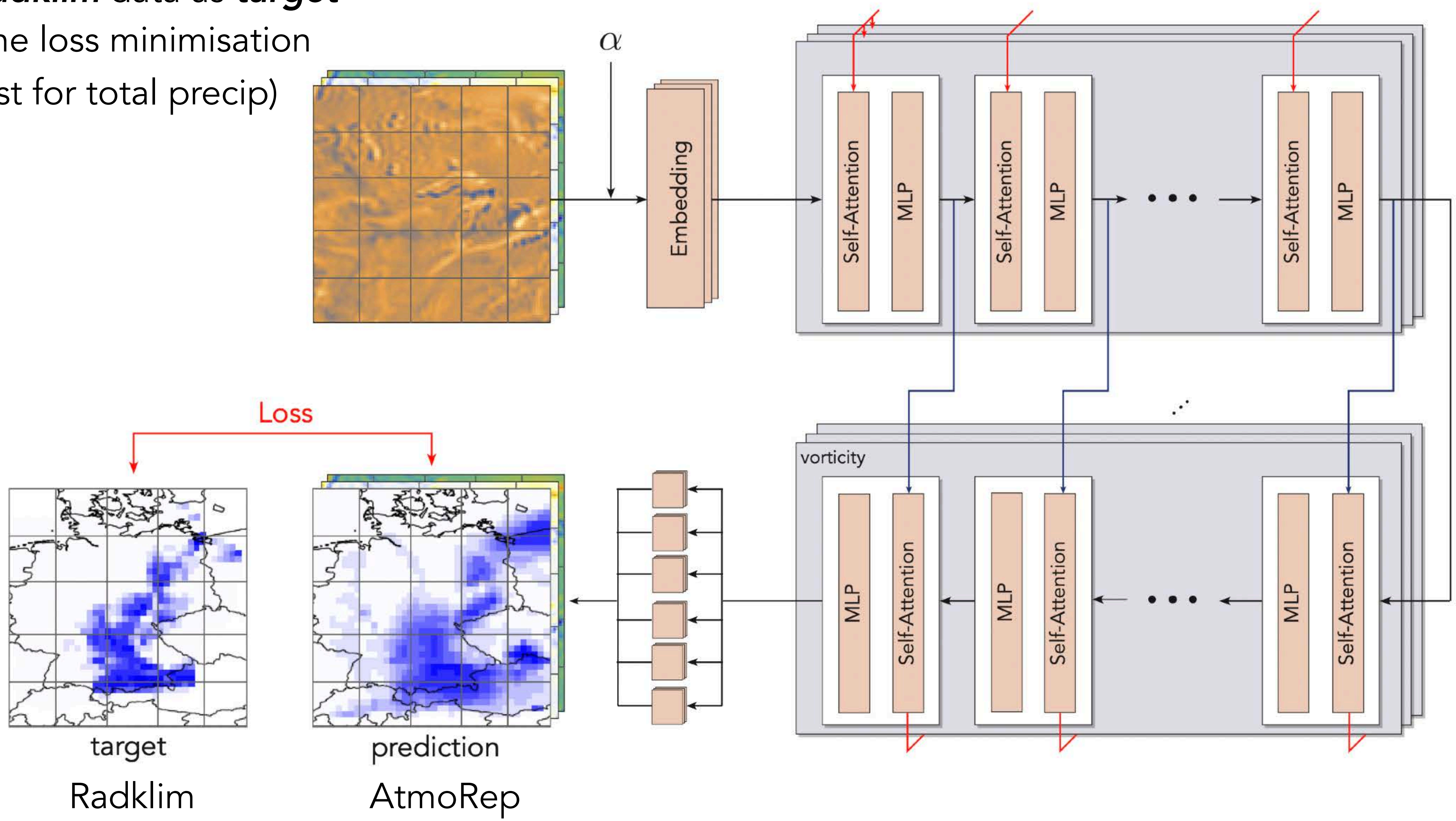


Ilaria Luise, CERN - ilaria.luise@cern.ch

Bias corrections

Precipitation rates are known to be suboptimal in ERA5
Use RADKLIM radar data to fine-tune the precipitation rates in AtmoRep

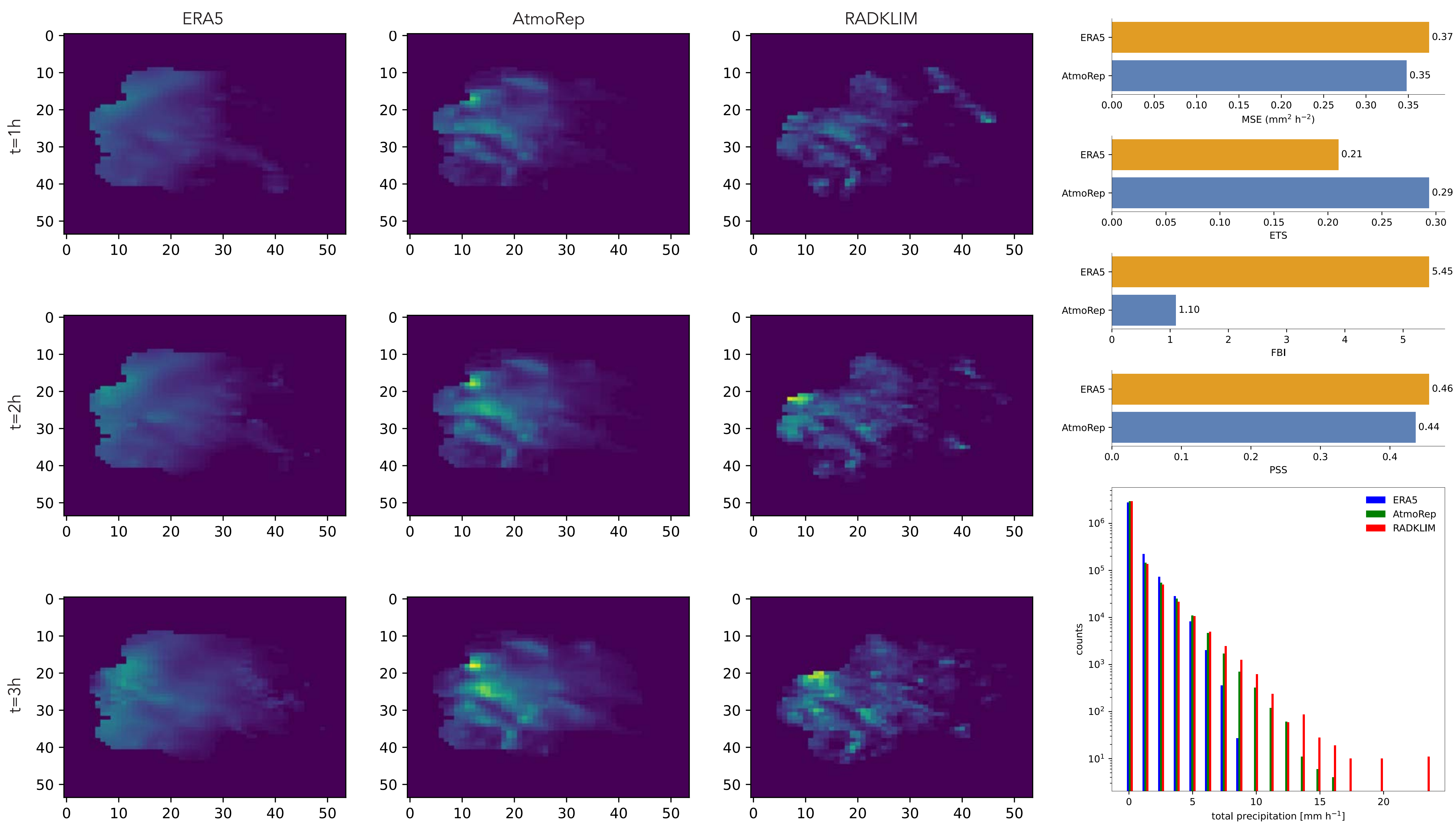
Use ***Radklim*** data as ***target*** for the loss minimisation (just for total precip)



Ilaria Luise, CERN - ilaria.luise@cern.ch

Bias corrections: Results

Precipitation rates are known to be suboptimal in ERA5
Use RADKLIM radar data to fine-tune the precipitation rates in AtmoRep

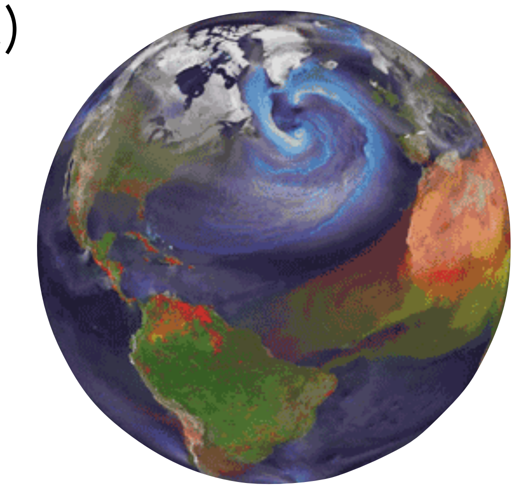


Ilaria Luise, CERN - ilaria.luise@cern.ch

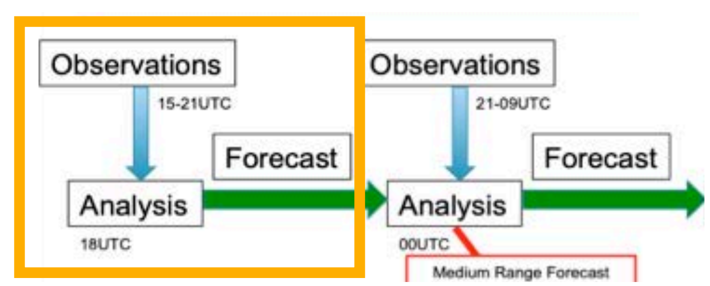
Outlook & Future plans

AtmoRep: next plans

- **Multimodality:** train on multiple datasets at the same time (ERA5, CERRA..)
  - Can we extrapolate to regions with same climatology?
- **Medium range weather forecasting:** roll-out mechanism
  - Use diffusion models maybe?
  - Can we use a smaller network for pre-training?
- **Climate projections:** the HClimRep project will build on top of AtmoRep for climate projection studies



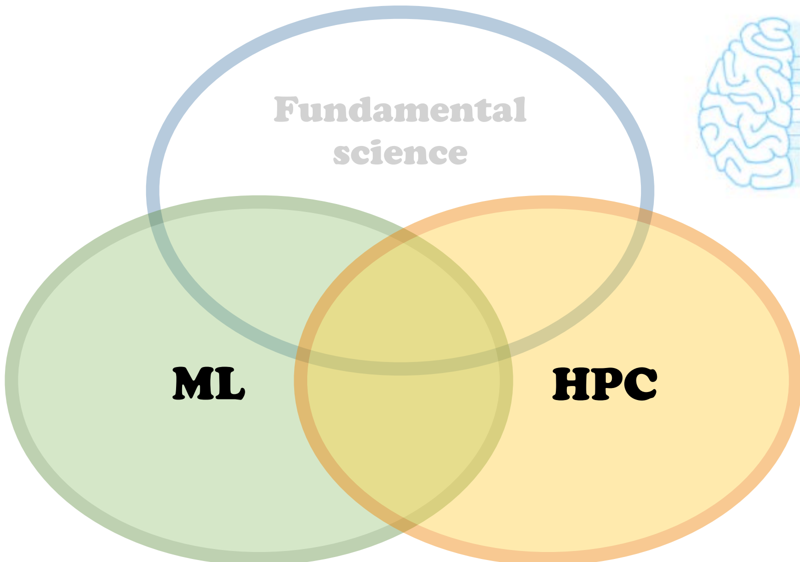
Can we re-think the full system?



- Large ensemble generations
- Correct biases through ML
- Coupled ocean+land+atmo systems
- Learn directly from observation

Ilaria Luise, CERN - ilaria.luise@cern.ch

Future challenges



Scaling:

- **Efficiently scaling distributed training to larger models**
- Develop the software infrastructure and model architecture suitable for such big models

Accessibility:

- **Deployment of the models on the cloud**
- we need an integration of the HPC centers to provide **seamless access** and data movement in the background (example: Google Cloud)

Maintenance:

- **How to integrate new incoming data**
- How to **expand** to new fields/variables without fully retraining the model each time?

From Martin Schultz - ESA-ECMWF workshop 2024

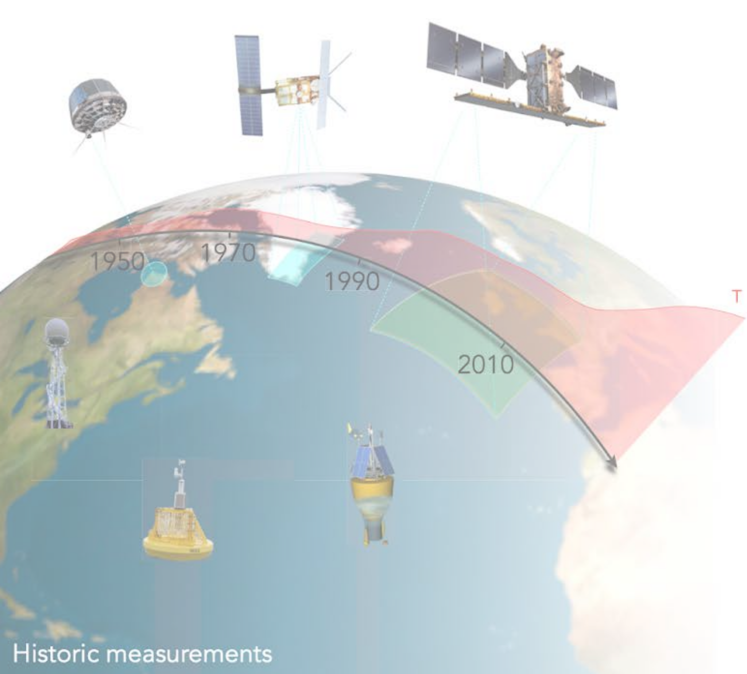
Conclusions

AtmoRep: First prototype of a multi-purpose model for Earth system applications

The **model is available and testable** on the current applications:
nowcasting, downscaling, temporal interpolation and precipitation corrections.

More infos:

- Code is available on GitHub: link
- More infos on the website: www.atmorep.org
- **Pre-print on ArXiv: link**

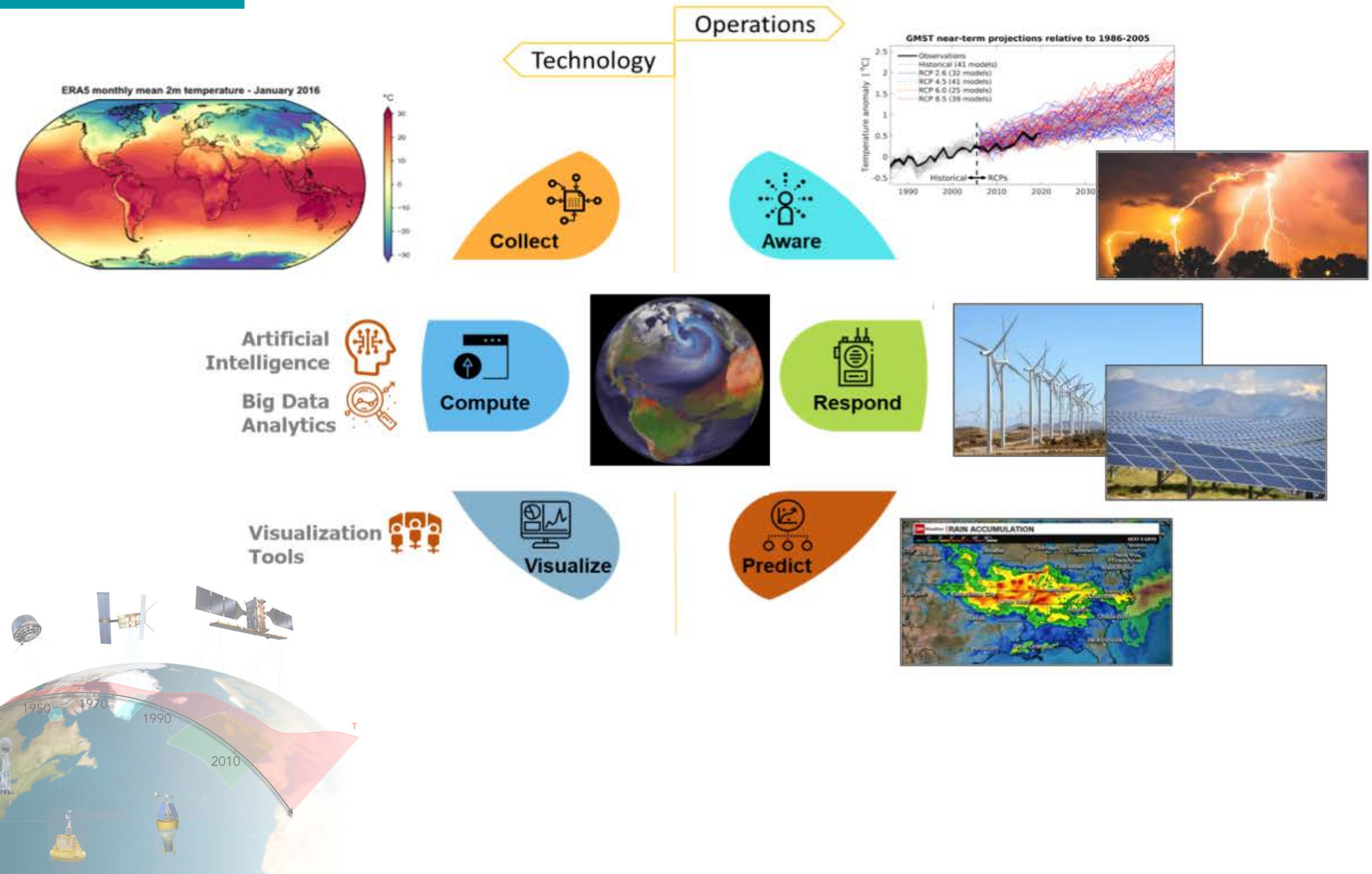


.. and some long term plans:

- How to integrate "raw" observations?
- Coupled atmosphere+ocean system?

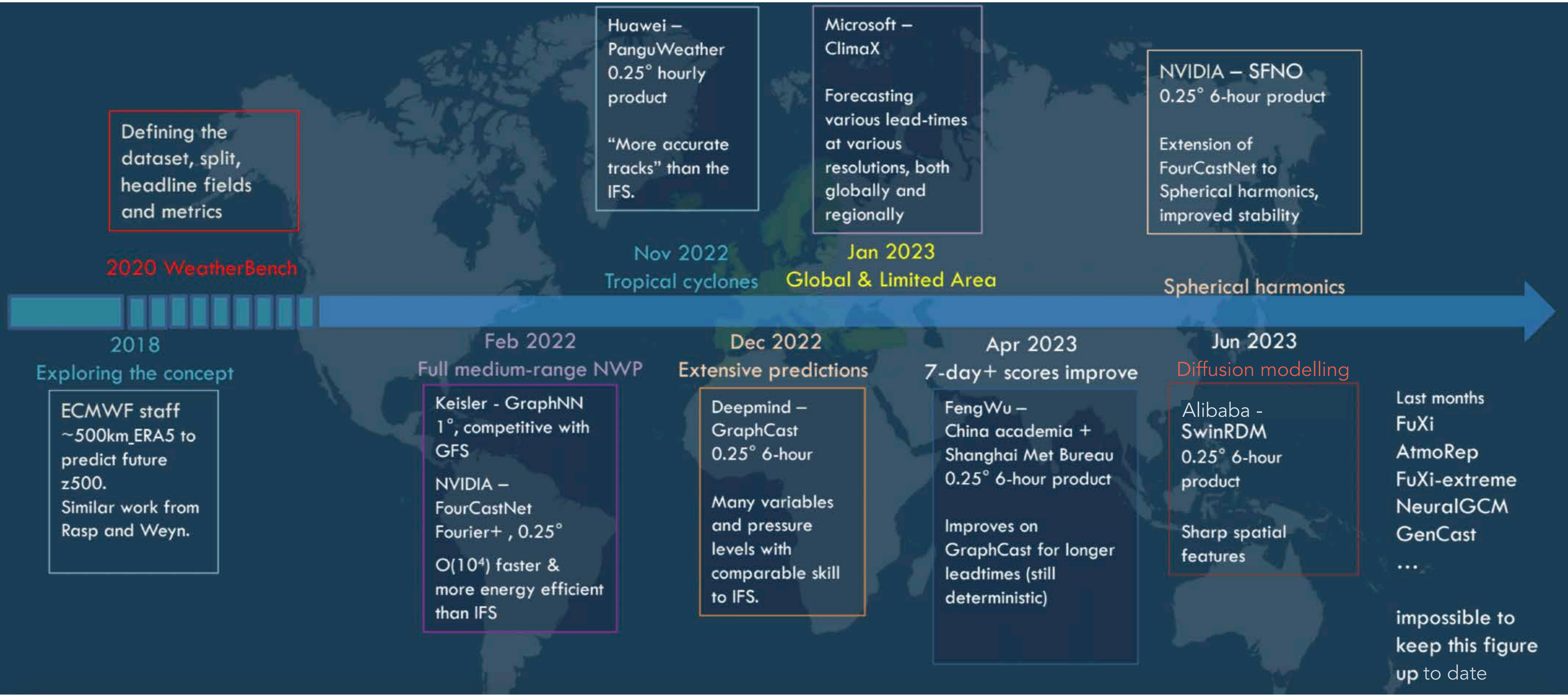
Ilaria Luise, CERN - ilaria.luise@cern.ch

Backup



Operations
Technology
ERA5 monthly mean 2m temperature - January 2016
°C
GMST near-term projections relative to 1986-2005
Observations
Historical (41 models)
RCP 2.6 (32 models)
RCP 4.5 (41 models)
RCP 6.0 (25 models)
RCP 8.5 (39 models)
Temperature anomaly [°C]
Historical
RCPs
1990
2000
2010
2020
2030
Collect
Aware
Artificial Intelligence
Big Data Analytics
Compute
Respond
Visualization Tools
Visualize
Predict
RAIN ACCUMULATION
1950
1970
1990
2010

Timeline of the breakthrough



From Mihai Alexe - ESA-ECMWF workshop 2024

The CERN Environmental initiative

CIPEA CERN Innovation Programme on Environmental Applications

15 projects supported in 2022

RENEWABLE AND LOW-CARBON ENERGY	**IVAC-RED: Insulation Vacuum of SC Cables for Renewable Energy Distribution**
CLIMATE CHANGE AND POLLUTION CONTROL	**EMP²: environmental modelling and prediction platform**
CLEAN TRANSPORTATION AND FUTURE MOBILITY	**CERN Clean Cool: Clean cooling and heating systems coupled with plug&play thermal storage batteries**
SUSTAINABILITY AND GREEN SCIENCE	**Optical interrogator for Fiber-Optic Sensors in Sustainable Agriculture**

check all the supported projects at: https://kt.cern/environment/CIPEA

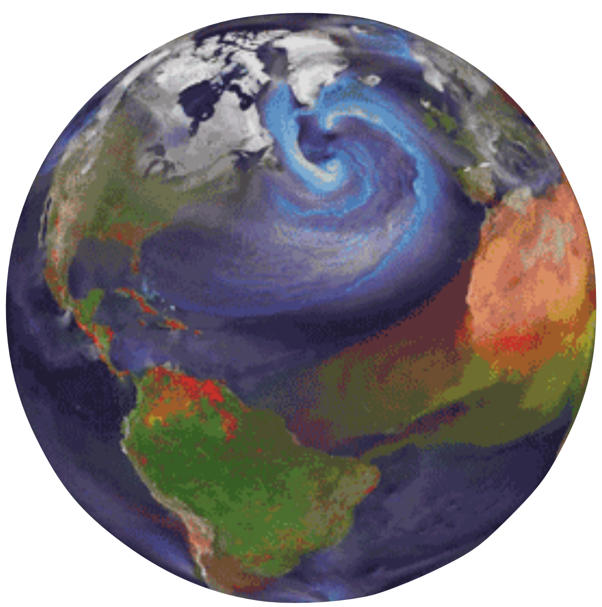
Ilaria Luise, CERN - ilaria.luise@cern.ch

One last word on accessibility

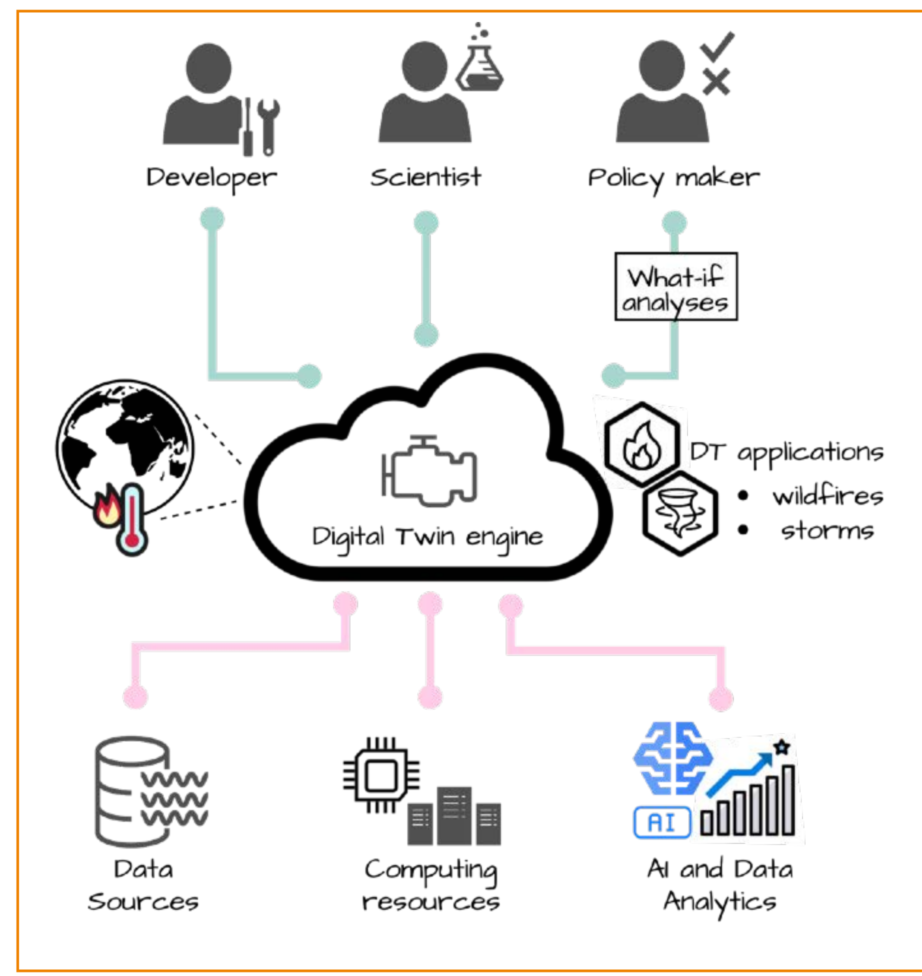
Accessibility: How can local communities make use of such complex models?

<u>**Concrete example:**</u>

A local administration needs to select a location for a wind farm



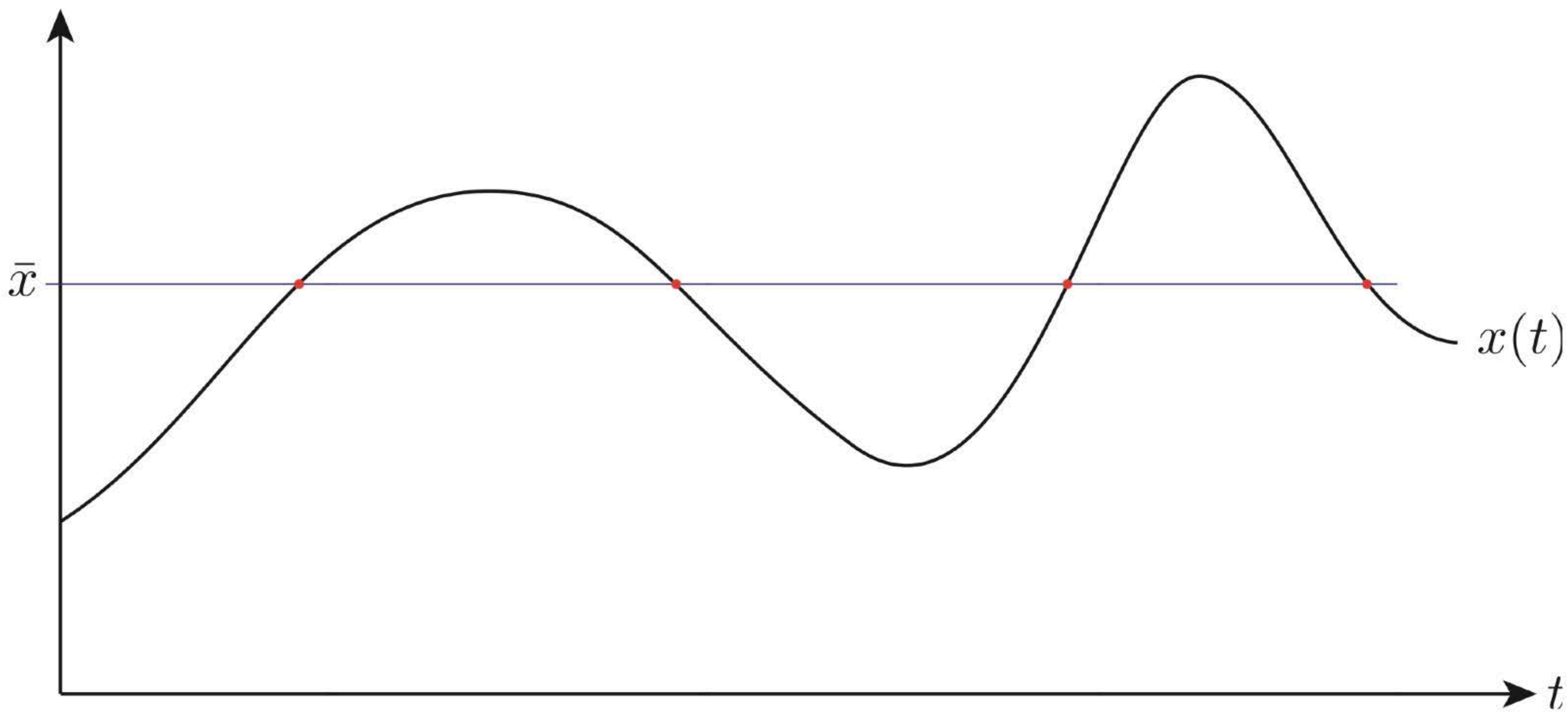
interTwin



Goal:

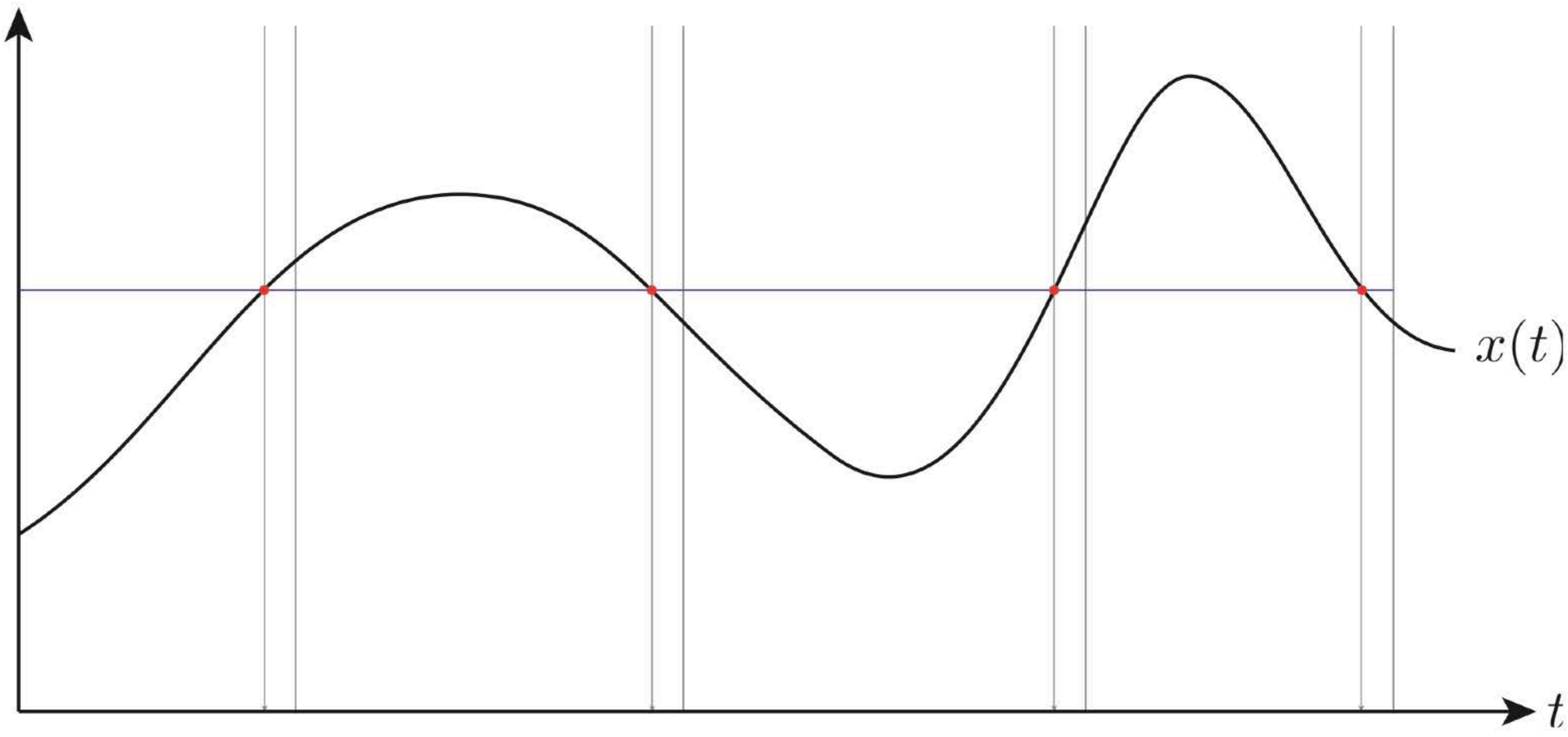
Bridge the gap between HPCs and users

Learning a distribution

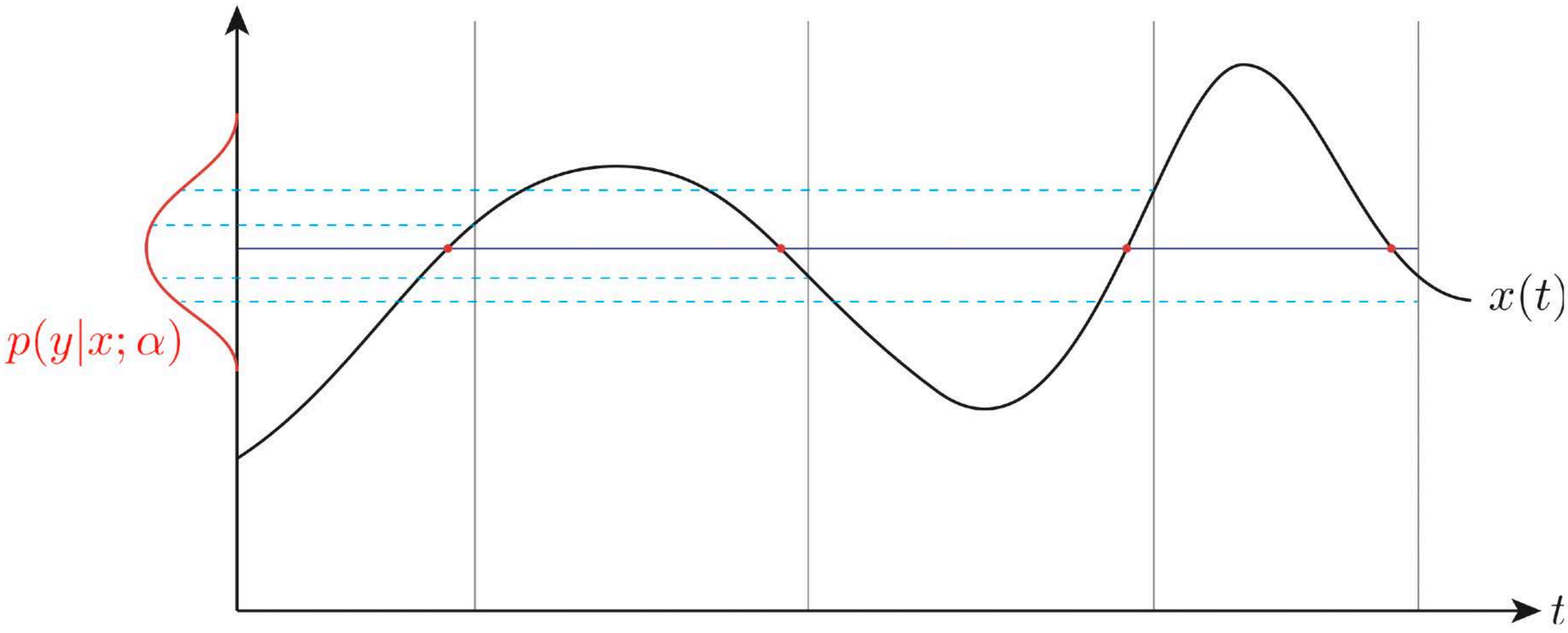


Ilaria Luise, CERN - ilaria.luise@cern.ch

Learning a distribution



Learning a distribution



Ilaria Luise, CERN - ilaria.luise@cern.ch

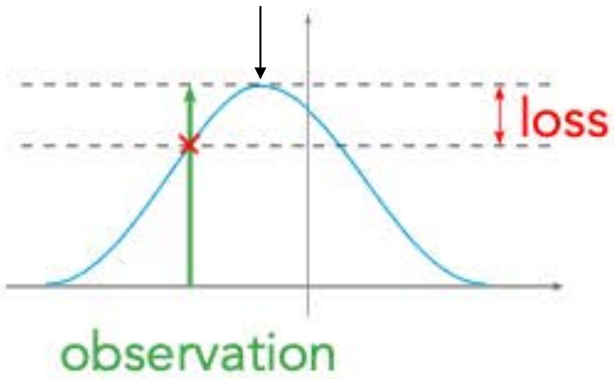
Statistical loss

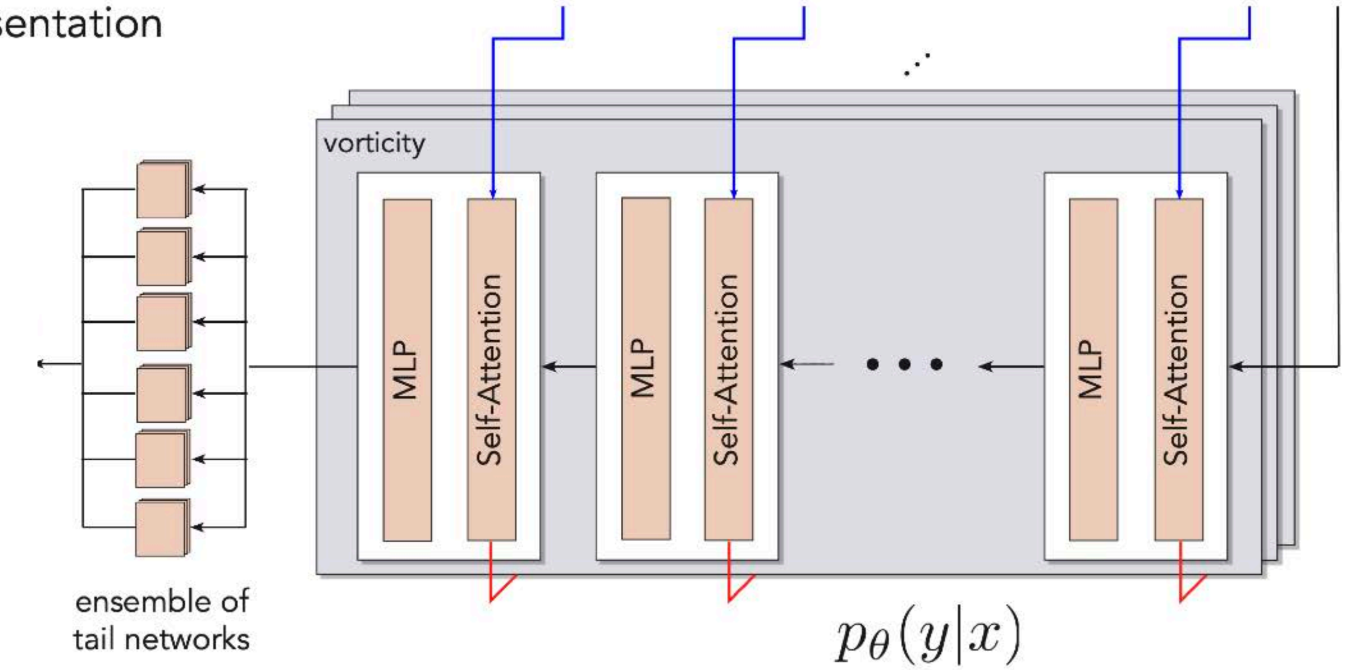
First
statistical momentum fit

end-to-end training encourages
statistical representation

Mean of the tail-nets

Data distrib from the tail-nets

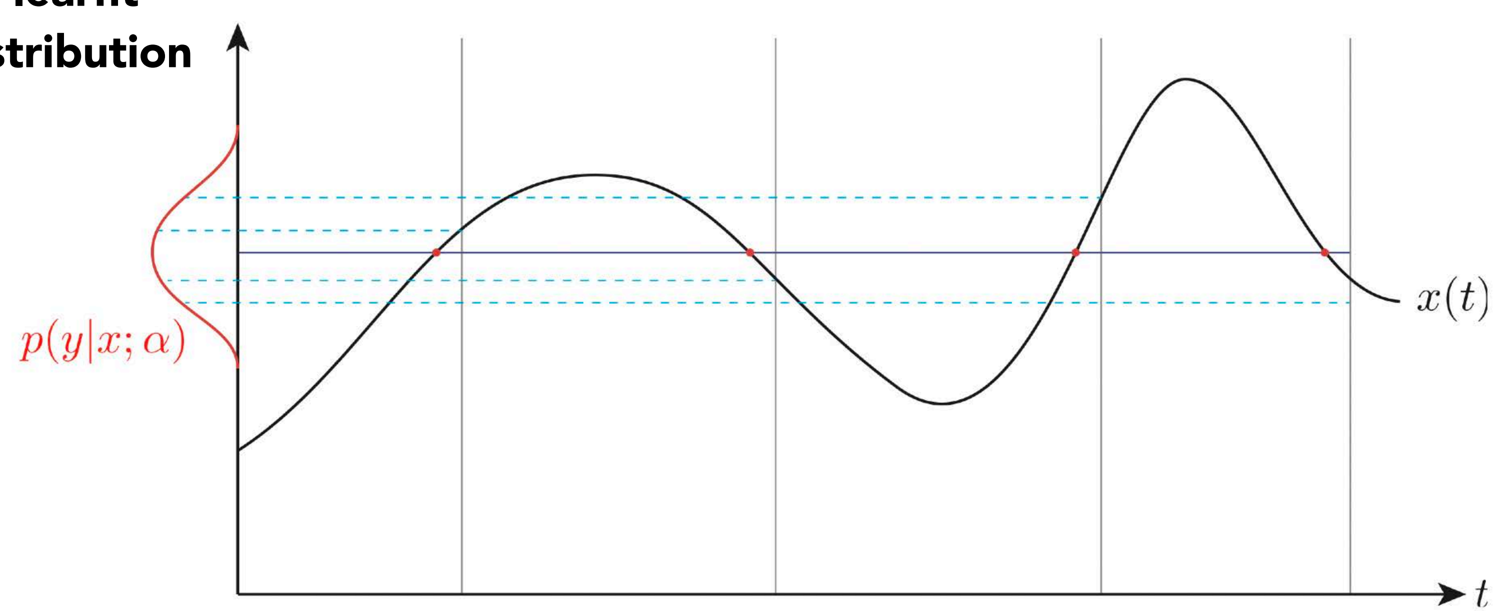




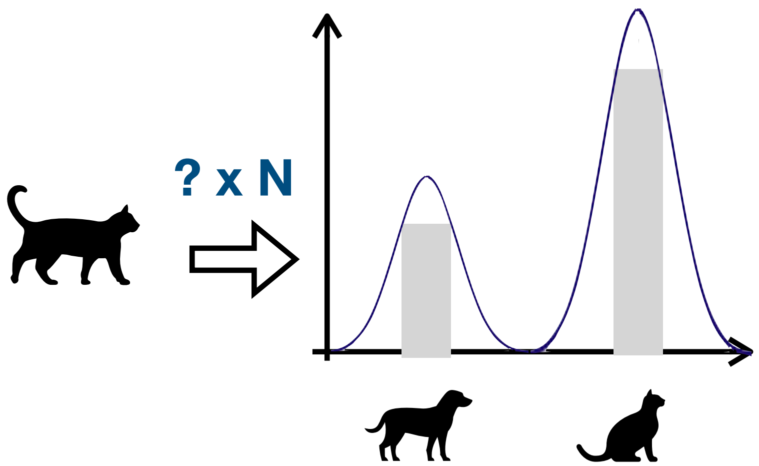
(Ground truth)

Ensemble of tail networks generates N predictions for each pixel

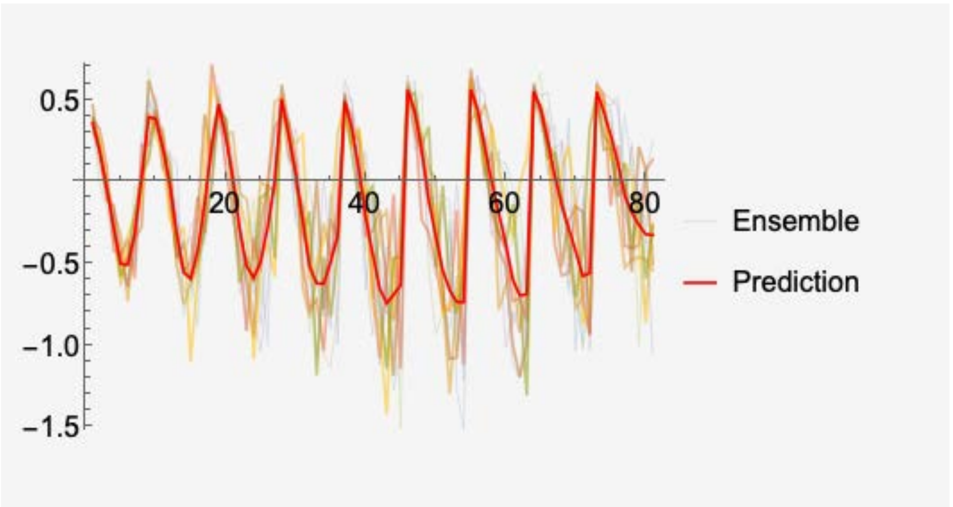
Sample from the learnt spatio-temporal distribution

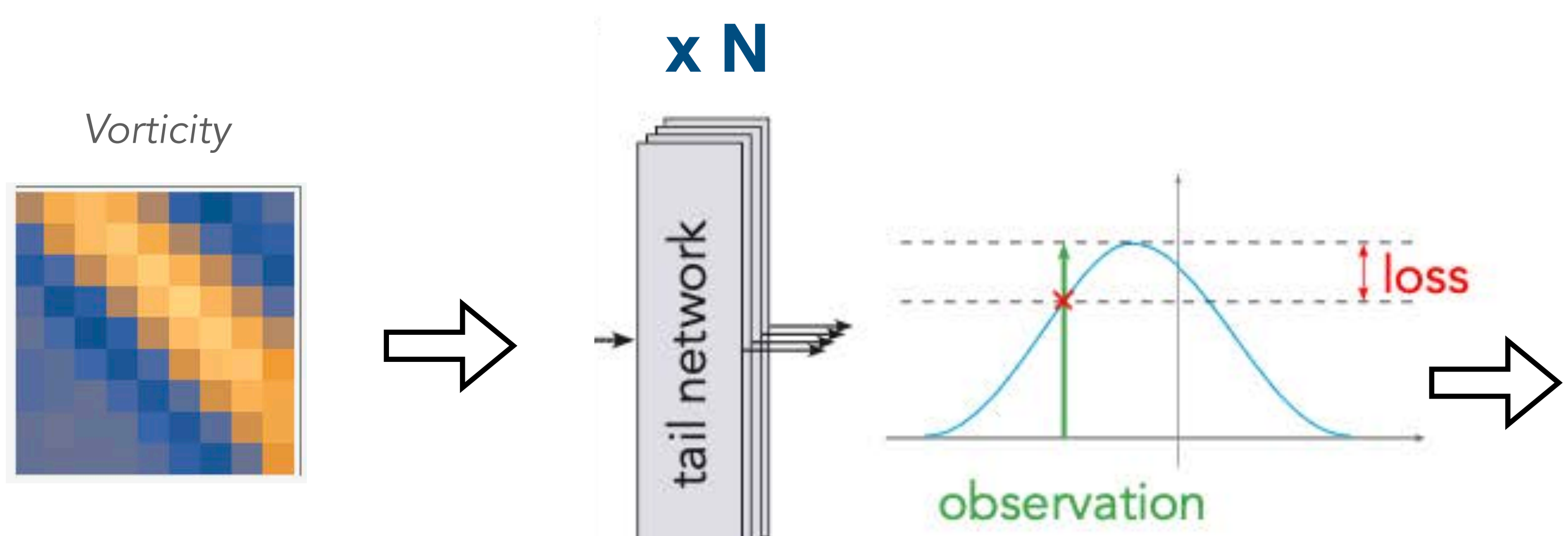


Key ingredients: statistical loss



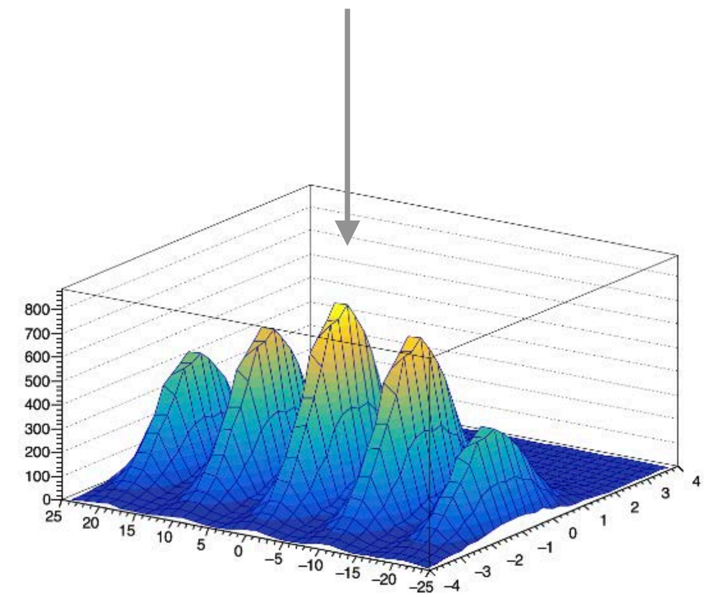
Inspired by
cross entropy loss in
classification problems





statistical interpretation: measure the difference between the pdf of the ML classification model and the predicted distribution

Ensemble of tail networks generates N predictions for each pixel



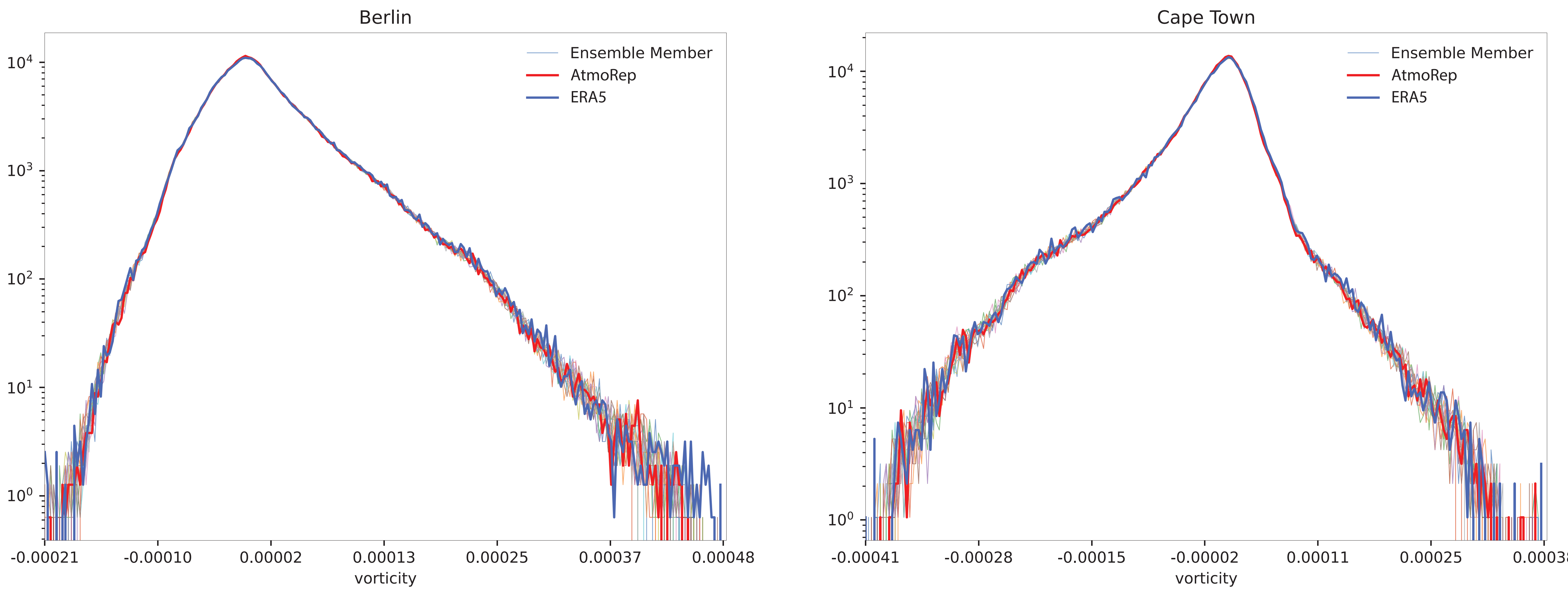
Interpret as probability distribution for each pixel
(assumed Gaussian)

Loss: Minimize the difference between the mean of the distribution and the true value

Ilaria Luise, CERN - ilaria.luise@cern.ch

Results: specific locations

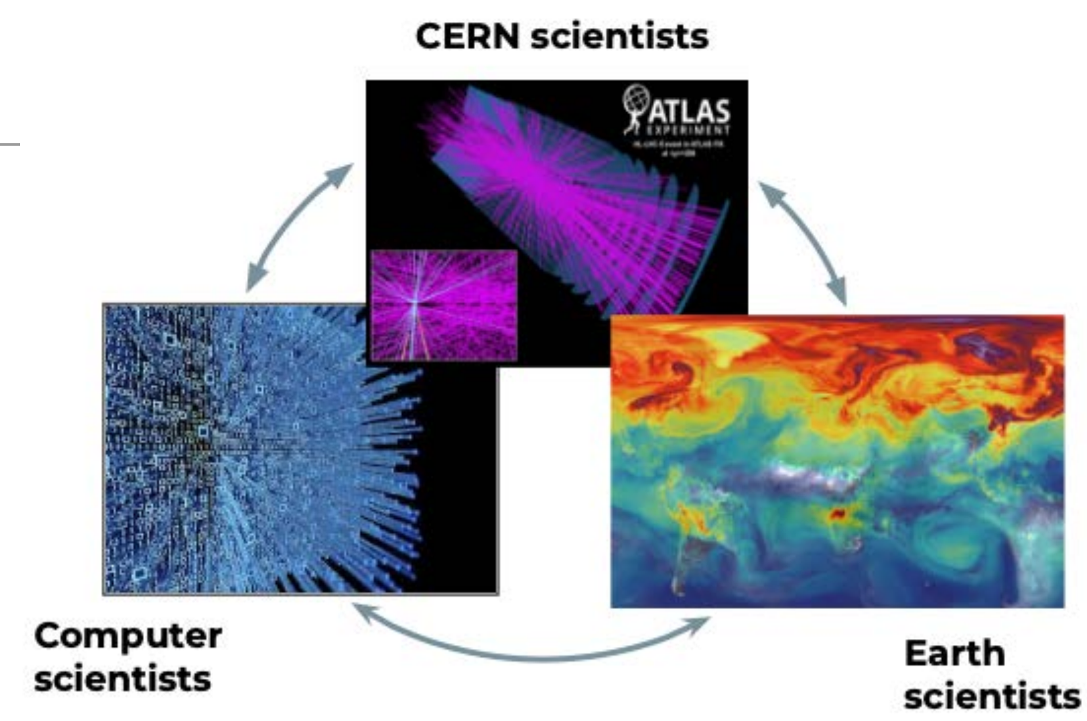
Vorticity, random samples, 2018



Ilaria Luise, CERN - ilaria.luise@cern.ch

AtmoRep: Introduction

Solve common scientific challenge(s) in high-energy physics and weather/climate science using AI/ML



Common challenges:

Explore the potential of unsupervised learning for scientific applications
Earth system science: eg. early detection of extreme events
CERN: eg. anomaly detection

Condense dataset information in a compact representation
eg. condense the info in a few GB rather than TB

→ Adapt the concept of representation learning used in LLMs to physics phenomena

Model complex, nonlinear phenomena and improve current simulations
Earth system science: eg. better understand convection phenomena
CERN: eg. particle-jet showers reconstruction

→ Physics = uncertainties
Need for a stochastic approach

Common Goal:
Use unsupervised learning to build a data-driven model that encapsulates complex physics phenomena

Ilaria Luise, CERN - ilaria.luise@cern.ch