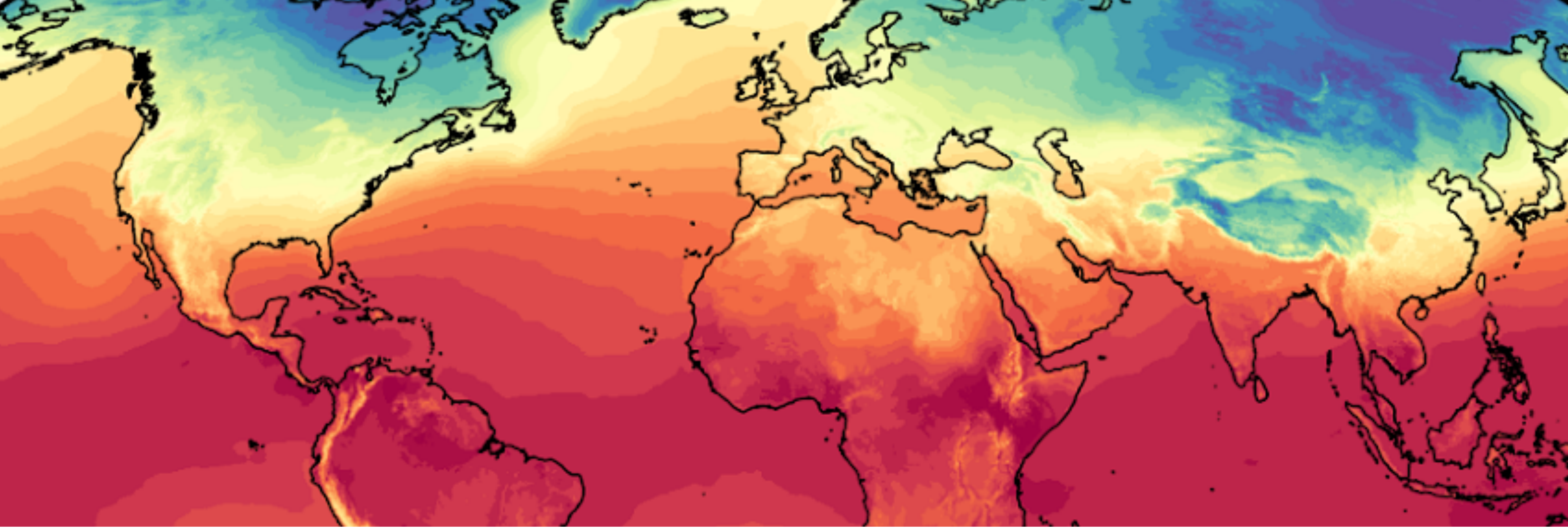


AtmoRep:
A stochastic model of atmosphere dynamics using large scale representation learning

Christian Lessig, Ilaria Luise, Martin Schultz,
Michael Langguth, Scarlet Stadtler, Bing Gong

ClimateChangeAI Workshop | NeurIPS 2024

CERN | CIPEA CERN Innovation Programme on Environmental Applications

ECMWF

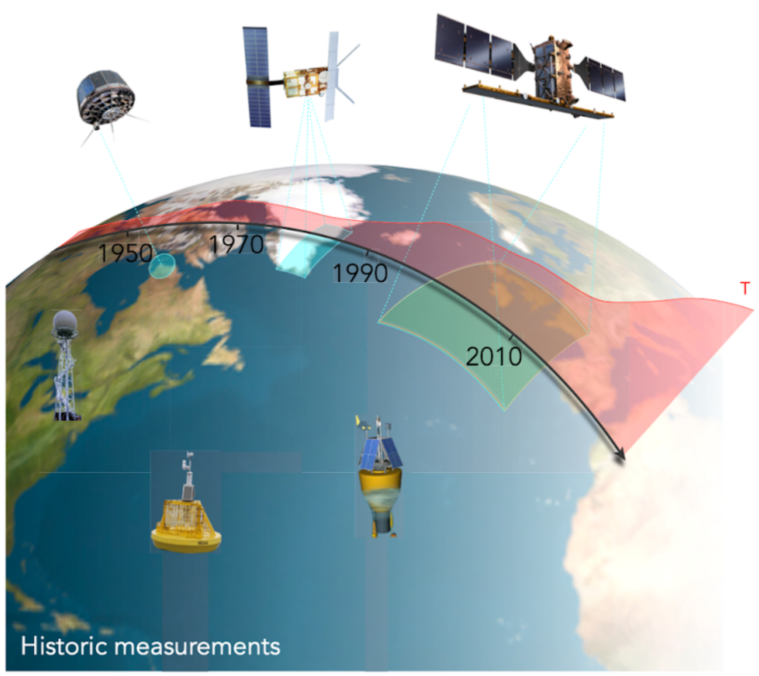
JÜLICH Forschungszentrum

A foundation model for Atmospheric dynamics

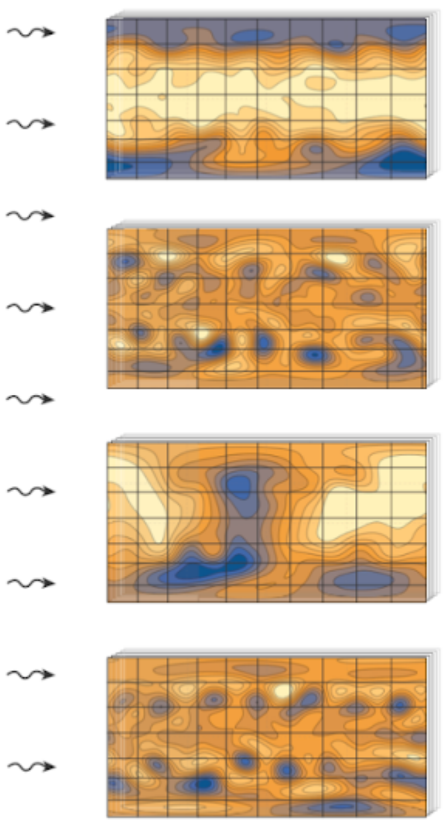
The spatio-temporal (4D) evolution of a dynamical system can be summarised as

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

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



Observations



50 TB

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

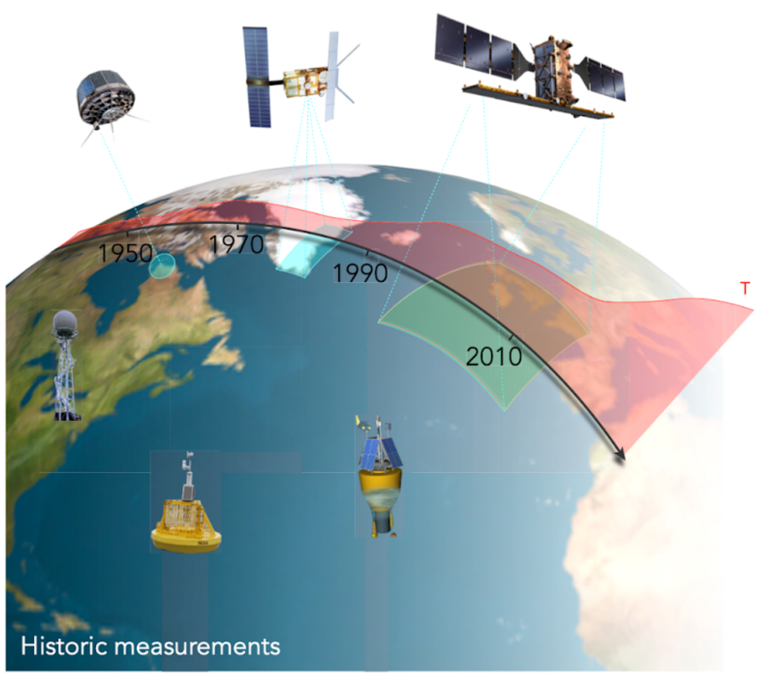
A foundation model for Atmospheric dynamics

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

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

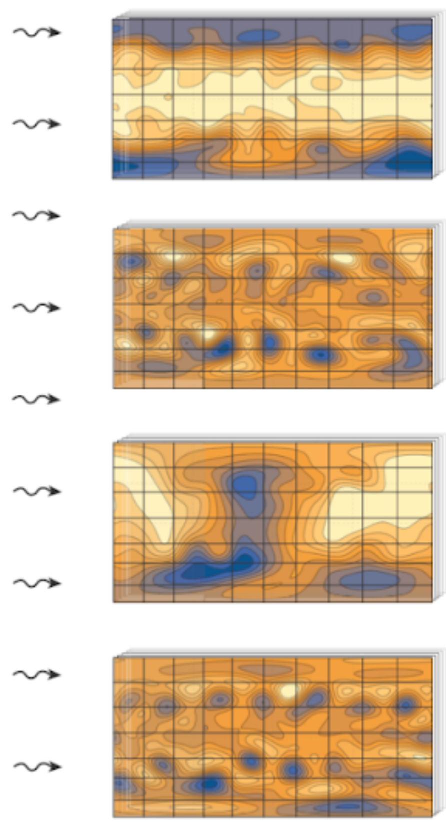
foundation model:
neural network that models the data distribution for a specific domain

Spatio-temporal representation of atmospheric dynamics

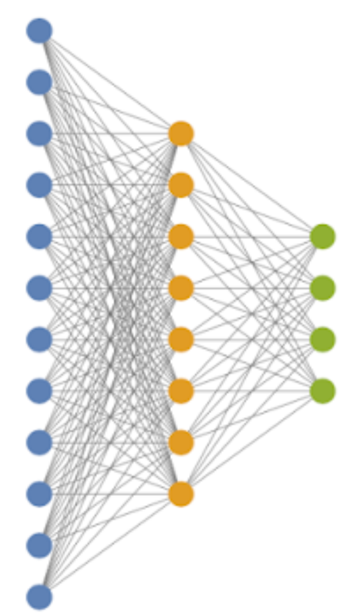


Observations

ERA5 reanalysis

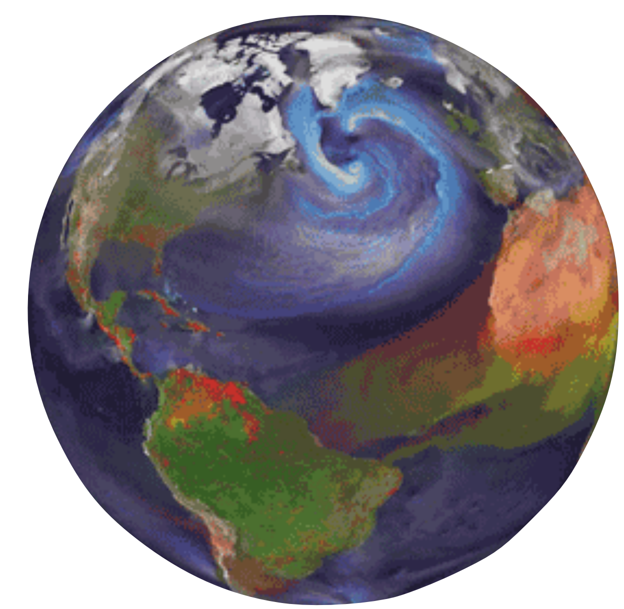


50 TB



large scale machine learning

Transformers architecture



Model
given by the trained neural network

R&D at Juelich SSC:
$4\text{x}10^6$ GPU hours granted in 2023

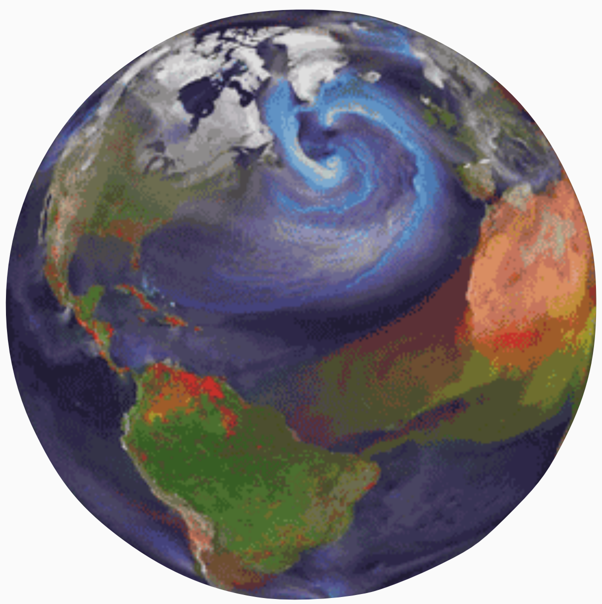
JÜLICH Forschungszentrum

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

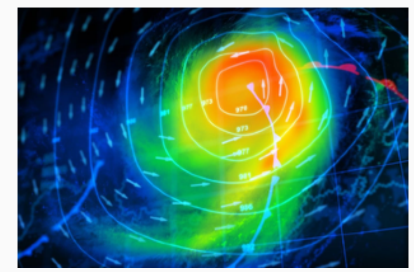
Spatio-temporal representation of atmospheric dynamics



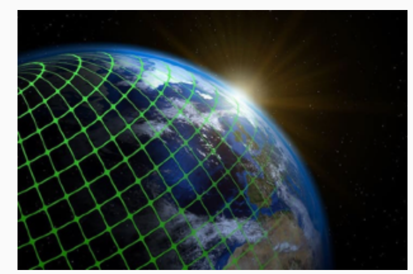
Model
given by the trained neural network

Task dependent Adaptable smaller networks

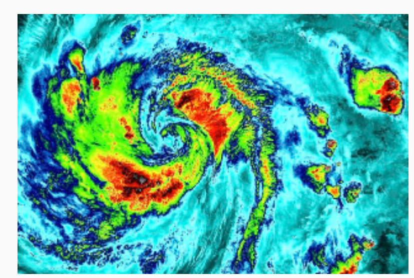
Adaptation



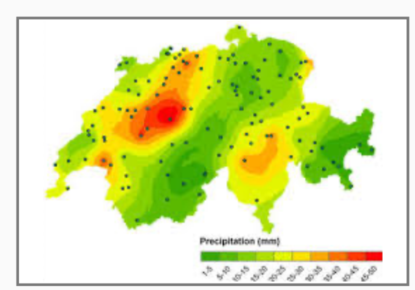
Weather predictions ✓



Downscaling ✓



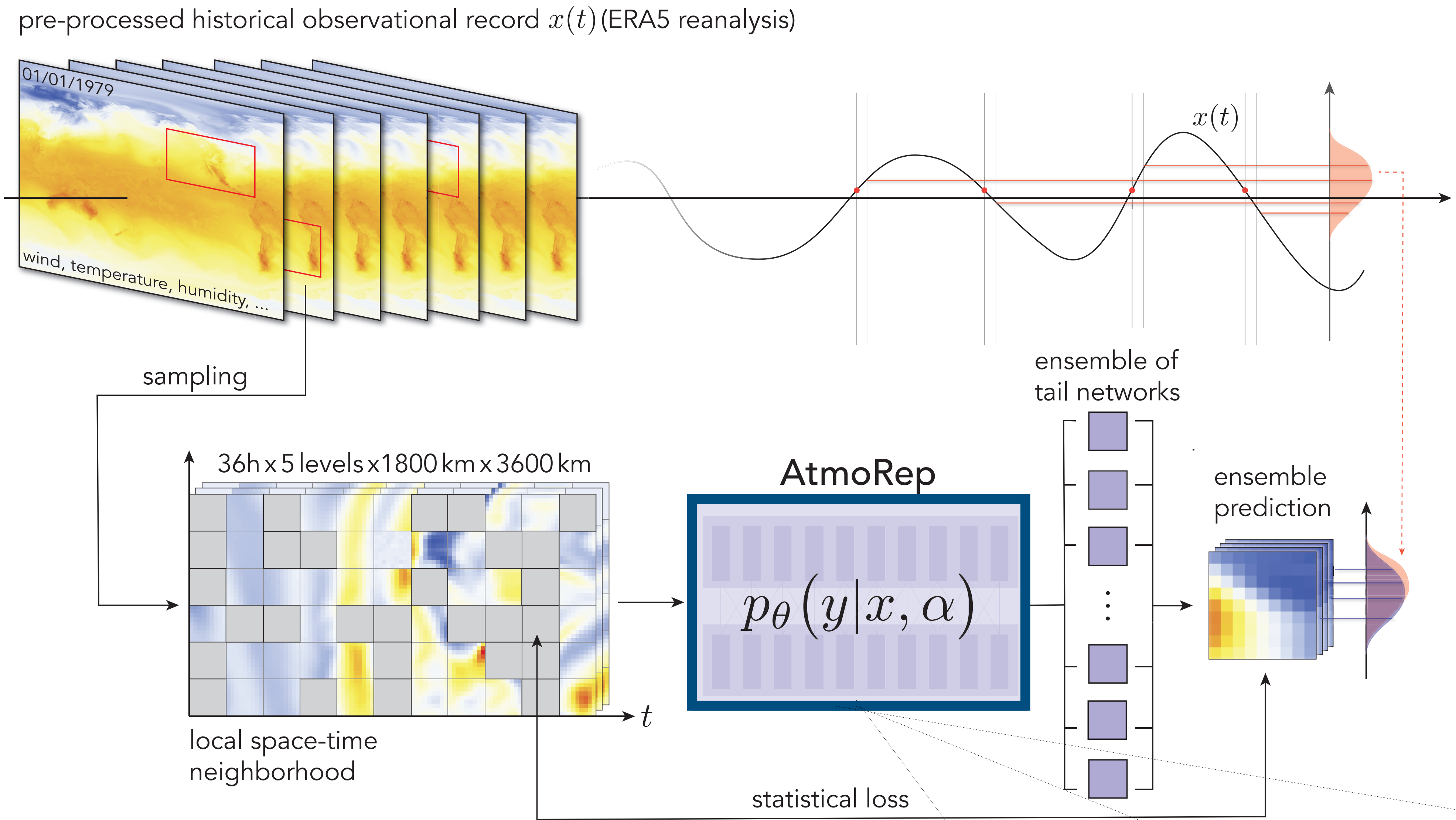
Bias corrections ✓



Spatio-temporal Interpolations (WIP)

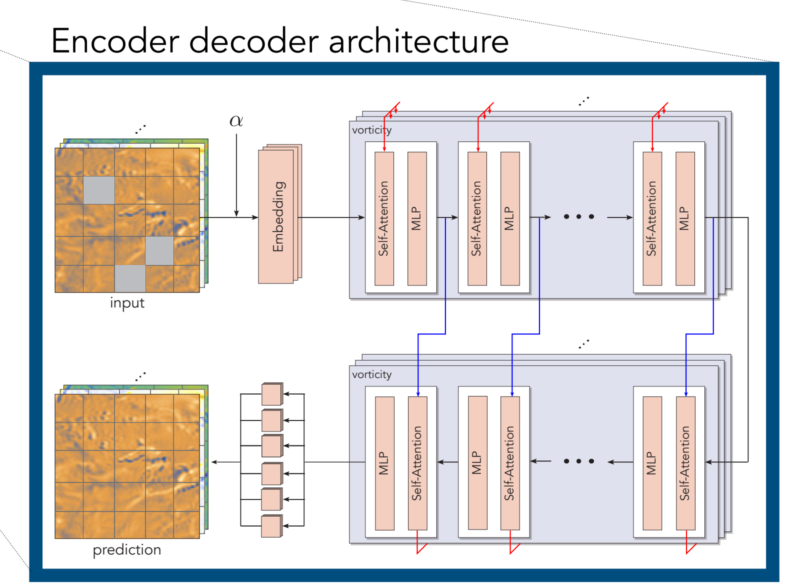
Ilaria Luise, CERN - ilaria.luise@cern.ch

The network architecture



New stochastic approach

ensemble predictions with 16 members



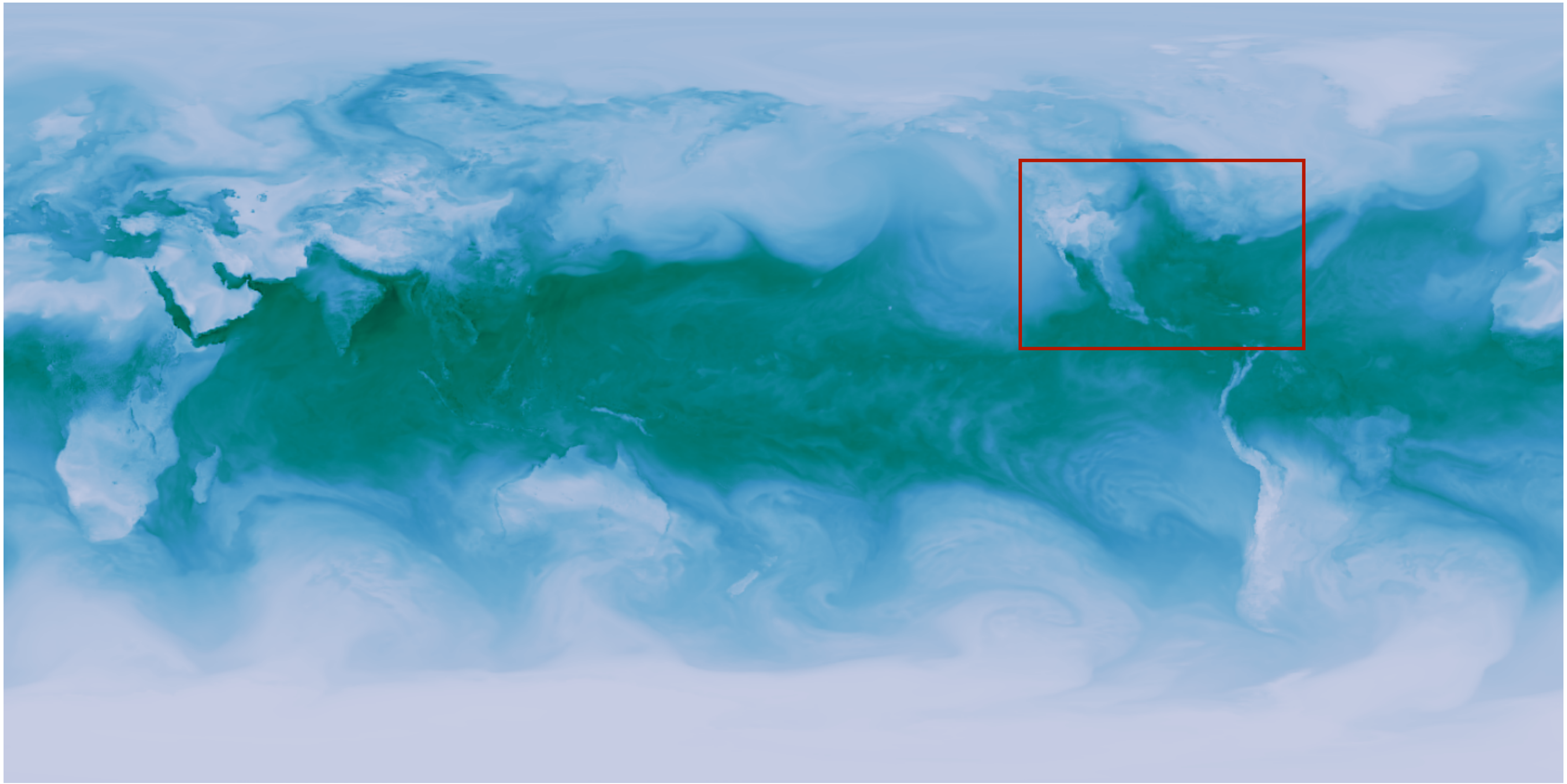
Approximate the 4-Dim PDF of the process using a Transformers-based network with 3.5 billion parameters

Short term weather forecasting

Zero-shot applications

Results: Target - ERA5

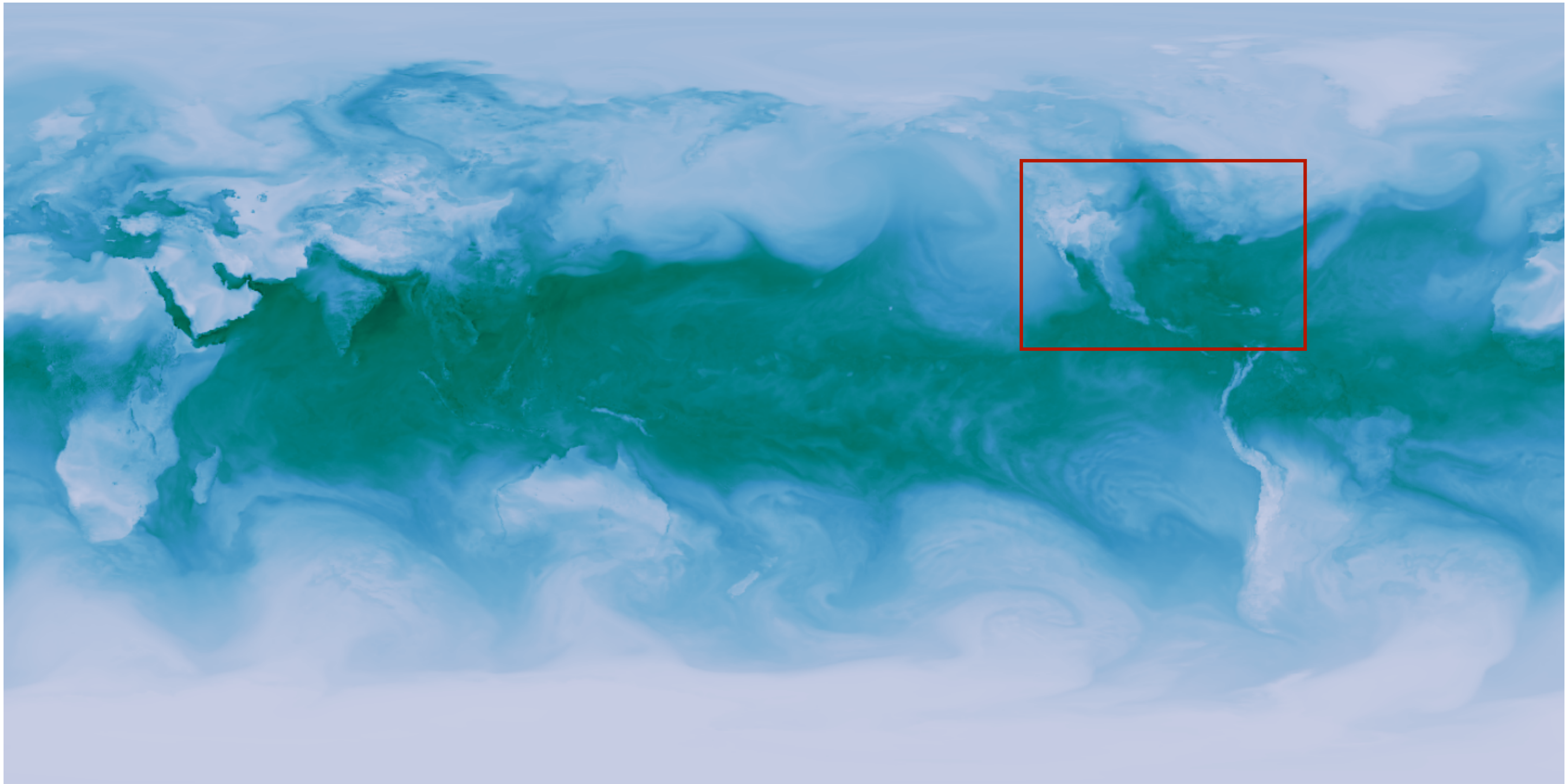
specific humidity, June 15th 2018 13:00 UTC



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

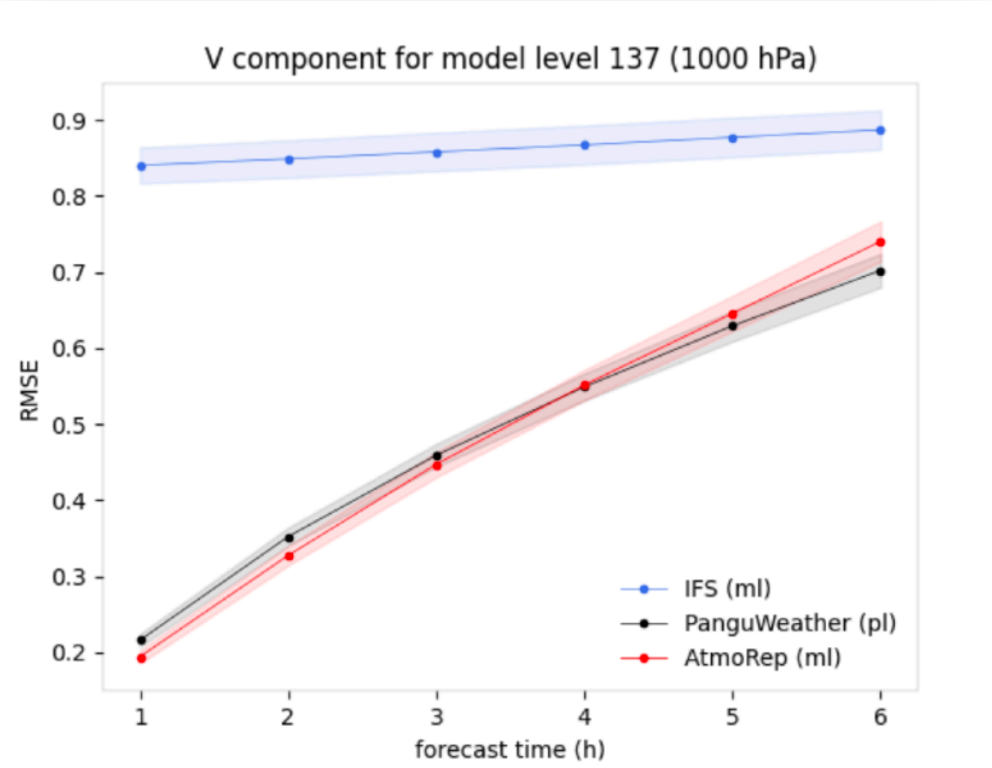
Results: Prediction - AtmoRep

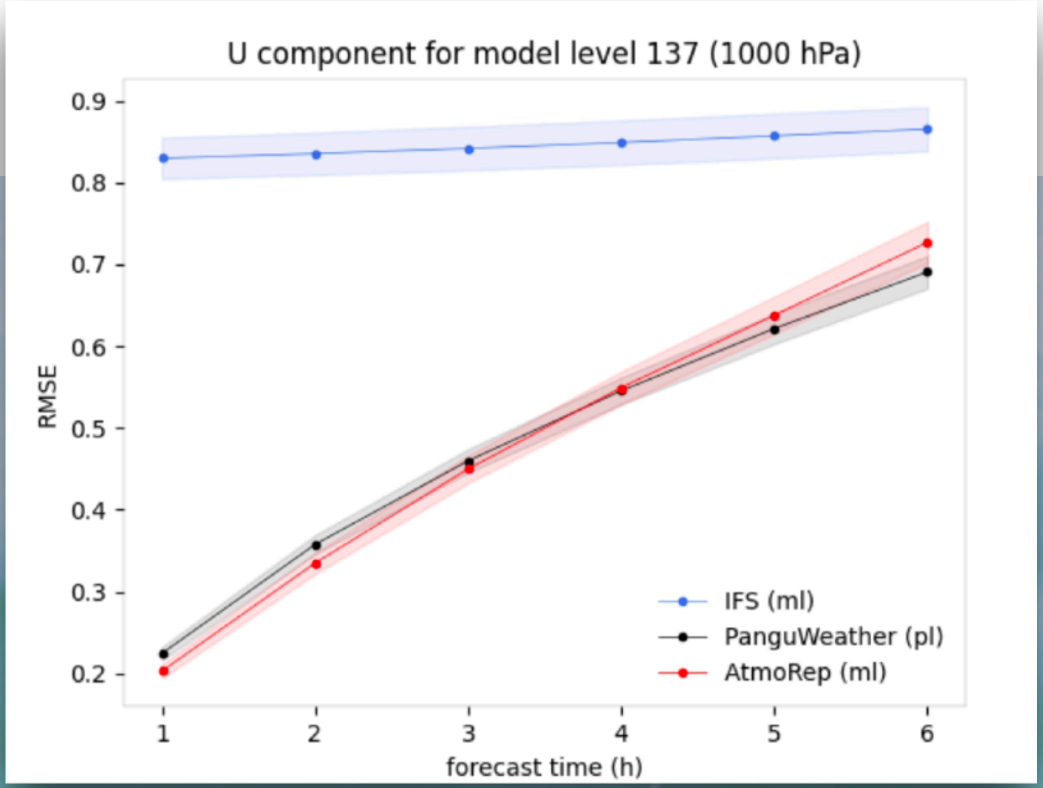
specific humidity, June 15th 2018 13:00 UTC

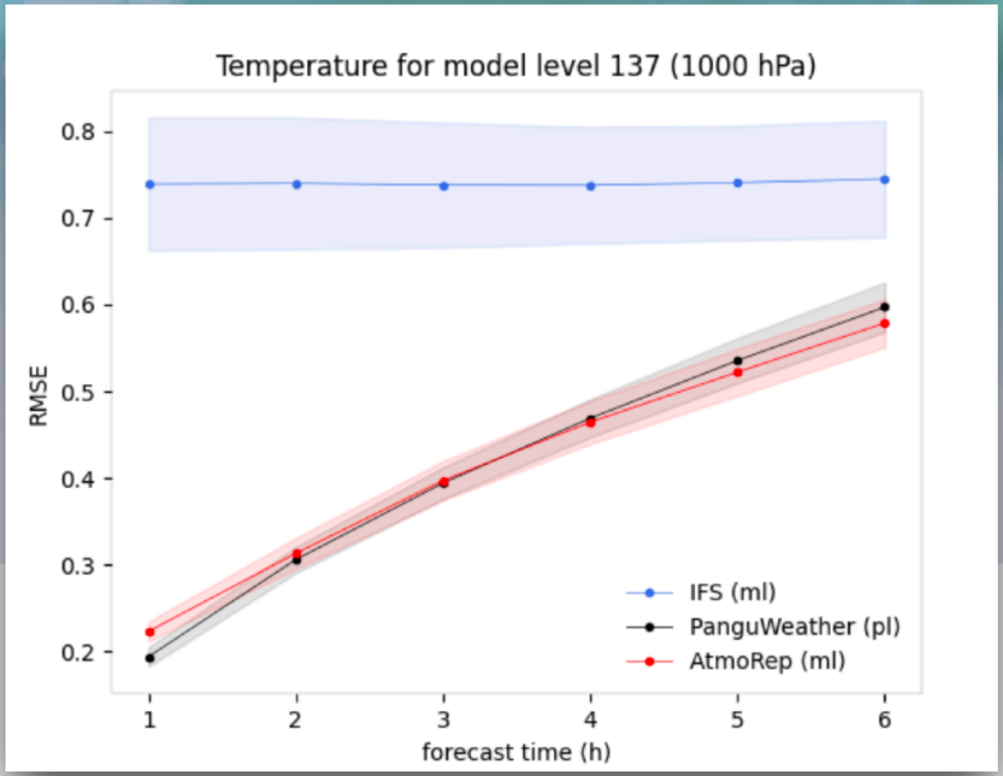


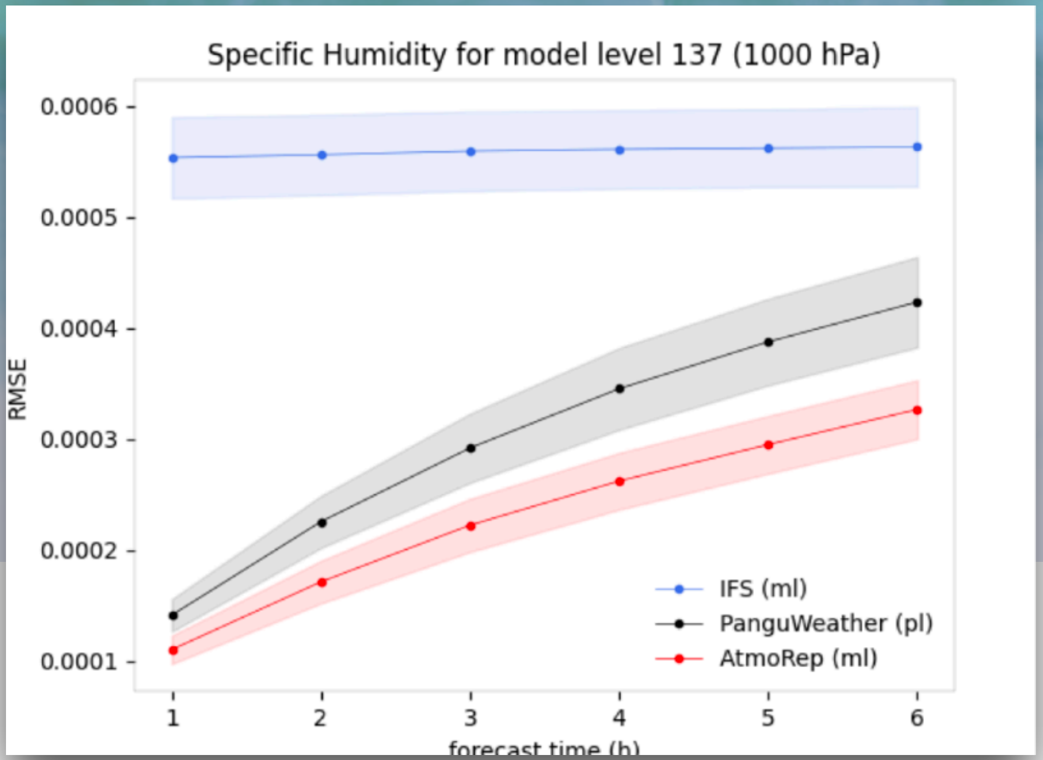
Ilaria Luise, CERN - ilaria.luise@cern.ch

Results: Prediction - AtmoRep







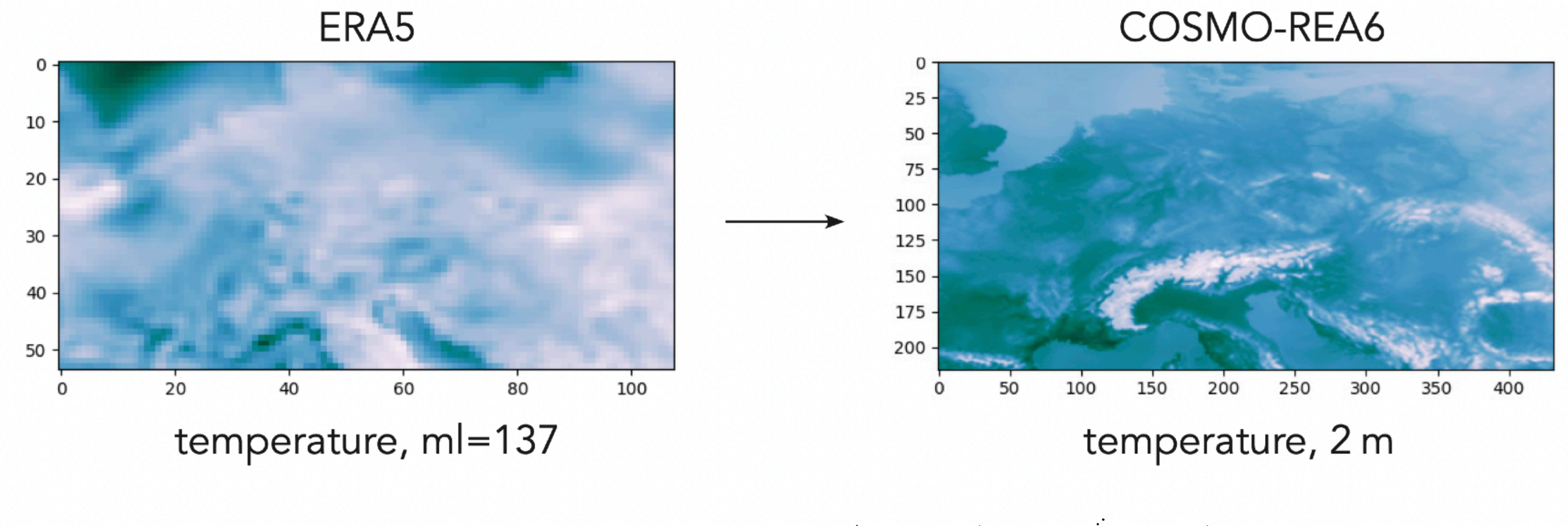


Ilaria Luise, CERN - ilaria.luise@cern.ch

Fine tuning on real data

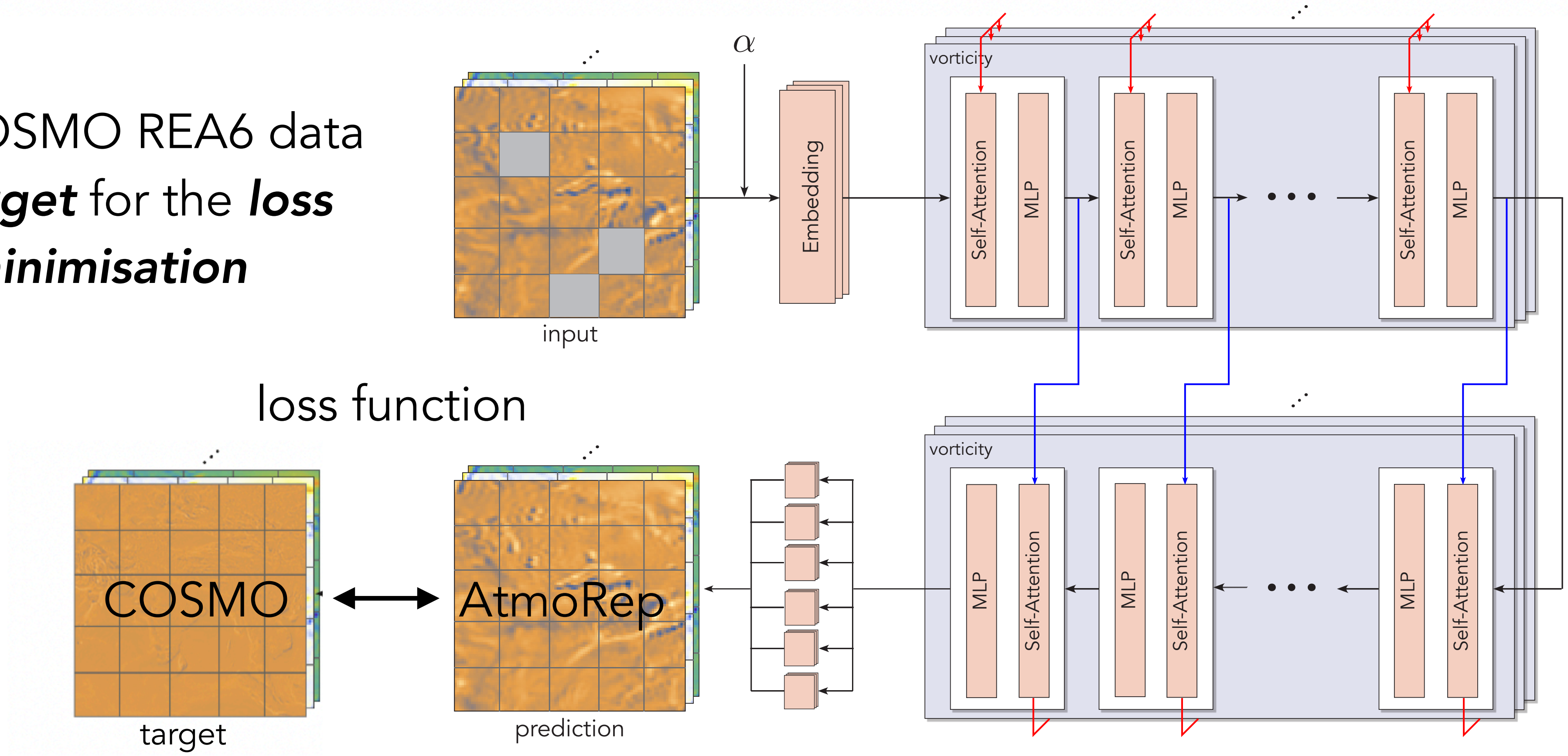
Data driven precipitation corrections & downscaling

Downscaling



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

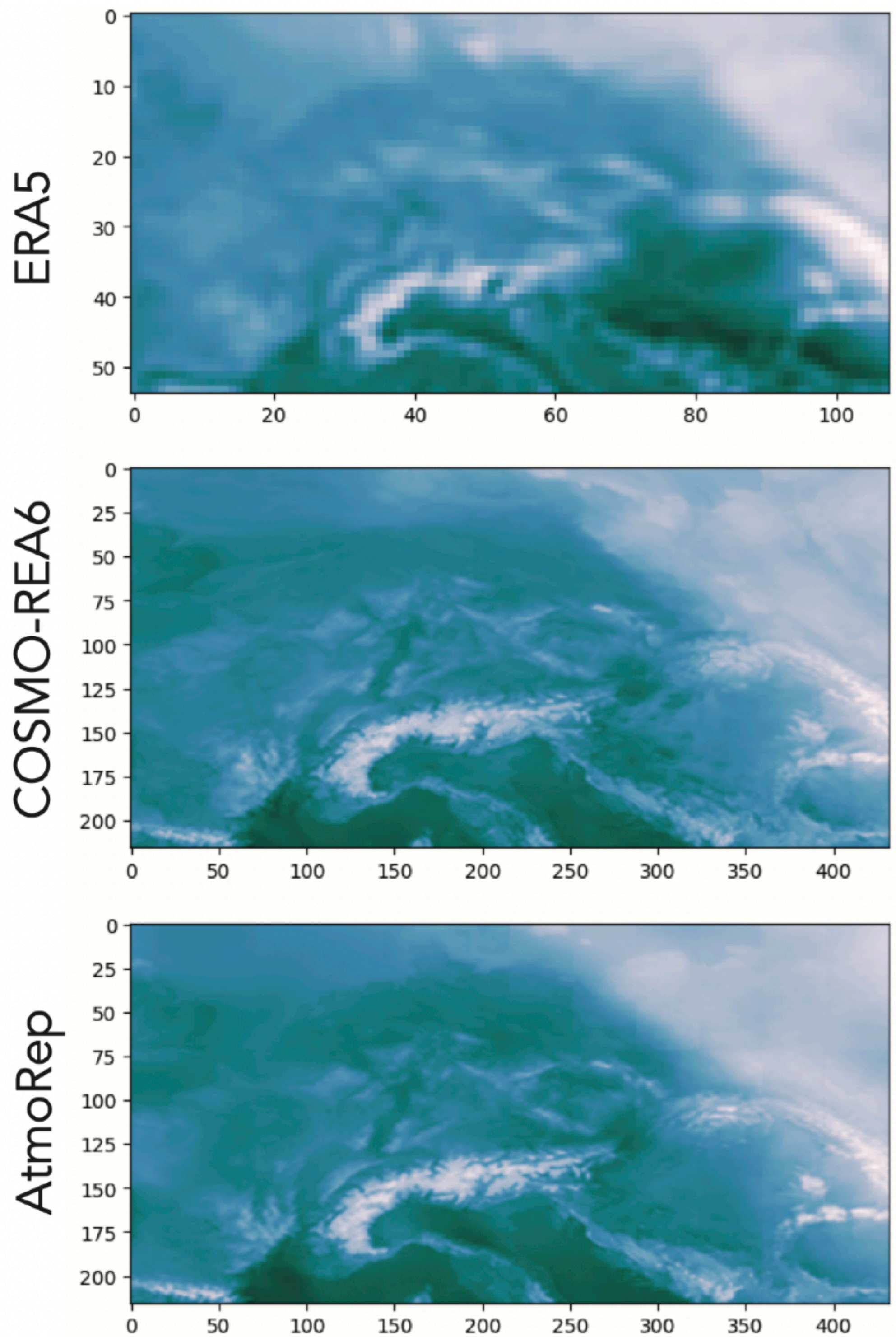
loss function



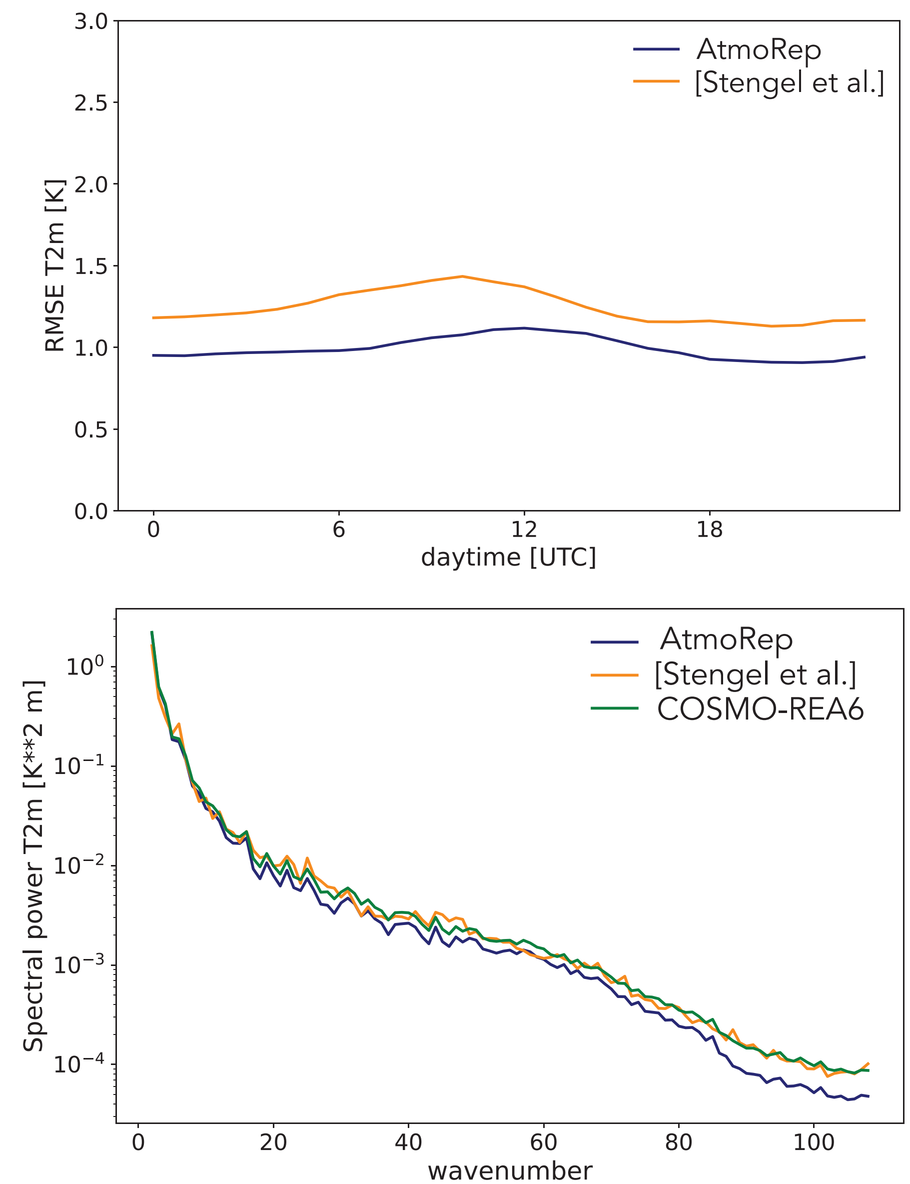
Ilaria Luise, CERN - ilaria.luise@cern.ch

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:

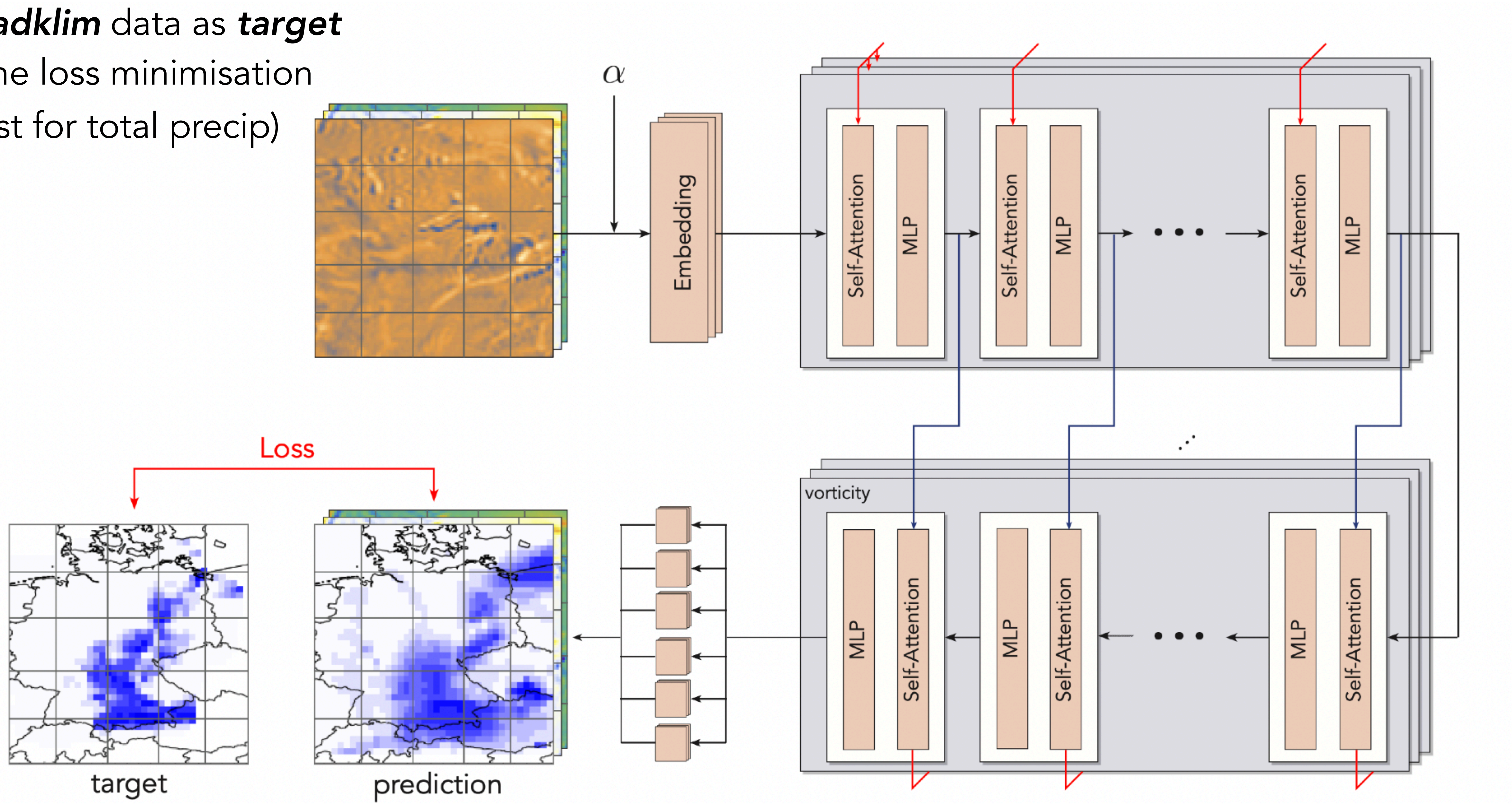


Ilaria Luise, CERN - ilaria.luise@cern.ch

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)



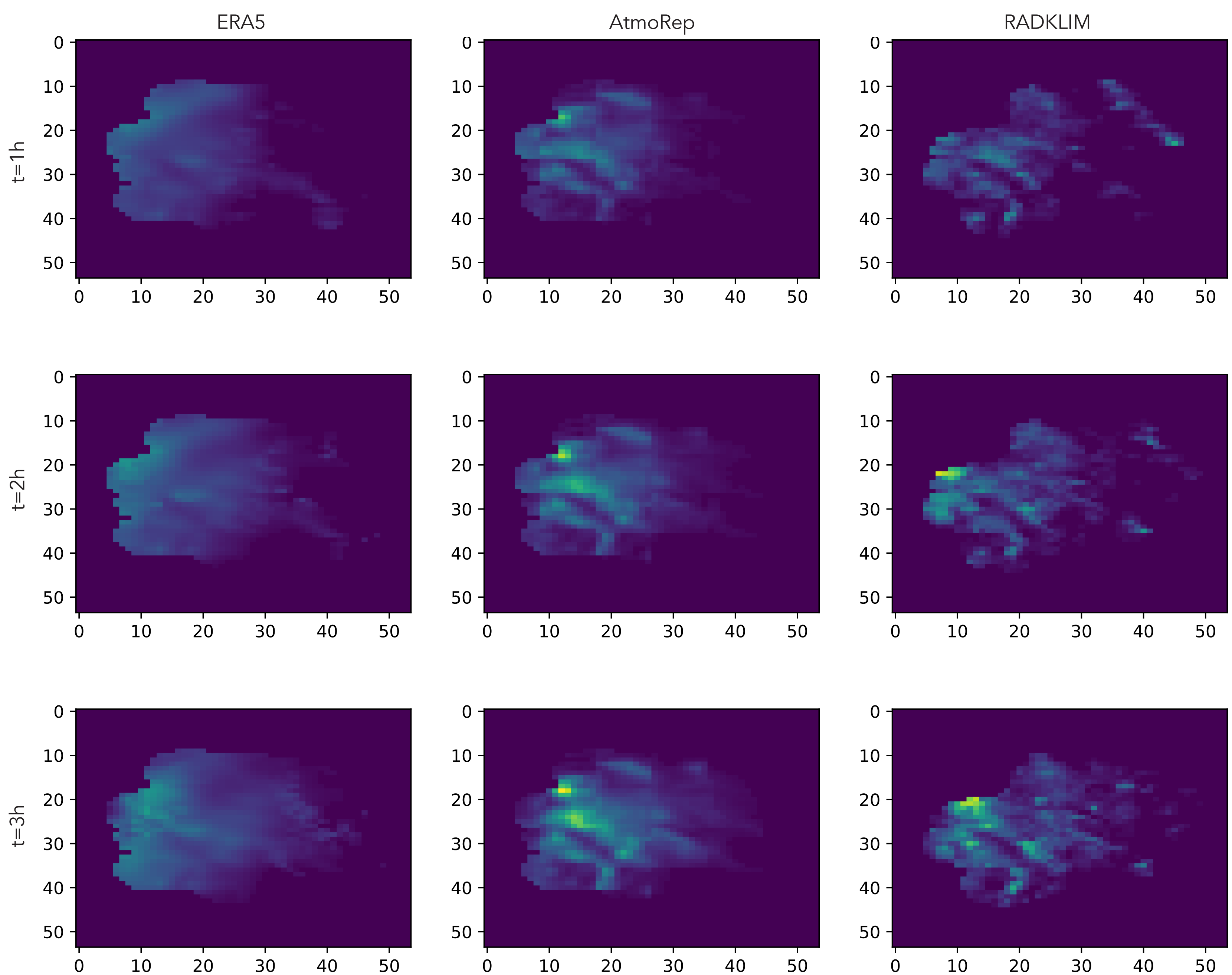
Radklim

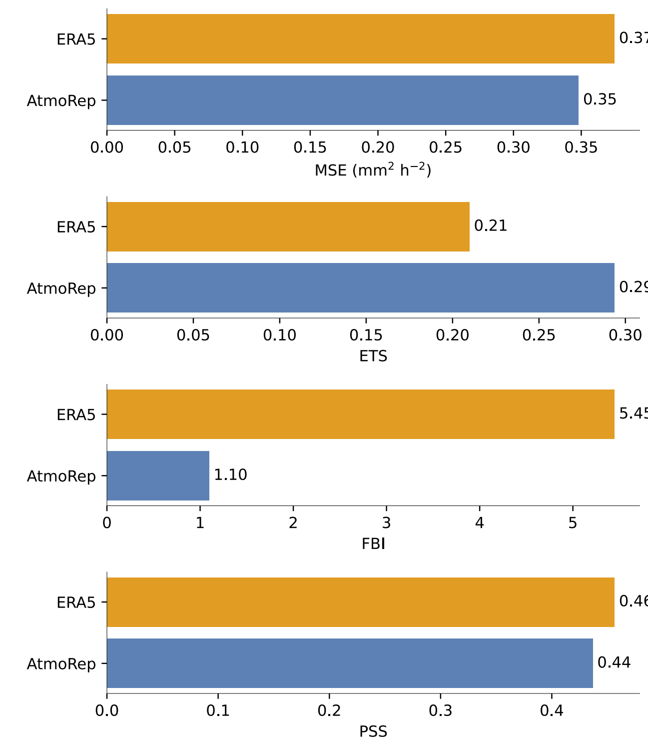
AtmoRep

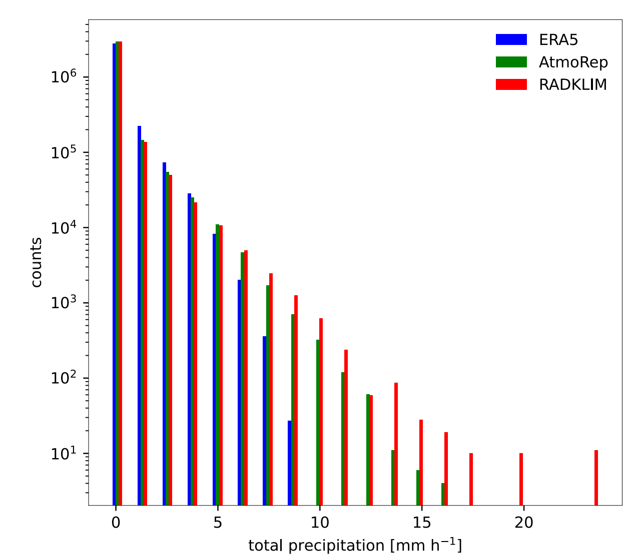
Ilaria Luise, CERN - ilaria.luise@cern.ch

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

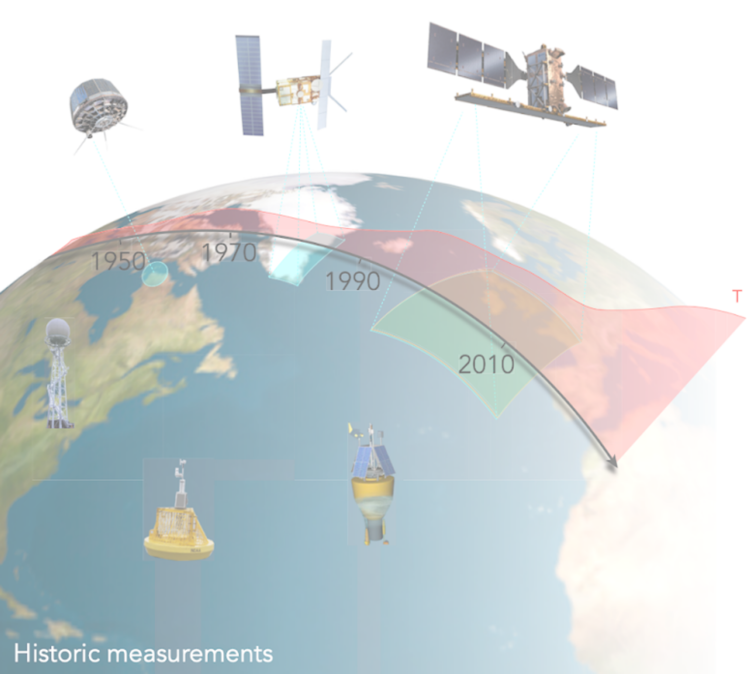
Conclusions

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 will be available open source before the end of the month!
- More infos on the website: www.atmorep.org
- **Pre-print on ArXiv: link**
- Submitted to Nature and accepted for review

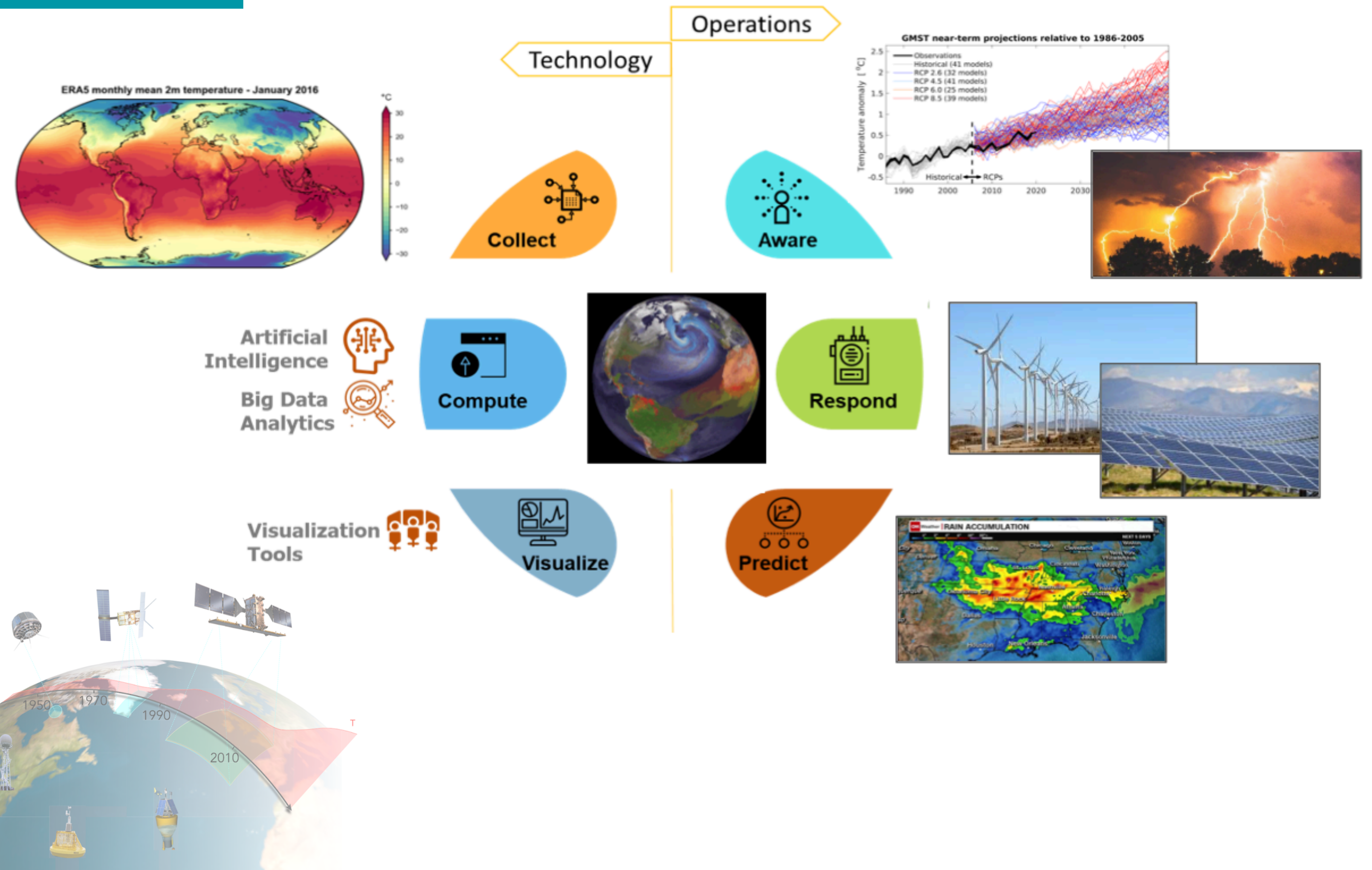


.. and some long term plans:

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

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

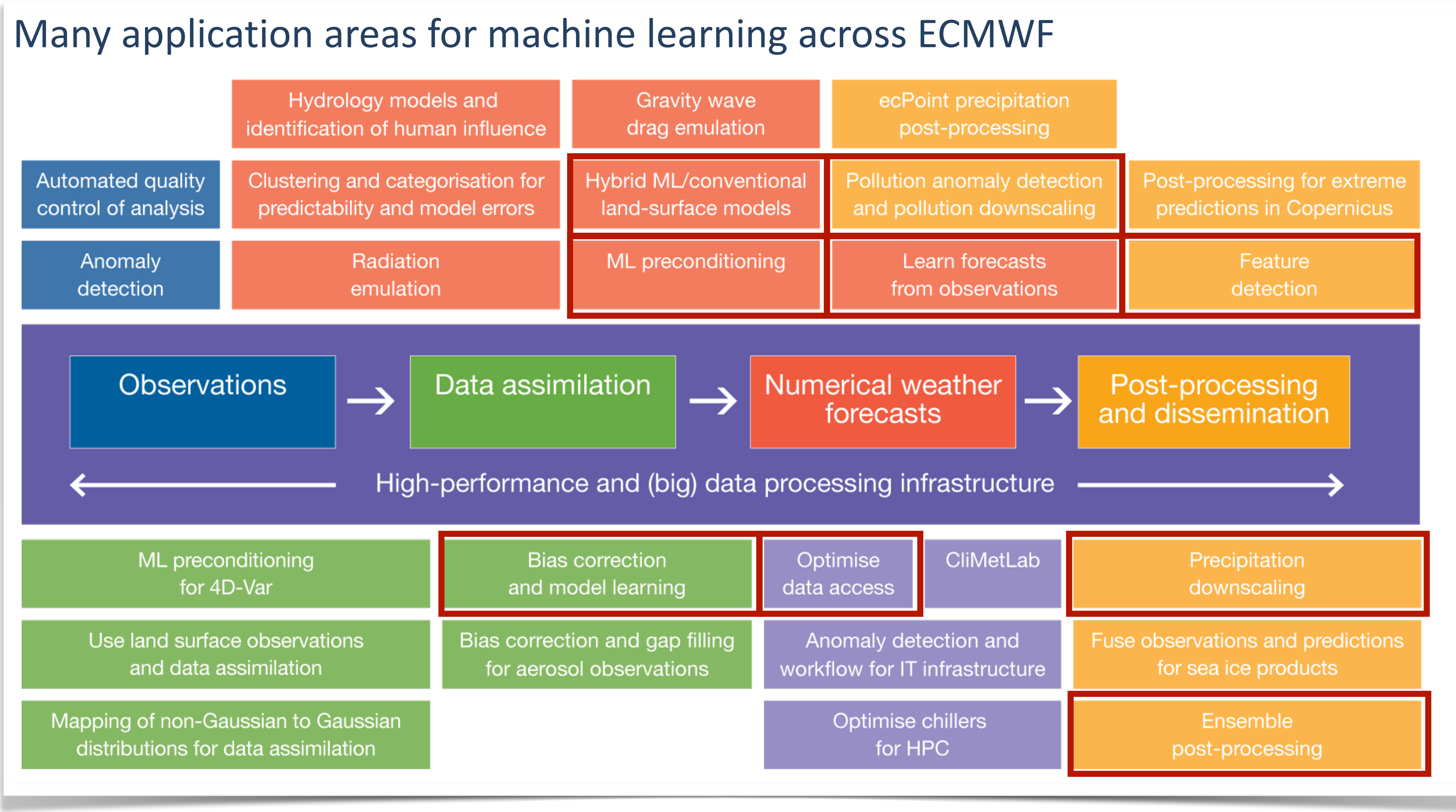
Backup



Technology
Operations
ERA5 monthly mean 2m temperature - January 2016
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
Collect
Aware
Artificial Intelligence
Big Data Analytics
Compute
Respond
Visualization Tools
Visualize
Predict
RAIN ACCUMULATION
1950
1970
1990
2010

A full set of possible applications

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

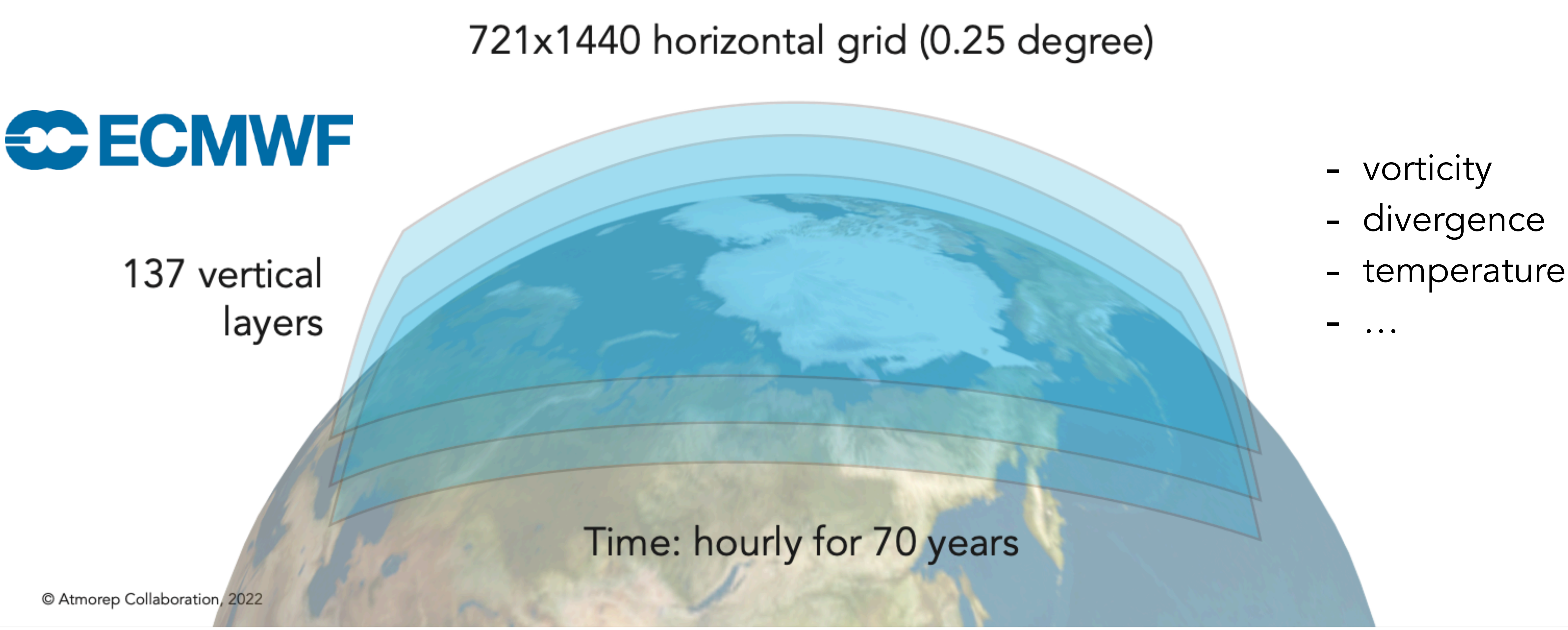


The dataset

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

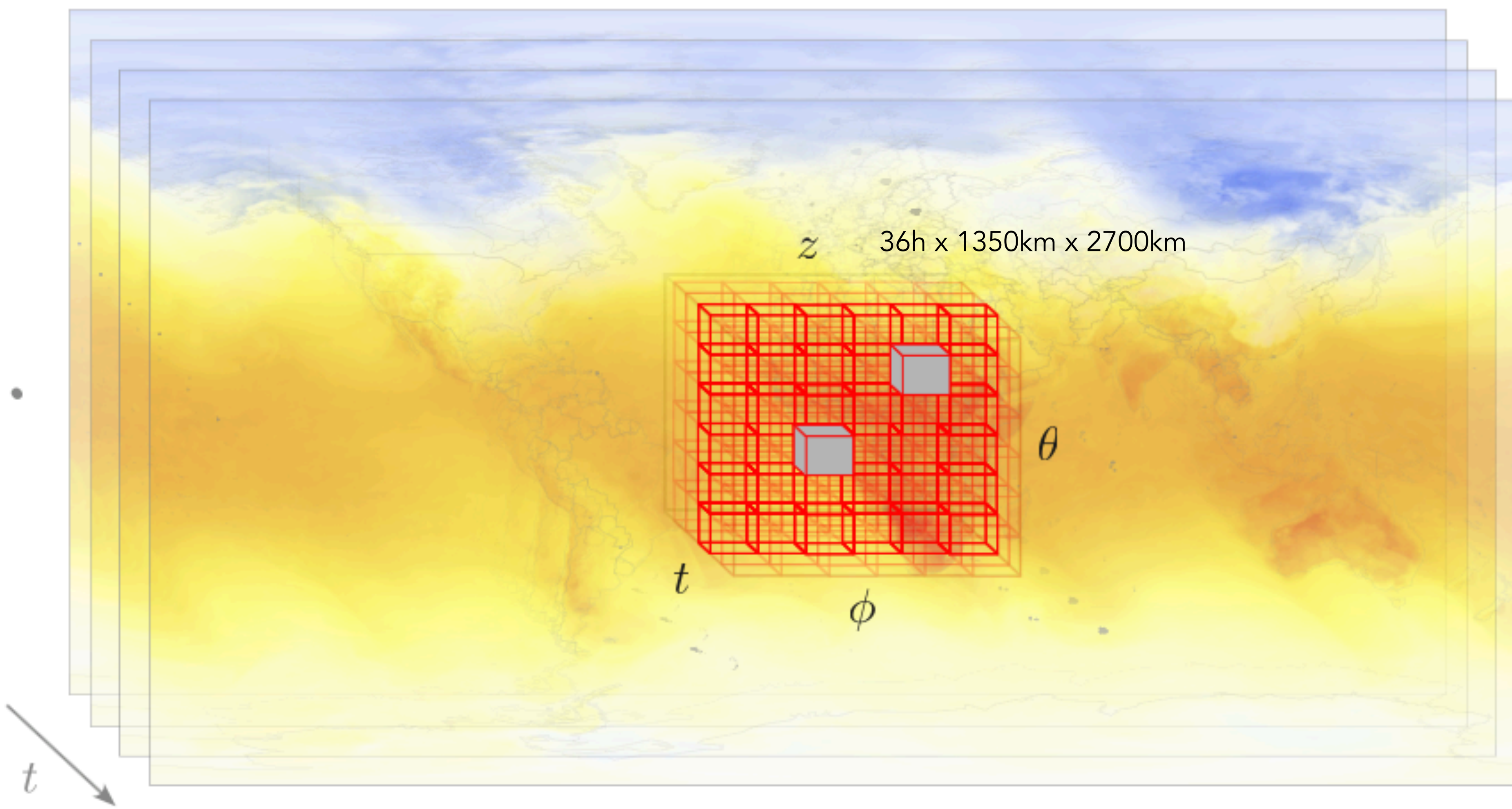


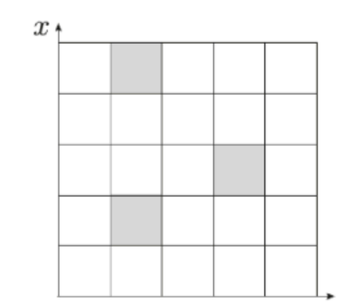
Ilaria Luise, CERN - ilaria.luise@cern.ch

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

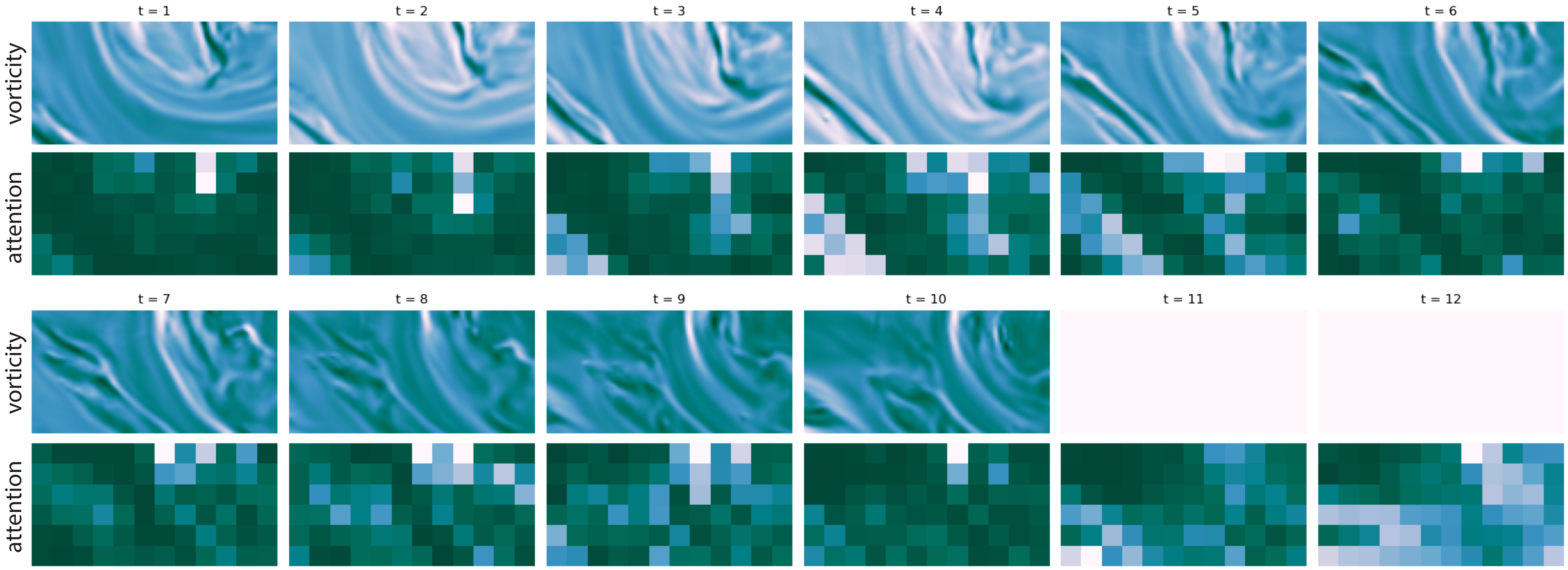
Mask random tokens within the hyper-cube and try to predict them back

visually: learn representation dynamics through interpolation

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

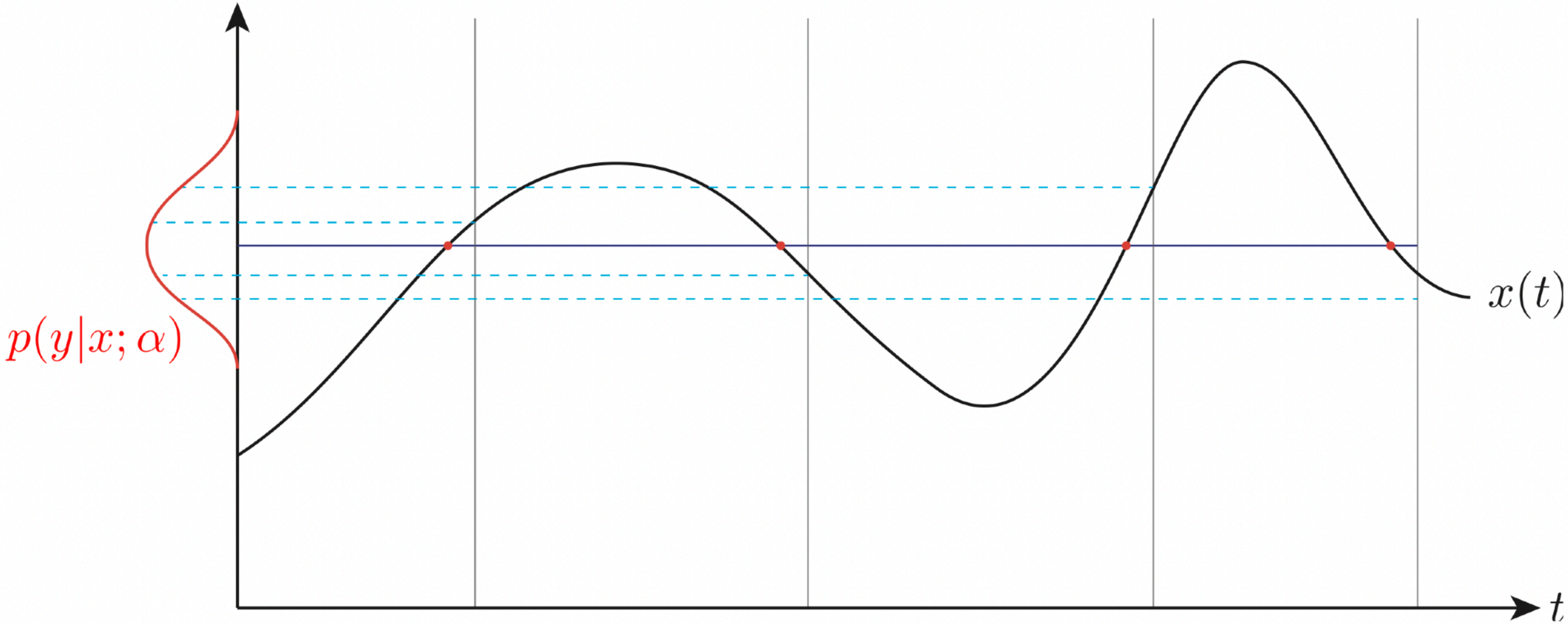
Ilaria Luise, CERN - ilaria.luise@cern.ch

Attention maps and interpretability



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

Ensemble variability



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

EMP² vs ClimaX: differences & similarities

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	