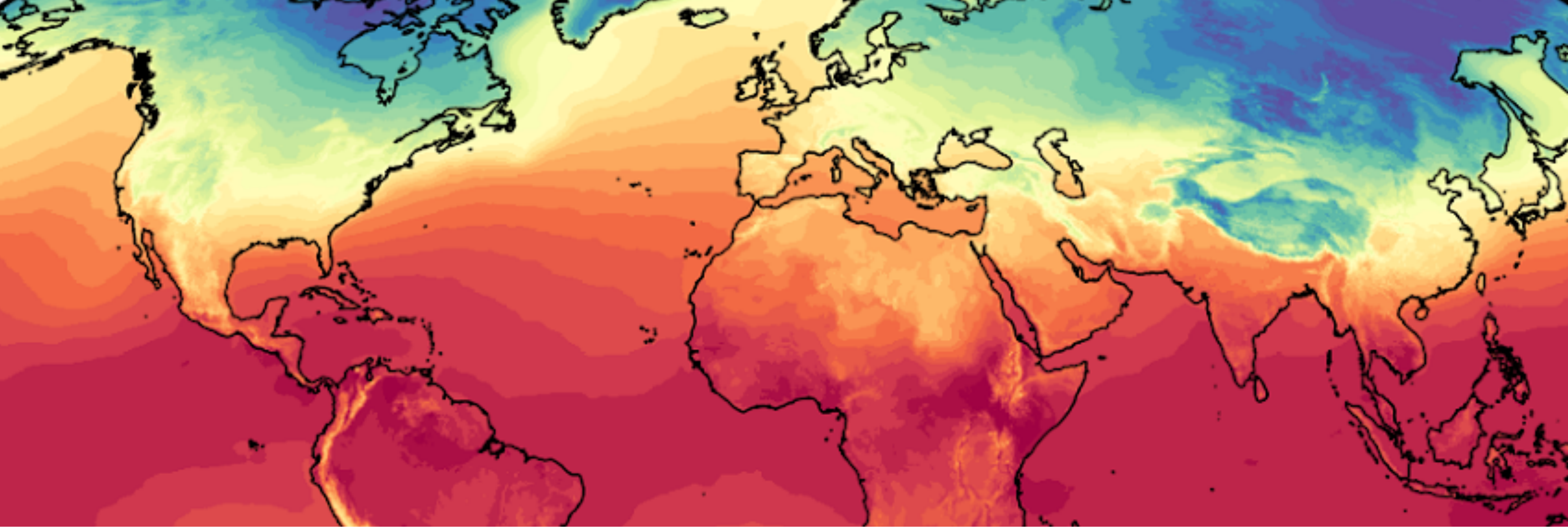


Building Foundational Models for Environmental Modelling and Prediction

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

ACAT 2024 | March 2024 - Stony Brook

CERN | CIPEA CERN Innovation Programme on Environmental Applications

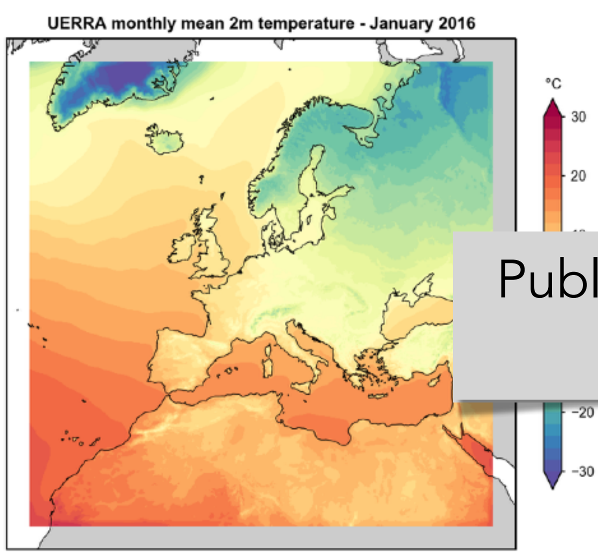
ECMWF

JÜLICH Forschungszentrum

Motivation and scientific challenge

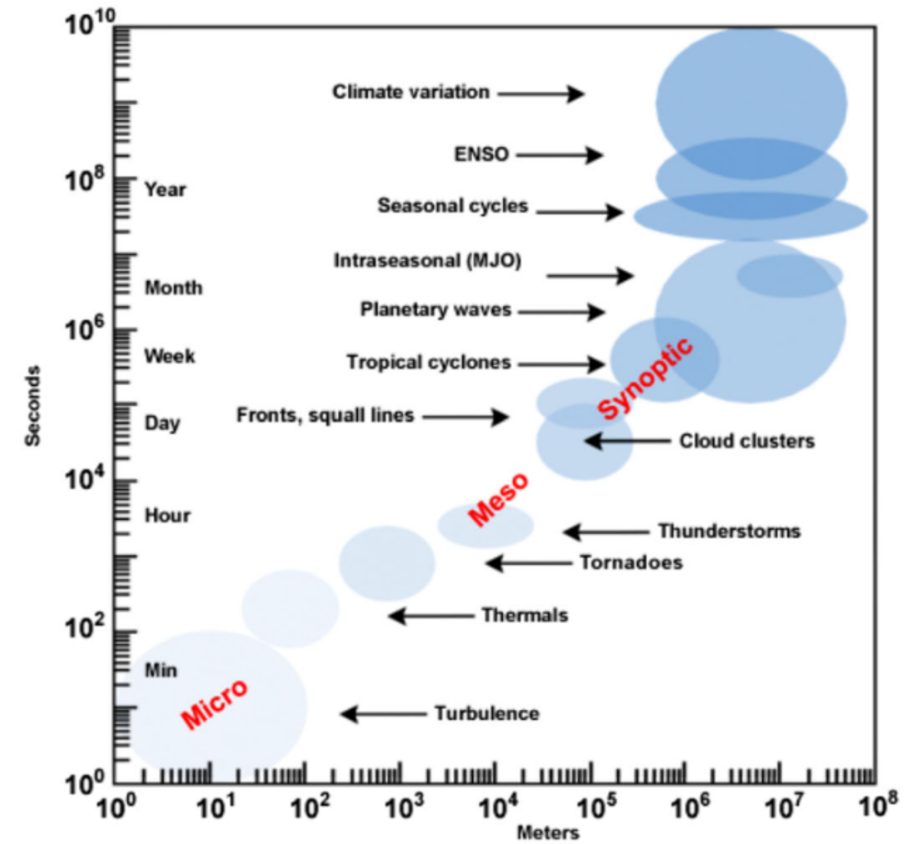
Atmosphere:

- Complex phenomena involving multiple scales
- No complete classical model to simulate the dynamics
- Very large amounts of **observational** data available



Publicly available datasets
eg. ERA5

Developments in hardware
(GPU clusters)
and software (exa-scale ML)



From V. M. Galfi, V. Lucarini, F. Ragone, and J. Wouters. Applications of large deviation theory in geophysical fluid dynamics and climate science. La Rivista del Nuovo Cimento, 44(6):291–363, 2021.

We have hundreds of TB of available atmospheric observations.

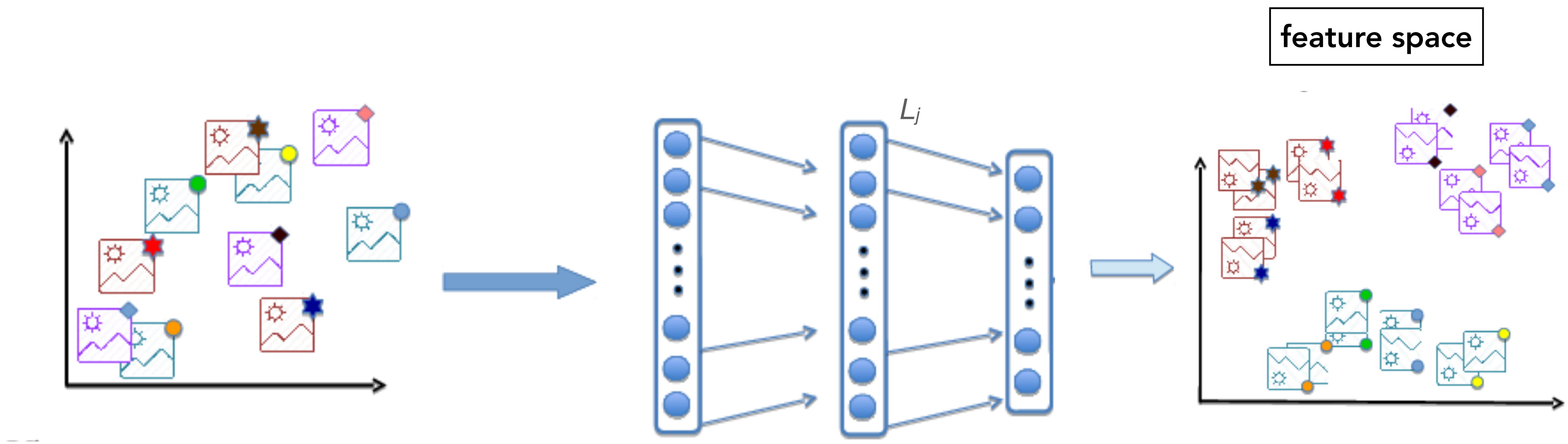
Can we use the information in these datasets for the next generation of improved weather and climate models?

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

Key ingredient: Representation learning

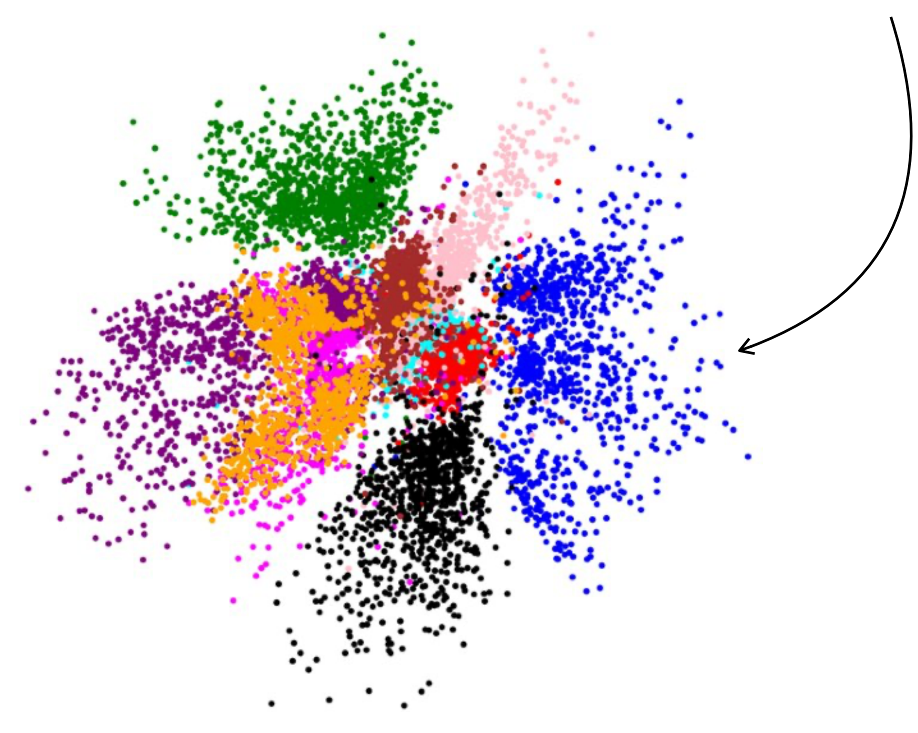
Representation learning:

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



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

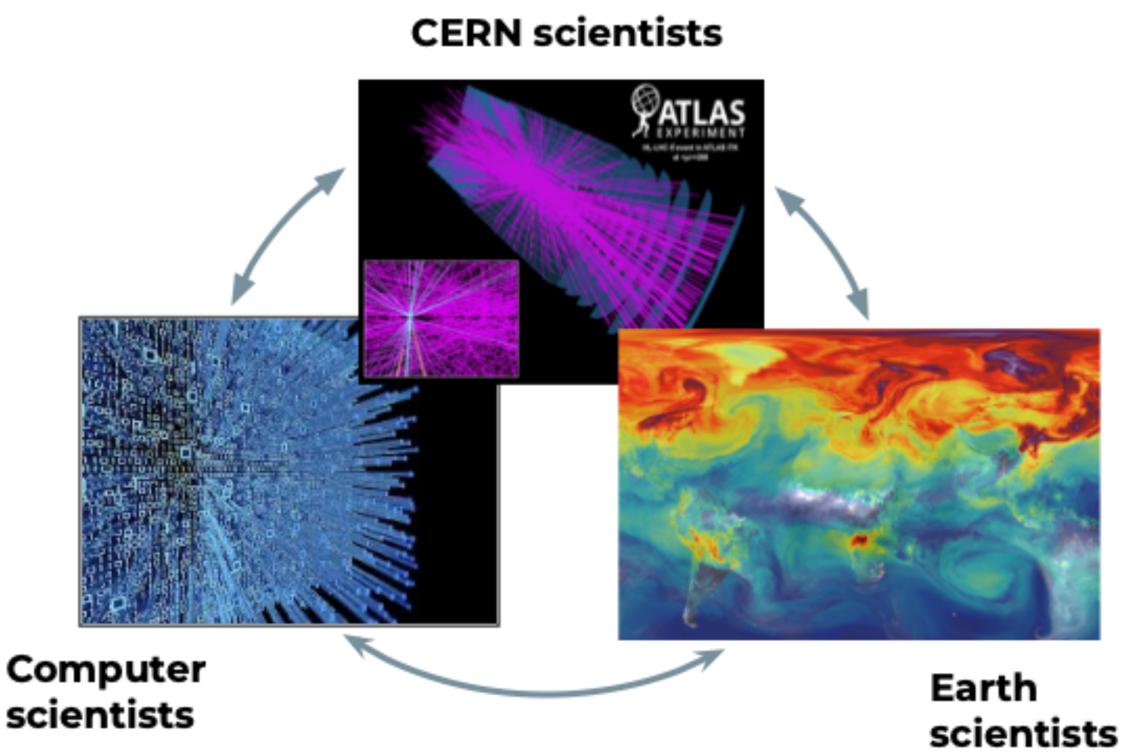
Effective "encoding" of the data useful for many applications



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

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

AtmoRep: Introduction



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

Model complex, nonlinear phenomena and improve current simulations

Access multi-scale dependencies of a given process

Earth science: eg. better understand convection phenomena

CERN: eg. particle-jet showers reconstruction

Condense dataset information in a compact representation

better handle the information in downstream applications.

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

Explore potential of unsupervised learning for scientific applications

Extract new information directly from data

eg. learn unknown correlation patterns

Earth science: eg. early detection of extreme events

CERN: eg. anomaly detection

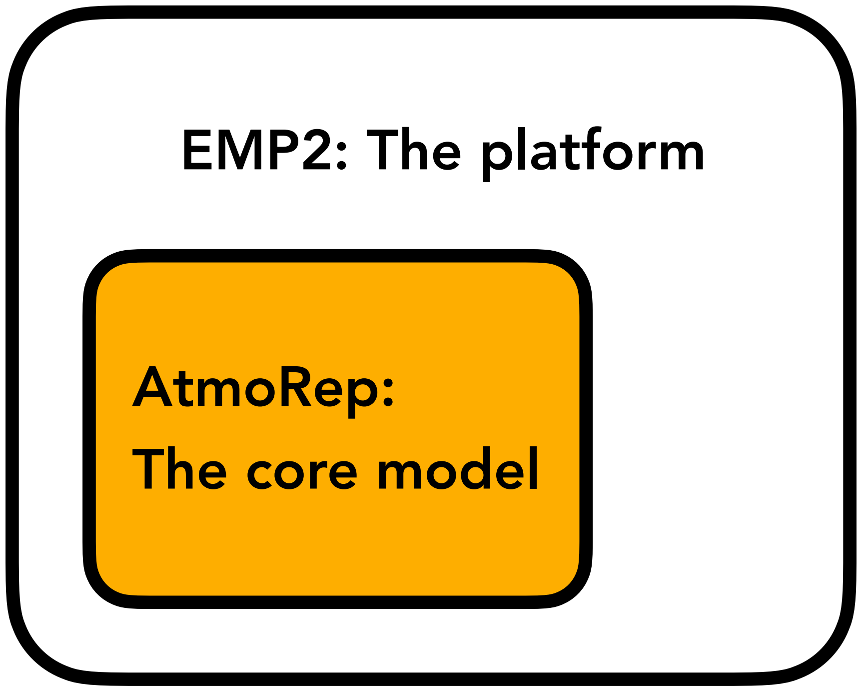
Common Goal:

Develop a proof of concept of representation learning for scientific applications based on observations

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

AtmoRep:
A foundation model for the atmosphere

The core model



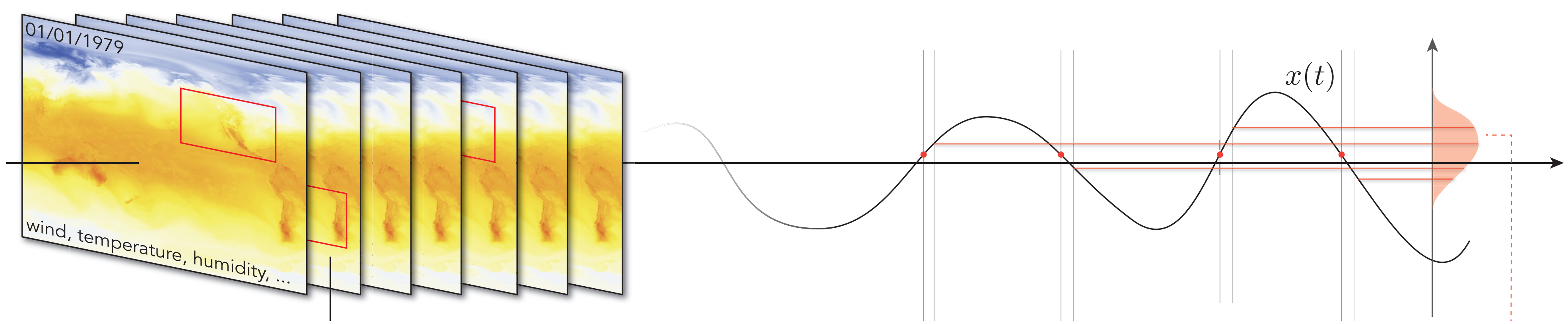
Key ingredient: What is a foundation model for us?

The spatio-temporal (4D) evolution of a dynamical system can be summarised as

Probability of getting the state y given the initial state x and the auxiliary info α

$$p(y \mid x, \alpha)$$

Auxiliary info: position, absolute time etc..



Training

The distribution can be approximated by a large neural network

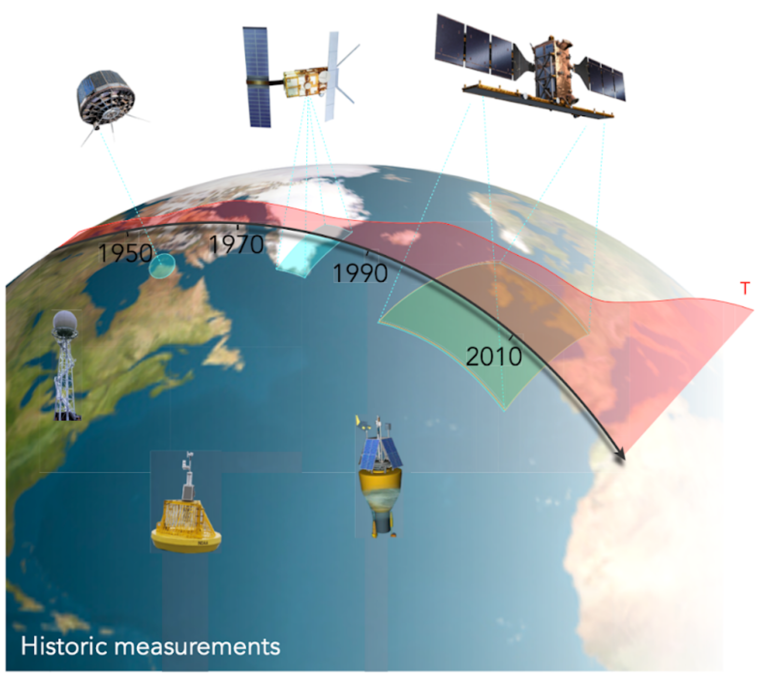
$$p(y \mid x, \alpha) \approx p_\theta(y \mid x, \alpha)$$

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

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

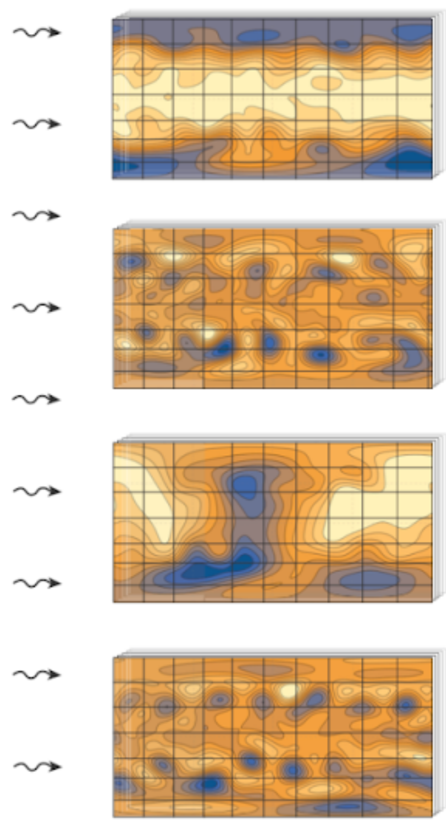
The project in a nutshell

A machine-learning based global environmental model trained on terabytes of observational data



Observations

ERA5 reanalysis

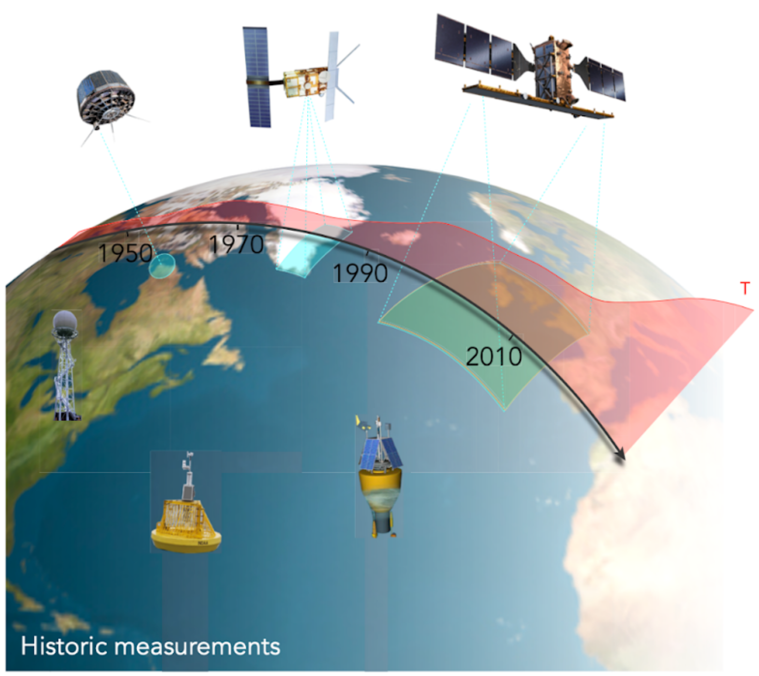


50 TB

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

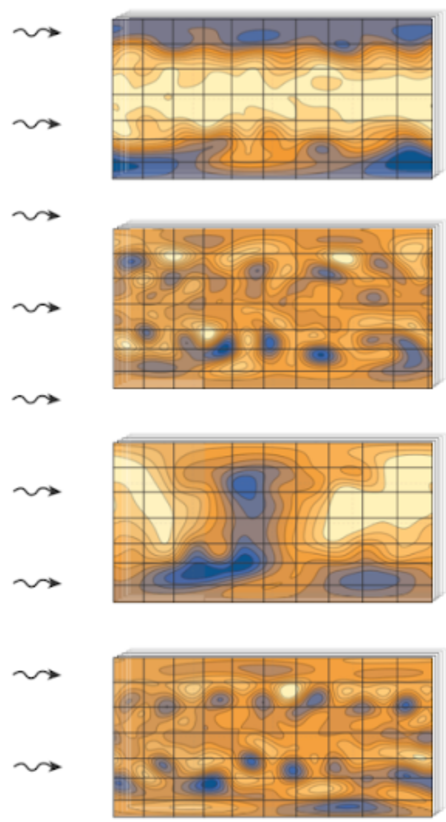
The project in a nutshell

A machine-learning based global environmental model trained on terabytes of observational data

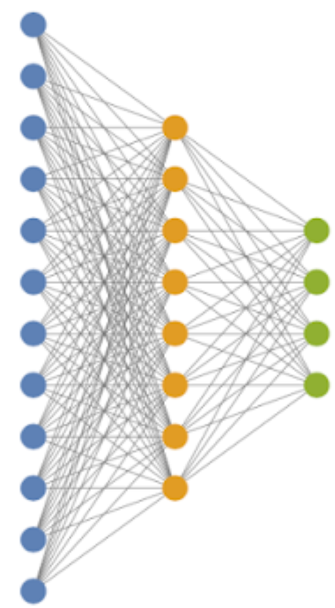


Observations

ERA5 reanalysis



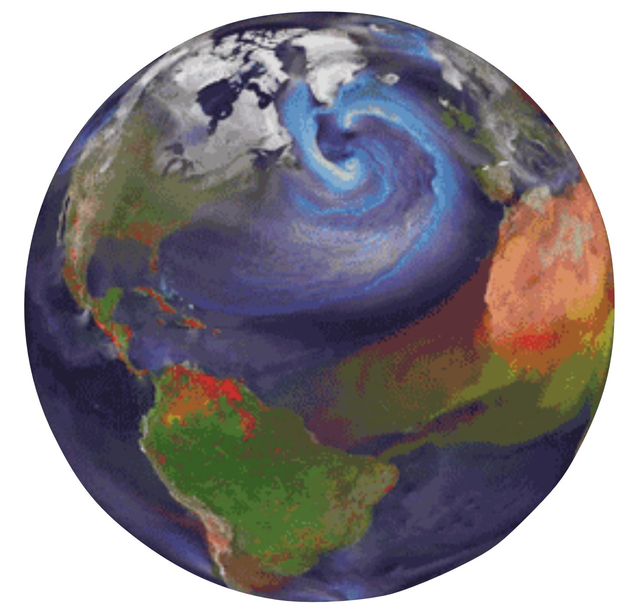
50 TB



large scale machine learning

Transformers architecture

Spatio-temporal representation of atmospheric dynamics



Model
given by the trained neural network

R&D at Juelich SSC:
4×10^6 GPU hours granted in 2023

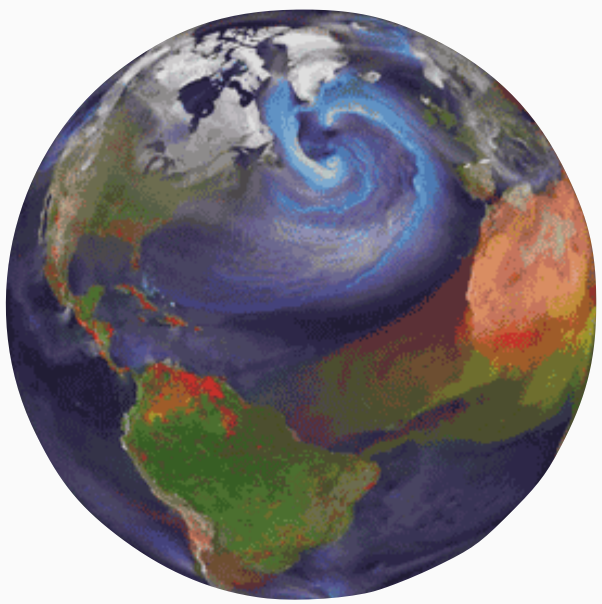
JÜLICH Forschungszentrum

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

Applications: one model for multiple purposes

Use the learned representation to improve the state-of-the-art of specific weather & climate-related scientific applications

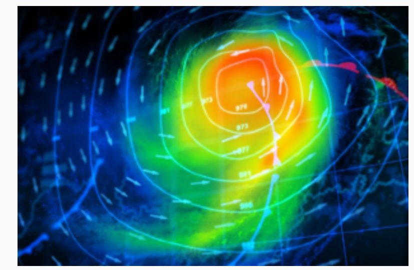
Spatio-temporal representation of atmospheric dynamics



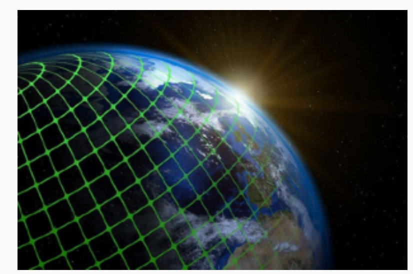
Model
given by the trained neural network

Task dependent Adaptable smaller networks

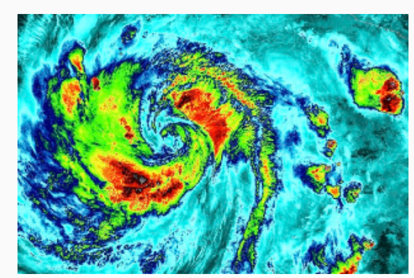
Adaptation



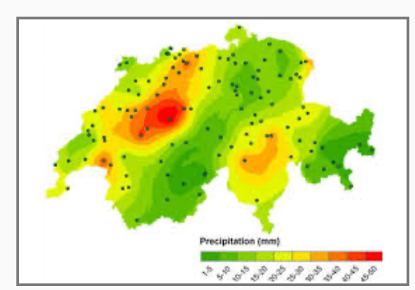
Weather predictions ✓



Downscaling ✓



Bias corrections ✓



Spatio-temporal Interpolations (WIP)

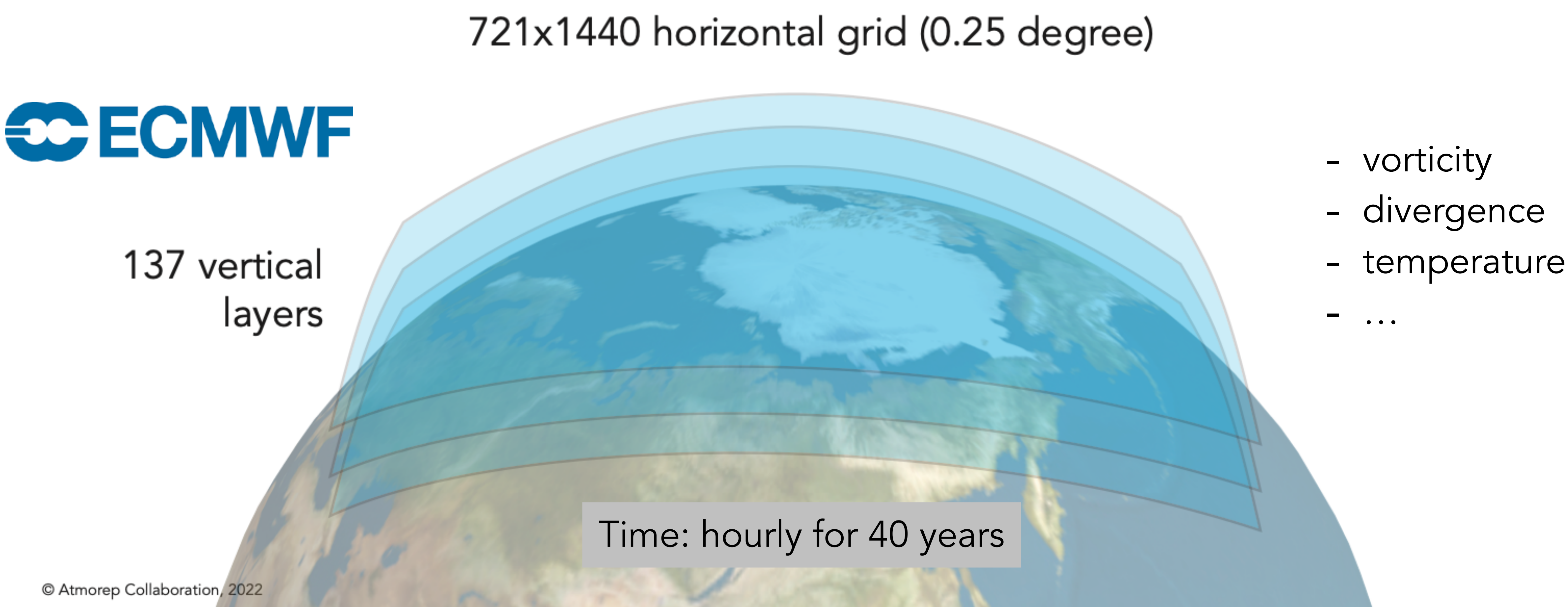
Ilaria Luise, CERN - ilaria.luise@cern.ch

The dataset

Publicly available pre-processed dataset of hourly spaced interpolated Earth observations: The ERA5 reanalysis from ECMWF

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

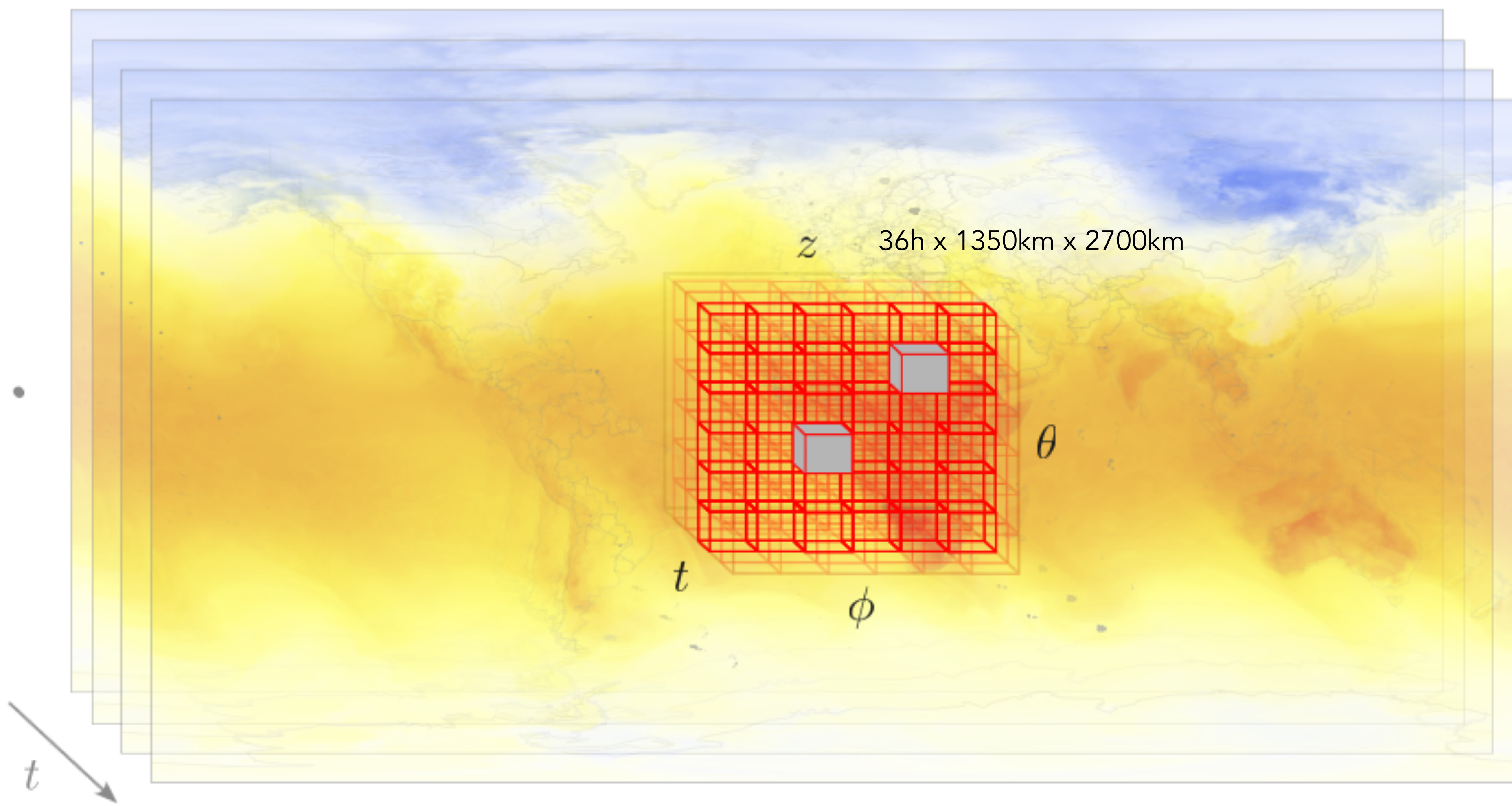


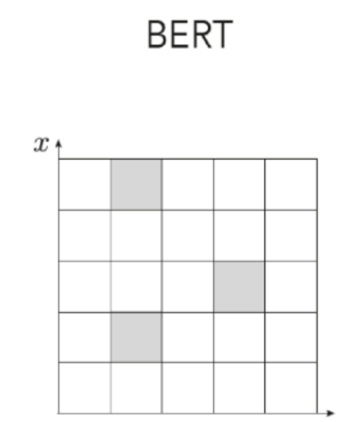
Ilaria Luise, CERN - ilaria.luise@cern.ch

Intermezzo: The training protocol

Use a variation of BERT masked language model from self supervised trainings in NLP

Random sampling of neighbourhoods for training → stochastic gradient descent





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

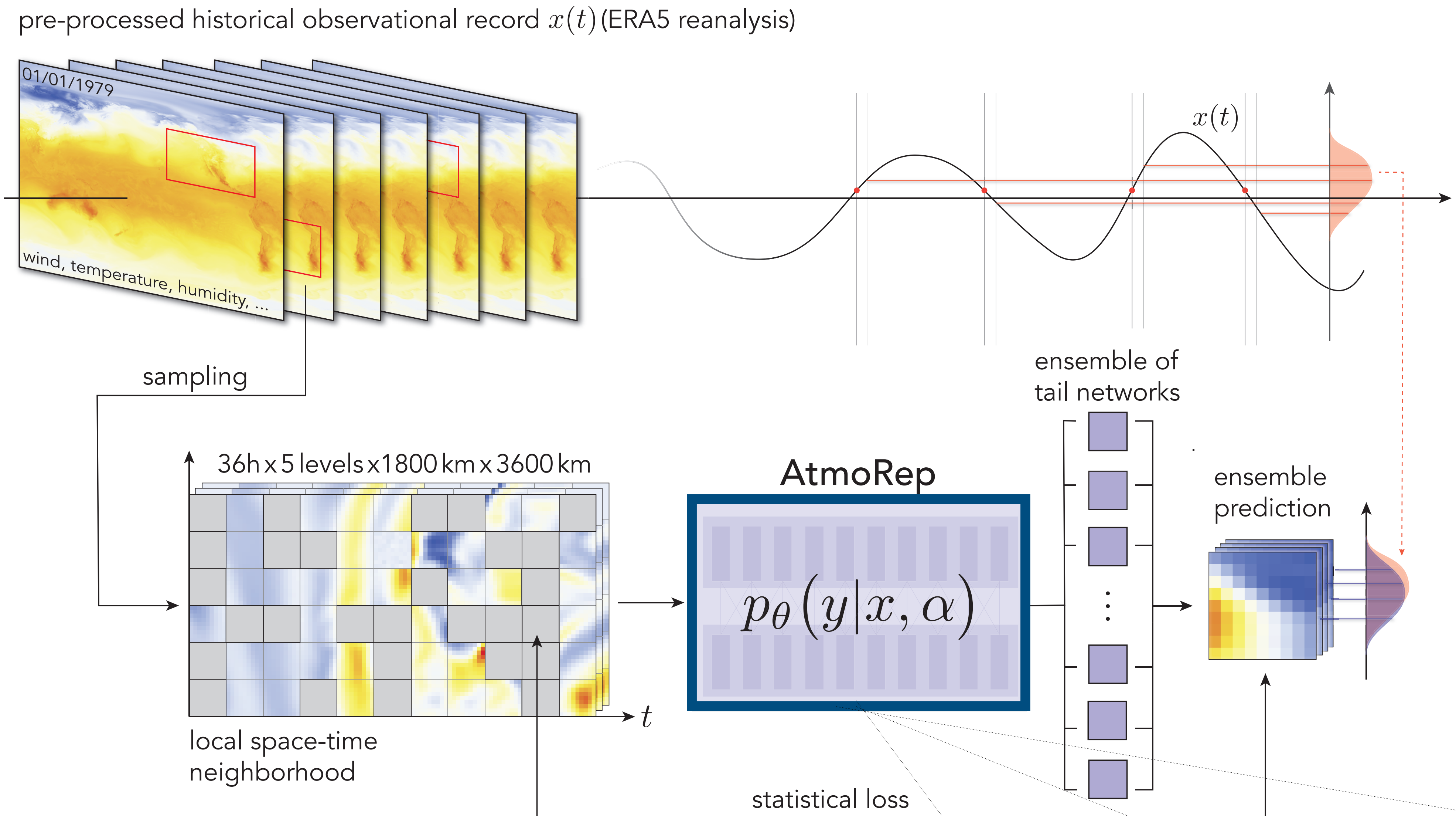
Mask random tokens within the hyper-cube and try to predict them back

visually: learn representation dynamics through interpolation

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

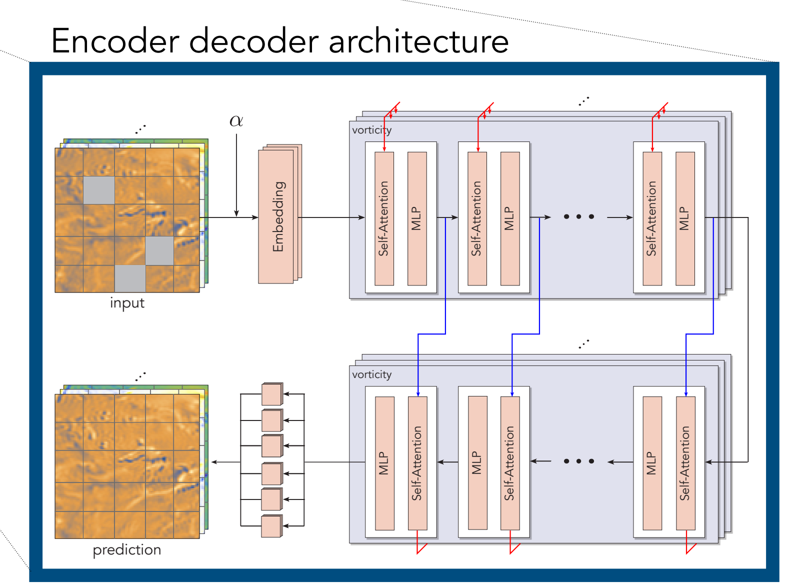
Ilaria Luise, CERN - ilaria.luise@cern.ch

The network architecture



New stochastic approach

ensemble predictions with 16 members



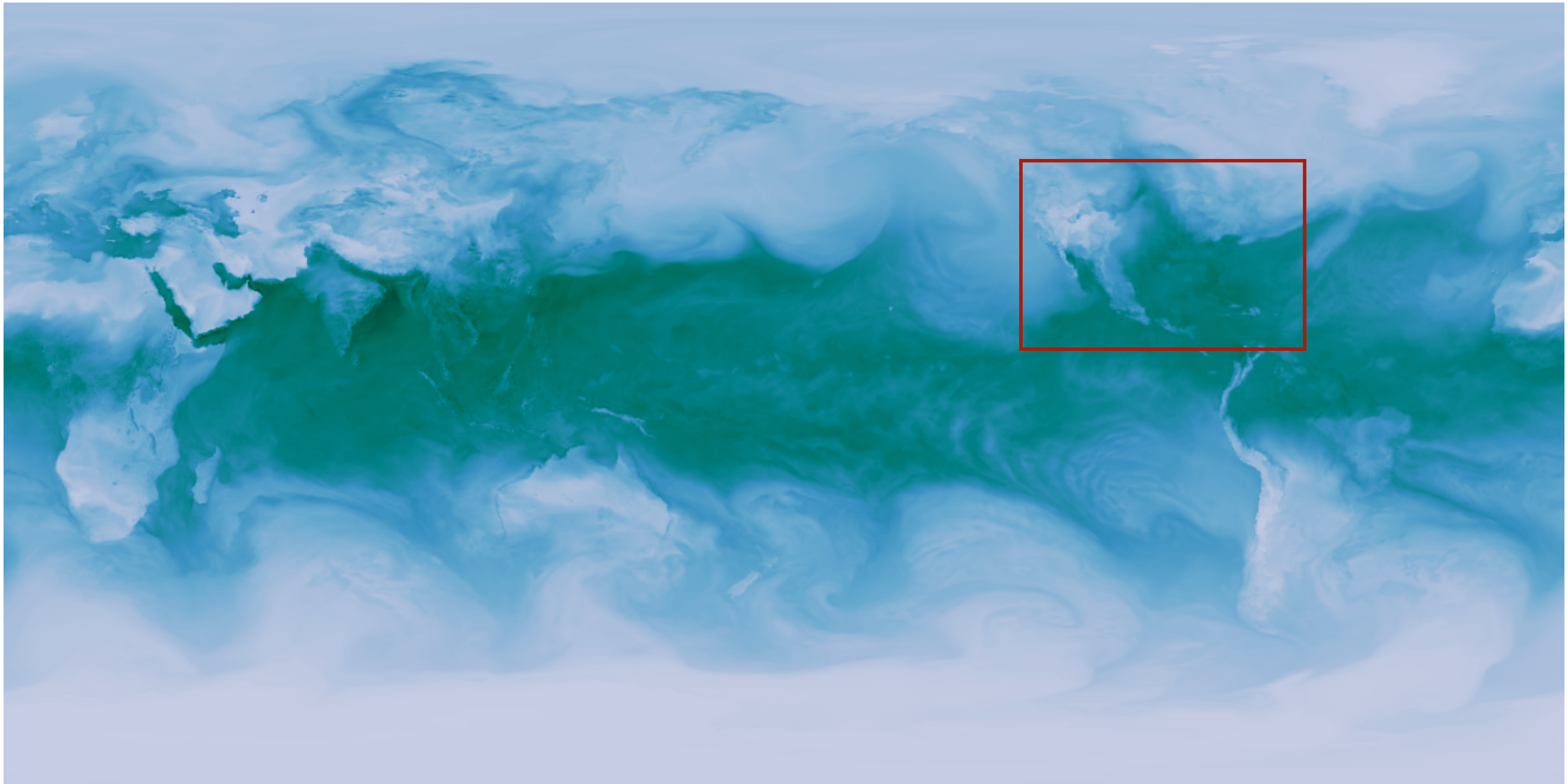
Approximate the 4-Dim PDF of the process using a Transformers-based network with 3.5 billion parameters

Short term weather forecasting

Zero-shot applications

Results: Target - ERA5

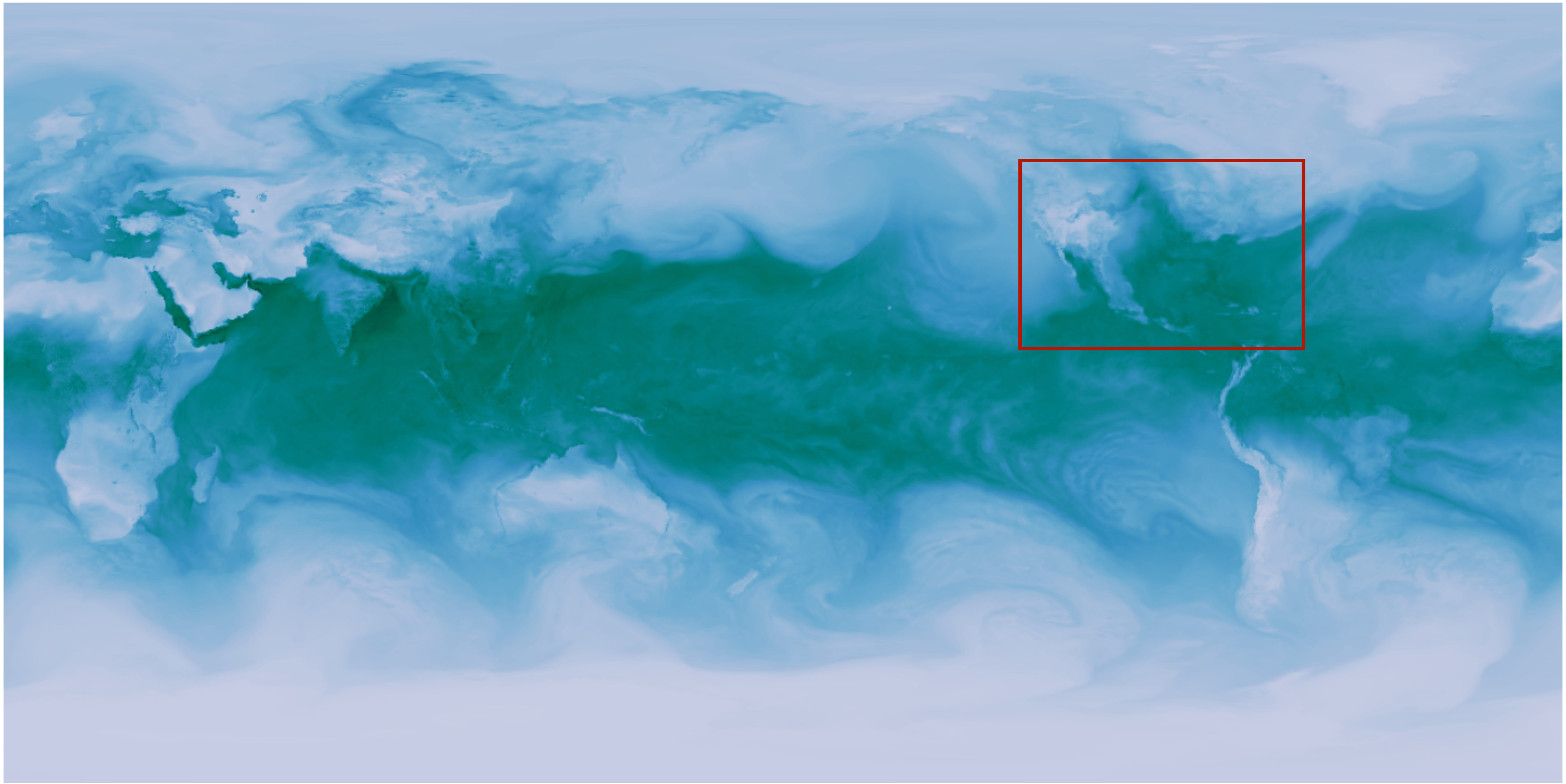
specific humidity, June 15th 2018 13:00 UTC



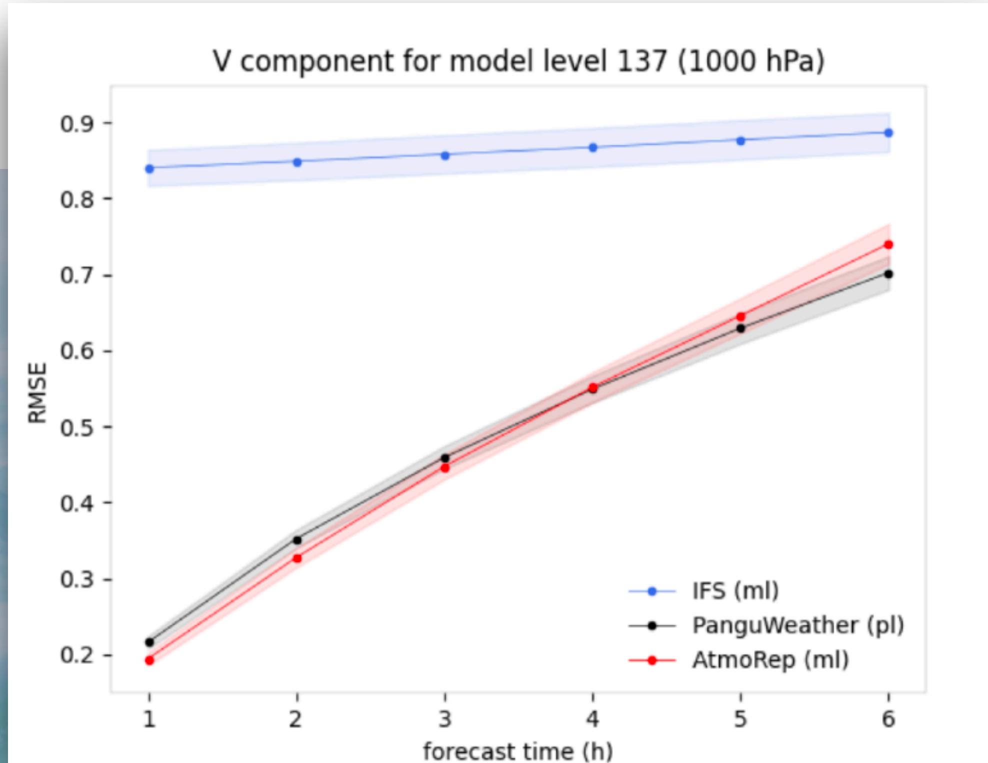
Ilaria Luise, CERN - ilaria.luise@cern.ch

Results: Prediction - AtmoRep

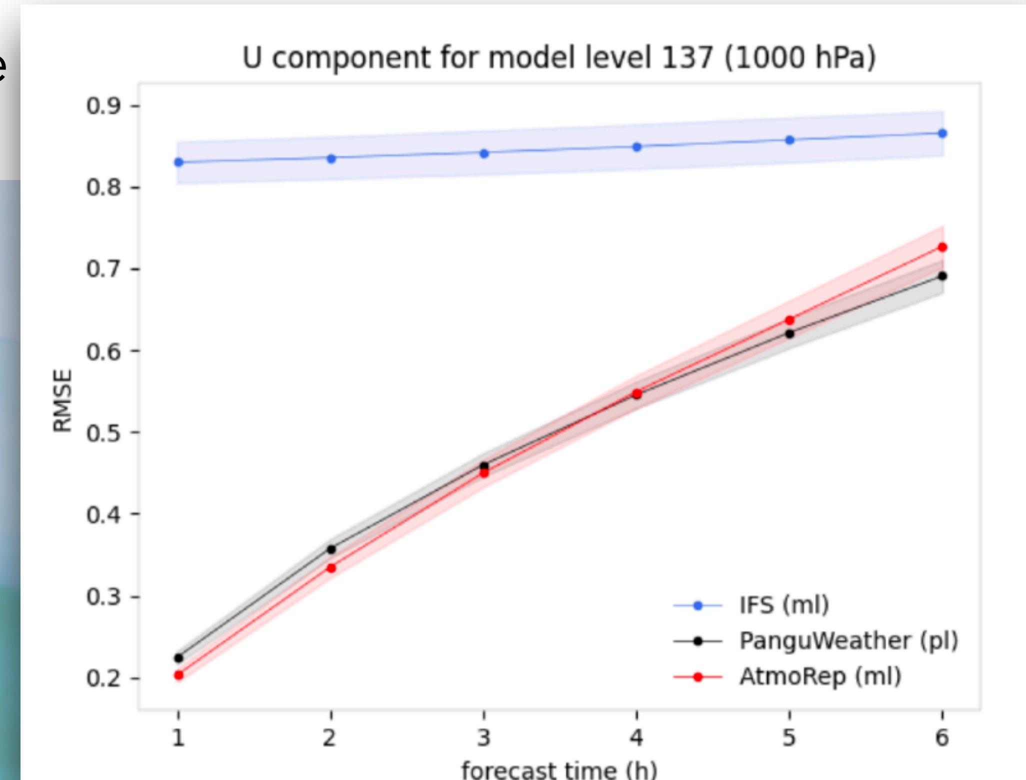
specific humidity, June 15th 2018 13:00 UTC

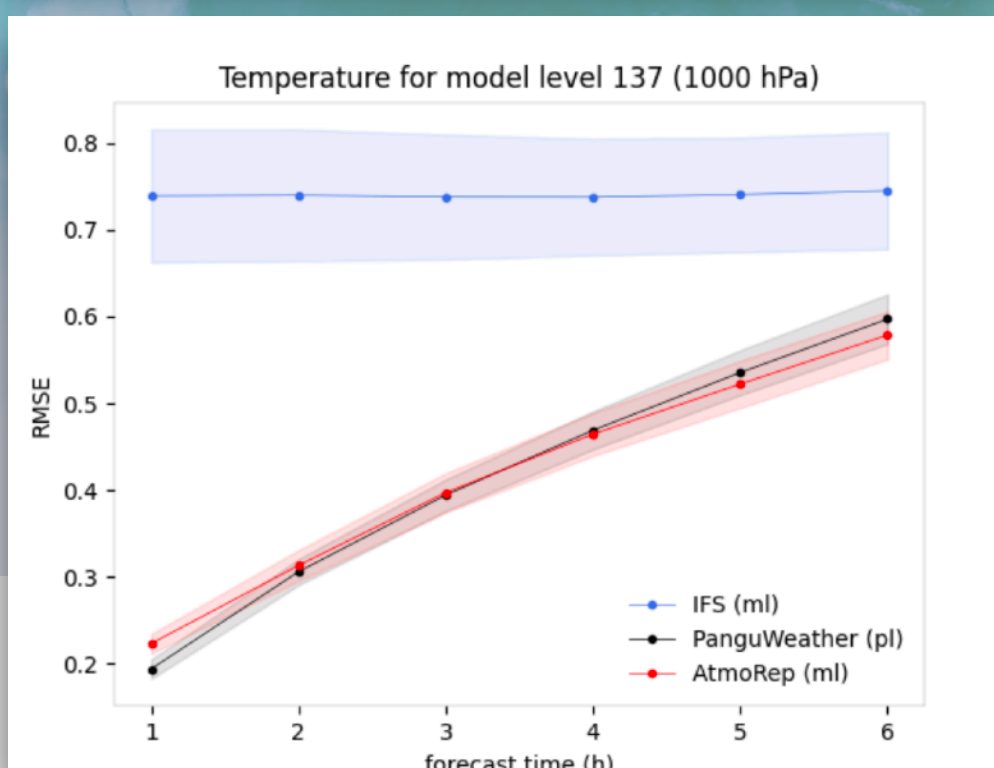


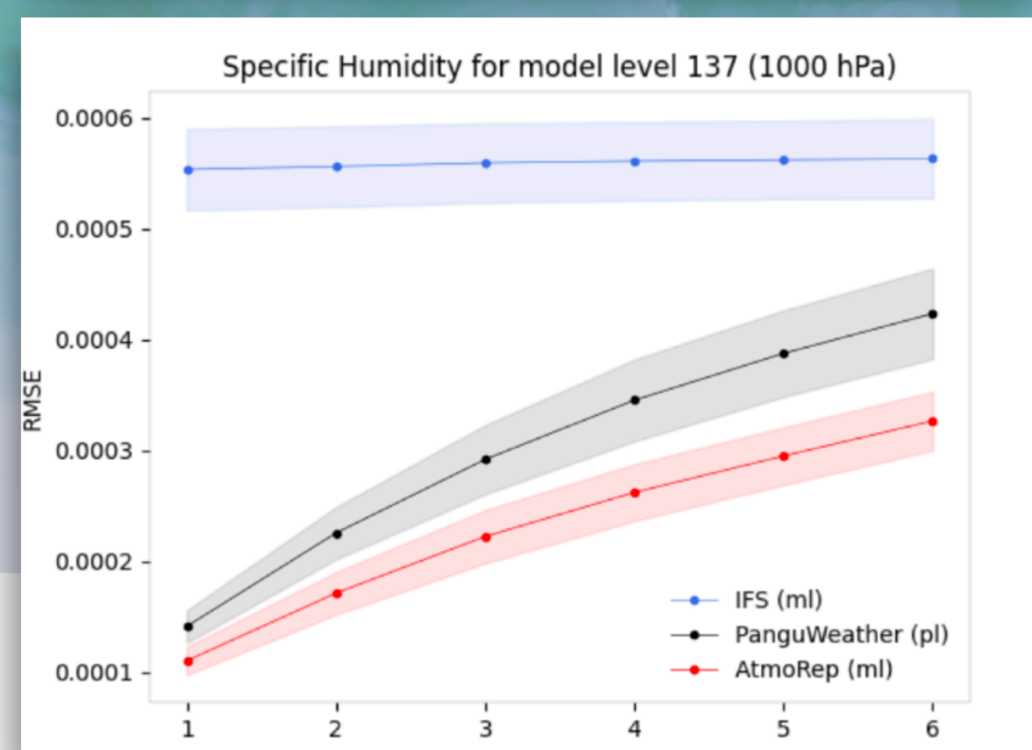
Results: Prediction - AtmoRep



y, June

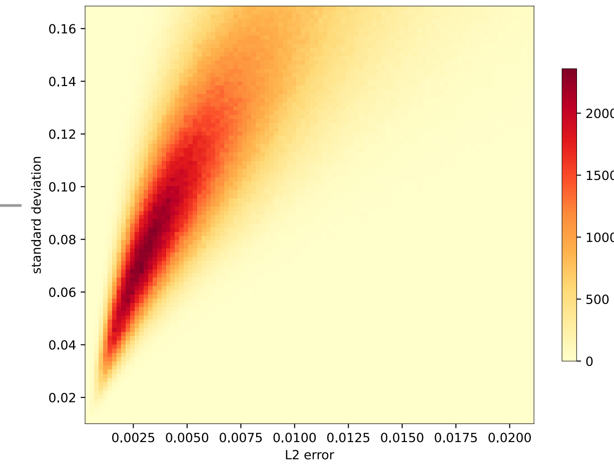




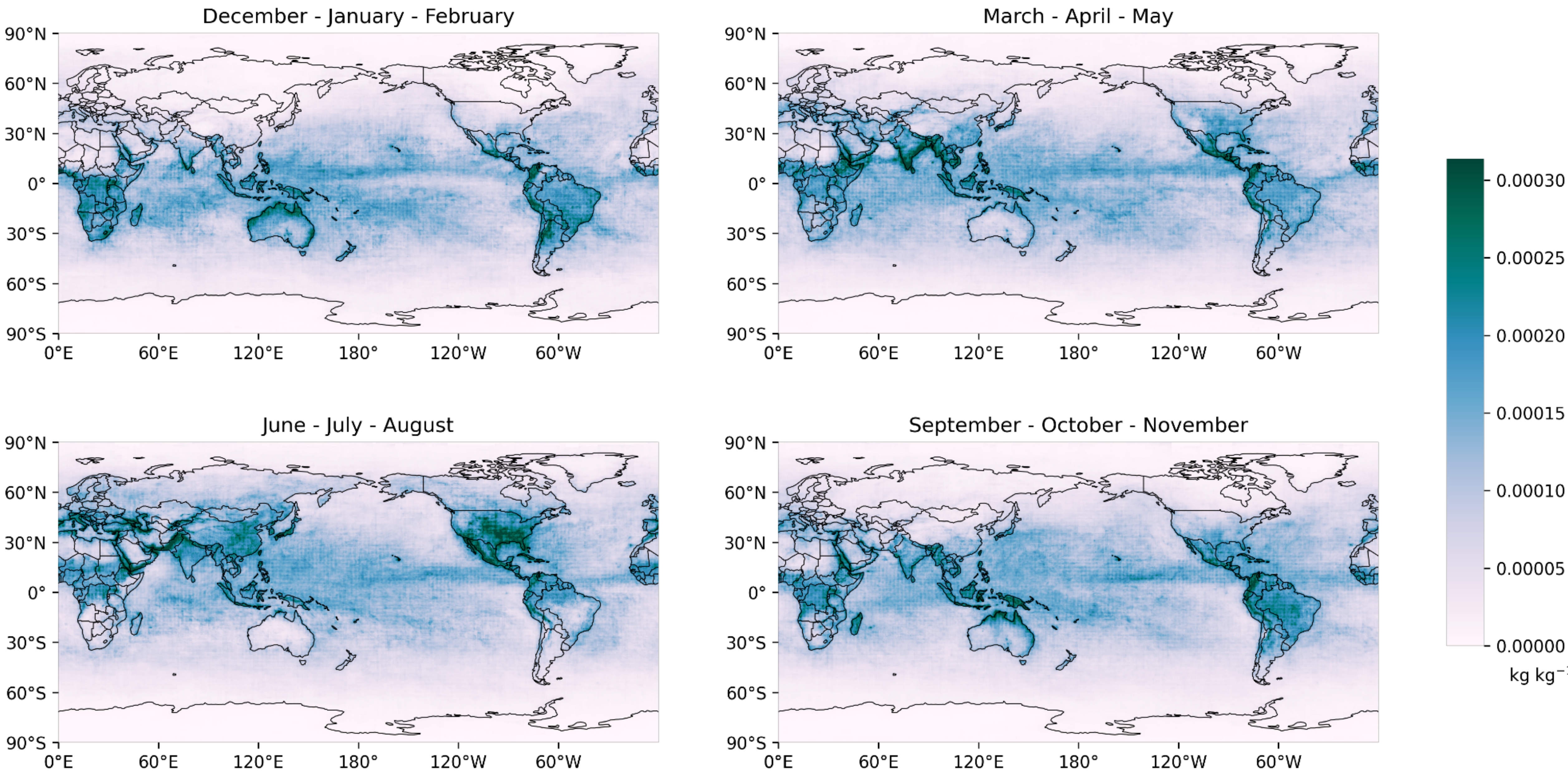


Ilaria Luise, CERN - ilaria.luise@cern.ch

Results: ensemble variability



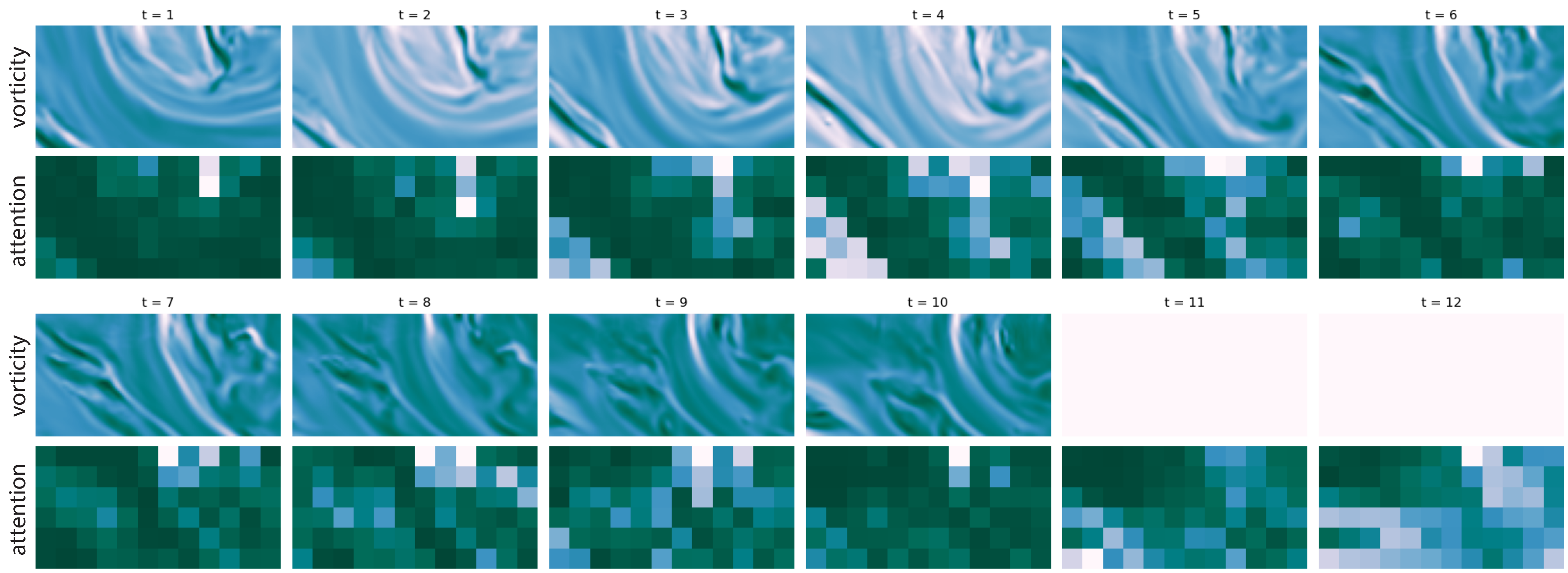
specific humidity: standard deviation of the ensemble



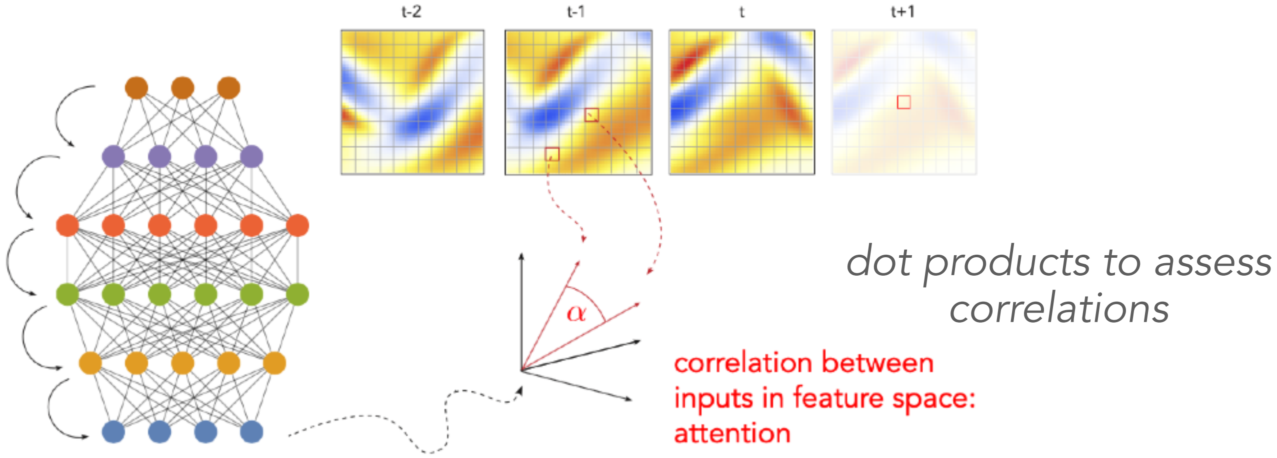
Ilaria Luise, CERN - ilaria.luise@cern.ch

Attention maps and interpretability

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



Attention:

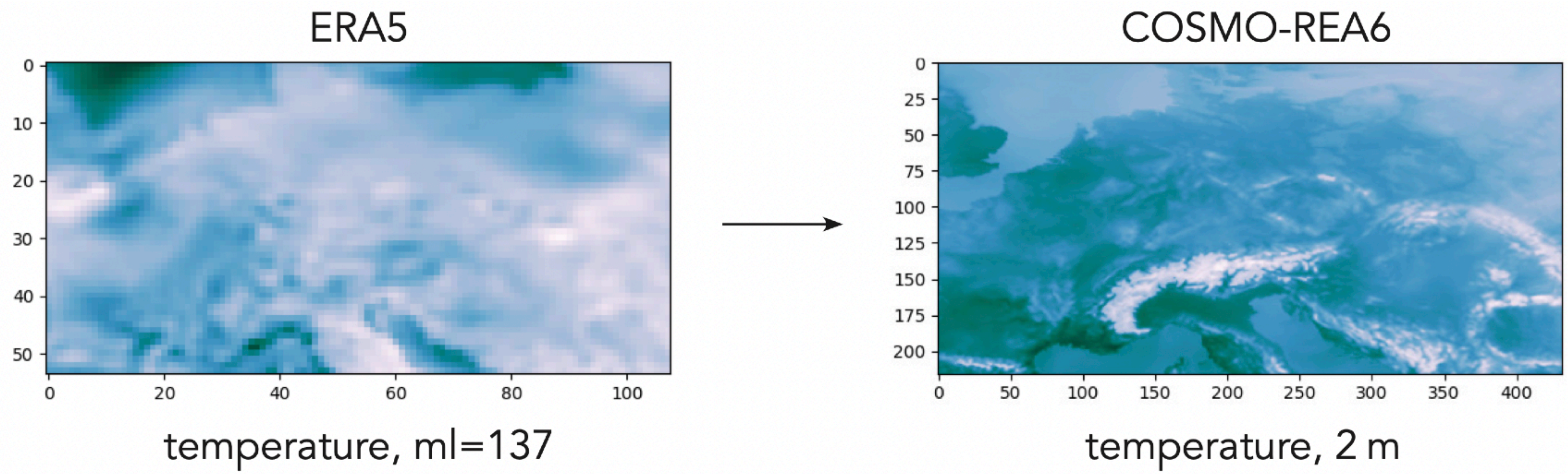


Ilaria Luise, CERN - ilaria.luise@cern.ch

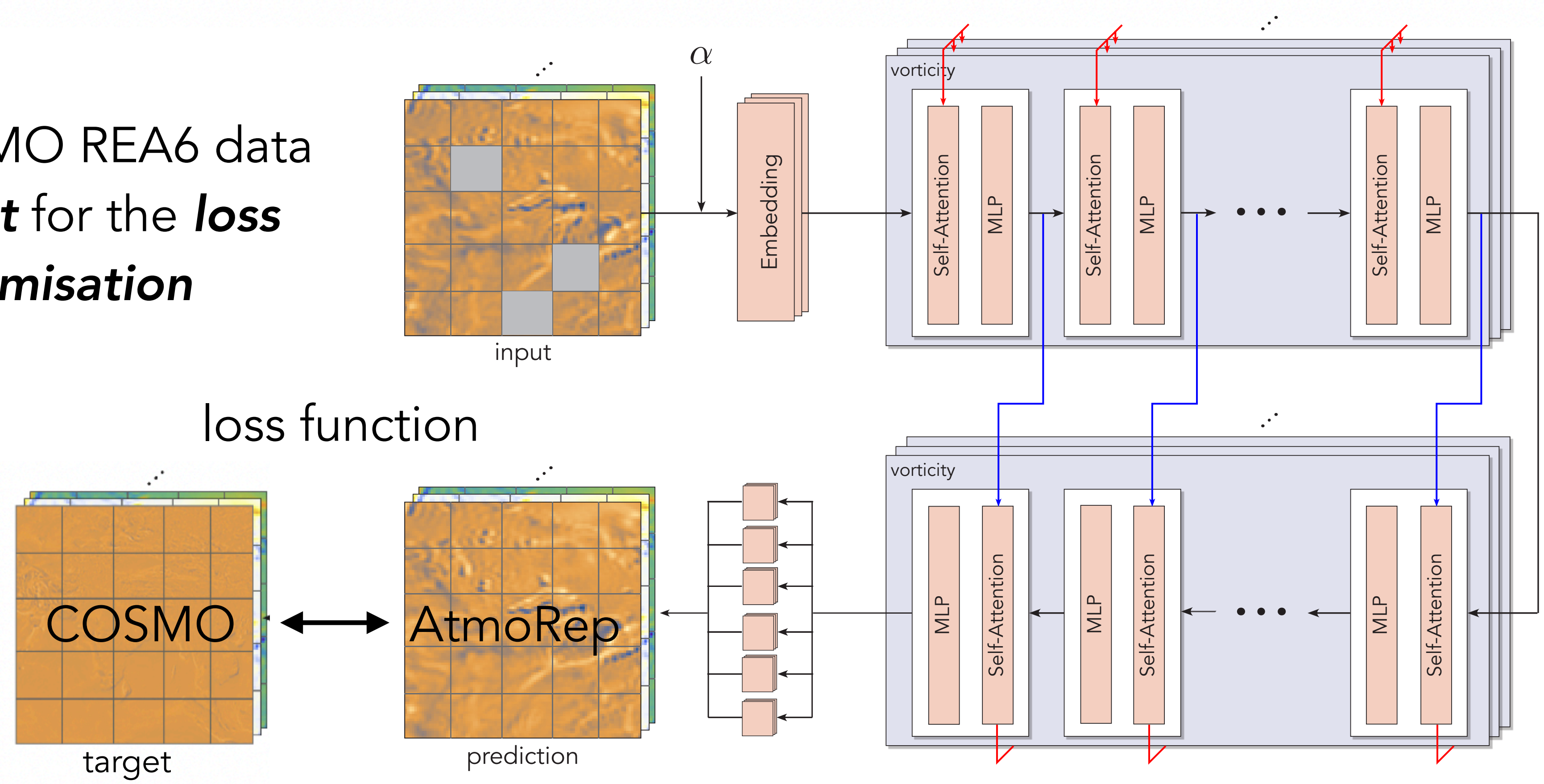
Fine tuning on real data

Data driven precipitation corrections & downscaling

Downscaling



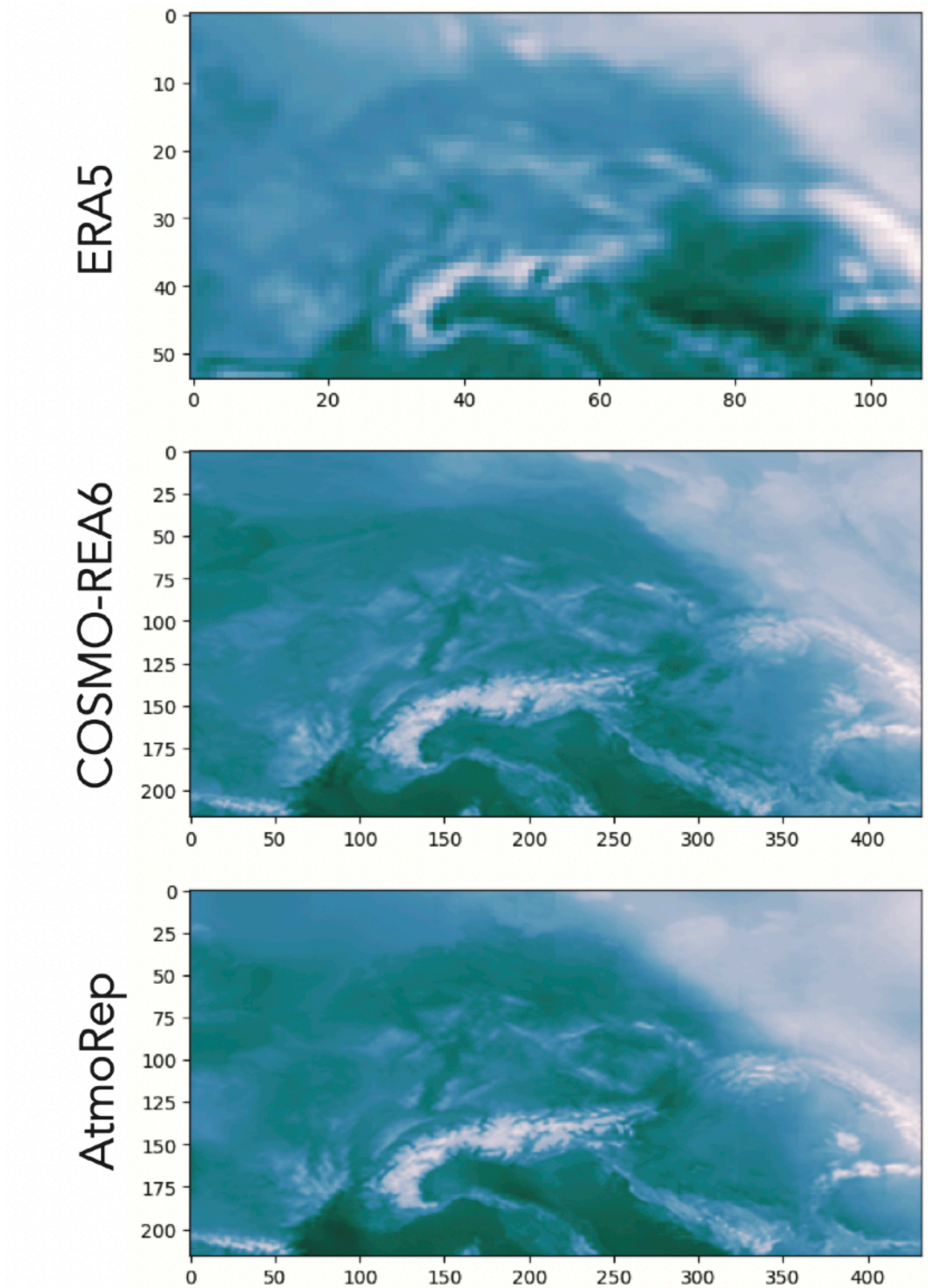
Use COSMO REA6 data as ***target*** for the ***loss minimisation***



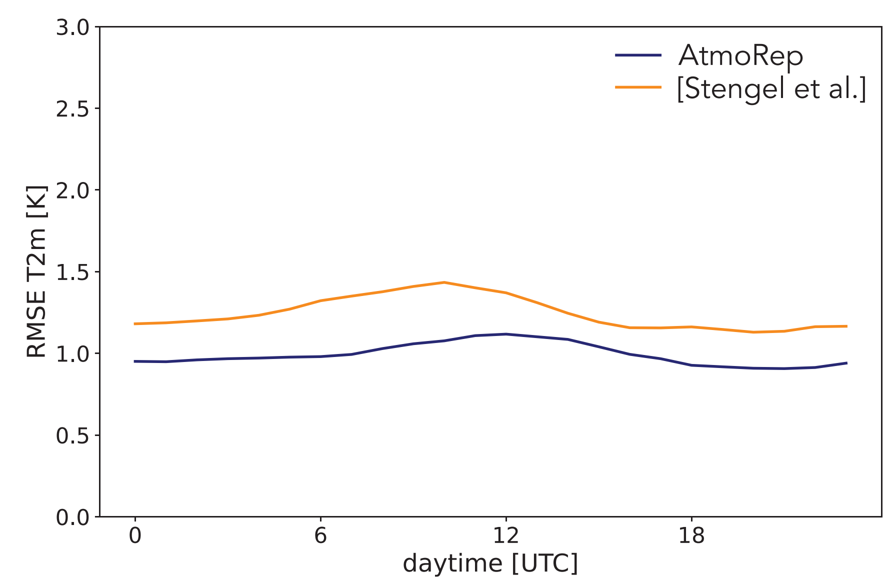
Ilaria Luise, CERN - ilaria.luise@cern.ch

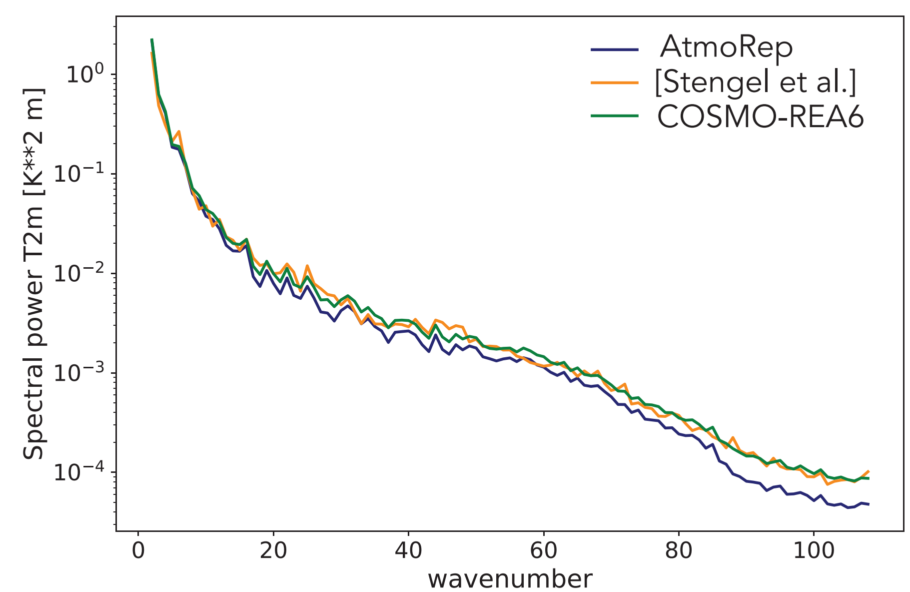
Downscaling

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



Comparison with a competing
AI-based model for downscaling:

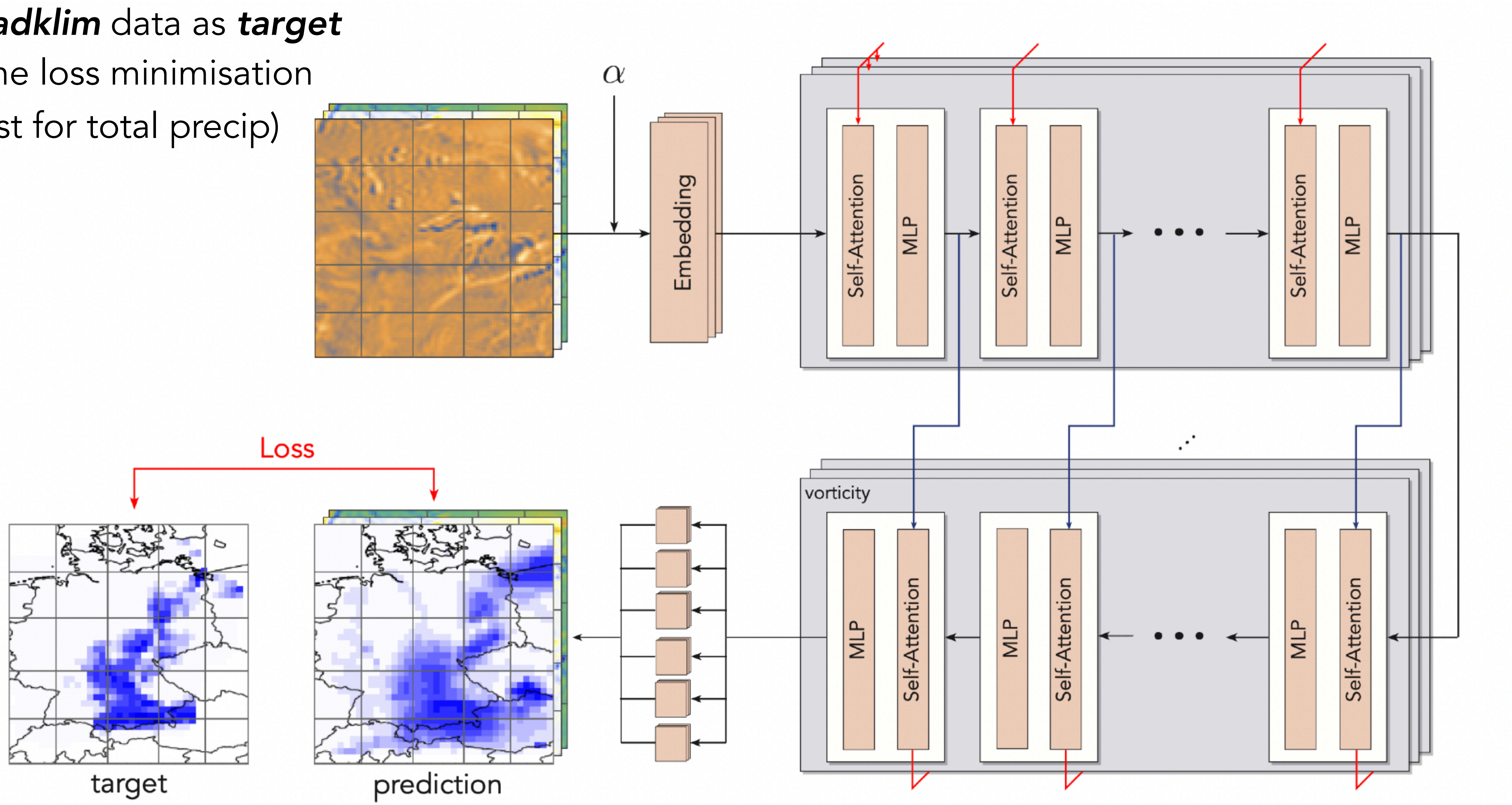




Bias corrections

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)



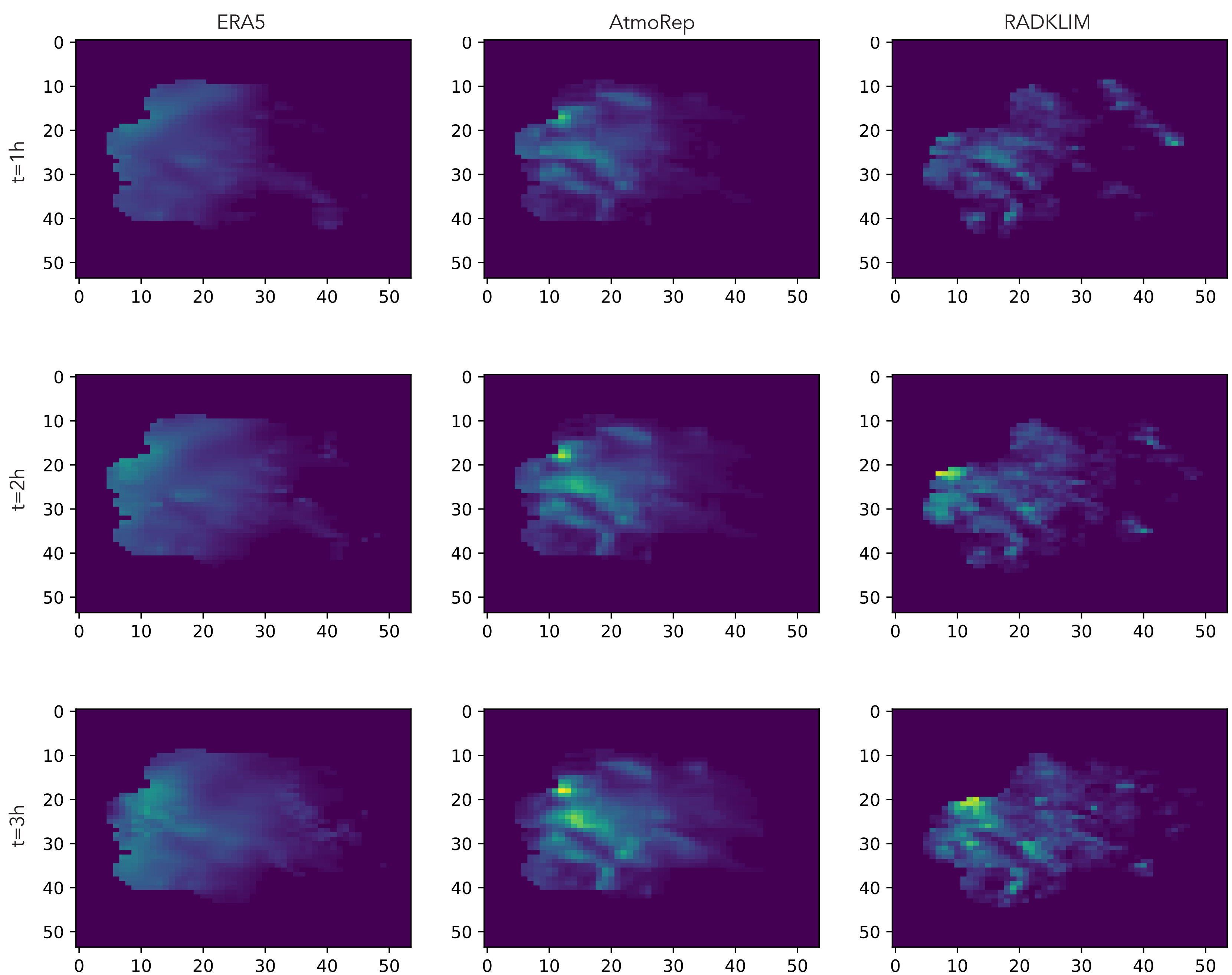
Radklim

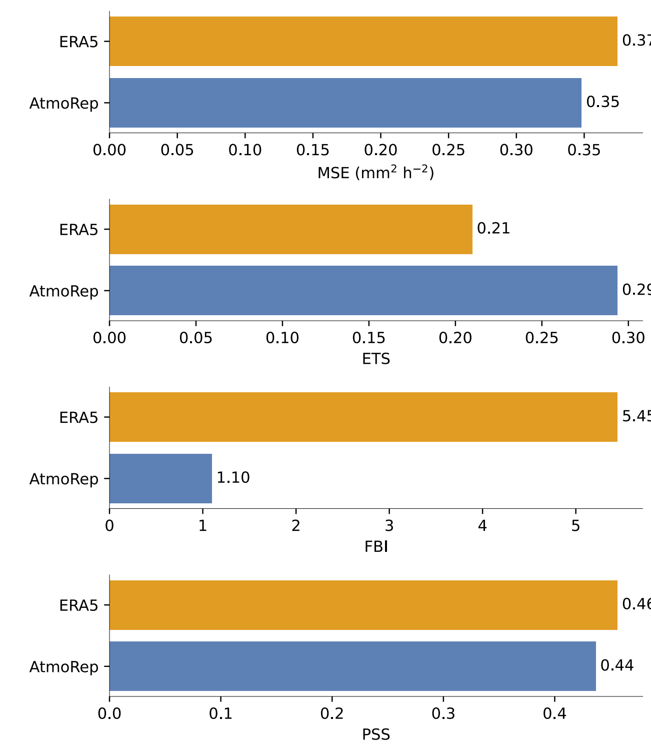
AtmoRep

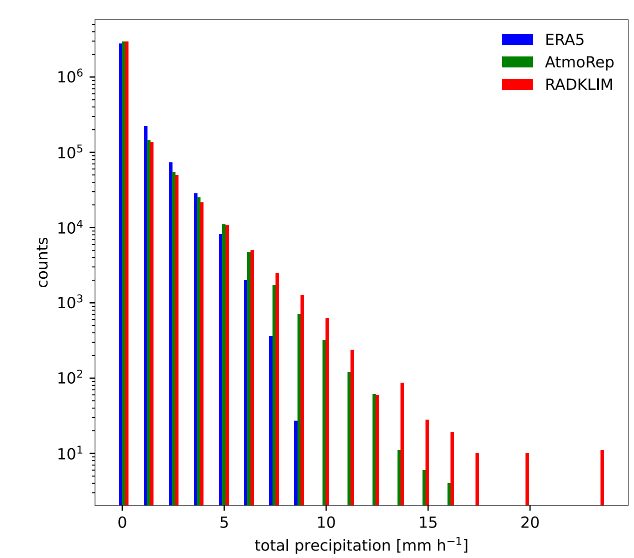
Ilaria Luise, CERN - ilaria.luise@cern.ch

Bias corrections: Results

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



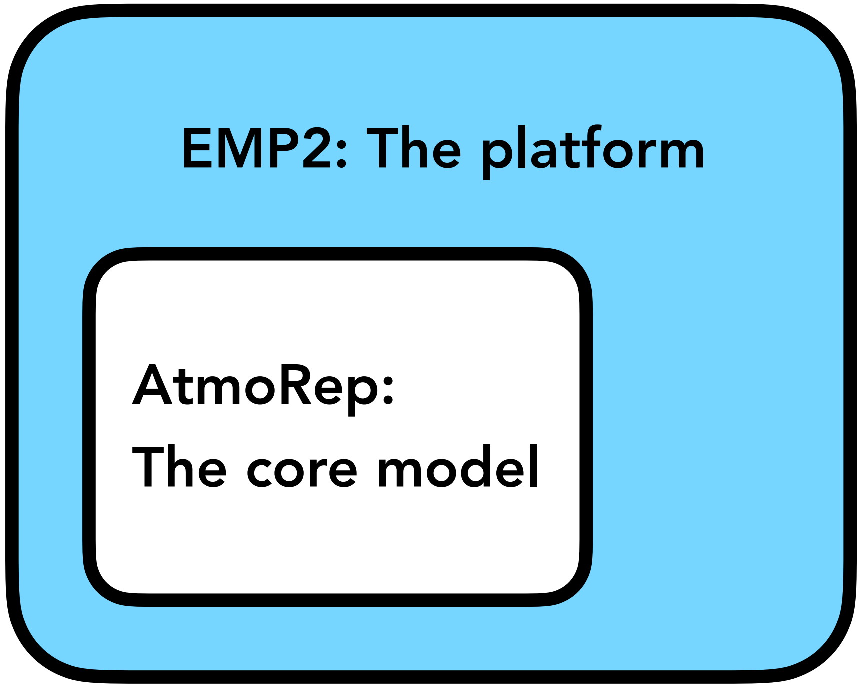




Ilaria Luise, CERN - ilaria.luise@cern.ch

Environmental modelling and prediction platform

The user-friendly platform

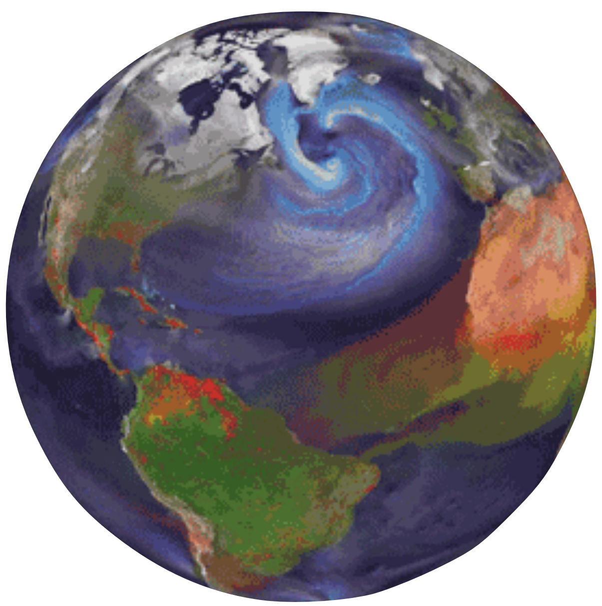


Generalisation from HPC centres to clouds

Future: develop the API & the user interface

Challenges: How foundation models interface with Digital Twin architectures?
Can we use foundation models as backbone for several digital twins
(e.g. core model = common atmospheric representation)?

Challenging part:
Close collaboration with the members of InterTwin & CS4OD projects at CERN

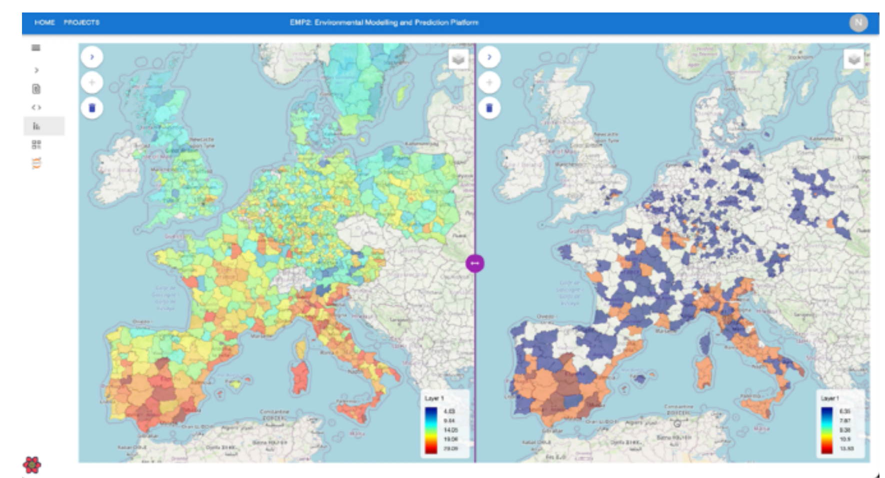


Onboard

(New CS4OD)

interTwin

Final product:
Prototype of a user oriented platform for environmental applications



Goal: test EMP² within a digital twin existing architecture.

EMP² will be implemented as one of the use cases to test the Digital Twin architecture developed through the InterTwin Project

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

Conclusions

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

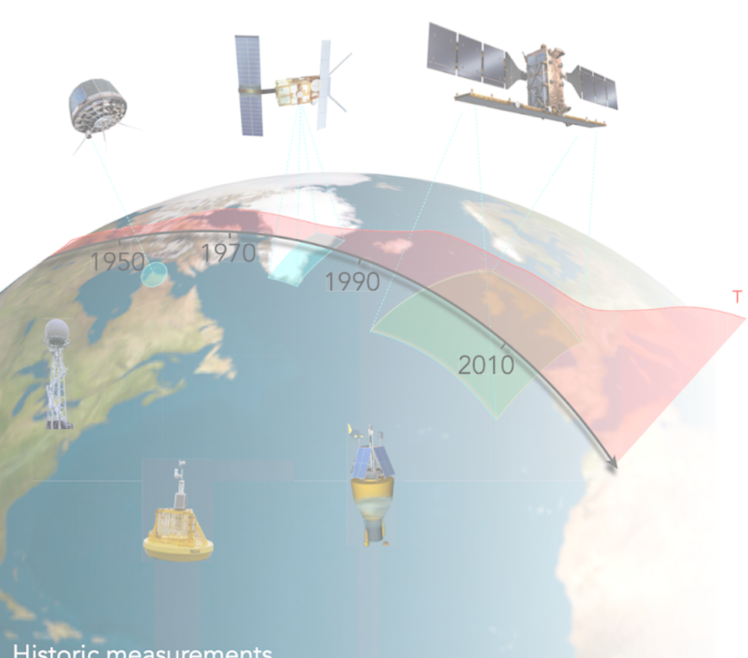
The **model is available and testable** on the current applications:
nowcasting, downscaling, temporal interpolation and precipitation corrections.

More infos:

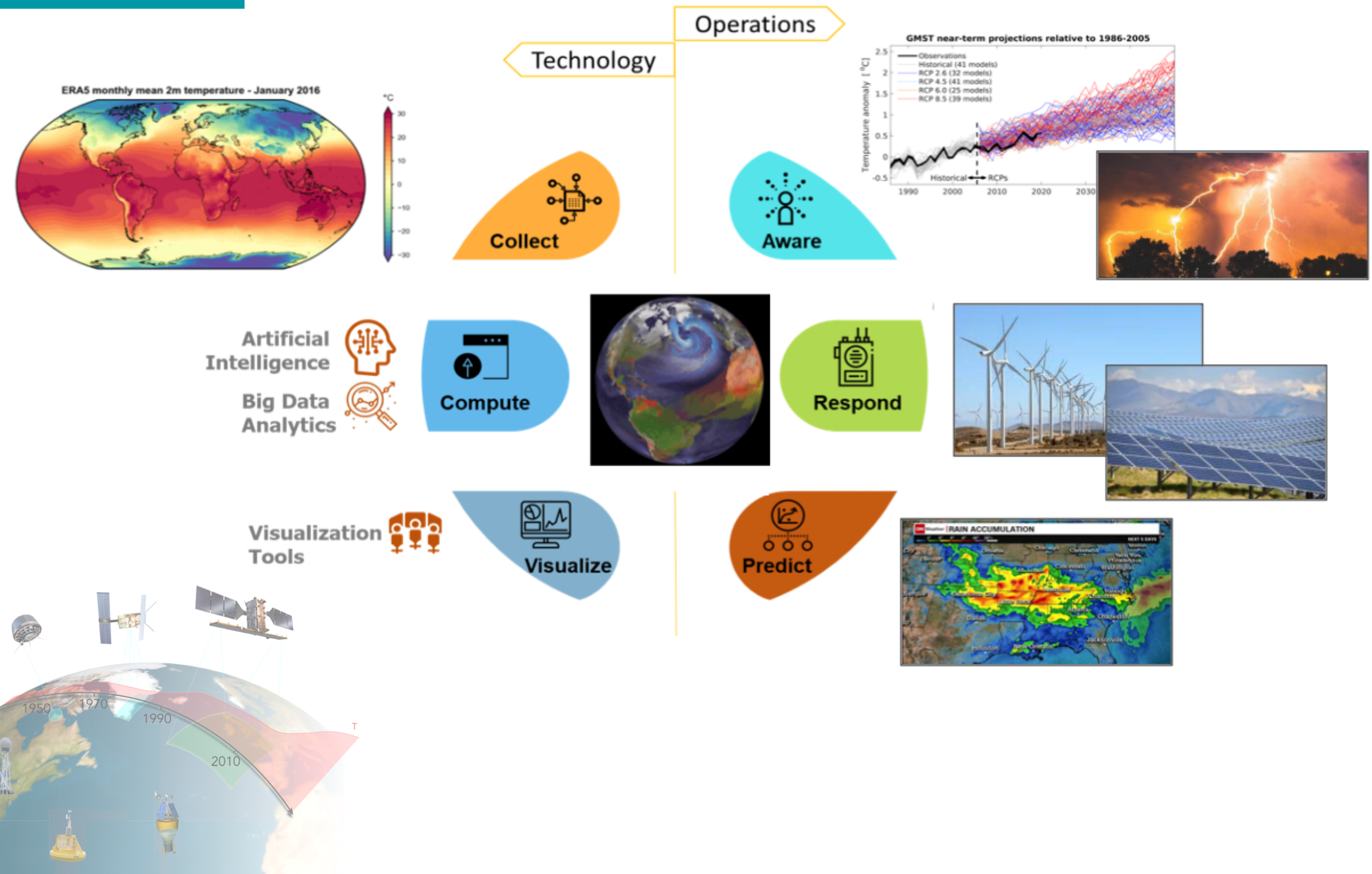
- Code is available on GitHub: link
- More infos on the website: www.atmorep.org
- **Pre-print on ArXiv: link**

.. and some long term plans:

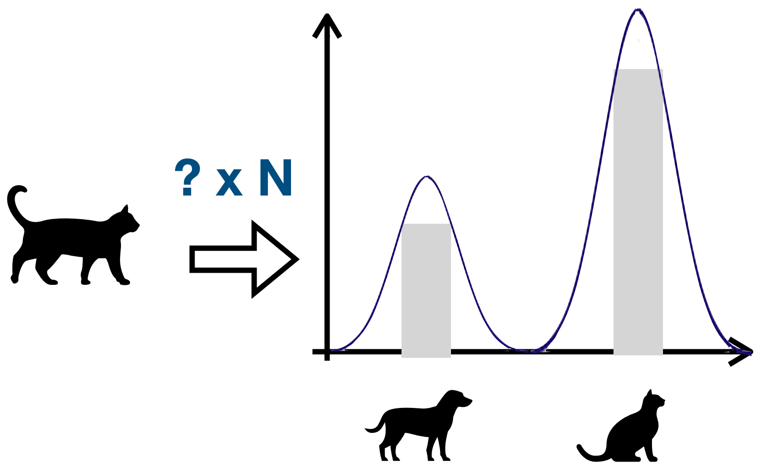
- How to integrate "raw" observations?
- Coupled atmosphere+ocean system?



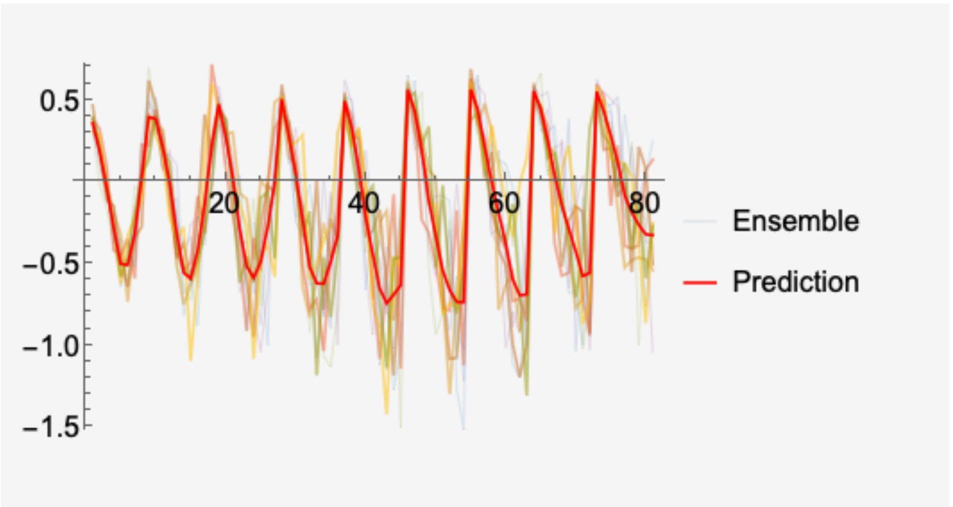
Backup

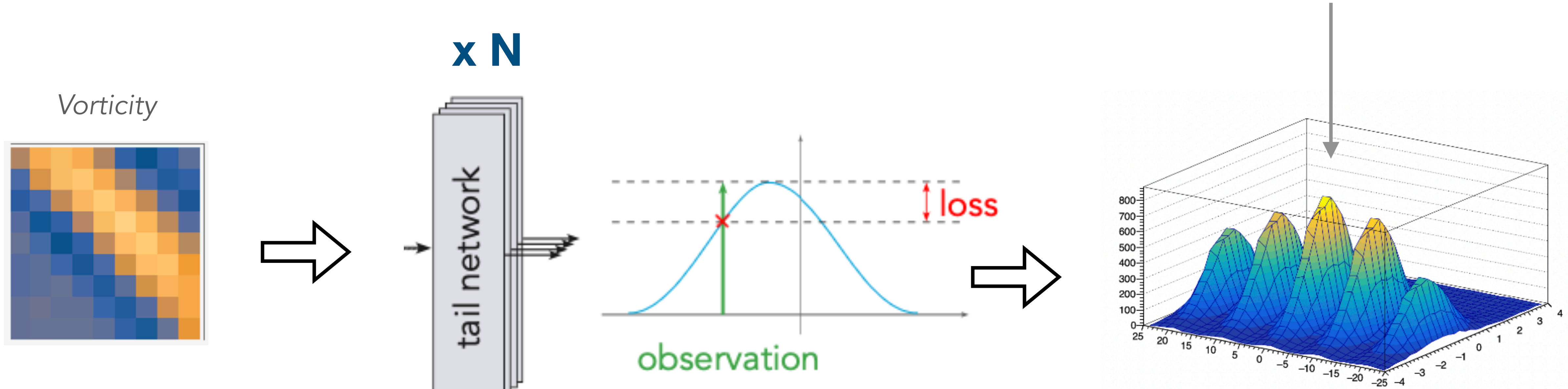

Technology
Operations
ERA5 monthly mean 2m temperature - January 2016
°C
GMST near-term projections relative to 1986-2005
Observations
Historical (41 models)
RCP 2.6 (32 models)
RCP 4.5 (41 models)
RCP 6.0 (25 models)
RCP 8.5 (39 models)
Temperature anomaly [°C]
Historical
RCPs
1990
2000
2010
2020
2030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

Key ingredients: statistical loss



Inspired by cross entropy loss in classification problems





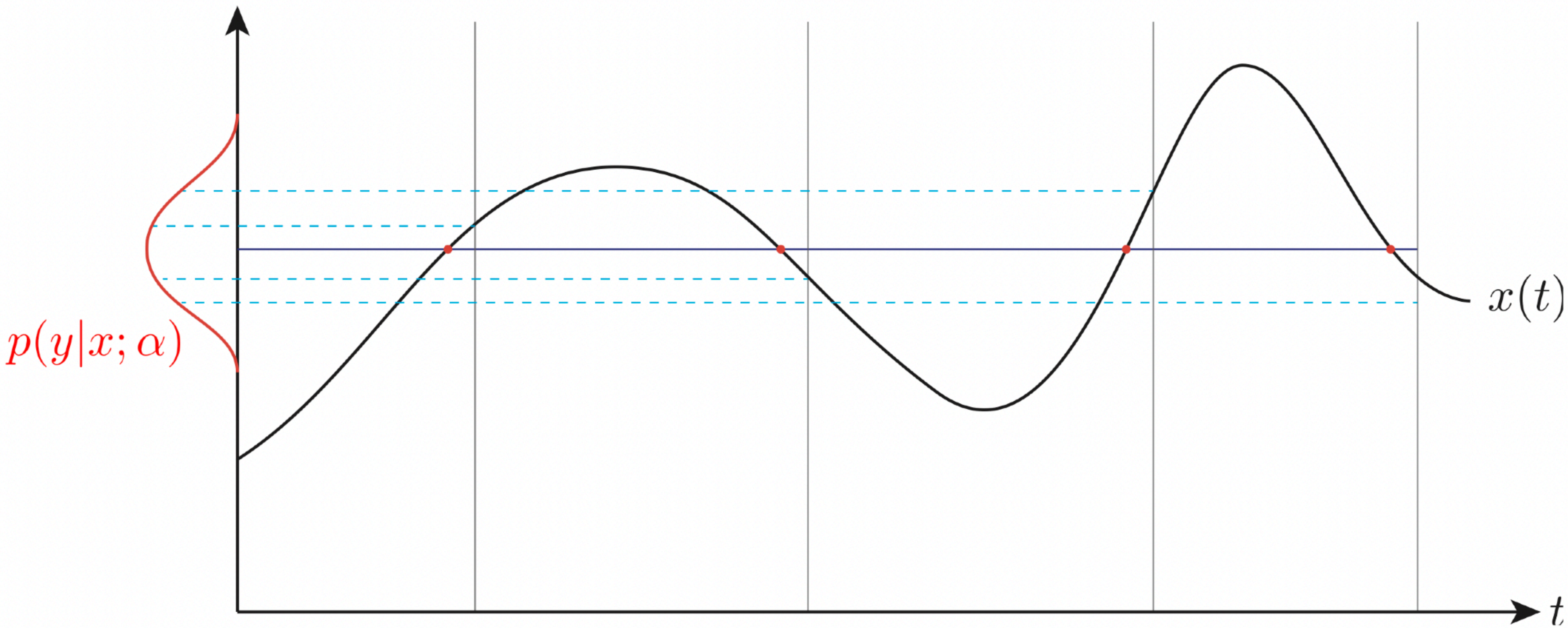
statistical interpretation: measure the difference between the pdf of the ML classification model and the predicted distribution

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

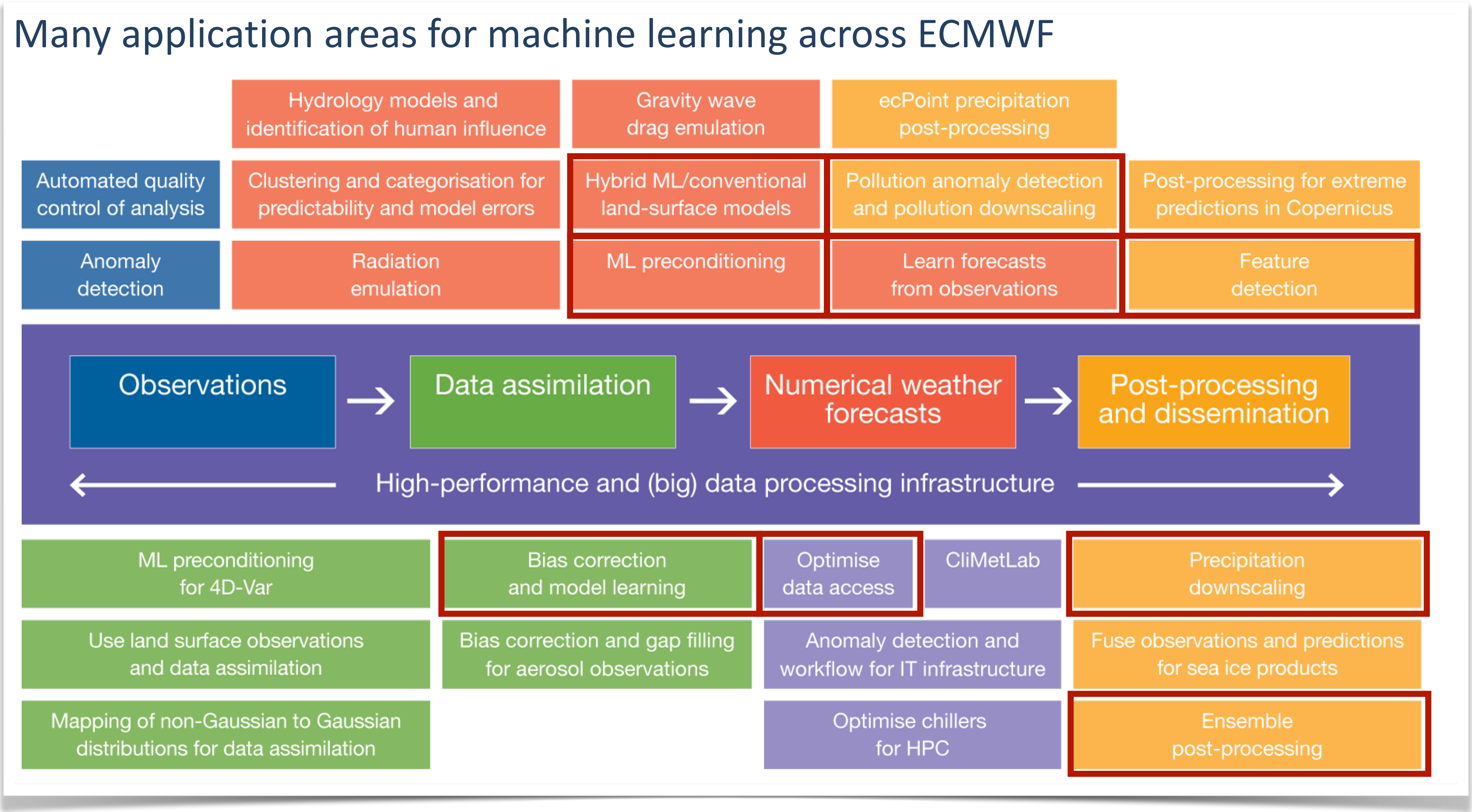
Ensemble variability



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

A full set of possible applications

From P. Dueben (ECMWF): https://events.ecmwf.int/event/232/attachments/963/1688/Presentation_slides.pdf



EMP² vs ClimaX: differences & similarities

Clima X

6 February 2023!

ClimaX:
A foundation model for weather and climate

Tung Nguyen*[1,3], Johannes Brandstetter[2], Ashish Kapoor[1],
Jayesh K. Gupta†[1], and Aditya Grover†[1,3]
[1]Microsoft Autonomous Systems and Robotics Research, [2]Microsoft Research AI4Science, [3]UCLA

ClimaX	EMP²	
Both are foundation models based on visual transformers!		
Investigating similar downstream applications		
Trained on a randomised forecasting objective Goal: reconstruct states in the future	BERT-style training adapted to scientific data reconstruct *masked tokens* within a random hypercube	→ The model is less "forecasting oriented"
using ERA5 on pressure level variables	using ERA5 on model level variables	→ This is what ECMWF uses: an eye into the integration within their systems.
deterministic predictions	**stochastic predictions** Model uncertainty quantification through newly defined statistical loss	
single transformer Concatenation of fields in the variable aggregation step	**stack of transformers** one transformer for each field, coupled with cross attention.	→ Plug and play approach: new fields can be easily integrated.
private company	public research	