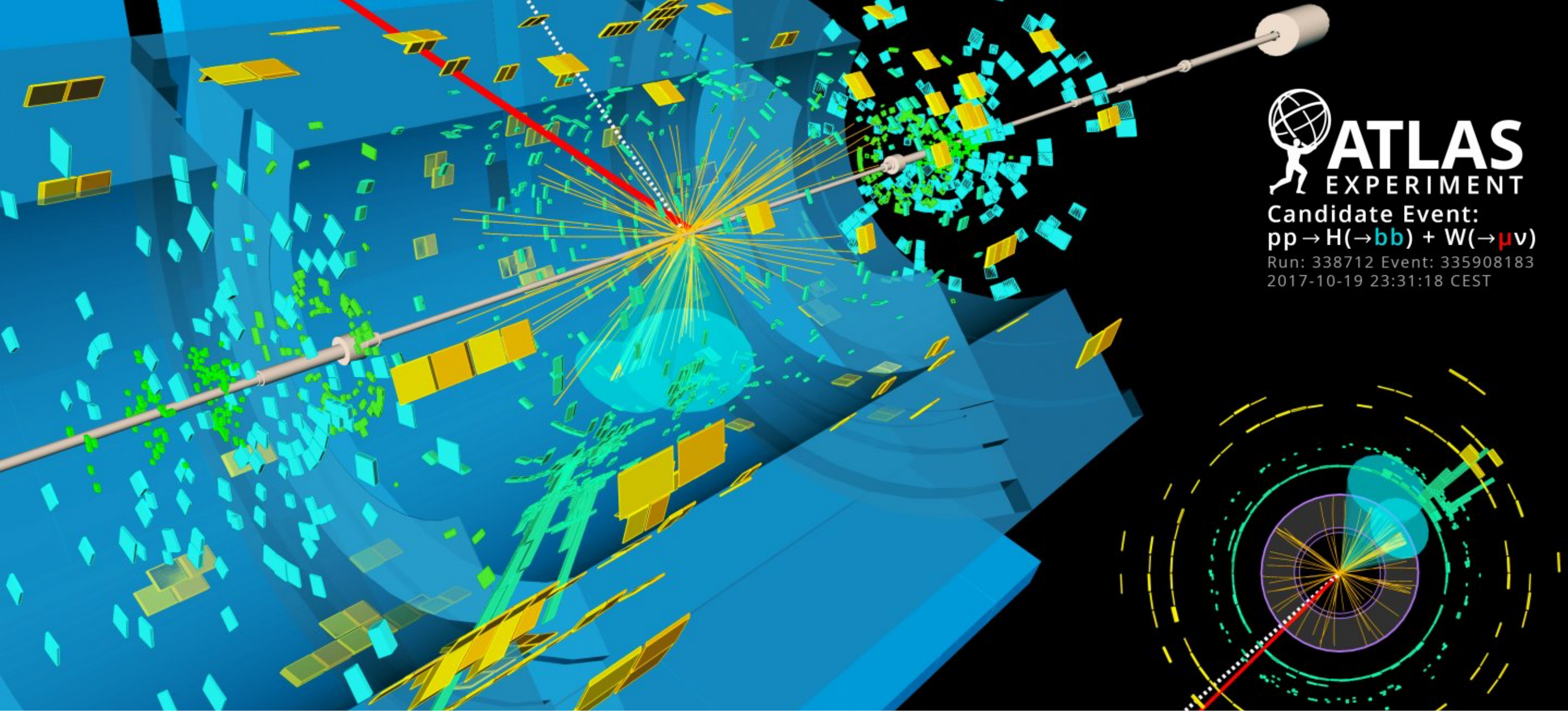


Advanced Topics:
Quantum computing &
ML in other scientific domains

Lecture 5

Sofia Vallecorsa | Ilaria Luise

Thematic CERN School of Computing on Machine Learning
18th October 2024

CERN

CERN openlab

Table of Content

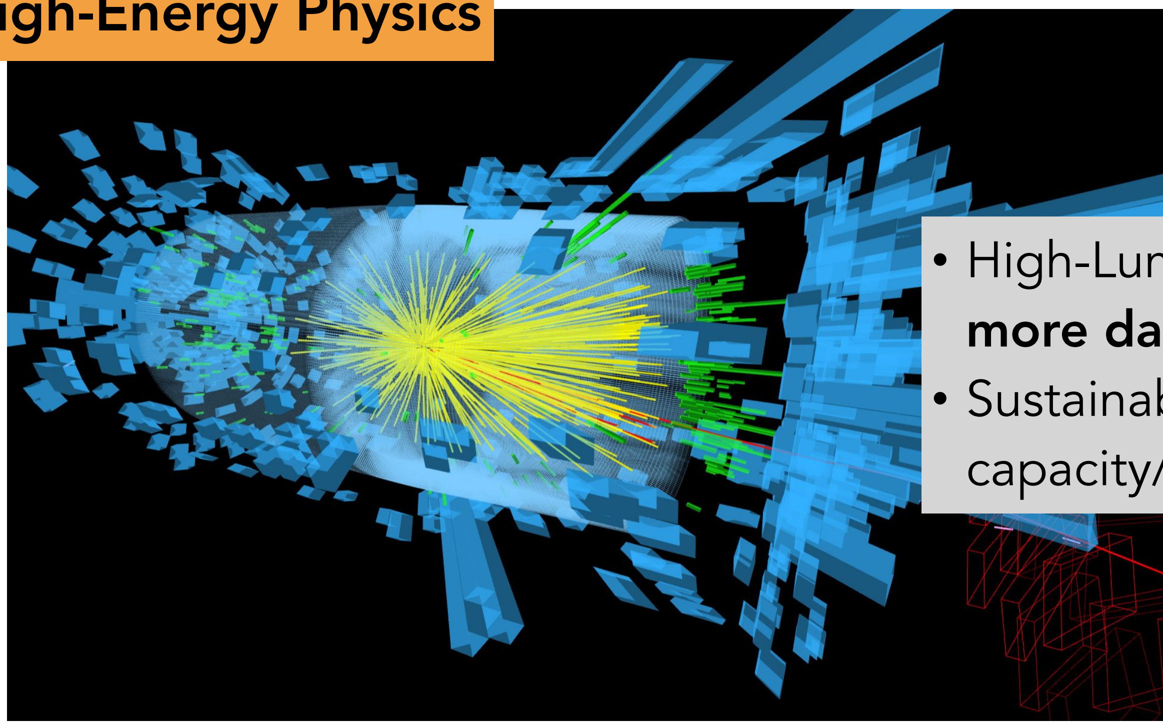
Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

Weather and climate

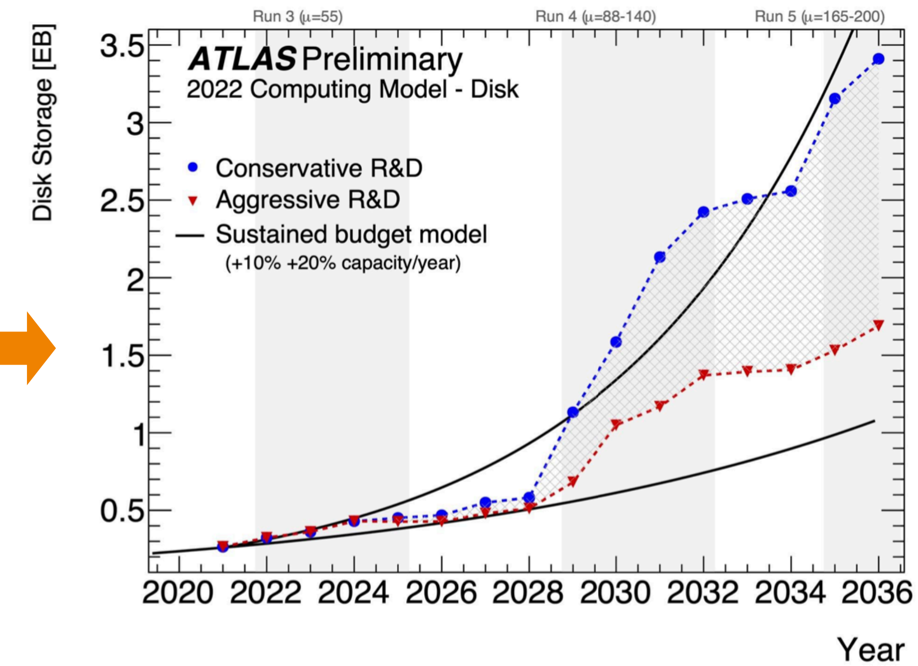
The AI revolution in weather and climate

The future of observational data

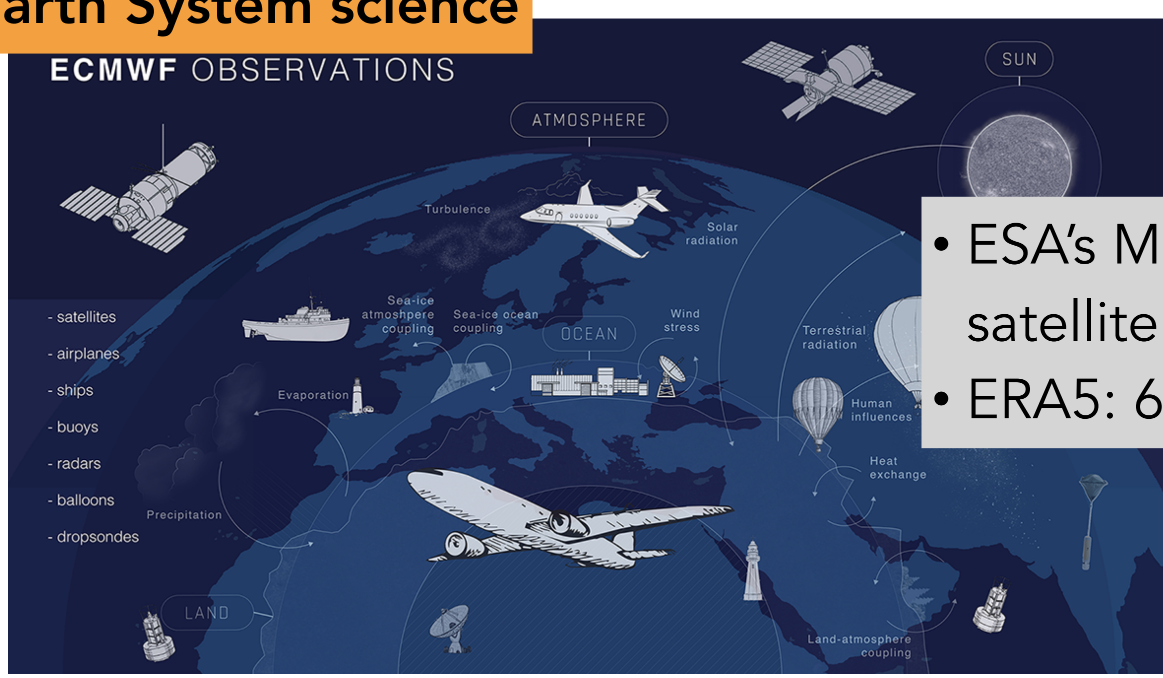
High-Energy Physics



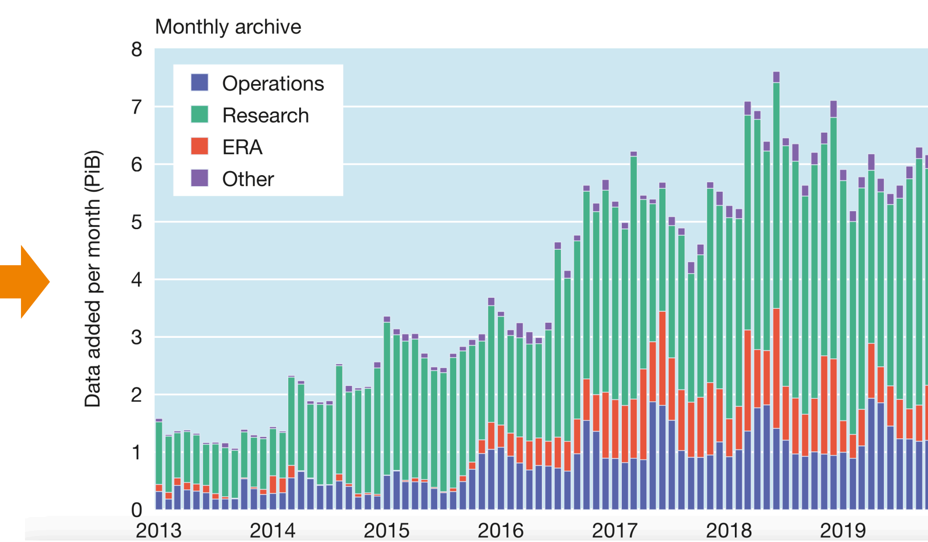
- High-Lumi LHC: **6 times more data by 2030**
- Sustainable model +20% capacity/year



Earth System science



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

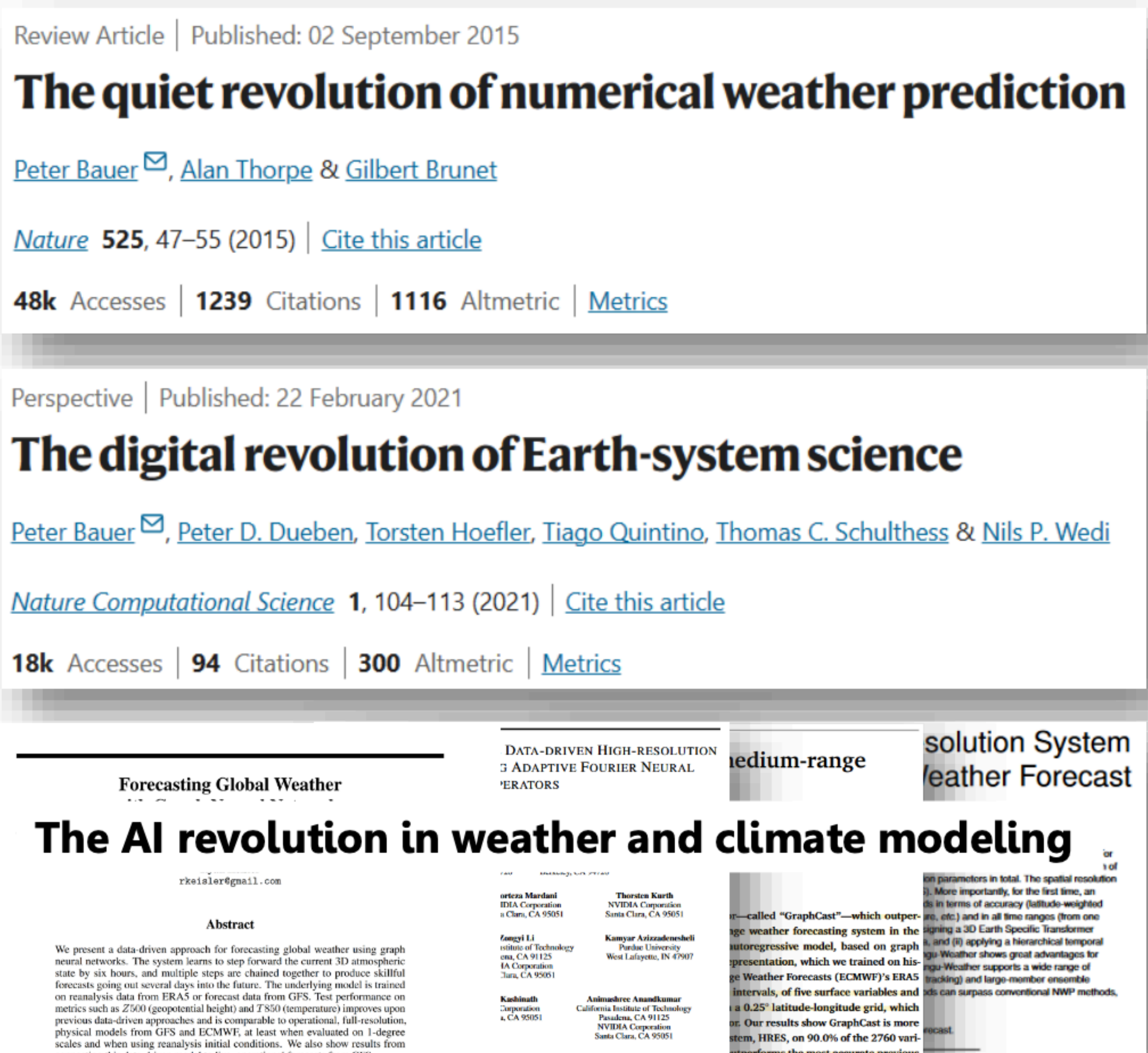


Need to find sustainable ways to store all these data

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

The first breakthrough: weather & climate

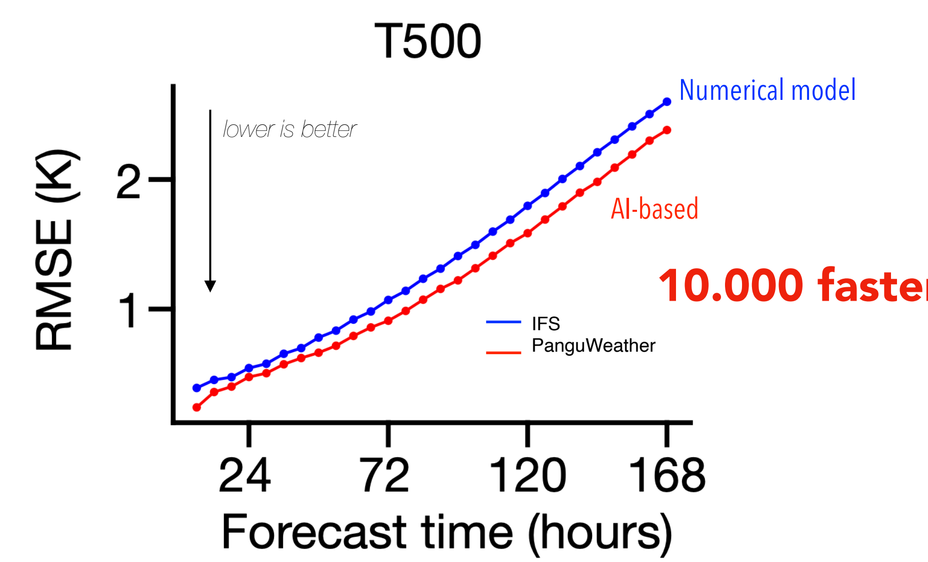
Large datasets:
First time that an AI-model trained on TBs of pre-processed observations outperforms the numerical models for a 10 day forecasts

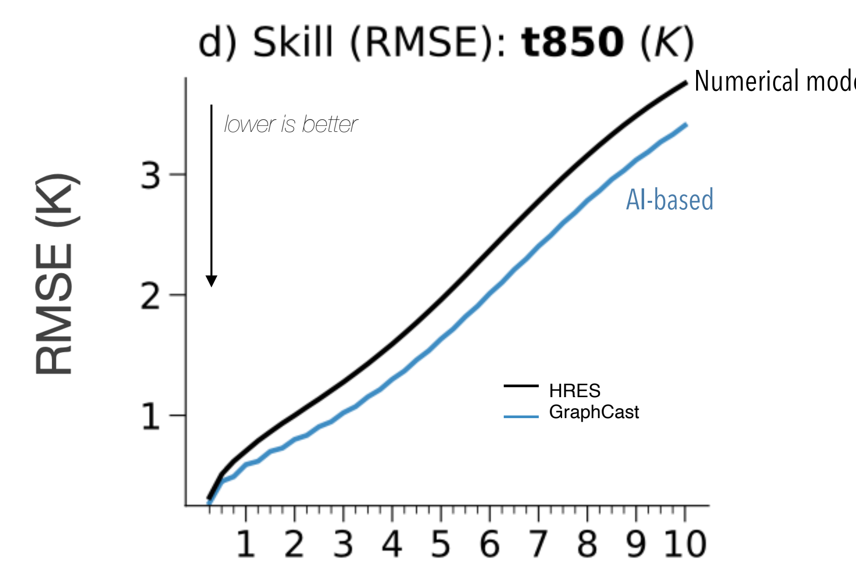


1960-2010

2005-2025

2022-





All these models have been trained on a *single* task: weather forecasting

From Martin Schultz - ESA-ECMWF workshop 2024

State of the art - numerical models: IFS

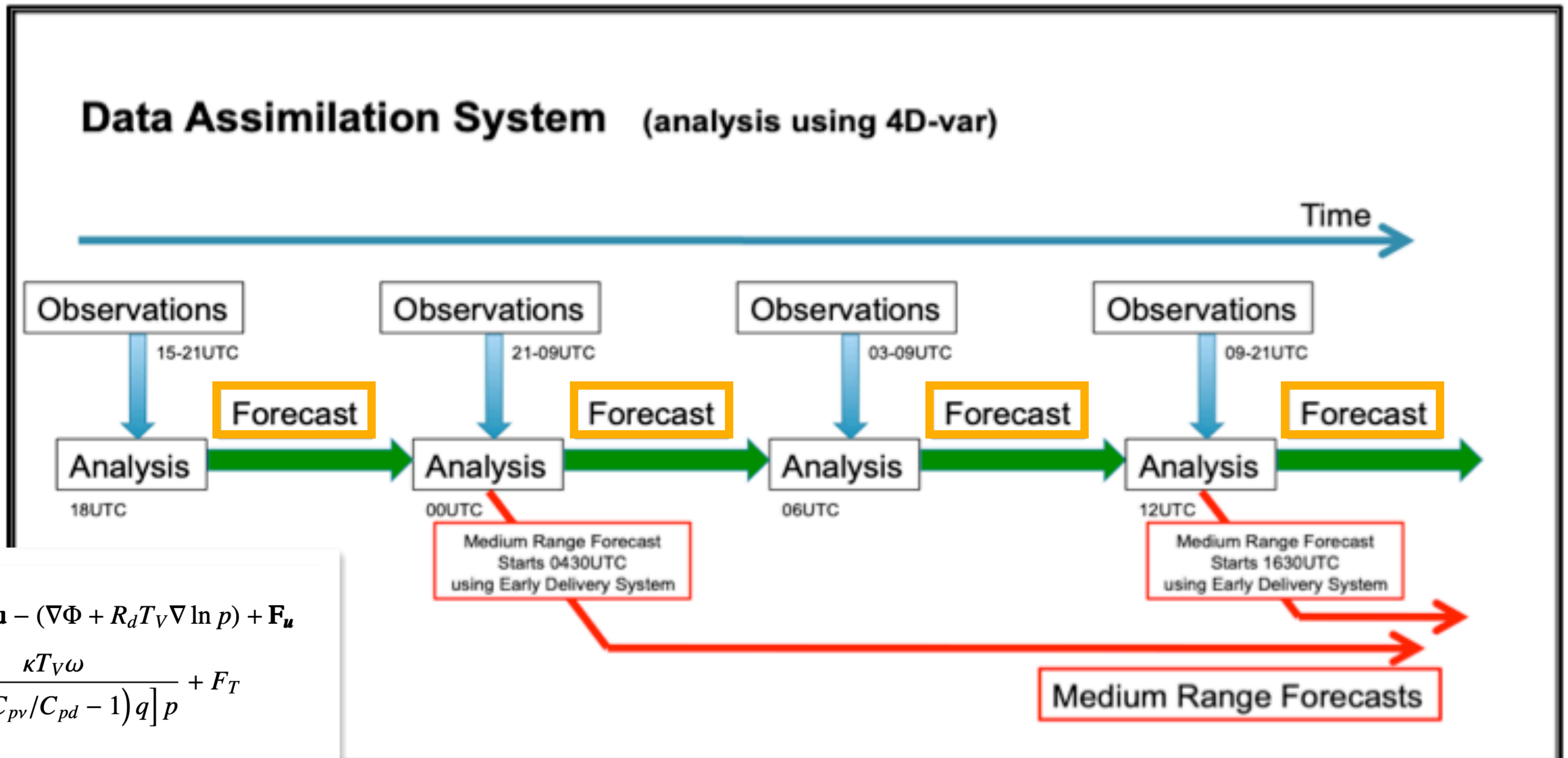
Europe: Integrated Forecasting System (IFS)

Language: **Fortran-based**

Length: **over 2 million lines**

US: Global Forecast System (GFS)

Japan: Global Spectral Model (GSM2303) …



$$\frac{d\mathbf{u}}{dt} = -f\mathbf{z}\times\mathbf{u} - (\nabla\Phi + R_d T_V \nabla \ln p) + \mathbf{F}_{\mathbf{u}}$$

$$\frac{dT}{dt} = \frac{\kappa T_V \omega}{\left[1 + \left(C_{pv}/C_{pd} - 1\right) q\right] p} + F_T$$

$$\frac{dq}{dt} = F_q$$

$$\frac{dq_c}{dt} = F_c$$

$$\frac{\partial}{\partial t}\left(\frac{\partial p}{\partial \eta}\right) + \nabla\cdot\left(\mathbf{u}\frac{\partial p}{\partial \eta}\right) + \frac{\partial}{\partial \eta}\left(\dot{\eta}\frac{\partial p}{\partial \eta}\right) = 0$$

Machine Learning approaches

Language: python + torch

Length: a few thousands lines

From Martin Schultz - ESA-ECMWF workshop 2024

Who? What? Where?

- PanguWeather

HUAWEI

- GraphCast
- NeuralGCM
- MetNet3

Google Research

DeepMind

- FourCastNet

NVIDIA

Microsoft Research

- ClimaX
- Aurora

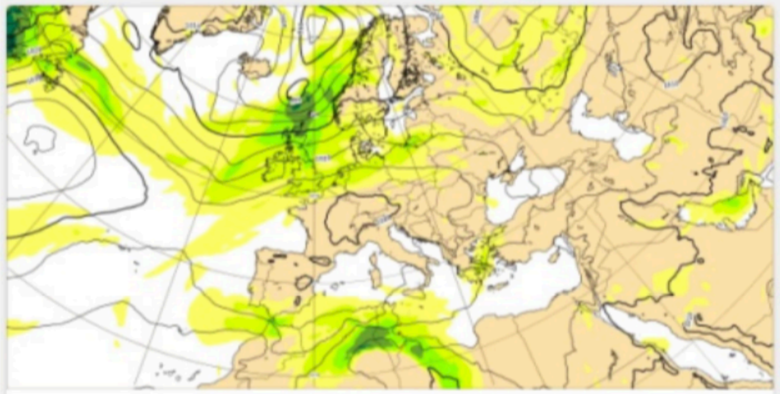
IBM

- Prithvi (earth observations)

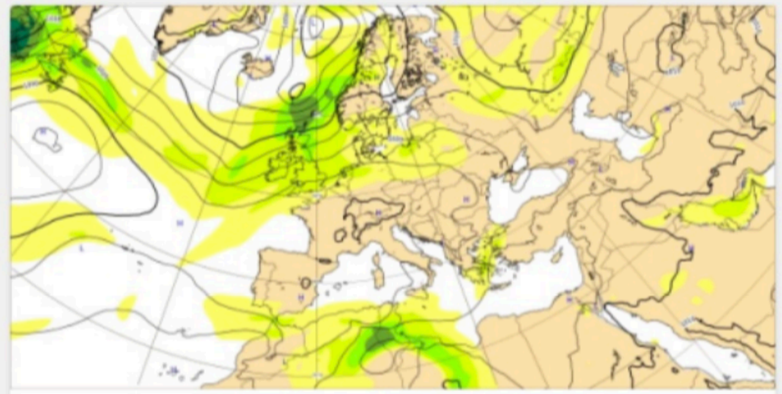
… but what about academia?
Are the research centres following the trend?

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

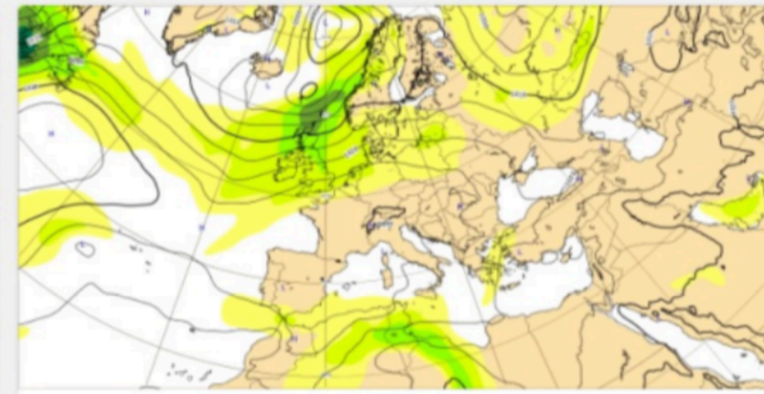
Real time AI-based forecasts @ ECMWF



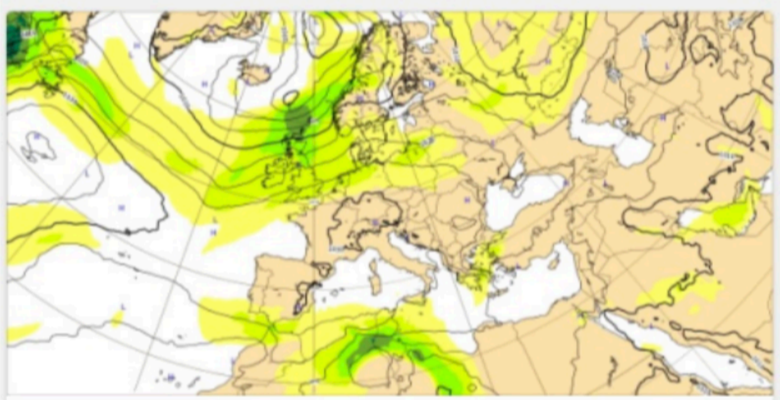
Mean sea level pressure and 850 hPa wind speed



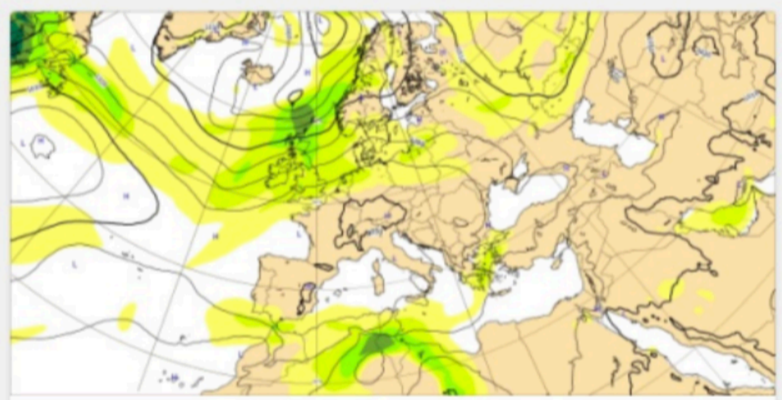
Experimental: AIFS (ECMWF) ML model: Mean sea level pressure and 850 hPa wind speed



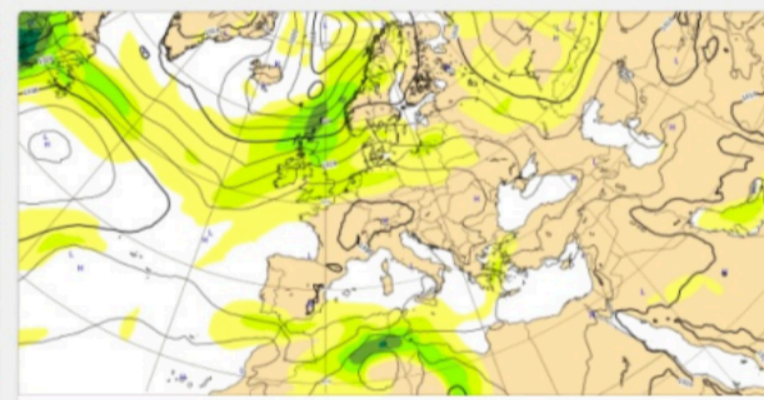
Experimental: FourCastNet ML model: Mean sea level pressure and 850 hPa wind speed



Experimental: FuXi ML model: Mean sea level pressure and 850 hPa wind speed



Experimental: GraphCast ML model: Mean sea level pressure and 850 hPa wind speed



Experimental: Pangu-Weather ML model: Mean sea level pressure and 850 hPa wind speed

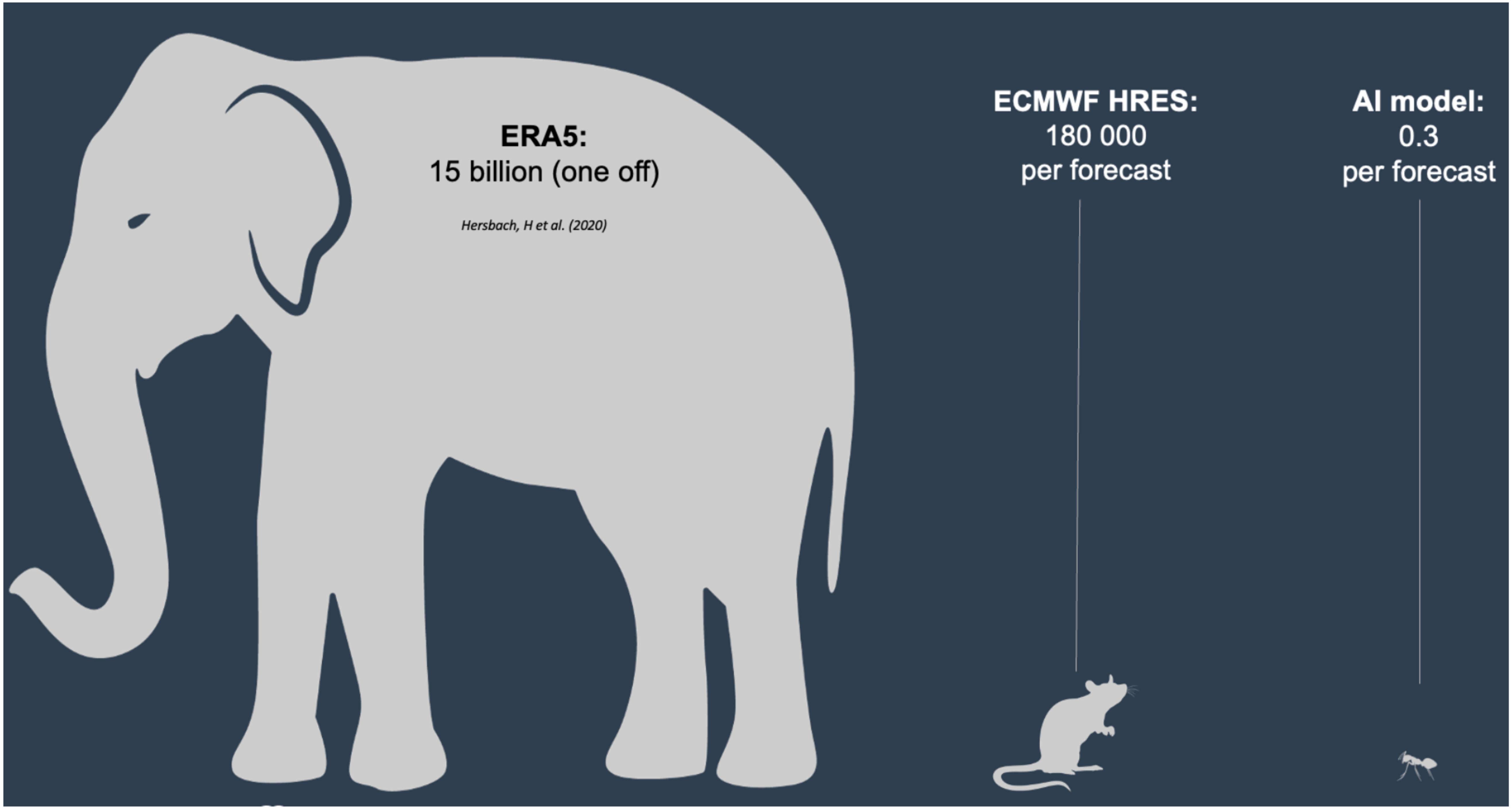
THE RISE OF DATA-DRIVEN WEATHER FORECASTING

A FIRST STATISTICAL ASSESSMENT OF MACHINE LEARNING-BASED WEATHER FORECASTS IN AN OPERATIONAL-LIKE CONTEXT

A PREPRINT V2

https://charts.ecmwf.int

Evaluation costs



Numerical model - processing time: hour(s)

AI-based model inference time: 1 A100 for 2 min for a 10 day forecast ➝ Ensemble: 50x

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

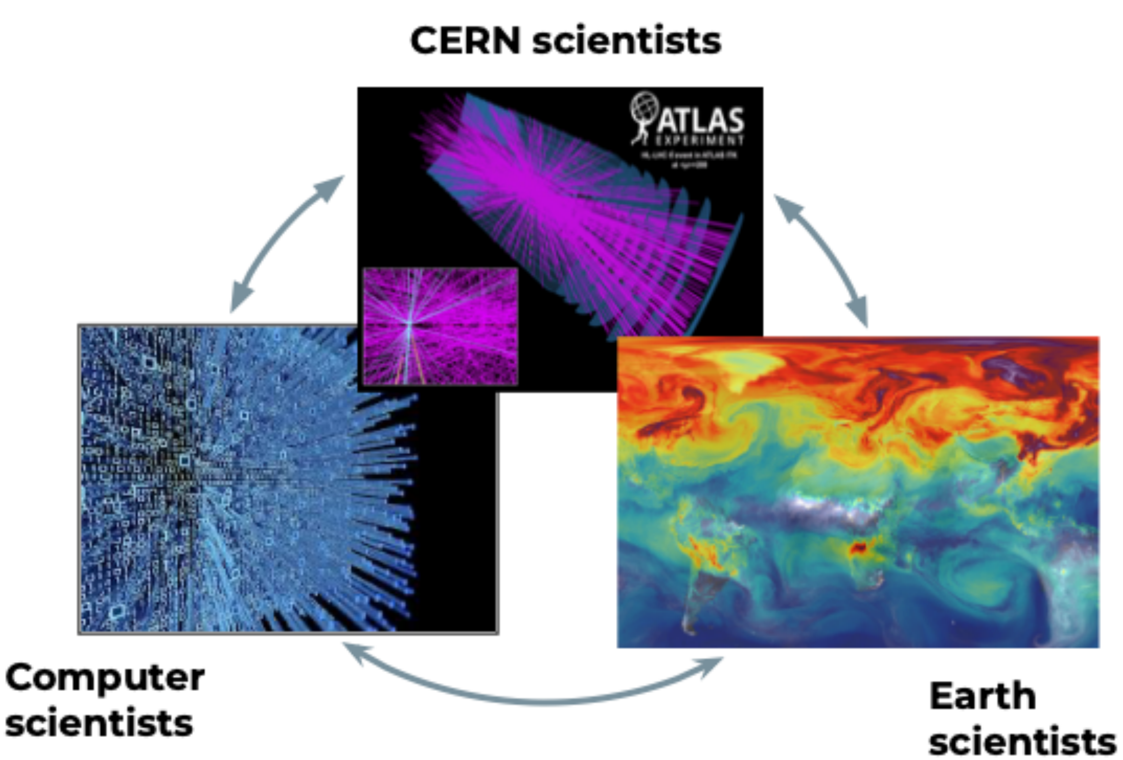
Can we go beyond?

Building foundation models for science

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**



Common challenges:

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:

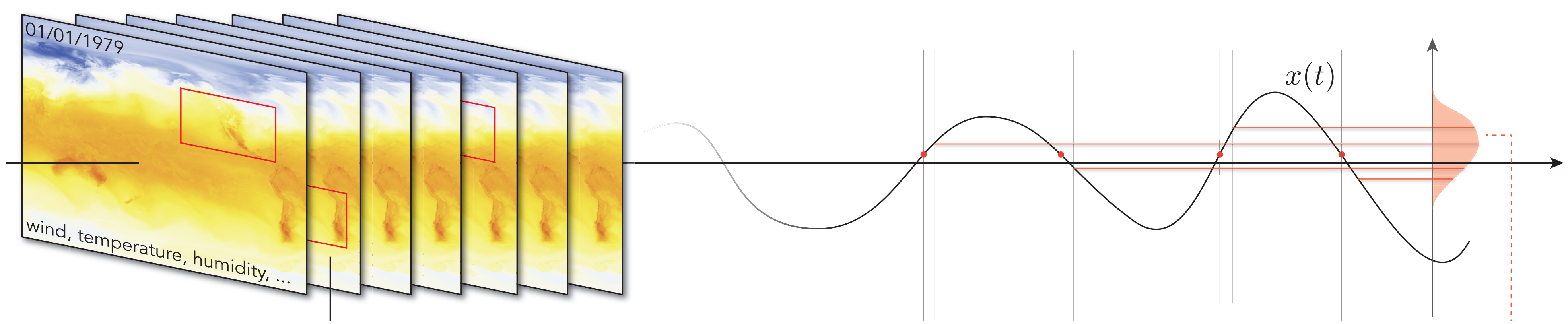
Use unsupervised learning to build a task-independent data-driven model to encapsulate complex physics phenomena

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 → $p(y \mid x, \alpha)$ ← *Auxiliary info: position, absolute time etc..*



Training

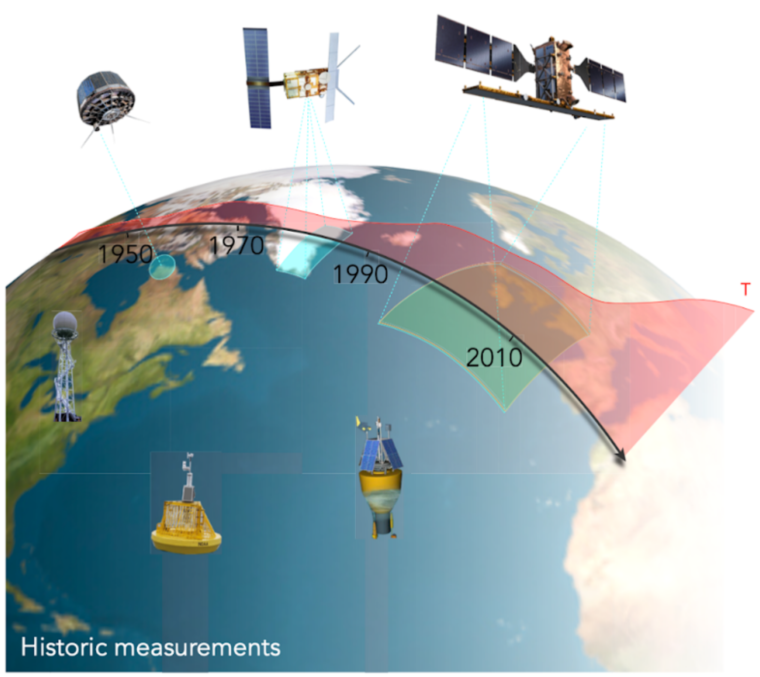
The distribution can be approximated by a large neural network

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

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

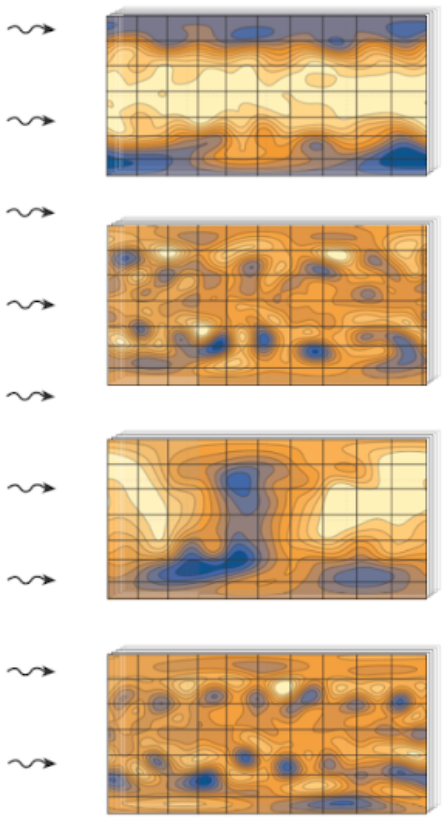
Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

The project in a nutshell



Observations

ERA5 reanalysis

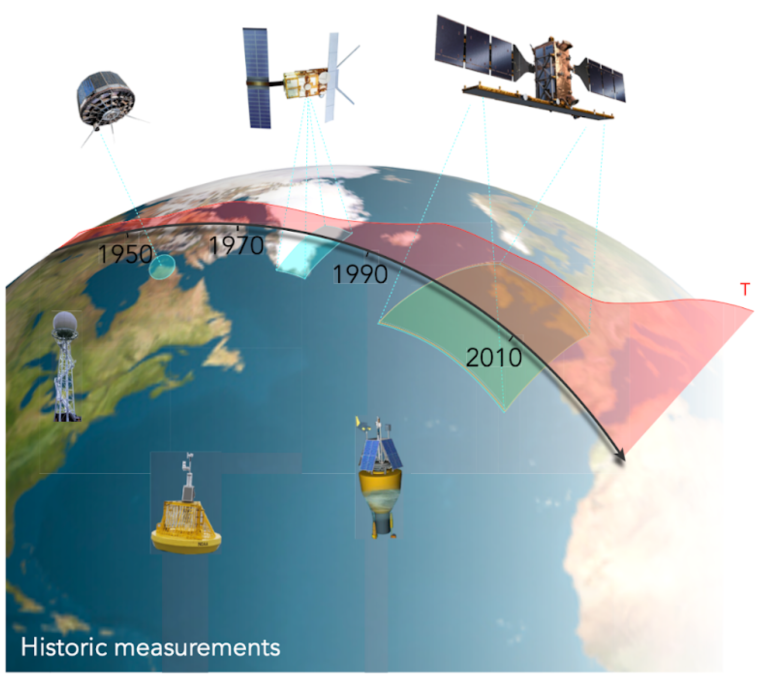


50 TB

Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

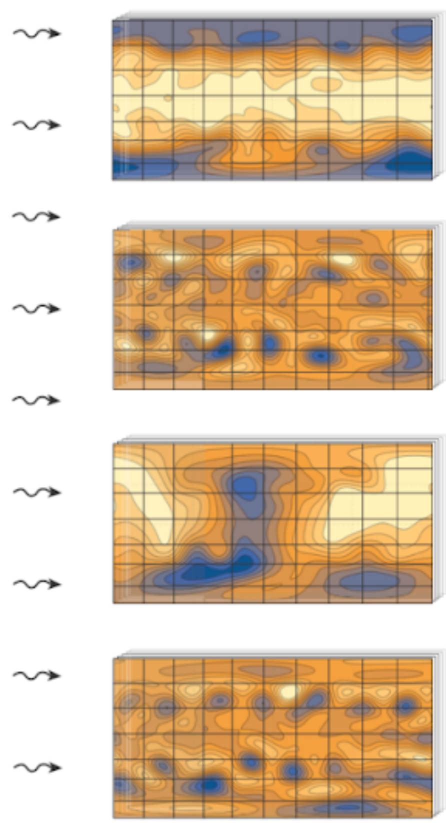
The project in a nutshell

Spatio-temporal representation of atmospheric dynamics

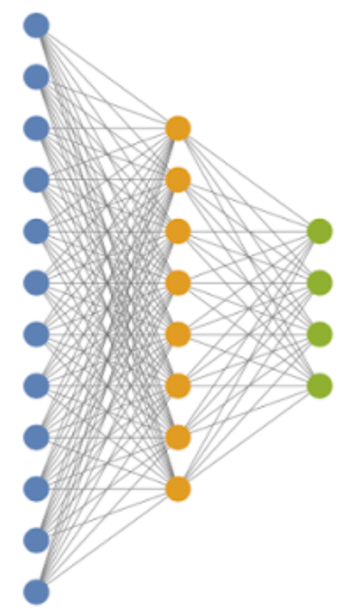


Observations

ERA5 reanalysis

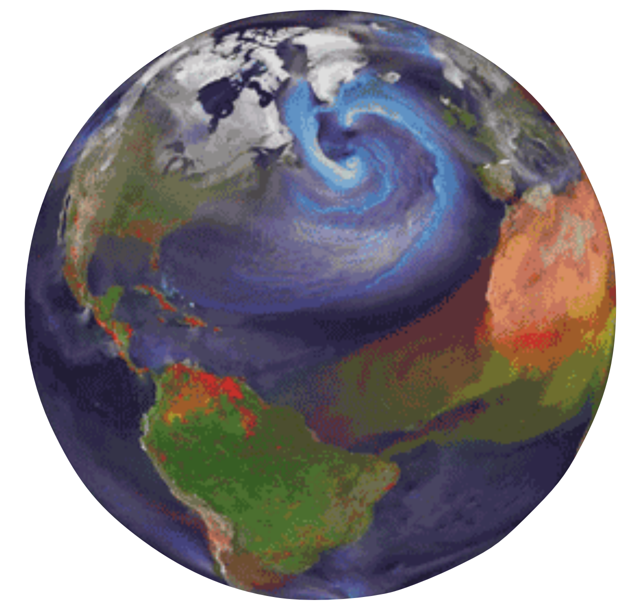


50 TB



large scale machine learning

Transformers architecture



Model
given by the trained neural network

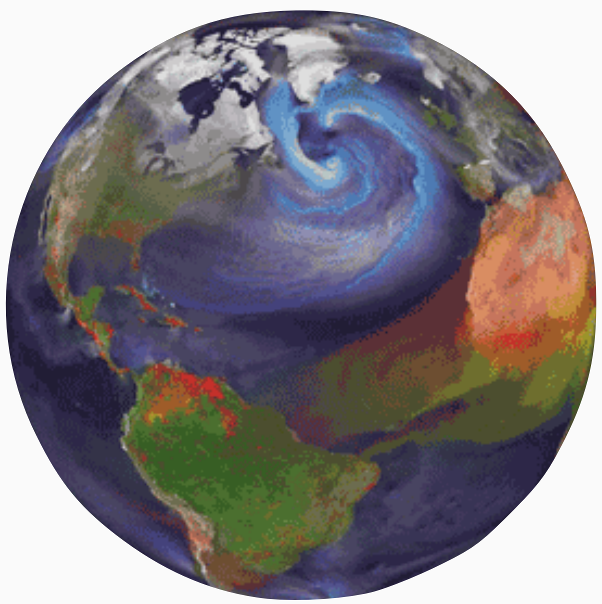
R&D at Juelich SSC:
4x10^{6} GPU hours granted in 2023

JÜLICH Forschungszentrum

Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

Applications: one model for multiple purposes

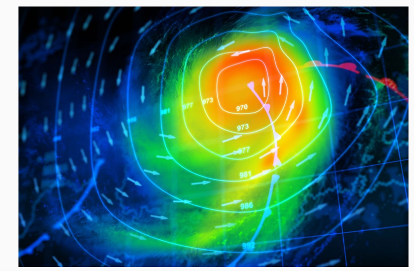
Spatio-temporal representation of atmospheric dynamics



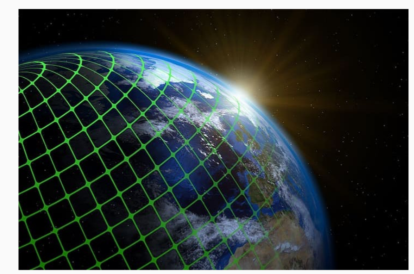
Model
given by the trained neural network

Task-dependent adaptable smaller networks

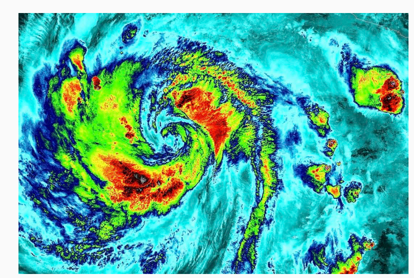
Adaptation



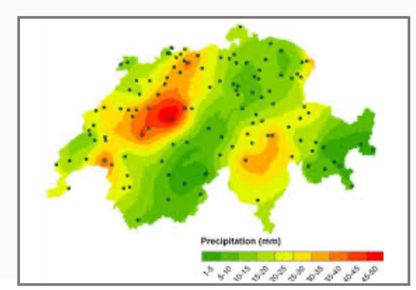
Weather predictions ✓



Downscaling ✓



Bias corrections ✓



Spatio-temporal Interpolations (WIP)

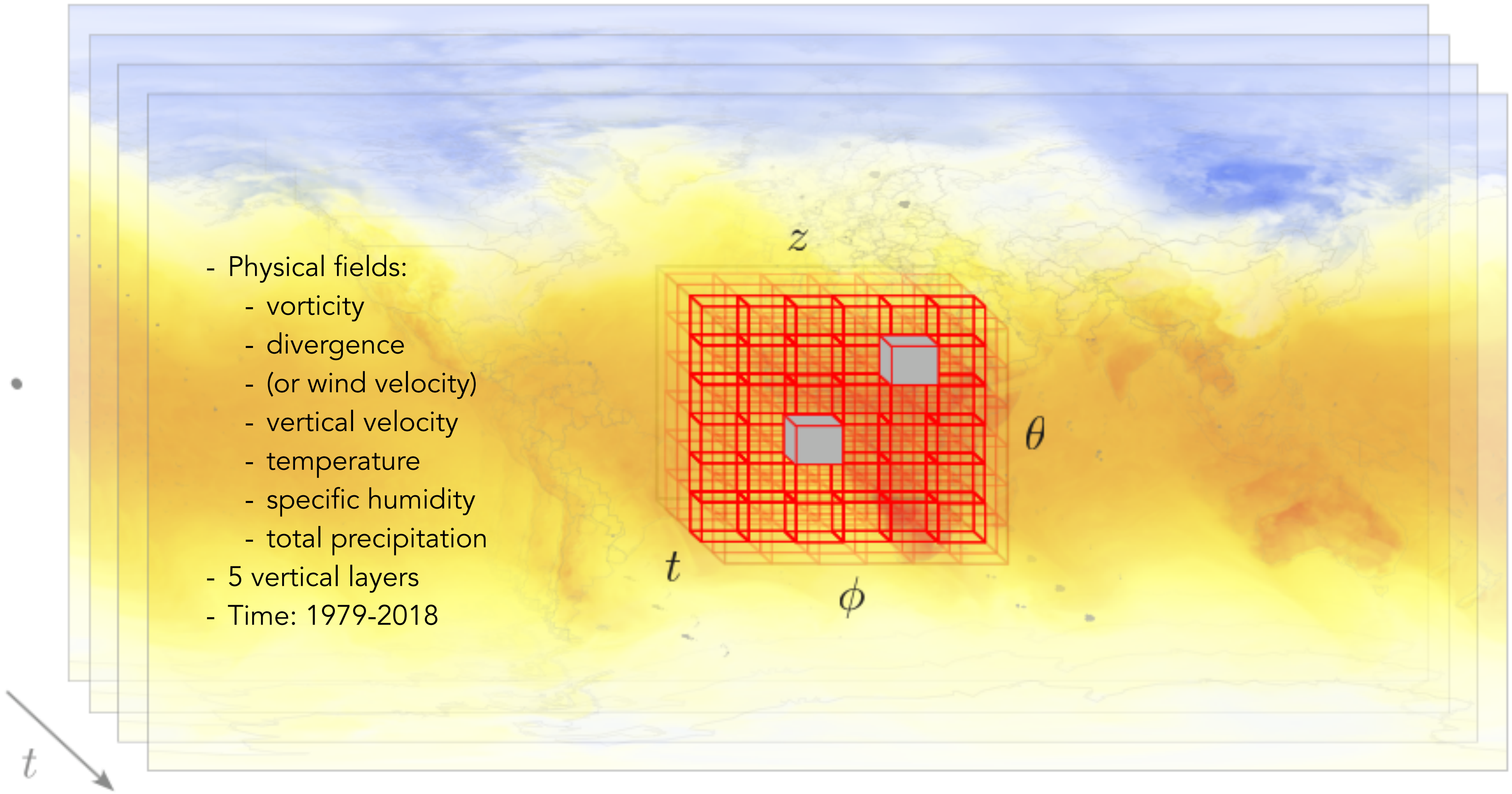
Physics-related applications = uncertainties
*Need for a **stochastic approach***

Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

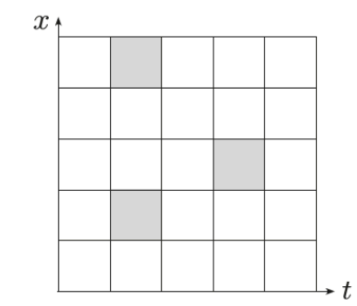
Key Ingredient: The training protocol

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

Random sampling of neighbourhoods for training



BERT



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

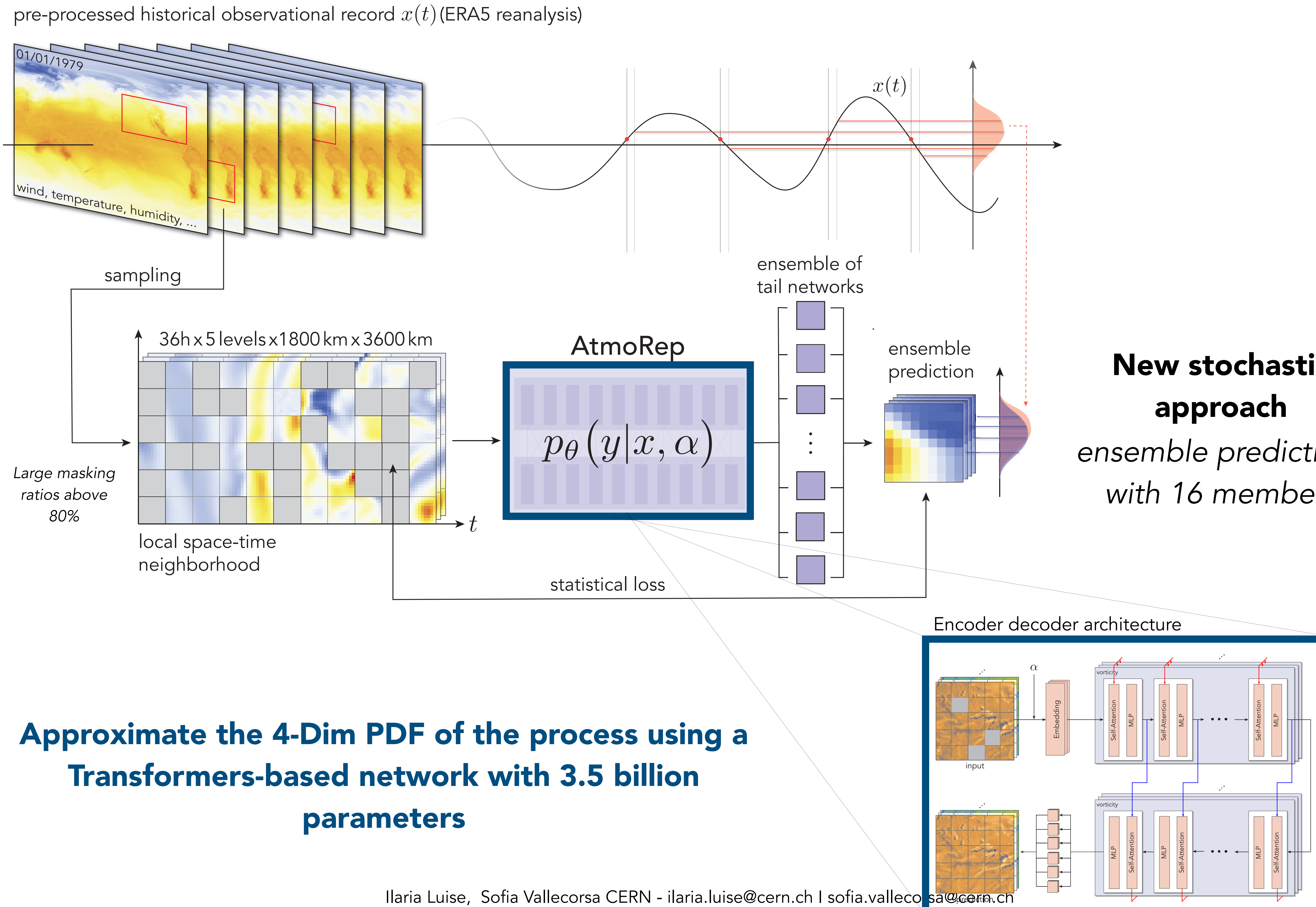
Mask random tokens within the hyper-cube and predict them

Large masking ratios above 80% using full masking, noise and climatology

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

Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@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

Task-specific fine-tuning

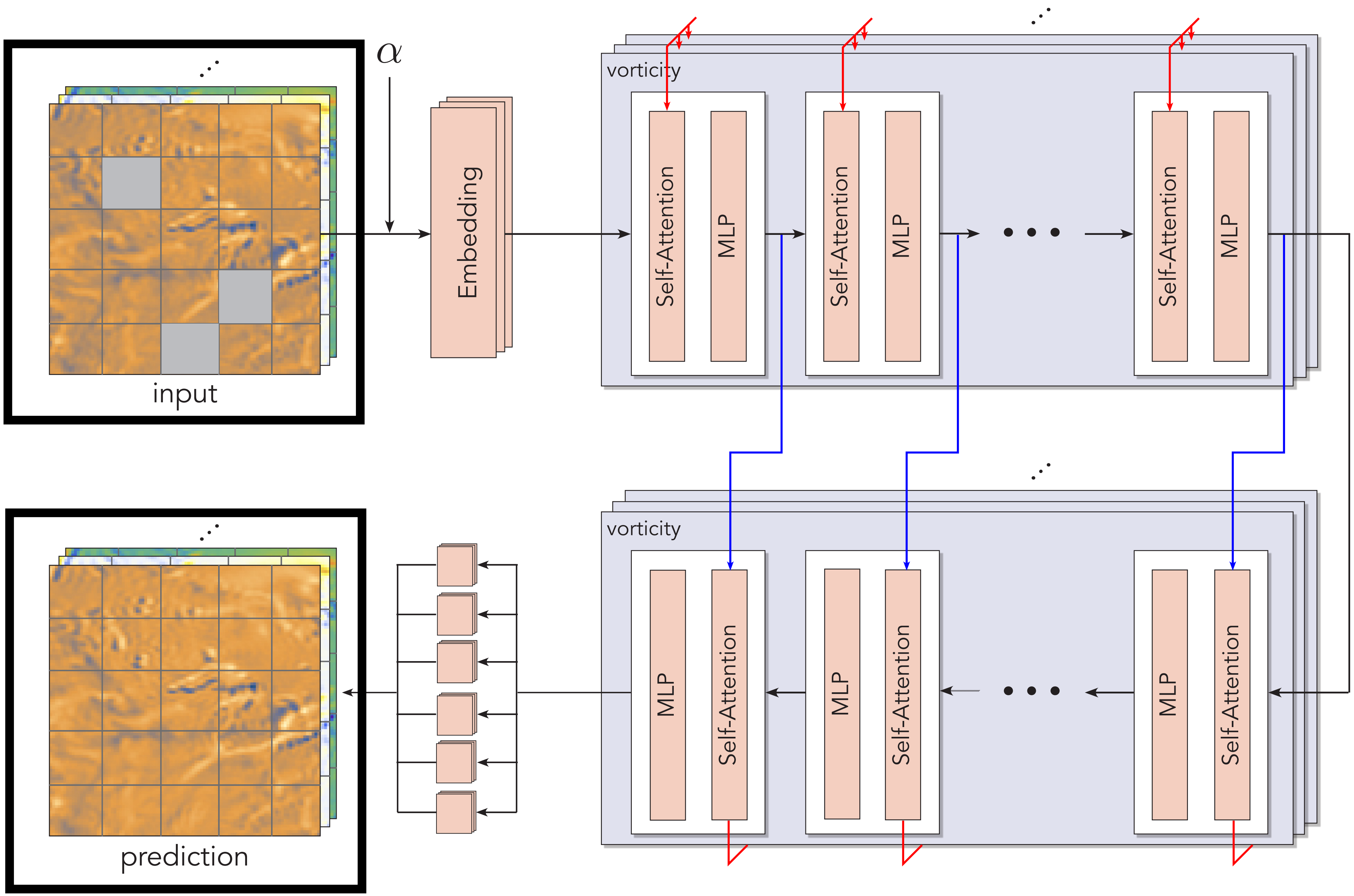
Goal: improve model performance for a specific task
e.g. forecasting, downscaling…

Examples:

e.g. fix masking scheme

OR

Change target dataset



Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

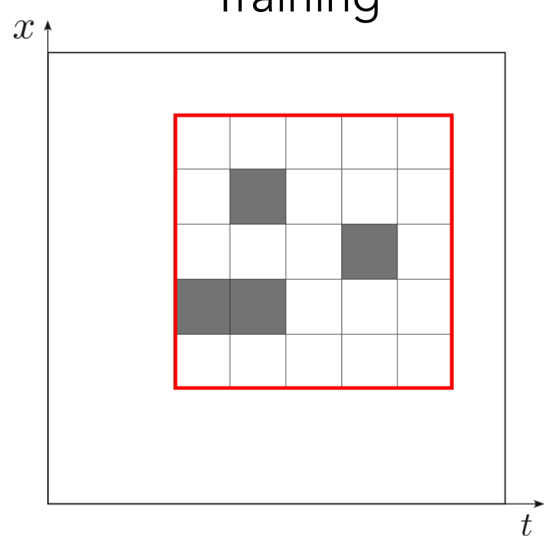
Task-specific fine-tuning

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

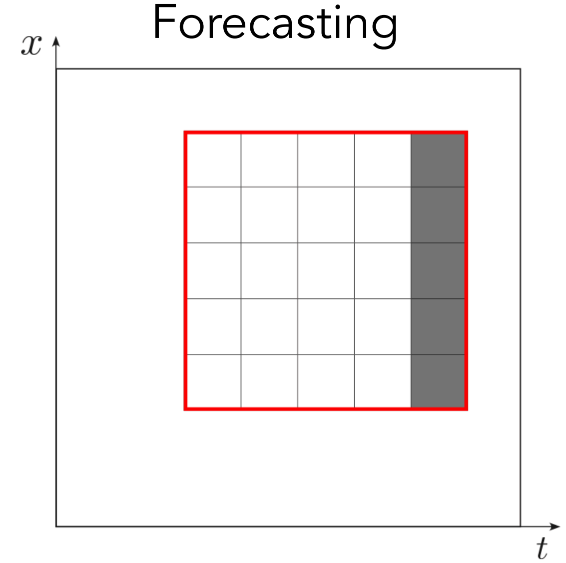
Examples:

e.g. fix masking scheme

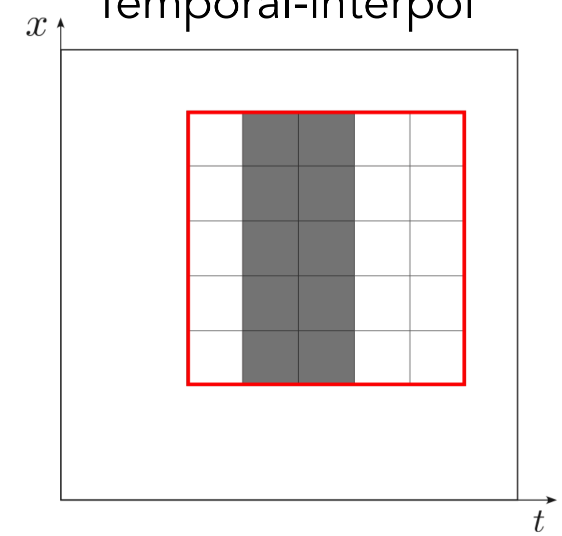
Training



Forecasting

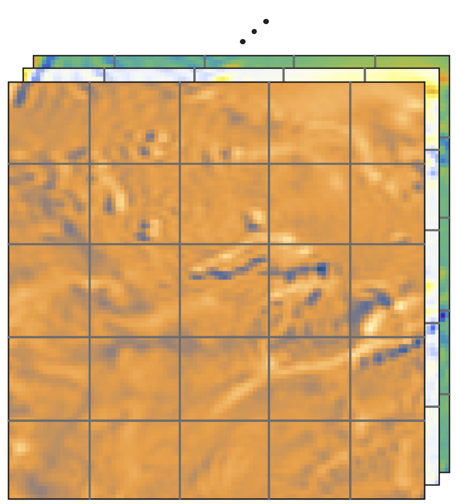


Temporal-interpol



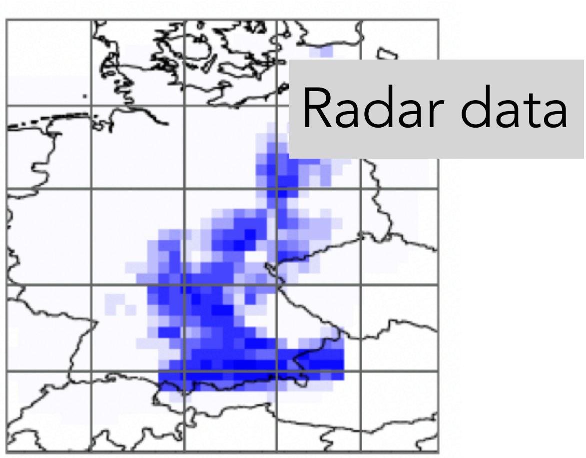
OR

Change target dataset



ERA5

Radklim data

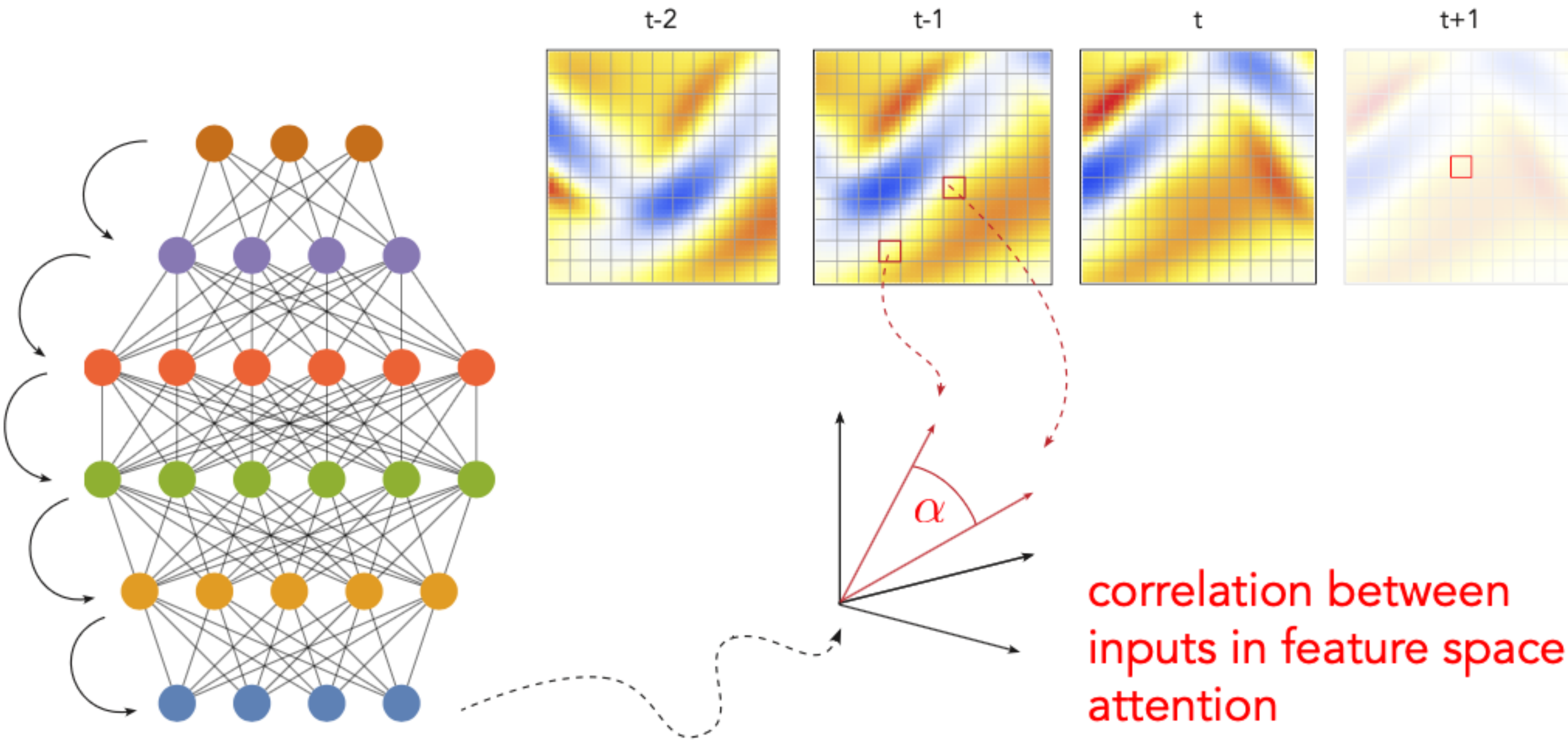


COSMO-REA6

Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

Attention maps and interpretability

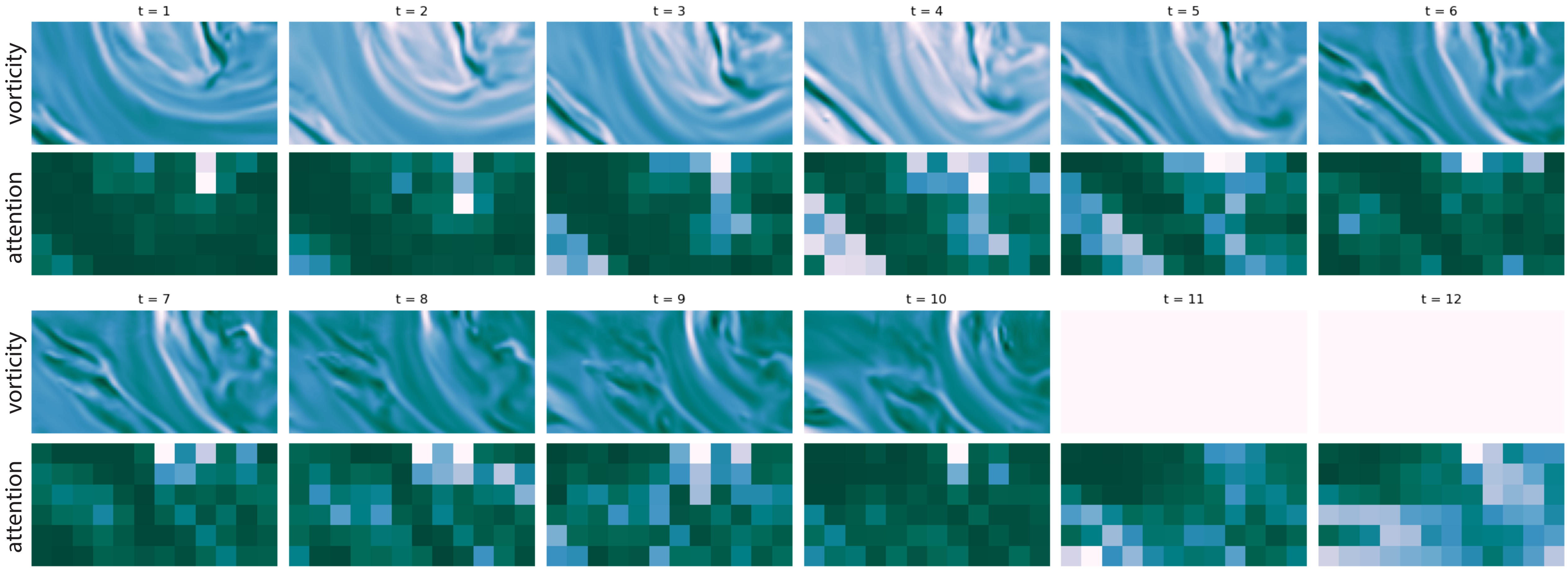
Inspect the self-attention mechanism:
can we identify physics phenomena (e.g. hurricane formation) before they are even created?



Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

Attention maps and interpretability

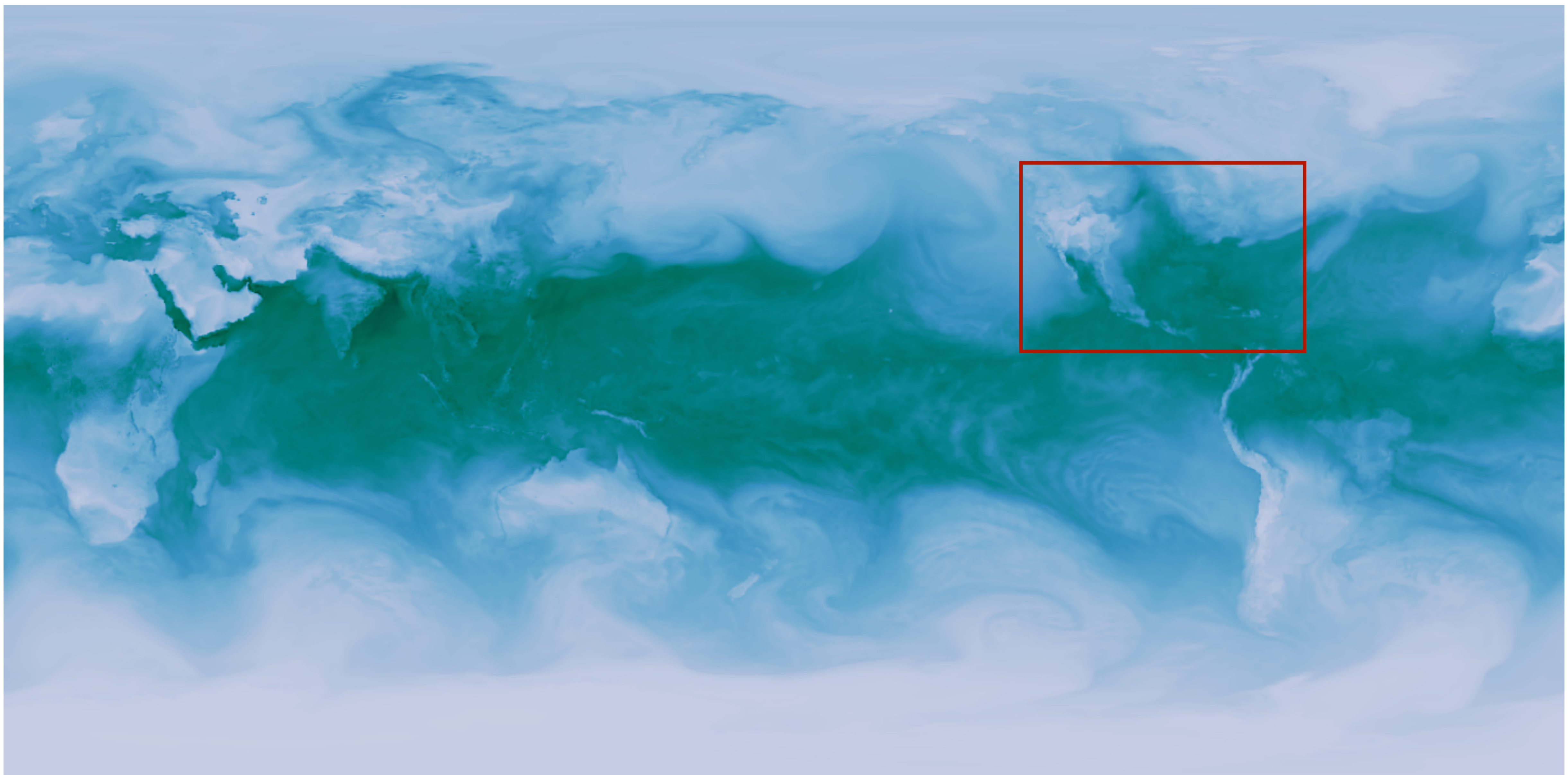
Inspect the self-attention mechanism:
can we identify physics phenomena (e.g. hurricane formation) before they are even created?



Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

Results: Target - ERA5

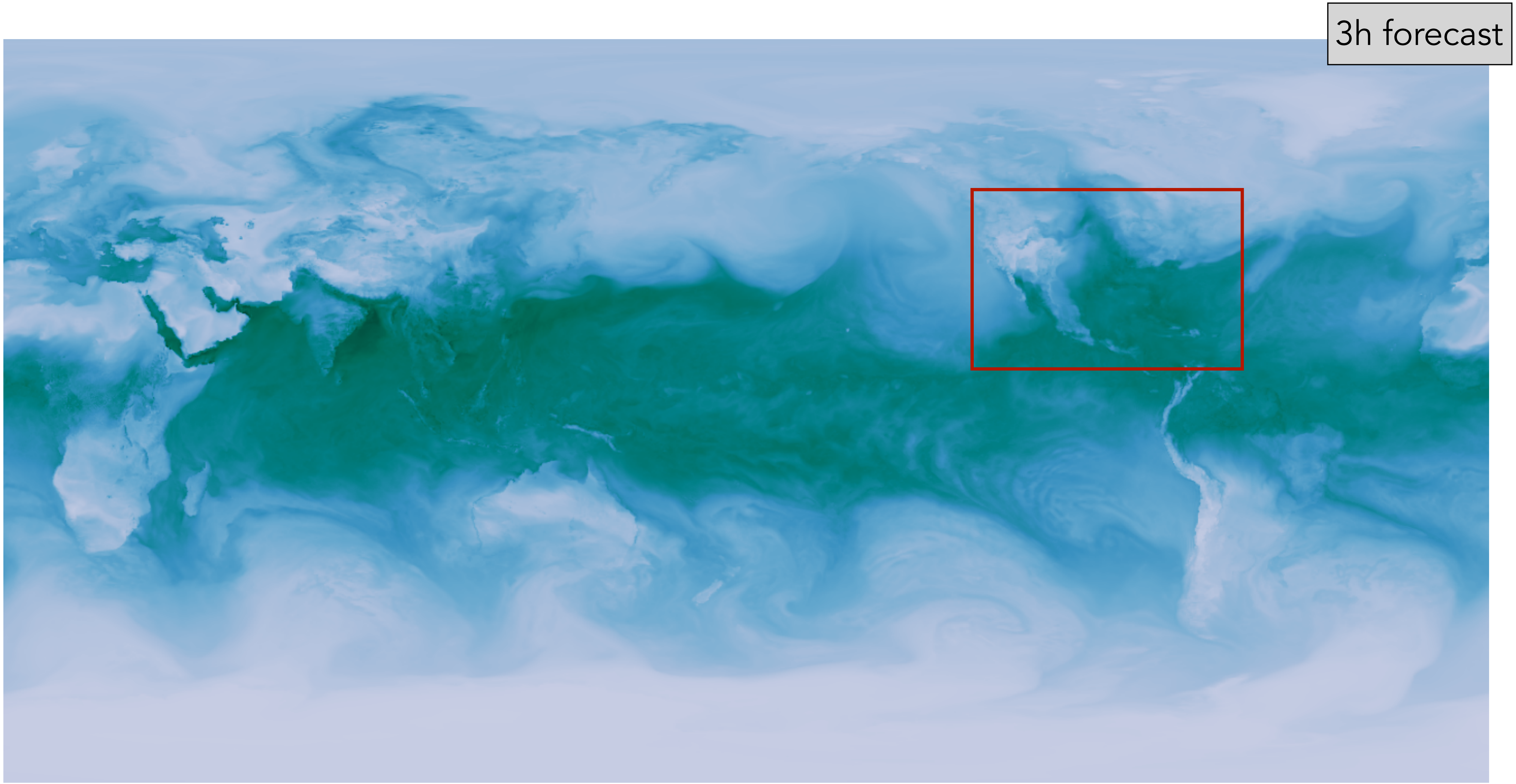
specific humidity, June 15th 2018 13:00 UTC



Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

Results: Prediction - AtmoRep

specific humidity, June 15th 2018 13:00 UTC

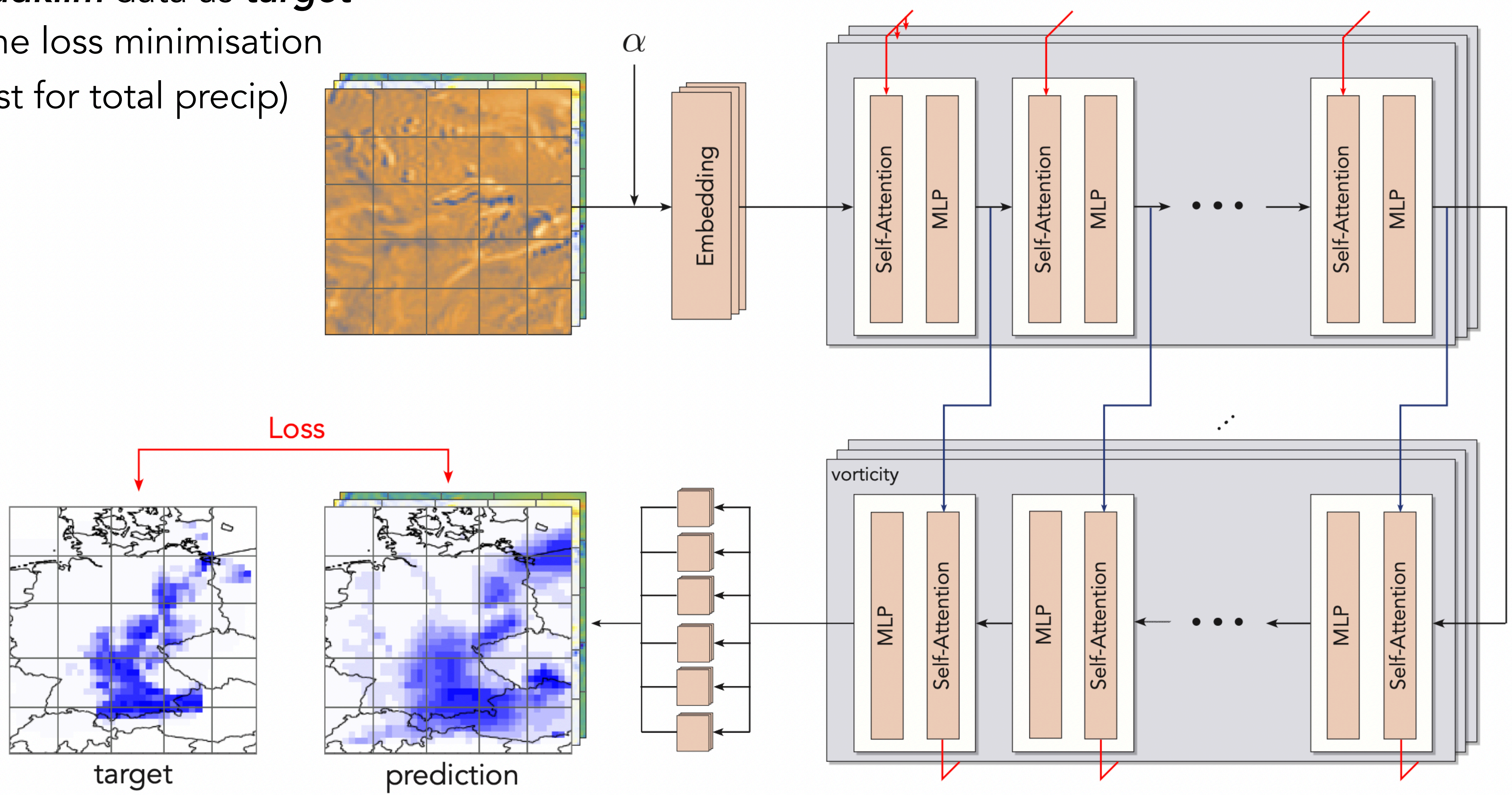


Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

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)



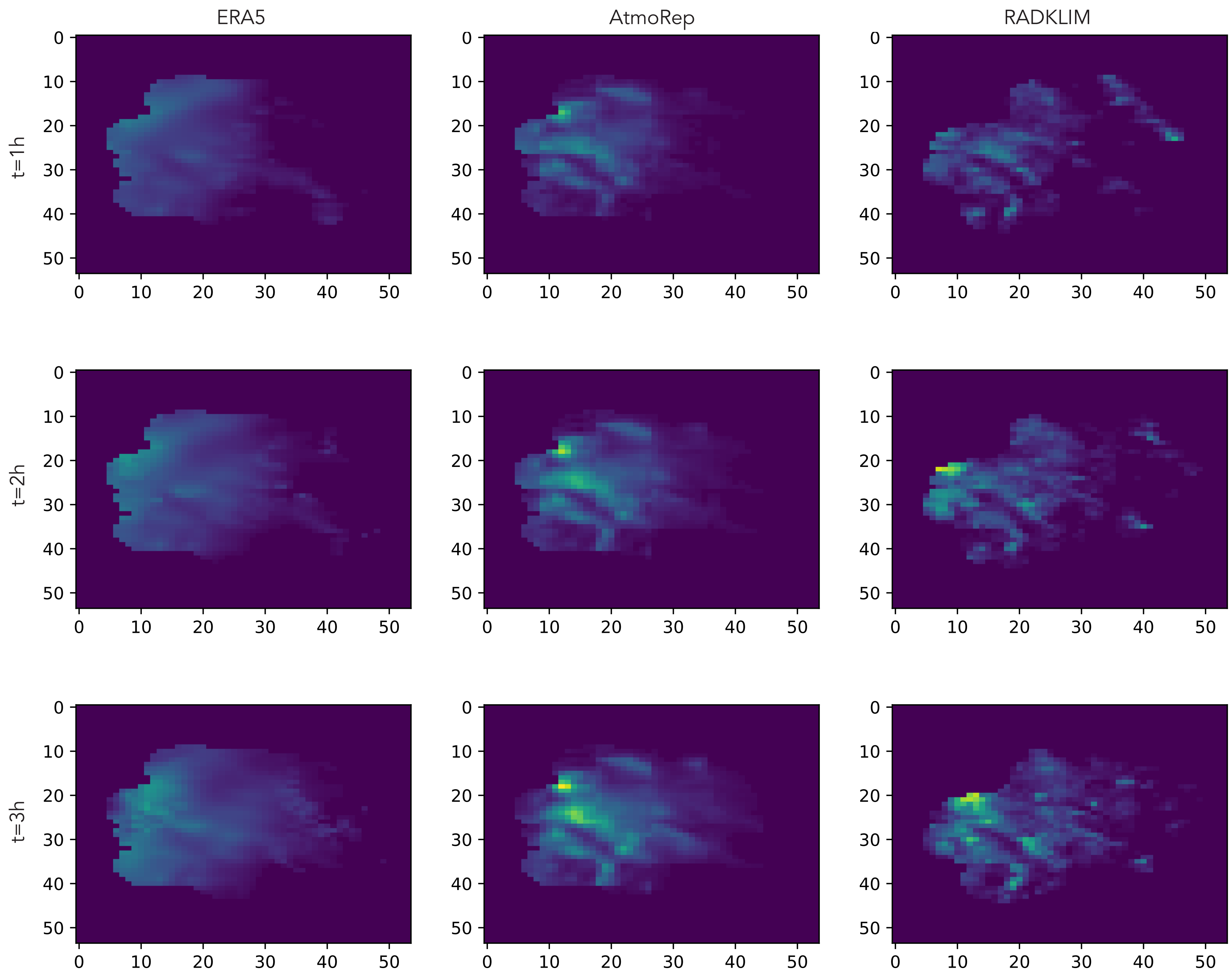
Radklim

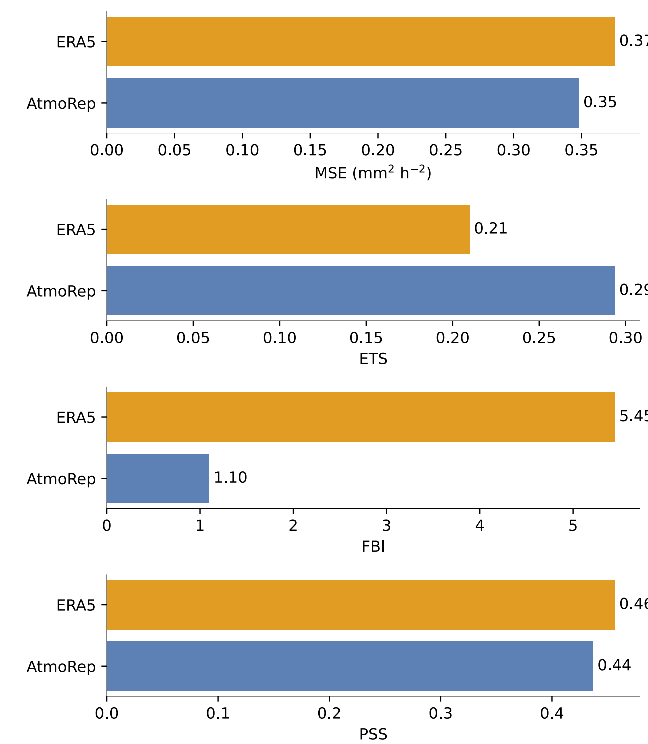
AtmoRep

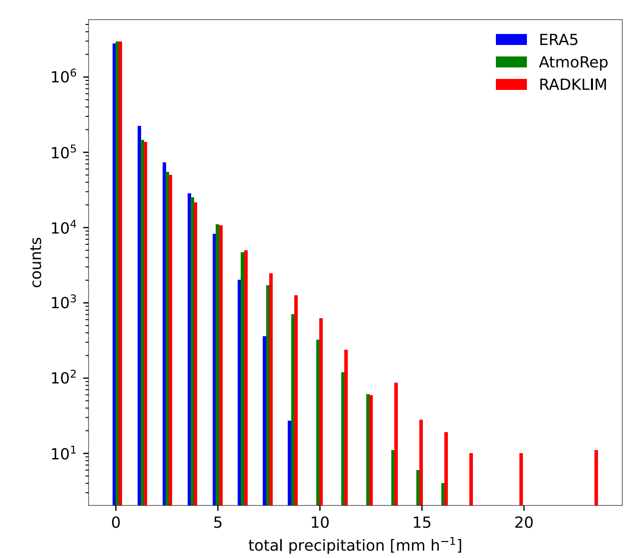
Ilaria Luise, CERN - ilaria.luise@cern.ch

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, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

Quantum Machine Learning

The CERN Quantum Technology Initiative was launched in 2020

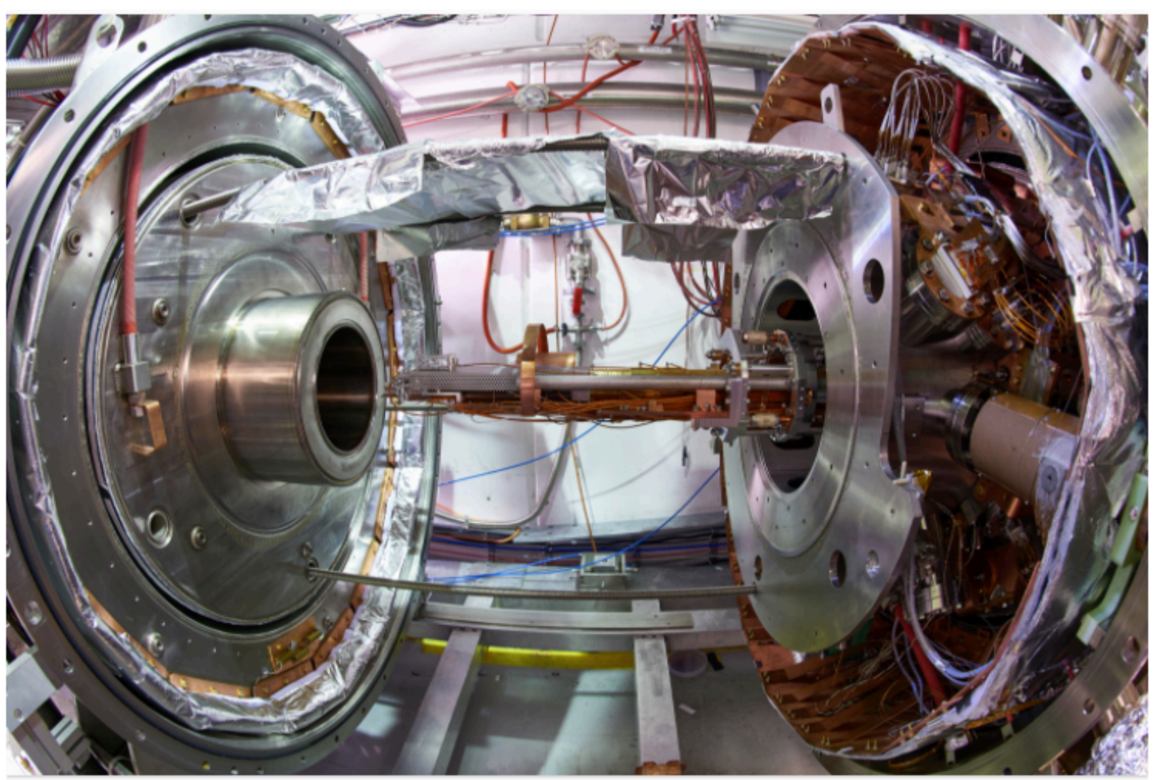
Understand the impact of quantum technologies in HEP

Voir en *français*

CERN meets quantum technology

The CERN Quantum Technology Initiative will explore the potential of devices harnessing perplexing quantum phenomena such as entanglement to enrich and expand its challenging research programme

30 SEPTEMBER, 2020 | By Matthew Chalmers



The AEgIS 1T antimatter trap stack. CERN's AEgIS experiment is able to explore the multi-particle entangled nature of photons from positronium annihilation, and is one of several examples of existing CERN research with relevance to quantum technologies. (Image: CERN)

Quantum simulation and HEP theory applications

Quantum Computing

Quantum Sensing

Quantum Communication

Quantum potential…

Source: McKinsey 2023
https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/quantum-technology-sees-record-investments-progress-on-talent-gap

Principles of quantum mechanics enhance computations

Superposition leads to parallelism

→ **exponential speedup?**

Entanglement

→ **non linear correlation and classical intractability?**

Operations (gates) are unitary transformations

→ **reversible computing?**

Output is the result of a quantum state measurement according to Born rule

→ **stochastic computation ?**

No-cloning theorem

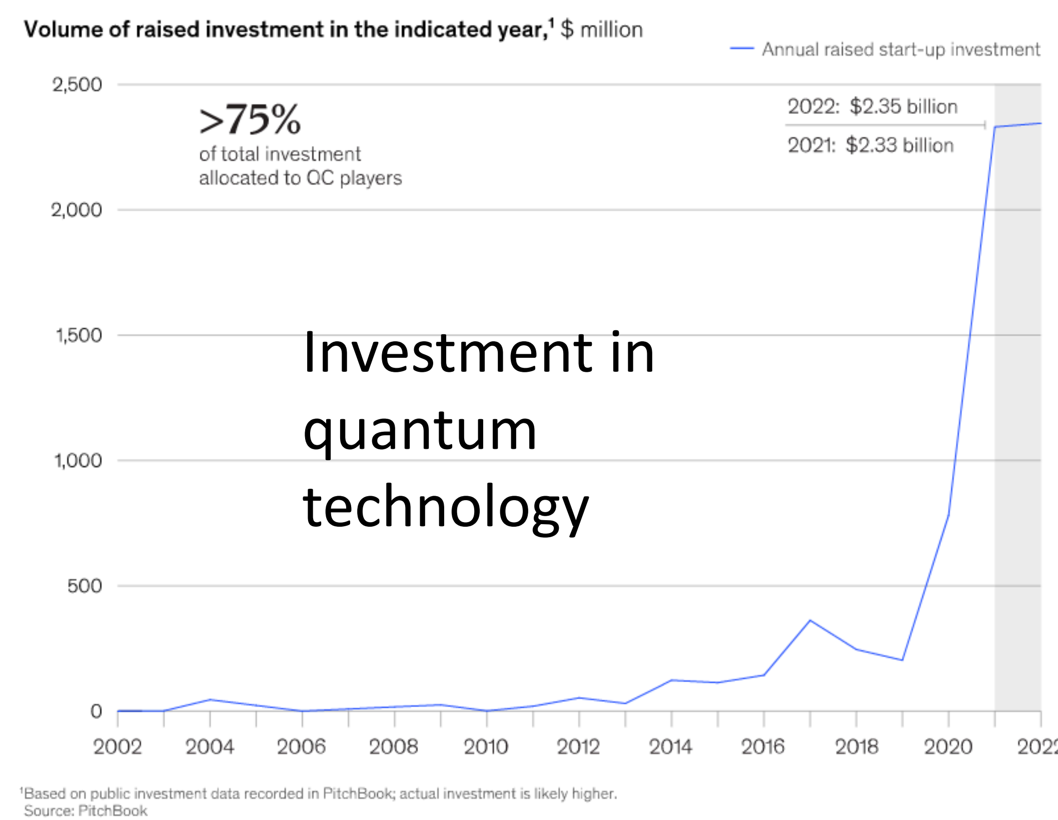
→ **information security**

Quantum state coherence and isolation

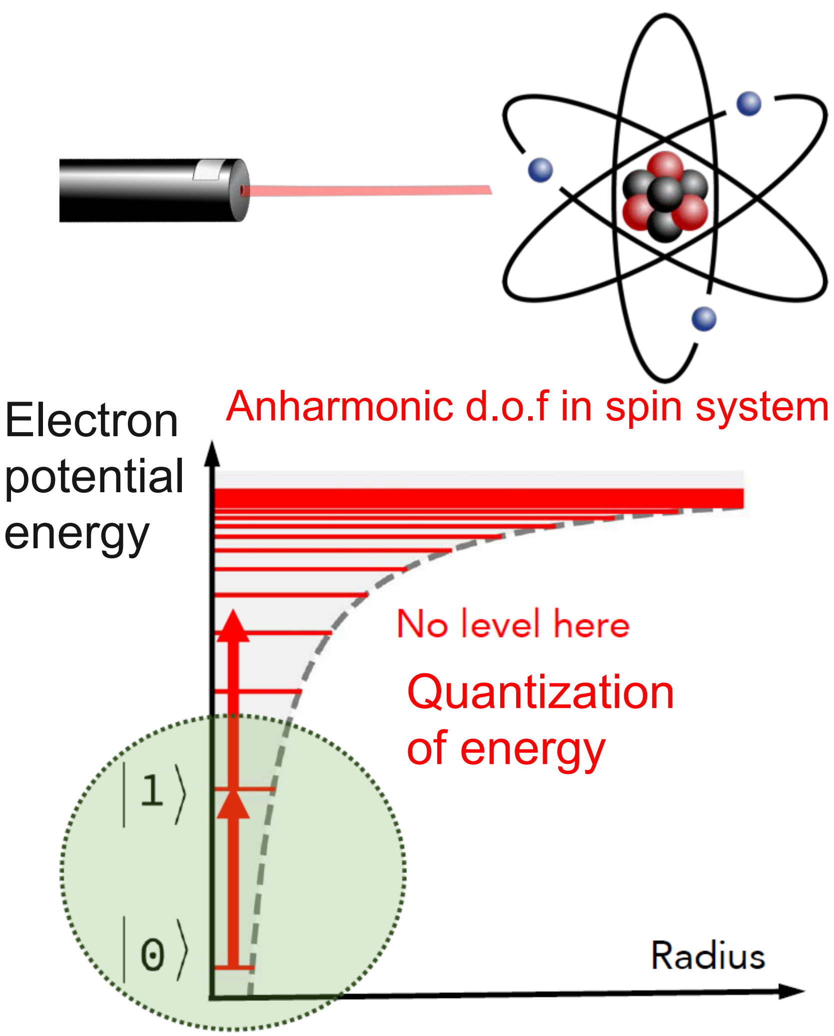
→ **computation stability and errors**

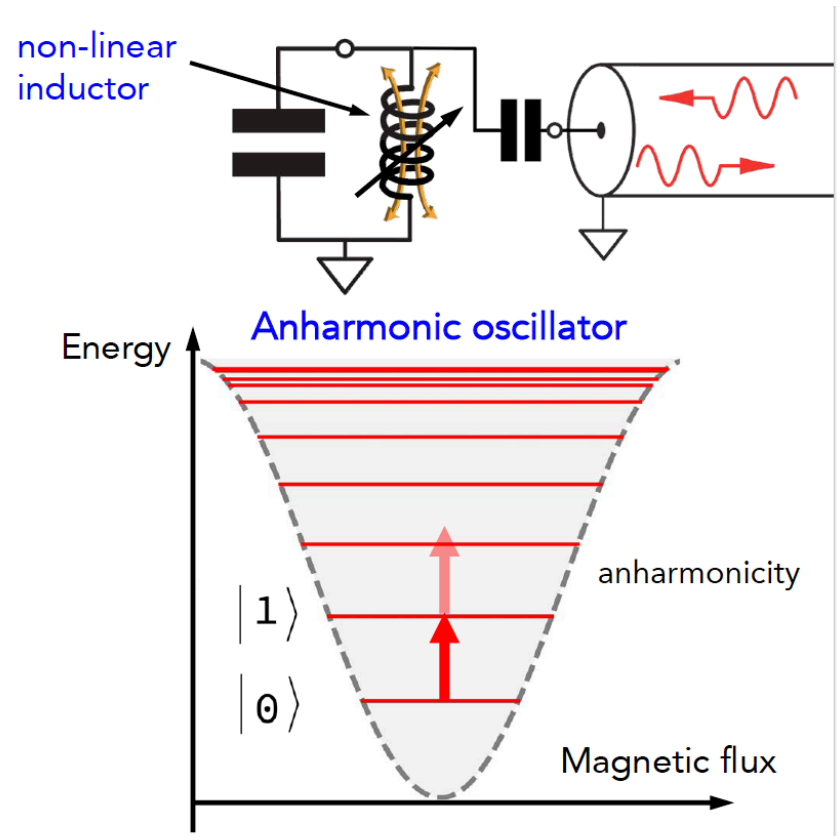
Qubit state collapses

→ **reproducibility?**



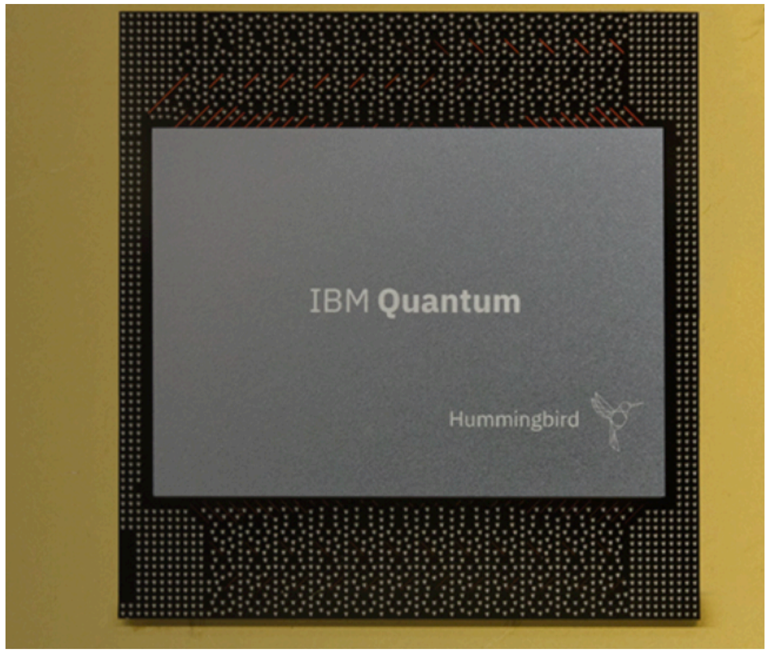
Simple qubits





Z. Minev, Qiskit Global Summer School 2020

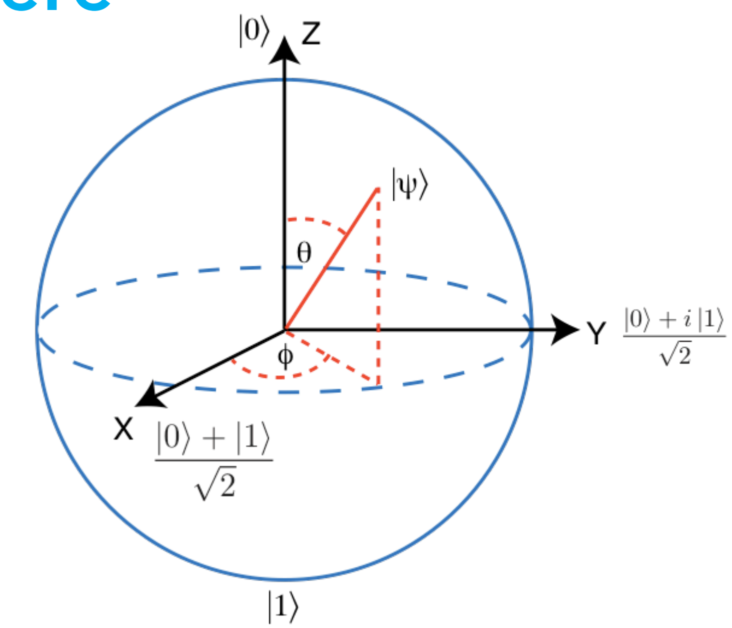
- In zero resistenze circuit, current will oscillate
- Microwaves can excite currents in a superposition of states



Ex. Google, IBM, …

Qubits and algorithms

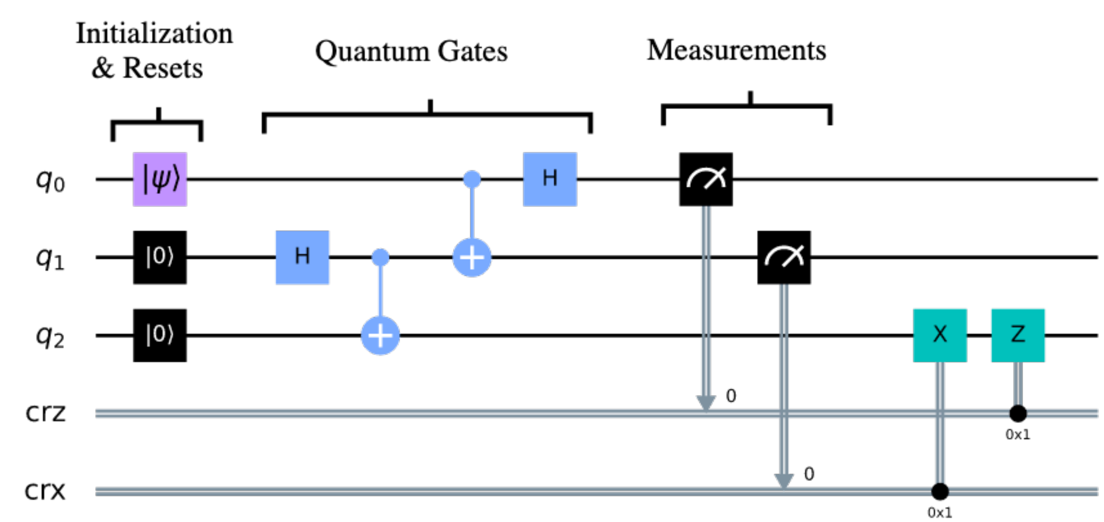
Bloch Sphere



$$|\psi\rangle = \cos\frac{\theta}{2}\,|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}\,|1\rangle$$

- Basic Unit of Quantum Computation
  - **Classical bits are binary** "0 or 1"
- Quantum Mechanics predicts **superposition states (exponential storage information)**
- **Dirac notation** $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$

- Operations are **unitary** matrices
  - Input and output states have the **same dimension**
  - Some classical gates (or, and, nand, xor…) **cannot be implemented directly**
  - Can **simulate** any classical computation with small overhead

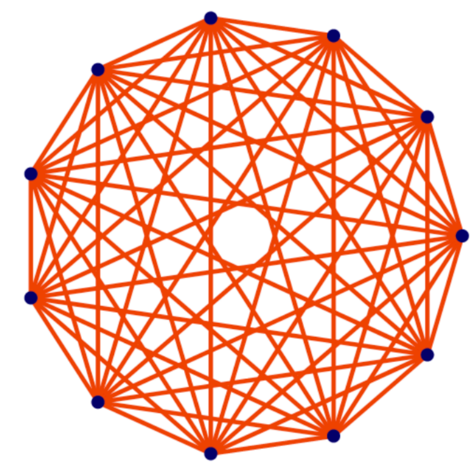


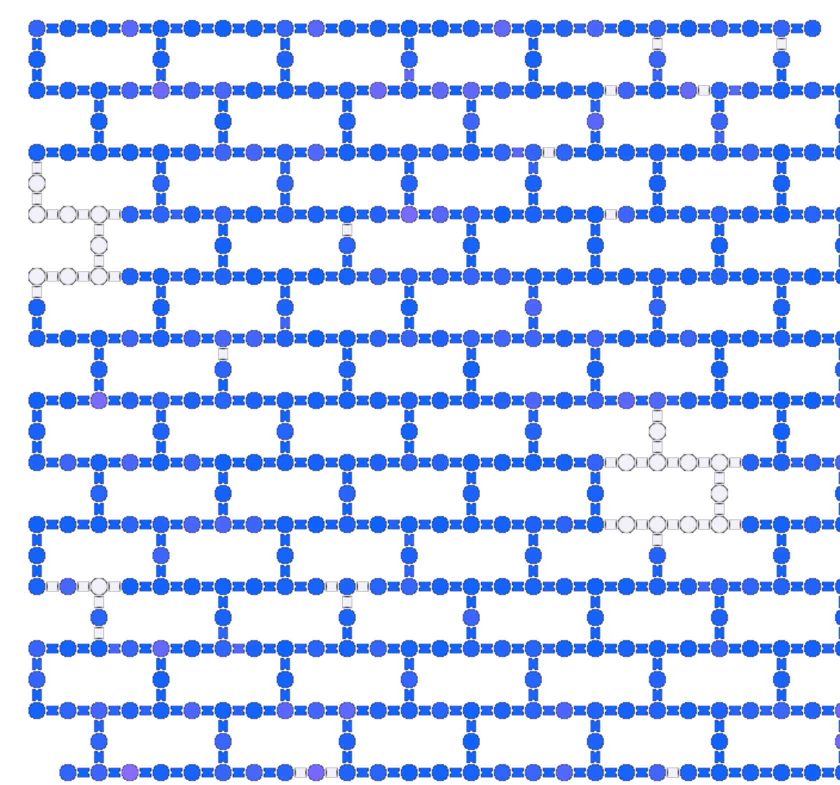
Interest in **multi level** representations: qutrits…

Noisy Intermediate-Scale Quantum devices

- Limitations in terms of **stability** and **connectivity**
  - **Circuit optimisation**
- **De-coherence**, measurement errors or gate level errors (**noise**)
  - Specific **error mitigation techniques**
  - Prefer algorithms **robust against noise**
- **Problem size**
- Initially integrated in **hybrid quantum-classical infrastructure (HPC + QC)**
  - **Quantum Processing Units** as new "hardware accelerators"

Trapped ion technology: *ionQ*
with all-to-all connectivity

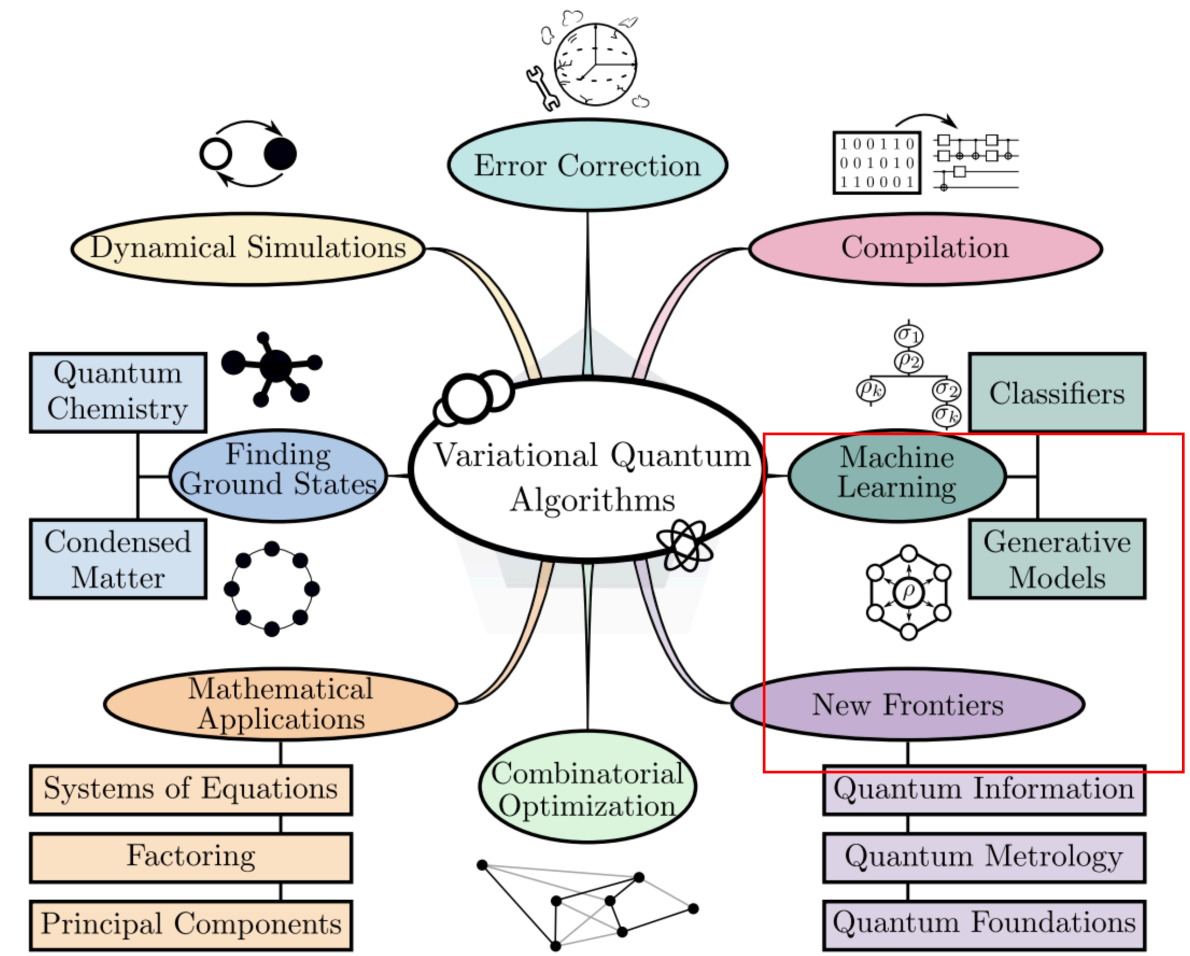




Superconducting qubits:
IBM Seattle

Quantum Machine Learning

Quantum Computing to improve ML



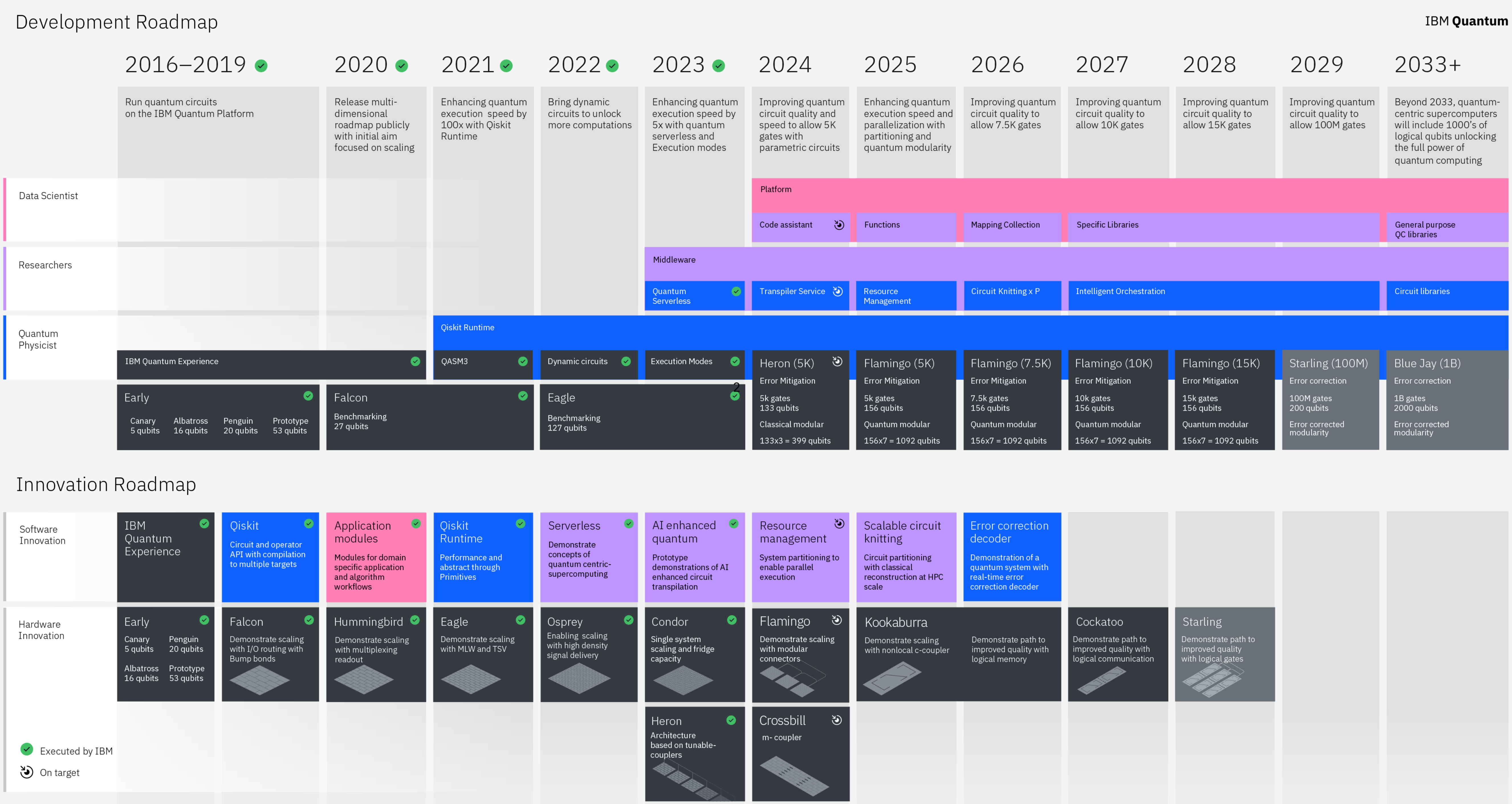
Cerezo, Marco, *et al.* *"Variational quantum algorithms."* Nature Reviews Physics 3.9 (2021)

- Speed-up and complexity
- Sample efficiency
- Representational power
- Energy efficiency???

- Evaluate performance on realistic use cases
- QPU as accelerators within classical infrastructure?

Study **classical intractability**:
Focus on quantum circuits that are **not efficiently simulable classically?**

IBM Quantum RoadMap



2024
2025
2026
2027
2028
Improving quantum circuit quality and speed to allow 5K gates with parametric circuits
Enhancing quantum execution speed and parallelization with partitioning and quantum modularity
Improving quantum circuit quality to allow 7.5K gates
Improving quantum circuit quality to allow 10K gates
Improving quantum circuit quality to allow 15K gates
Platform
Code assistant
Functions
Mapping Collection
Specific Libraries
Transpiler Service
Resource Management
Circuit Knitting x P
Intelligent Orchestration
Heron (5K)
Error Mitigation
5k gates
133 qubits
Classical modular
133x3 = 399 qubits
Flamingo (5K)
Error Mitigation
5k gates
156 qubits
Quantum modular
156x7 = 1092 qubits
Flamingo (7.5K)
Error Mitigation
7.5k gates
156 qubits
Quantum modular
156x7 = 1092 qubits
Flamingo (10K)
Error Mitigation
10k gates
156 qubits
Quantum modular
156x7 = 1092 qubits
Flamingo (15K)
Error Mitigation
15k gates
156 qubits
Quantum modular
156x7 = 1092 qubits
Resource management
System partitioning to enable parallel execution
Scalable circuit knitting
Circuit partitioning with classical reconstruction at HPC scale
Error correction decoder
Demonstration of a quantum system with real-time error correction decoder
Flamingo
Demonstrate scaling with modular connectors
Kookaburra
Demonstrate scaling with nonlocal c-coupler
Demonstrate path to improved quality with logical memory
Cockatoo
Demonstrate path to improved quality with logical communication
Starling
Demonstrate path to improved quality with logical gates

Quantum Machine Learning Lifecycle

The quantum advantage of many known QML algorithms is impeded by an input or output bottleneck

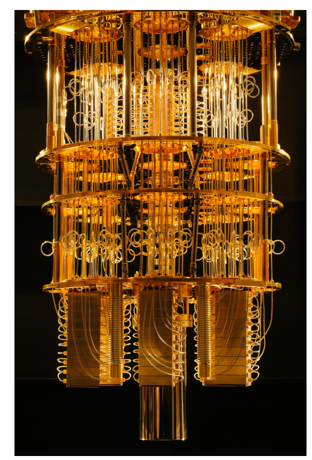
Data Preparation

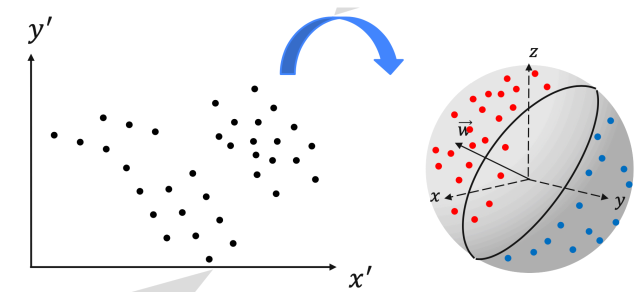
Model Definition

Model Training

Model Testing

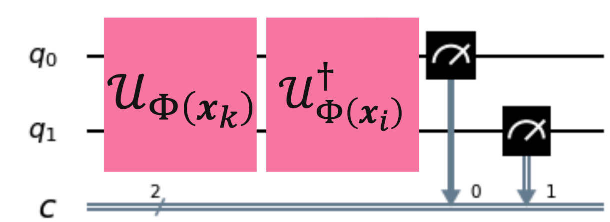
Model Interpretation



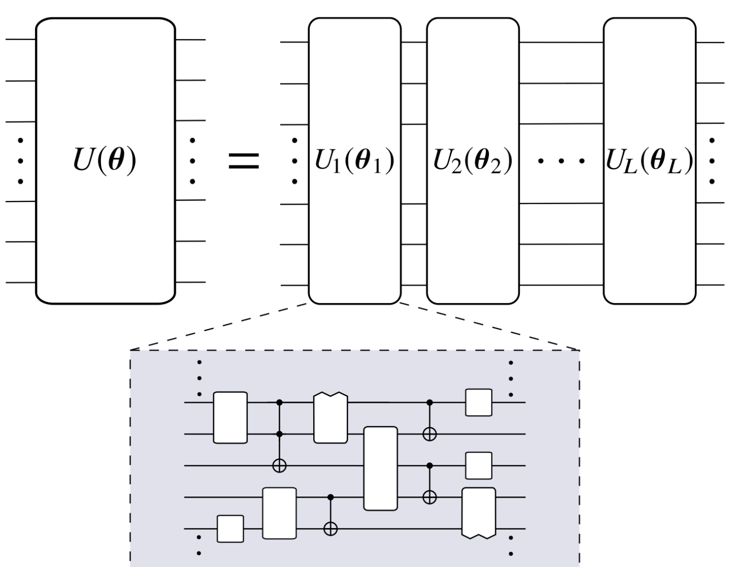


Data Reduction
Data Encoding
[1,2,3]

Read Out



Trainability (BP…)



[1] Robust data encodings for quantum classifiers, Ryan LaRose and Brian Coyle, Phys. Rev. A 102, 032420

[2] Quantum convolutional neural network for classical data classification, https://arxiv.org/pdf/2108.00661.pdf

[3] Quantum Support Vector Machines for Continuum Suppression in B Meson Decays, https://arxiv.org/abs/2103.12257

Models

Variational algorithms (ex. QNN)

Gradient-free or **gradient-based** optimization

Data Embedding can be **learned**

Ansatz design can leverage data symmetries[1]

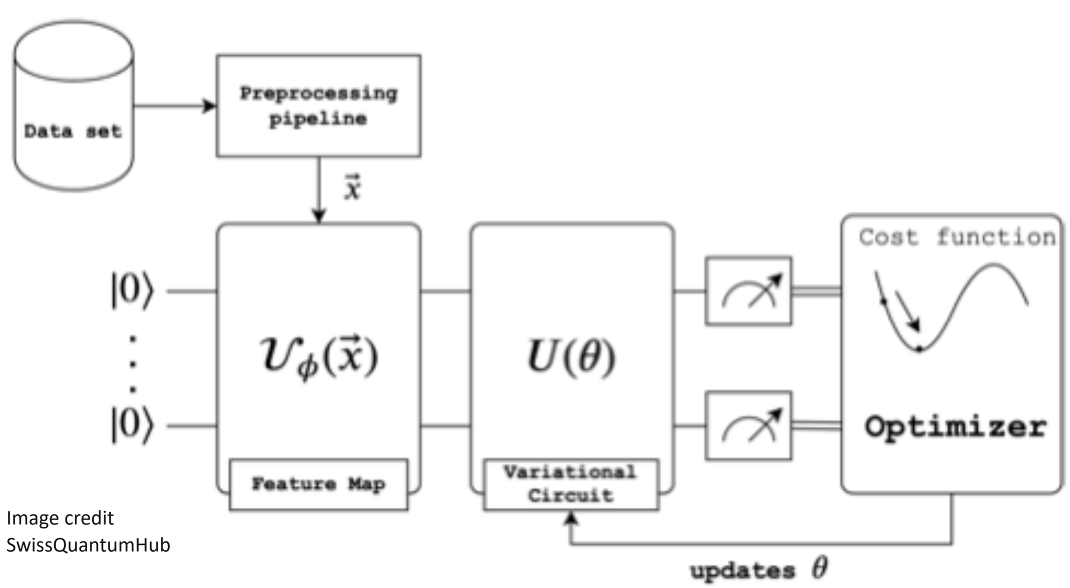


Image credit SwissQuantumHub

Kernel methods (ex. QSVM)

Feature maps as **quantum kernels**

Classical **kernel-based training** (**convex** losses)

Identify classes of kernels that relate to specific data **structures**[2]

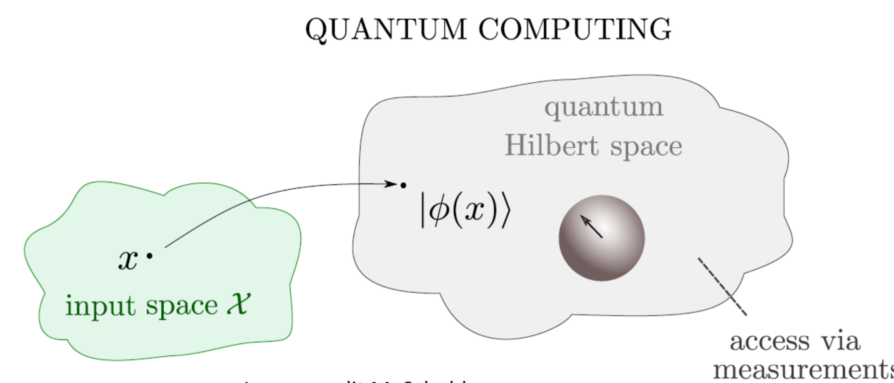


Image credit M. Schuld

Energy-based ML (ex. QBM)

Build networks of **stochastic binary units** and optimise their energy.
Quantum Boltzmann Machines has quadratic energy function that follows the Boltzmann distribution (Ising Hamiltonian)

[1] Bogatskiy, Alexander, et al. "**Lorentz group equivariant neural network for particle physics**." PMLR, 2020.

[3]Jerbi, Sofiene, et al. "**Quantum machine learning beyond kernel methods.**" *arXiv:2110.13162* (2021).

Parameter optimization

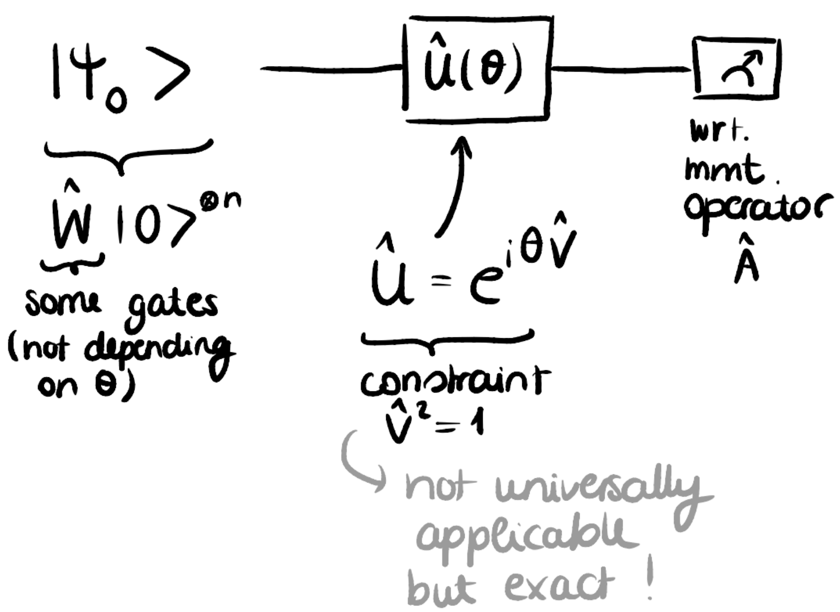
See C. Rieger's summer students lecture

The parameter-shift rule (gradient-based)

$$\theta \to \theta - \eta \nabla_\theta f$$

$$\langle \hat{A}(\theta) \rangle$$

Compute **partial derivative** of variational circuit parameter θ, alternative to analytical gradient computation and classical finite difference rule (numerical errors and resource cost considerations)



$$\Rightarrow \nabla_\theta \langle \hat{A} \rangle = u\left[\langle \hat{A}(\theta + \tfrac{\pi}{4u}) \rangle - \langle \hat{A}(\theta - \tfrac{\pi}{4u}) \rangle\right]$$

Evaluate Quantum Circuit twice at shifted parameters to compute gradient

Source:https://pennylane.ai/qml/demos/tutorial_stochastic_parameter_shift/

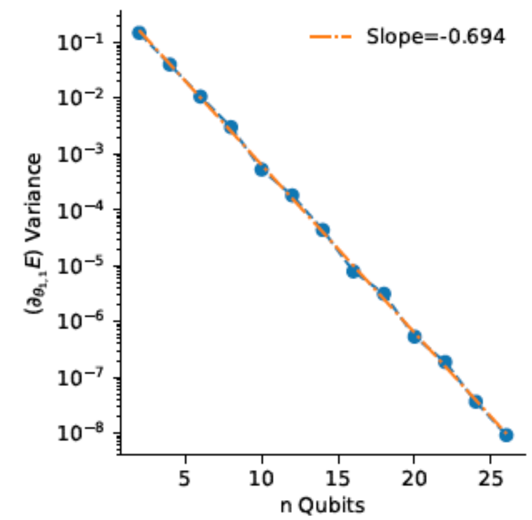
Gradients decay and Model Convergence

Classical gradients **vanish exponentially** with the number of layers (J.McClean *et al.*, arXiv:1803.11173)

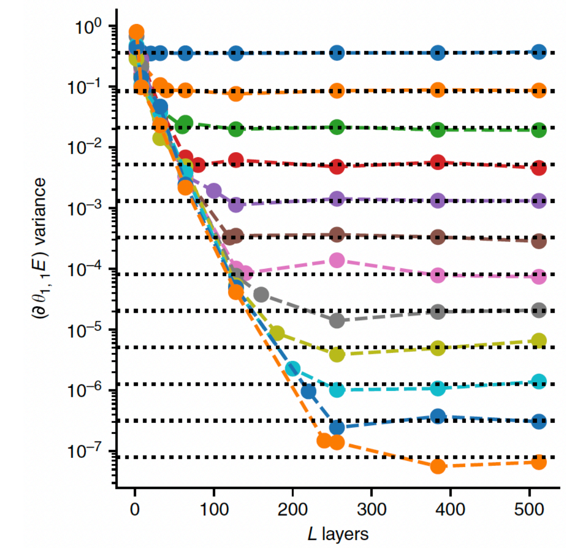
- Convergence still possible if gradients consistent between batches.

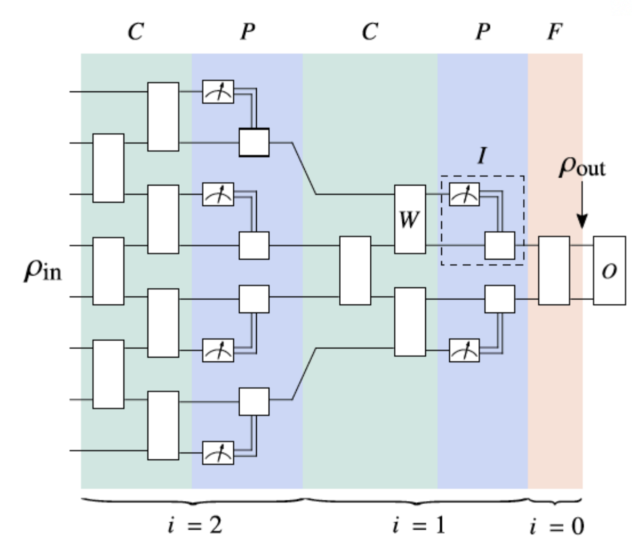
Quantum gradient **decay exponentially** in the number of qubits (number of graph paths is exponential in the number of gates)

- Random circuit initialization
- Loss function locality in shallow circuits (M. Cerezo *et al.*, arXiv:2001.00550)
- Ansatz choice: TTN, CNN (Zhang *et al.*, arXiv:2011.06258, A Pesah, *et al.*, *Physical Review X* 11.4 (2021): 041011.)
- Noise induced barren plateau (Wang, S *et al.*, Nat Commun 12, 6961 (2021))



J. McClean *et al.*, arXiv:1803.11173





QCNN: A Pesah, *et al.*, *Physical Review X* 11.4 (2021): 041011

Quantum Convolutional NN:
Convolution: general SU(4)
Pooling: reduces numer of qubits

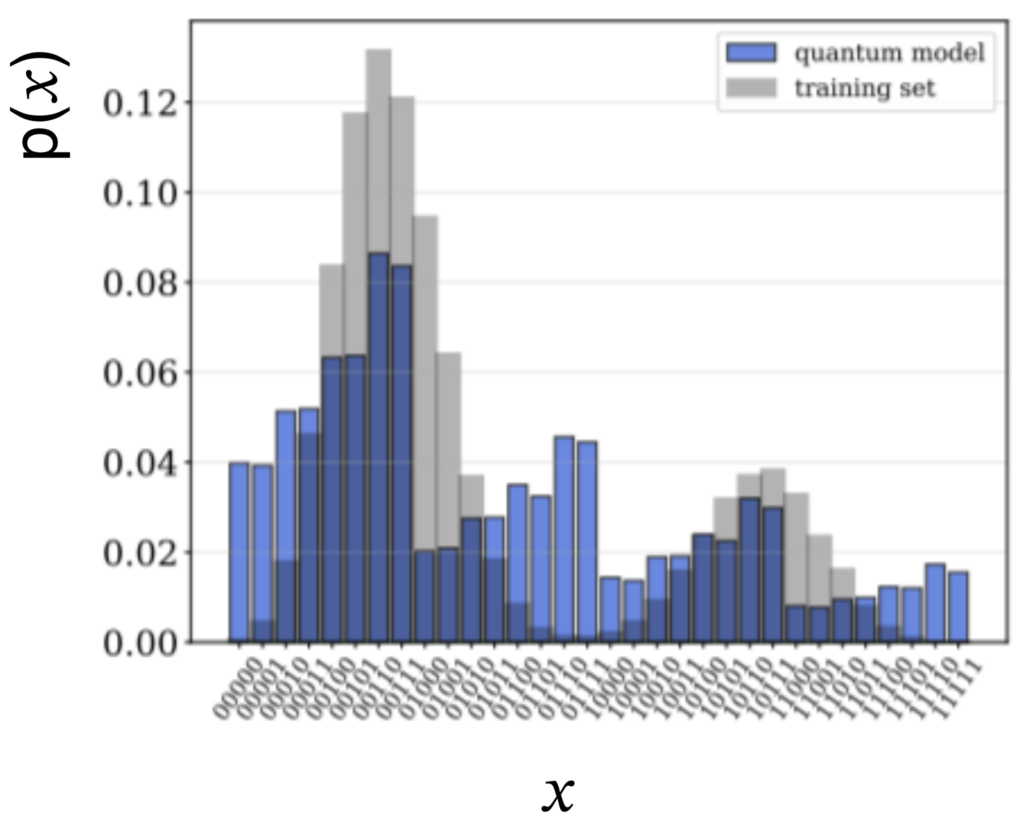
Large number of measurements: $1/\epsilon^2$ measurements to estimate a cost to precision ϵ

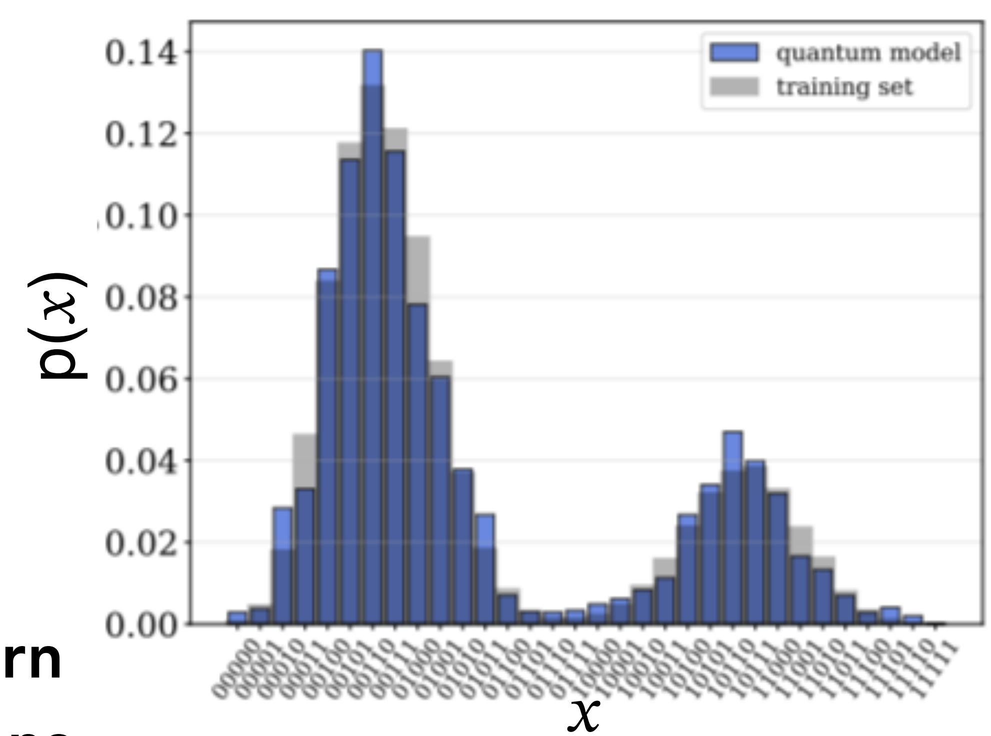
Quantum Generative Models

Learn probability distribution that best describes a data set

Quantum Circuit Born Machine

Sample variational pure state $|\psi(\theta)\rangle$ by projective measurement through **Born rule**: $\mathbf{p}_\theta(\mathbf{x}) = |\langle \mathbf{x} | \boldsymbol{\psi}(\boldsymbol{\theta}) \rangle|^2$.





Deeper circuits learn better representations..

Or don't they ??

Implicit and Explicit Models

Classified according to whether or not they **have access to the propability distribution function**

Explicit Models have access to PDF in polynomial time

- Use explicit losses that are defined by probabilities
- Ex. TN or autoregressive models

Explicit

$$\sum_{\boldsymbol{x}} f\left(\tilde{p}(\boldsymbol{x}), \tilde{q}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)$$

Ex. KL Divergence

$$D_{\mathrm{KL}}(P \| Q) = \sum_i P(i) \log\left(\frac{P(i)}{Q(i)}\right)$$

Implicit models do not have access to PDF. Can sample from it

- Use implicit losses built on samples
- Ex. GAN, QBM, VAE… QCBM…

Implicit

$$\underset{\boldsymbol{x},\boldsymbol{y}}{\mathbb{E}}\left[g(\boldsymbol{x}, \boldsymbol{y})\right]$$

Ex. MMD

$$\mathrm{MMD}(\mathbb{P}_r, \mathbb{P}_g) = \left(\mathbb{E}_{\substack{\mathbf{x}_r, \mathbf{x}_r' \sim \mathbb{P}_r, \\ \mathbf{x}_g, \mathbf{x}_g' \sim \mathbb{P}_g}}\left[k(\mathbf{x}_r, \mathbf{x}_r') - 2k(\mathbf{x}_r, \mathbf{x}_g) + k(\mathbf{x}_g, \mathbf{x}_g')\right]\right)^{\frac{1}{2}}$$

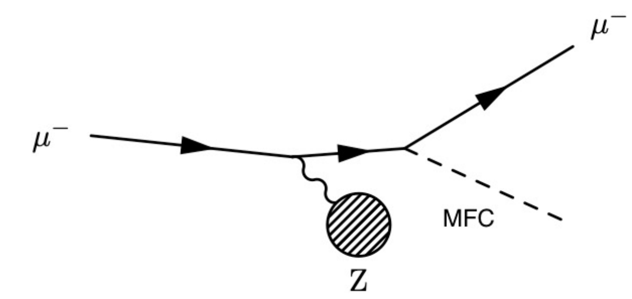
Strong impact on trainability!

M. Rudolph *et al.* , QTML 2023

Quantum Circuit Born Machine for Event Generation

Born machine:
Produces statistics according to Born's measurement rule using parametrized quantum circuit $|\psi(\theta)\rangle$

$$p_\theta(x) = \left| \langle x | \psi(\theta) \rangle \right|^2, x \in \{0,1\}^{3n}$$



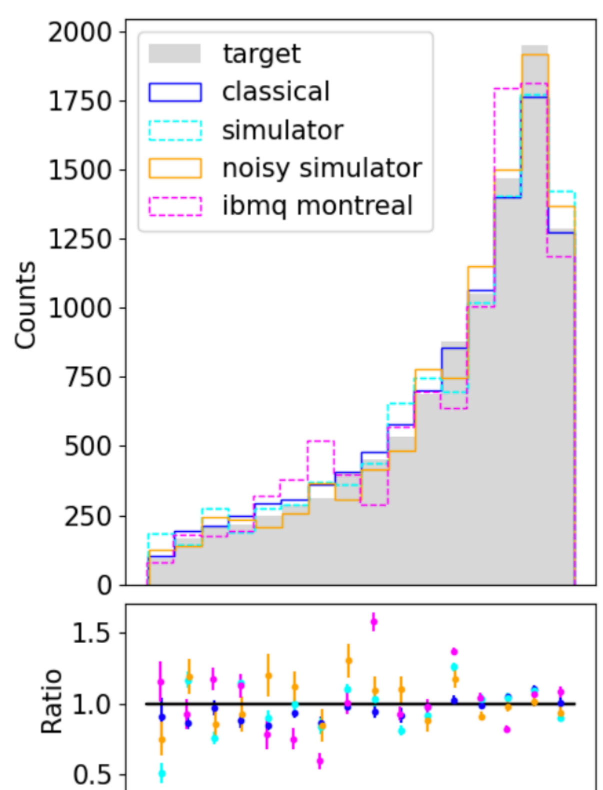
Muon fixed target scattering experiment

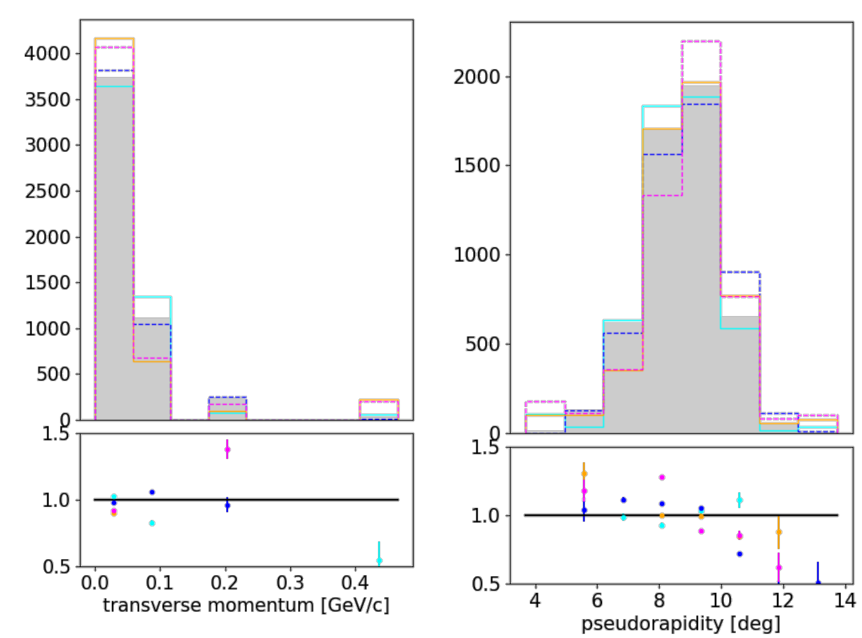
Train using Maximum Mean Discrepancy loss function:

$$\text{MMD(P,Q)} = \mathbb{E}_{\substack{X\sim P\\ Y\sim P}}\left[K(X,Y)\right] + \mathbb{E}_{\substack{X\sim Q\\ Y\sim Q}}\left[K(X,Y)\right] - 2\mathbb{E}_{\substack{X\sim P\\ Y\sim Q}}\left[K(X,Y)\right]$$

Gaussian kernel $K(x,y) = \exp\left(-\frac{(x-y)^2}{2\sigma}\right)$

Generate discrete PDFs (continuous in the limit increasing no. of qubits)

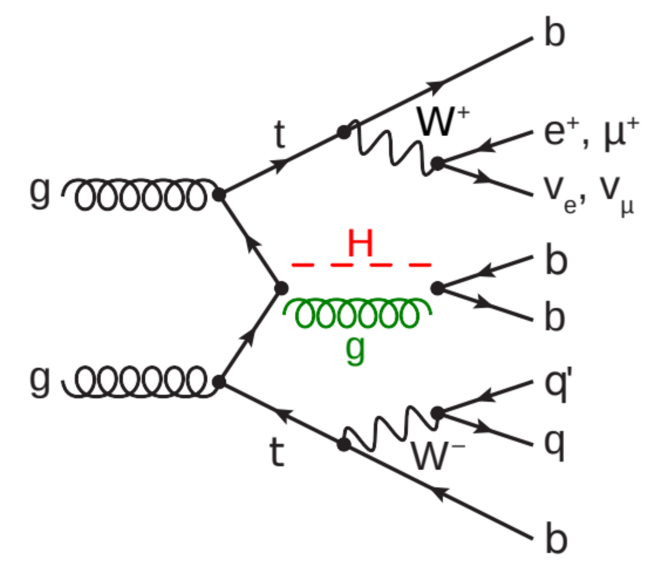




Kiss O., Grossi M. et all., **Conditional Born machine for Monte Carlo events generation**, *Phys. Rev. A* ***106****, 022612 (2022)*

A classification task

Analysis setup



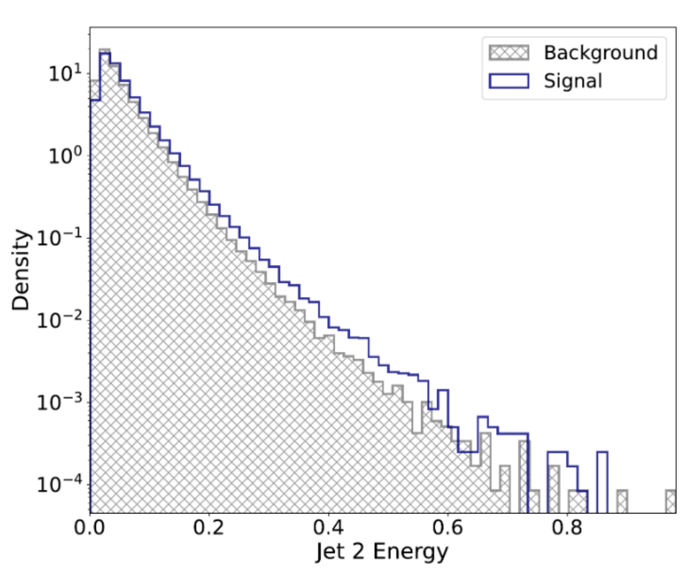
Analysis

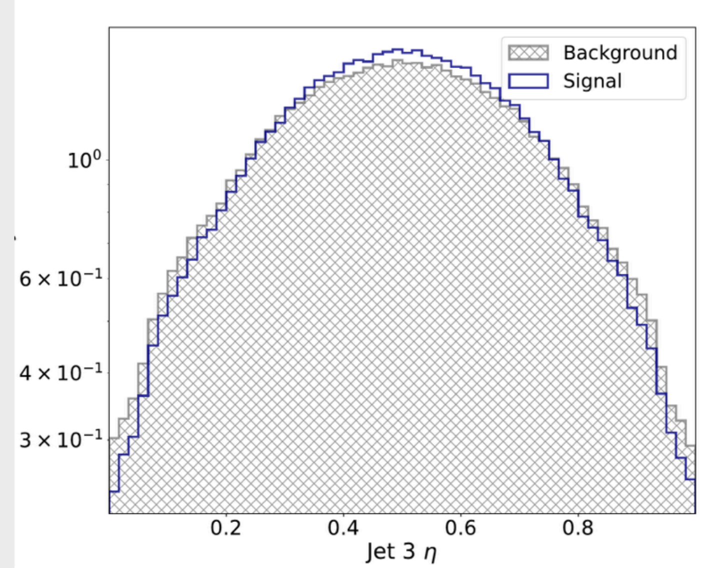
Discrimination of the signal over the overwhelming background

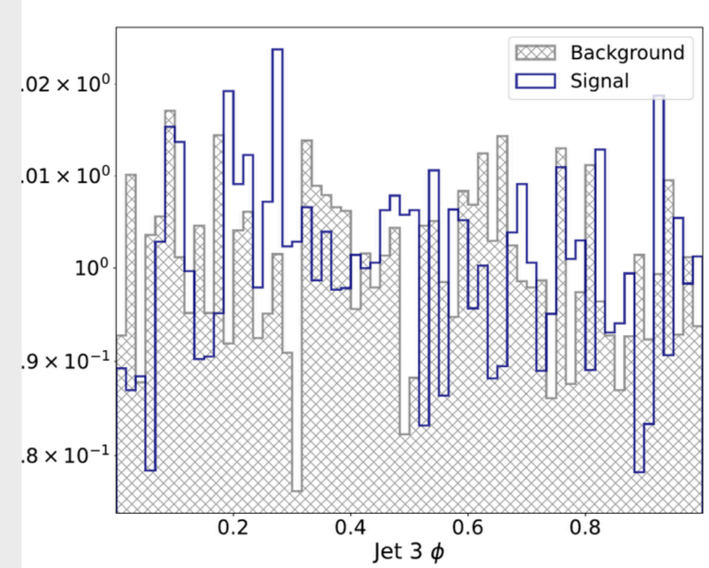
Features

- For the each jet we have 8 features: (pT,η,ϕ,E,b tag,px,py,pz)
- For MET we have 4 features: (pT,px,py,ϕ)
- For the lepton (electron or muon) we have 7 features: (pT,η,ϕ,E,px,py,pz)

#features = 8×7(jets)+7(1lepton)+4(MET) = 67





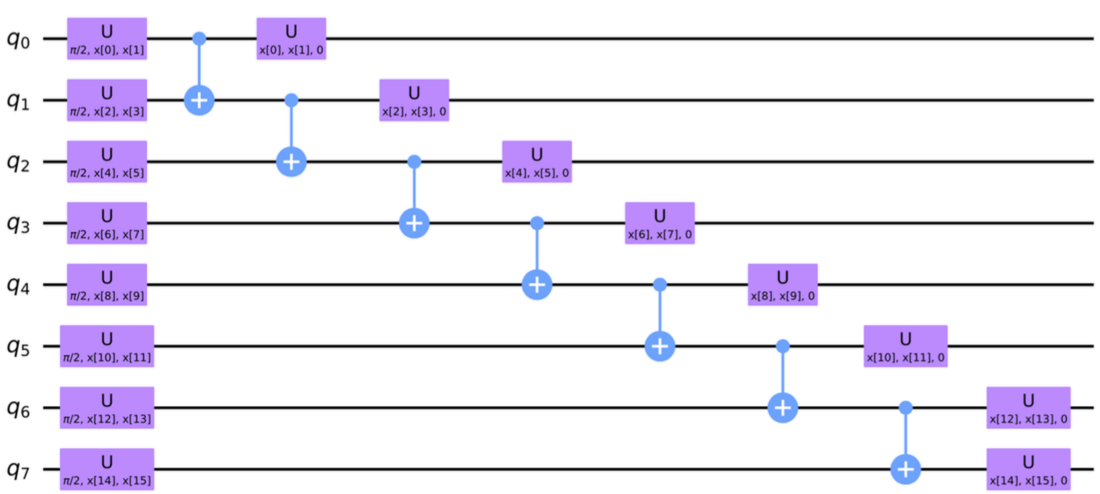


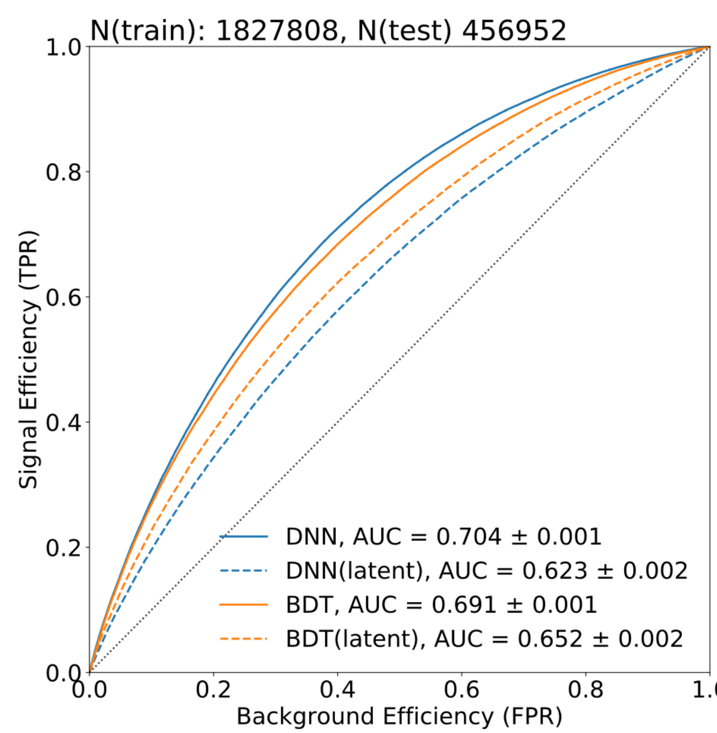
Quantum SVM for Higgs Classification

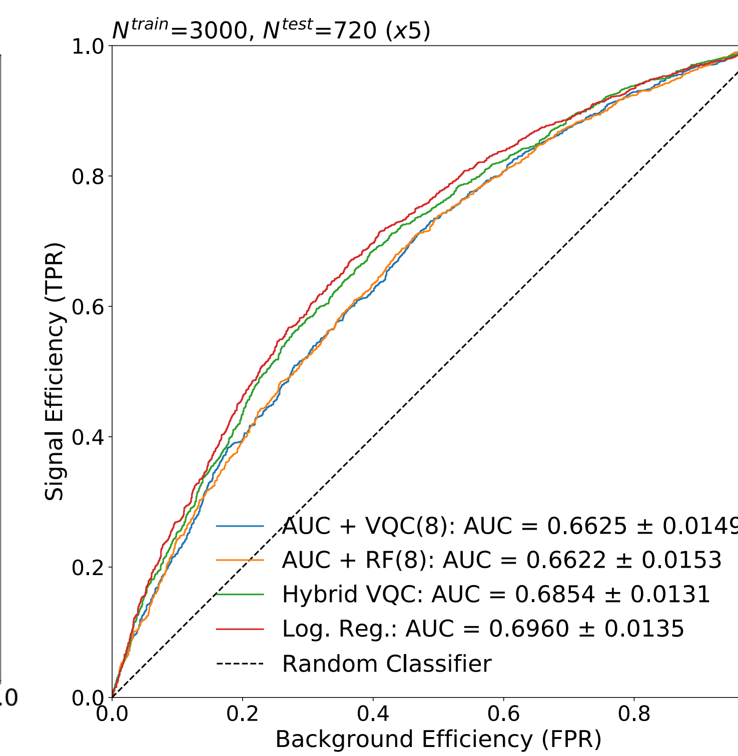
Input dimensionality reduction through an Auto-Encoder projects to a lower dimension latent space (8,16)

Feature selection + Model	AUC
AUC + QSVM	0.66 ± 0.01
PyTorch AE + QSVM	0.62 ± 0.03
AUC + SVM rbf	0.65 ± 0.01
PyTorch AE + SVM rbf	0.62 ± 0.02
KMeans + SVM rbf	0.61 ± 0.02

Feature selection + Model	AUC
AUC + QSVM	0.68 ± 0.02
AUC + Linear SVM	0.67 ± 0.02
Logistic Regression	0.68 ± 0.02



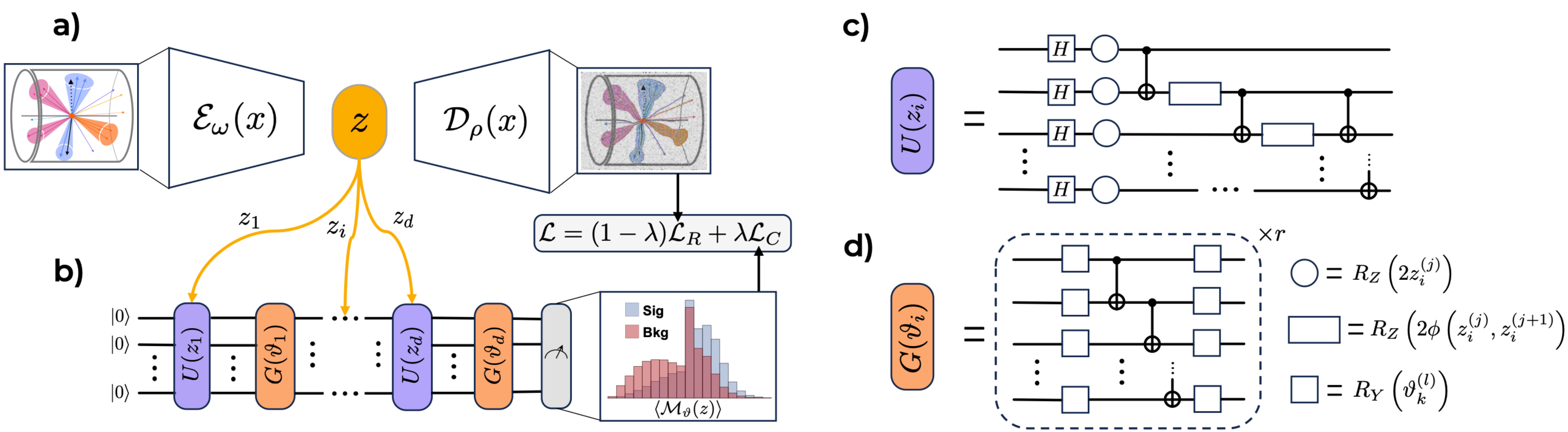




Data encoding circuit serving as feature map for the 8-qubit QSVM implementation.

https://www.nature.com/articles/s42005-024-01811-6

Guided Quantum Compression



Two hybrid quantum-classical strategies:

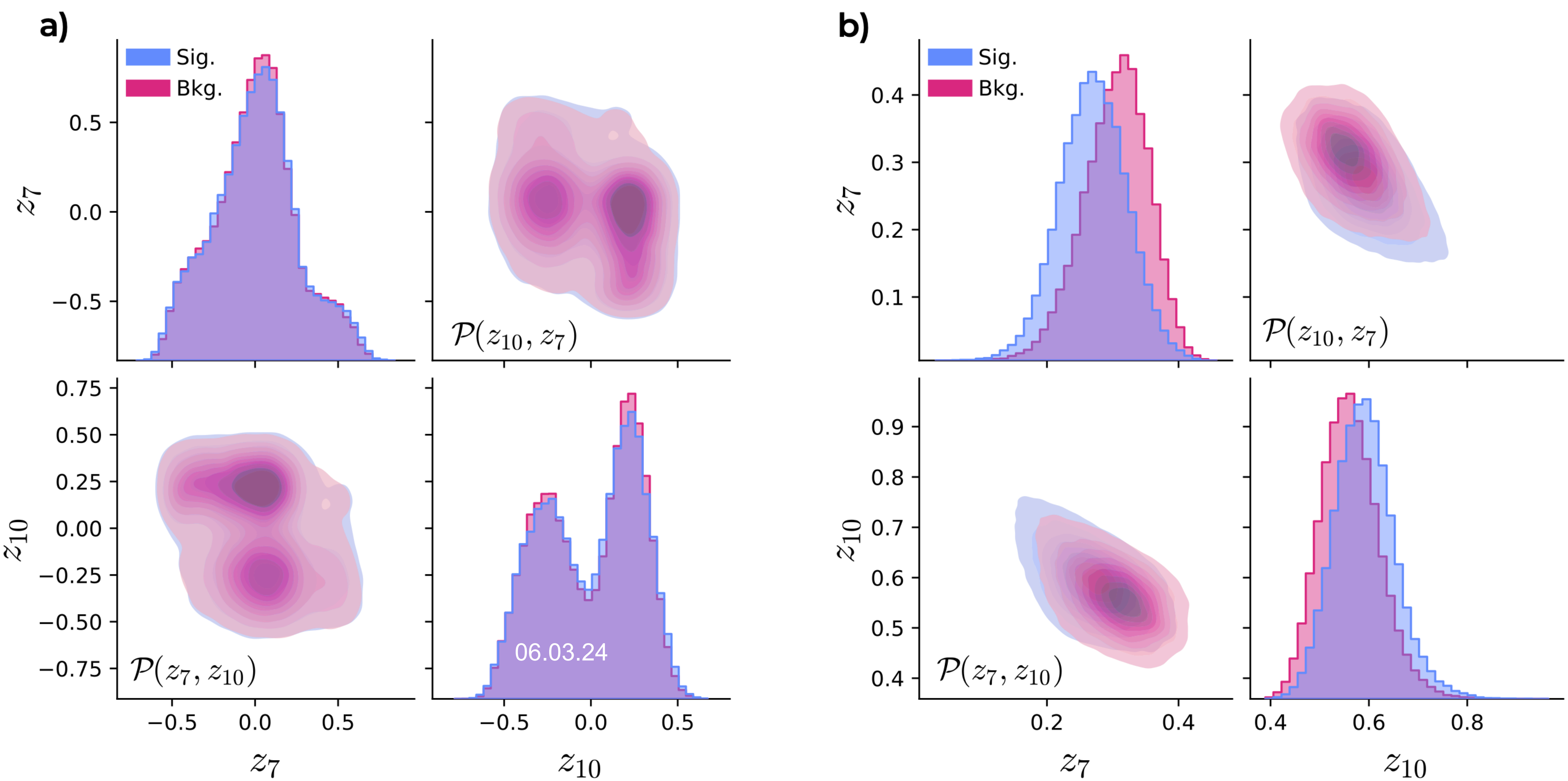
GQC: Joint training

2Steps: The data compression step is independently trained

V. Belis *et al.*, arxiv: 2402.0952

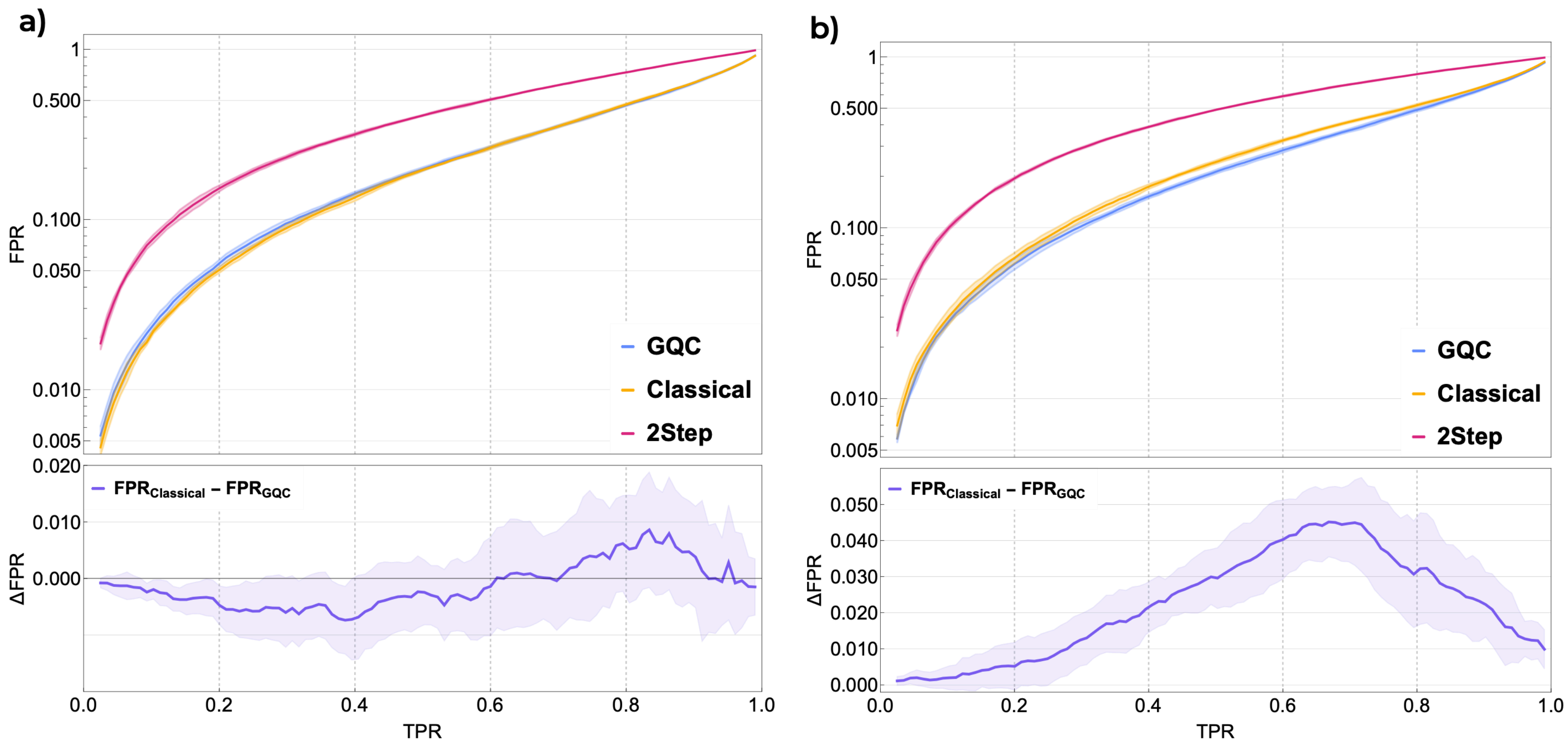
Latent Space Representation

2 Steps **GQC**



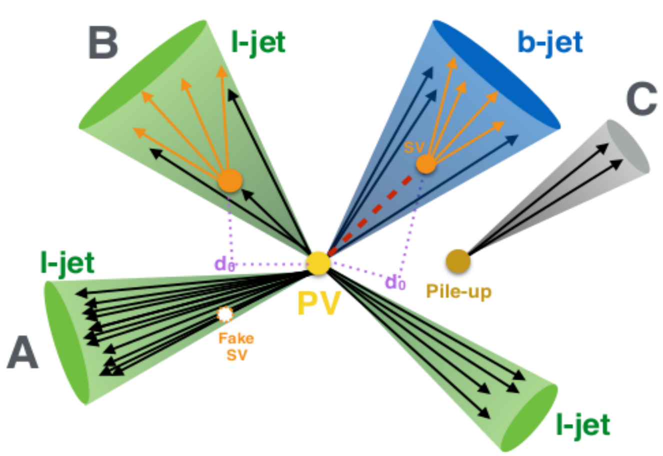
Results

Including b-tagging   No b-tagging



compression method has significant impact on the classifier performance.

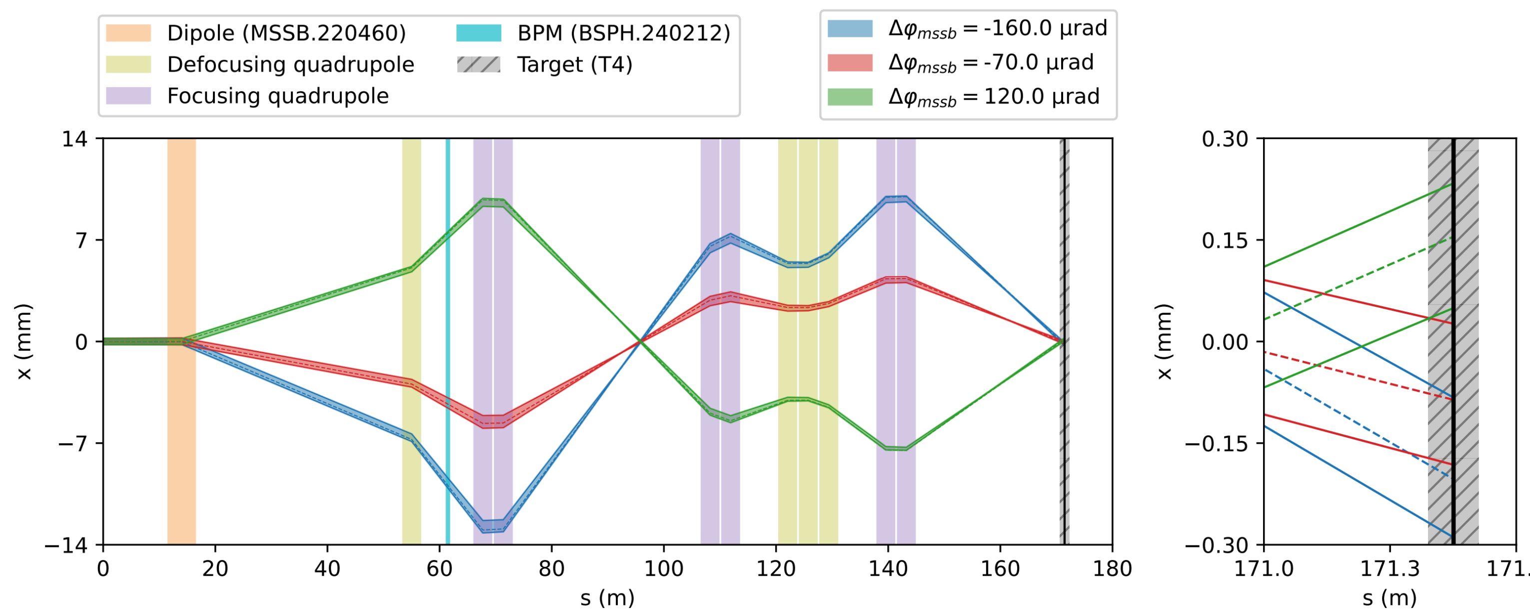
NB: b-tag features are high level features containing information about the quark content



Reinforcement Learning for steering CERN beams

1D beam steering inspired by the CERN North Area transfer line

- **Action: (discrete) deflection** angle
- **State: (continuous) BPM** position
- **Reward: integrated beam intensity on target**
- **Optimality**: fraction of states in which the agent takes the right decision



Approach based on Q-learning : the reward is estimated using a Q-function

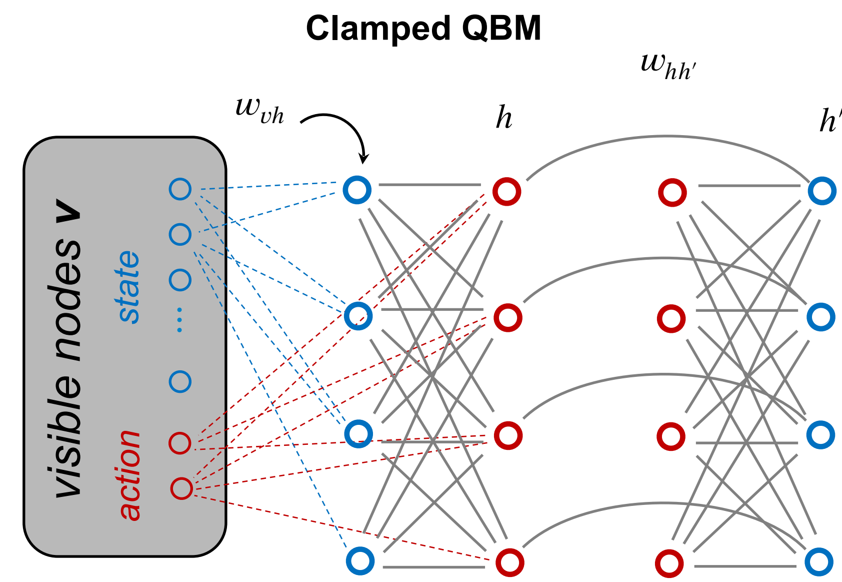
Quality of the Q-function drives performance

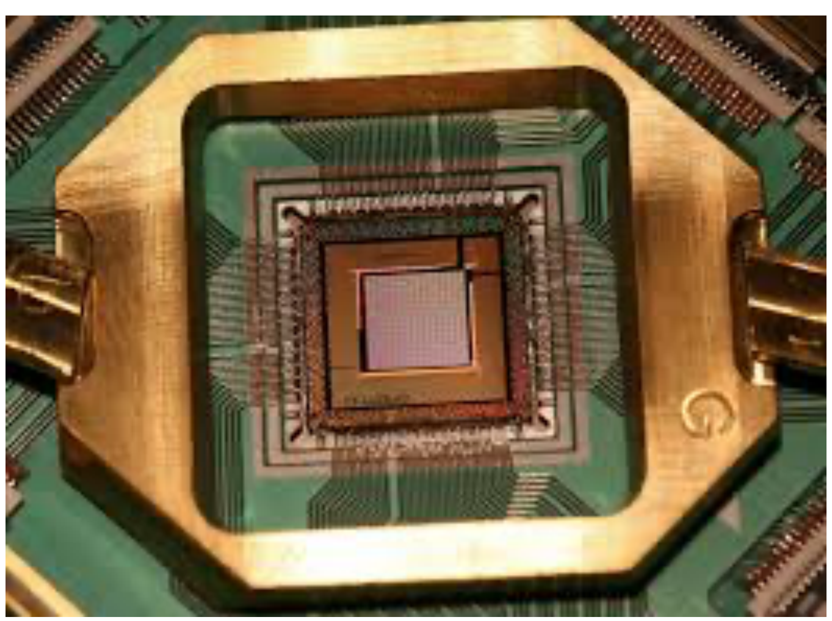
- **Classical Deep Q-learning (DQN)**
  Feed-forward neural net
- **Free-energy based RL (FERL)**
  Quantum Boltzmann machine (QBM)

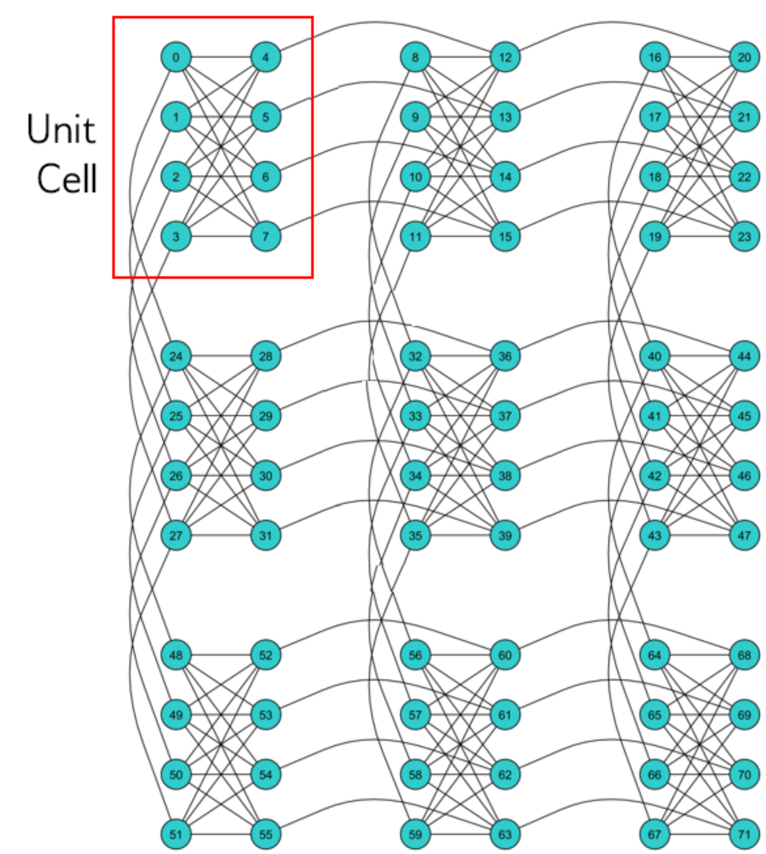
Schenk, M *et al.* Hybrid actor-critic algorithm for quantum reinforcement learning at CERN beam lines. *arXiv preprint arXiv:2209.11044.*, CHEP2023

Quantum Boltzmann Machines

- Create a network of qubits similar to the binary units in a classical Boltzmann machine
- Map network to the quantum computer graph
- Train by minimising the qubits system energy



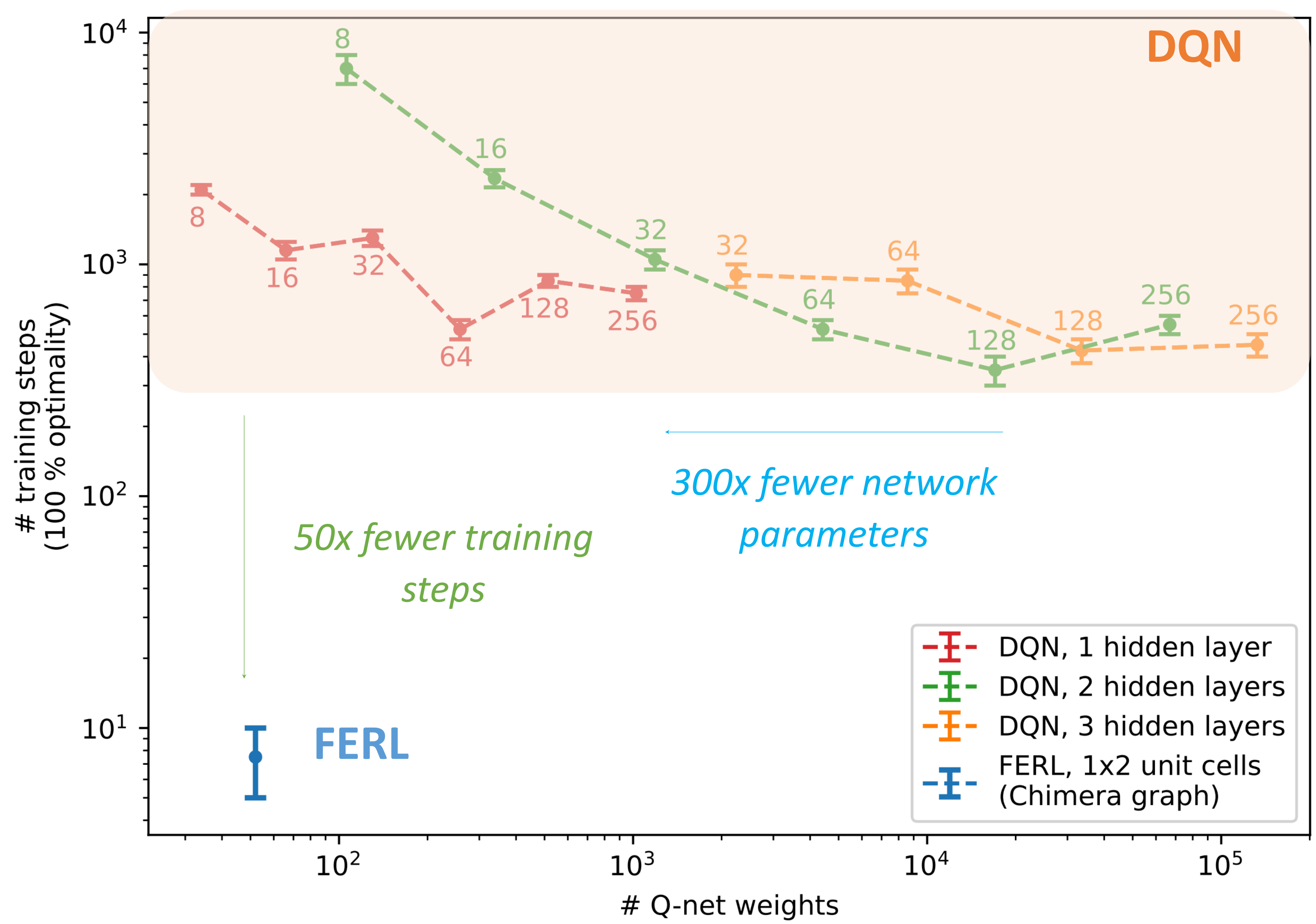




Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch

Free-energy based RL (FERL)

Sample-efficiency is really important given cost of training (beam time)



Schenk, M *et al.* Hybrid actor-critic algorithm for quantum reinforcement learning at CERN beam lines. *arXiv preprint arXiv:2209.11044.*, CHEP2023

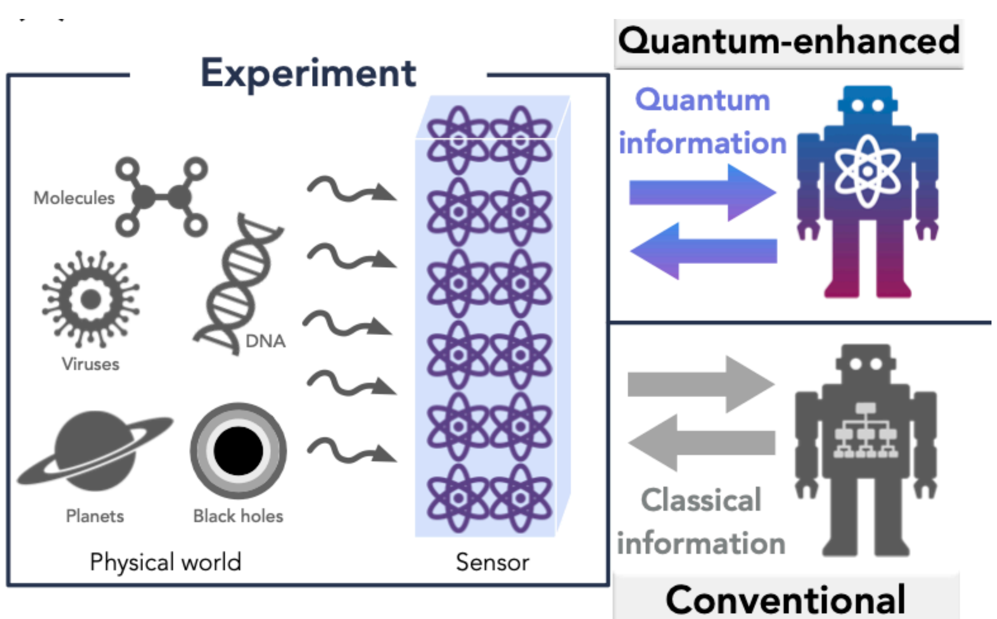
Classifying quantum data

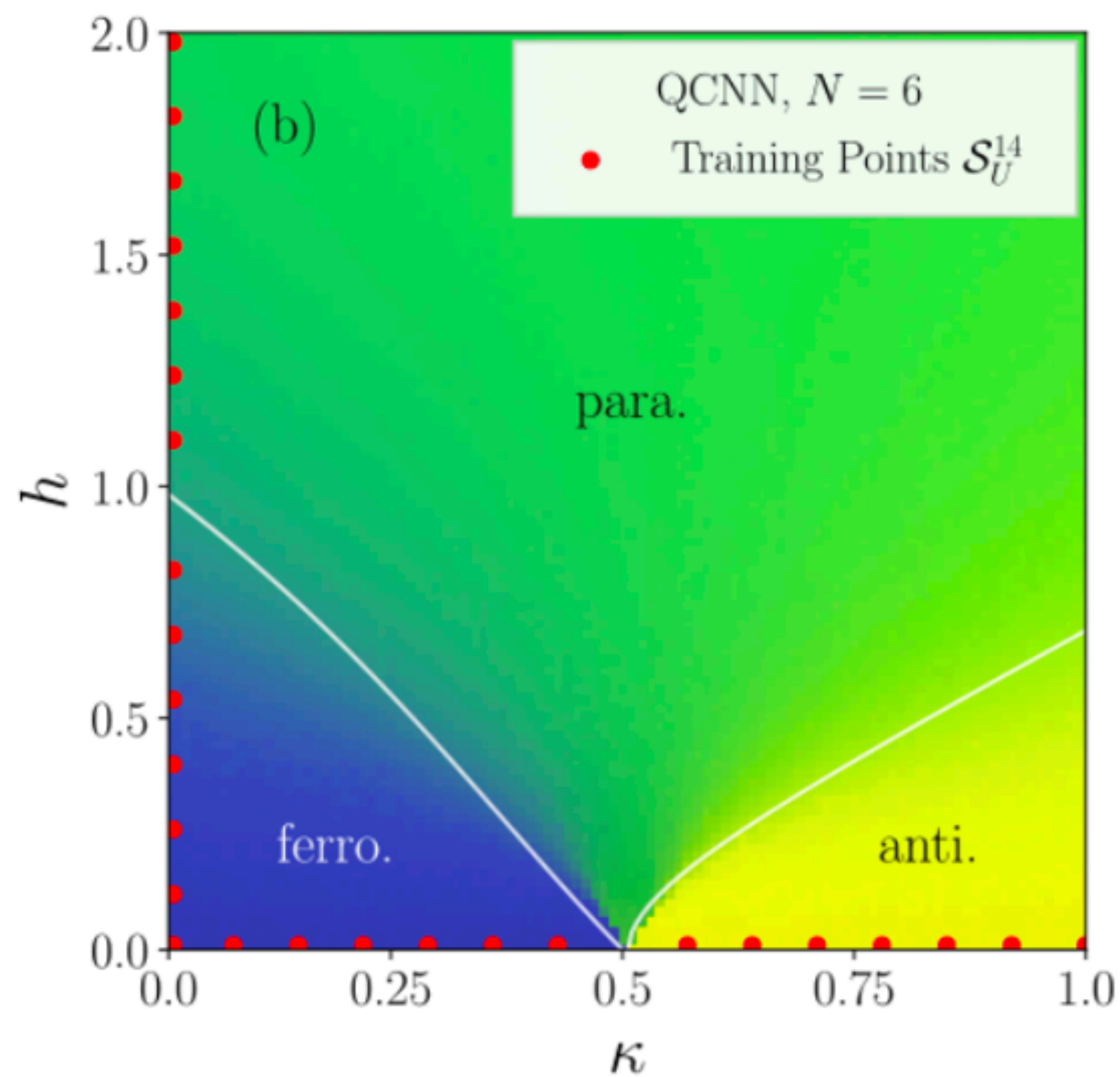
Saverio Monaco et al., **Quantum phase detection generalisation from marginal quantum neural network models**, arXiv:2208.08748v1.

Generate **quantum states** directly on the device

Train QCNN to **classify quantum states**

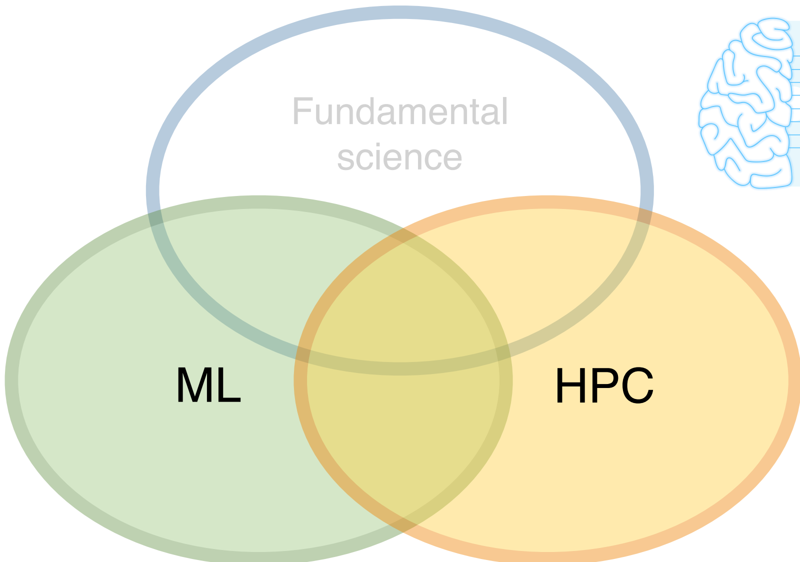
Use marginal datasets → **OOD generalization !**

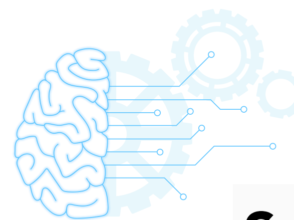




Backup

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:

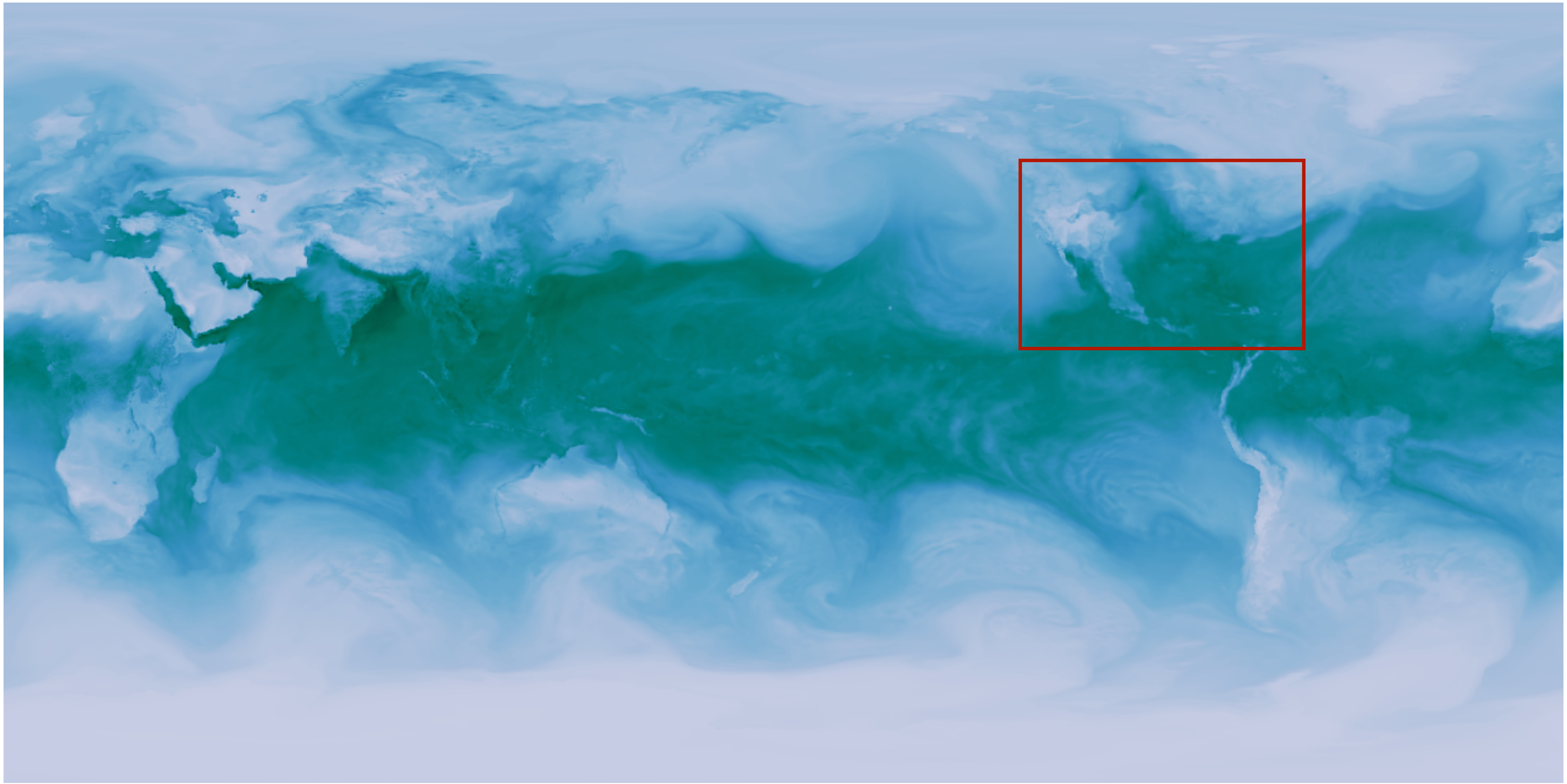
- **How to integrate new incoming data**
- How to **expand** to new fields/variables without fully retraining the model each time?

From Martin Schultz - ESA-ECMWF workshop 2024

Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

Results: Target - ERA5

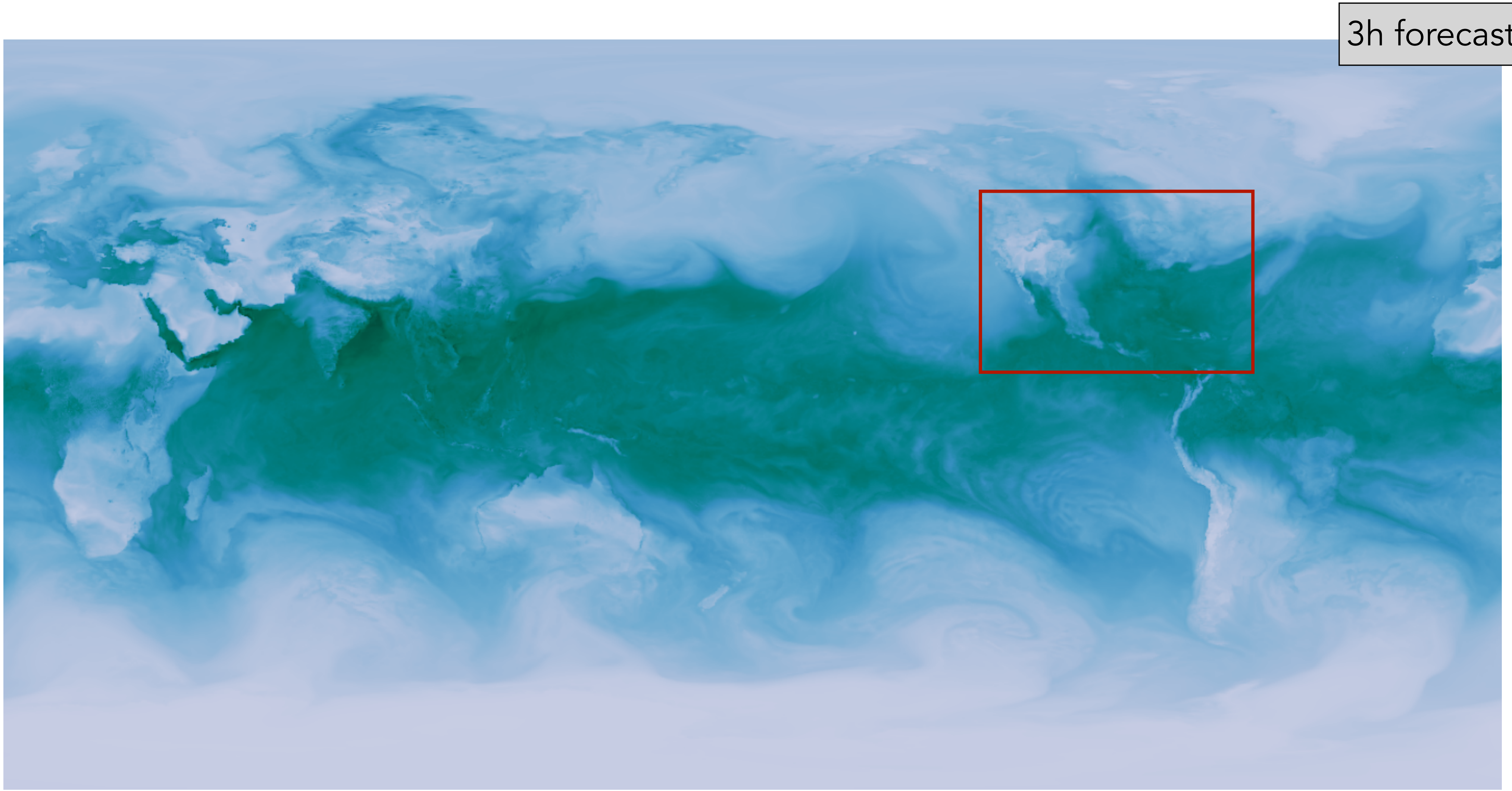
specific humidity, June 15th 2018 13:00 UTC



Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

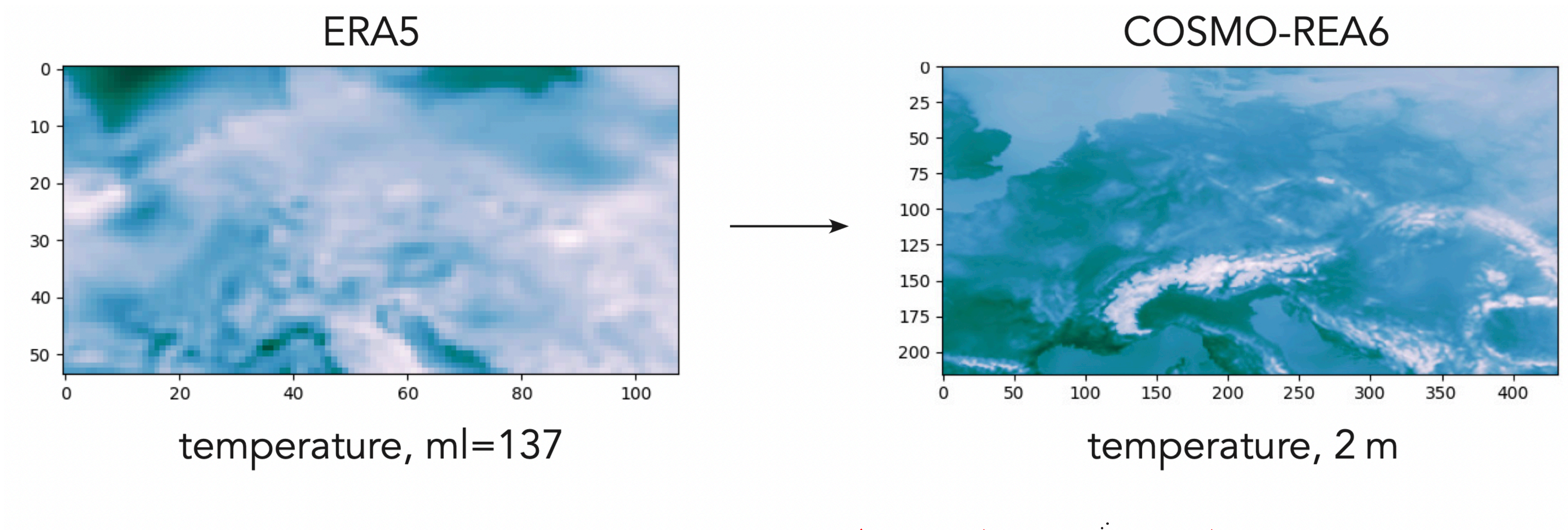
Results: Prediction - AtmoRep

specific humidity, June 15th 2018 13:00 UTC



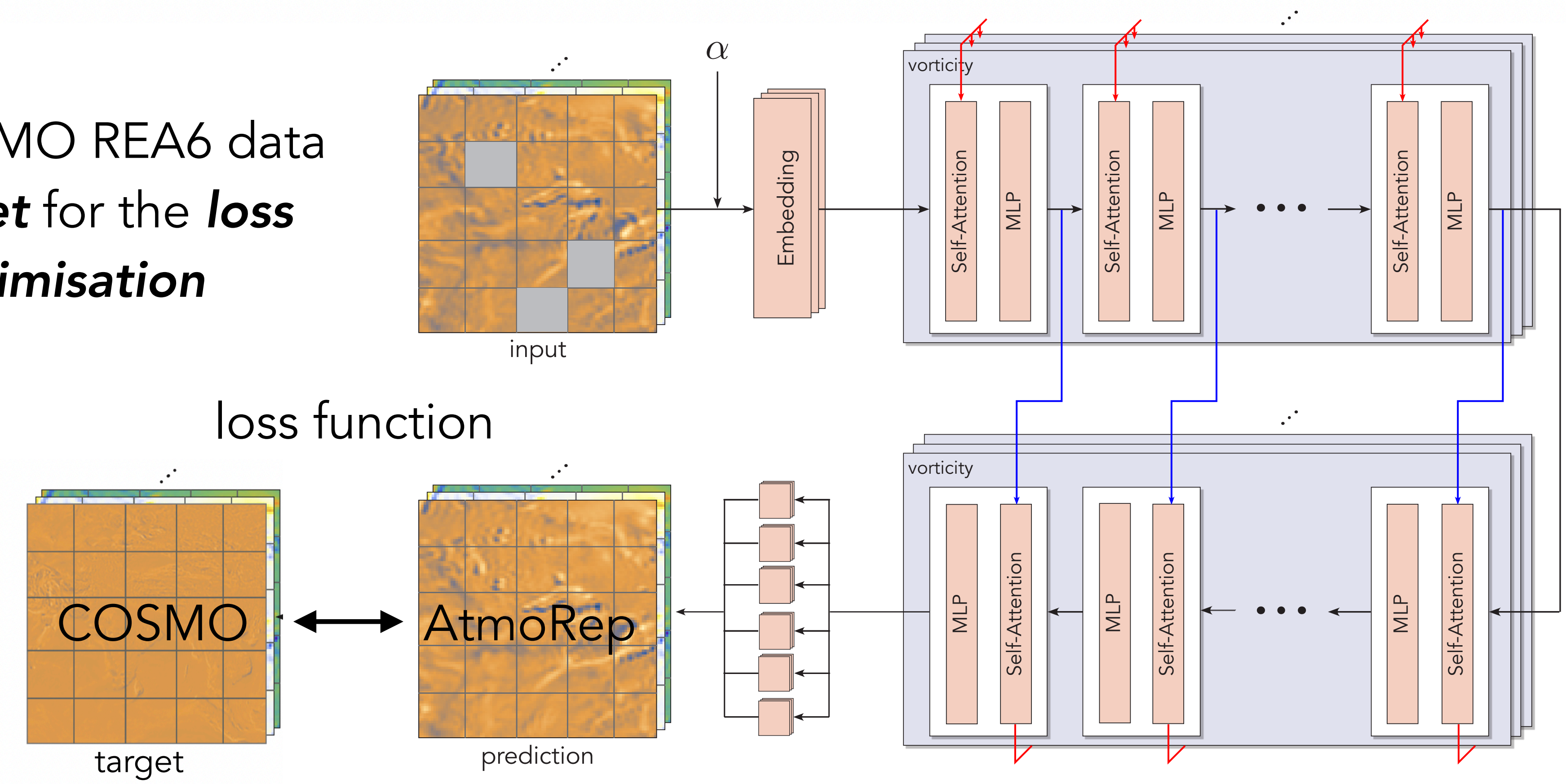
Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

Downscaling



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

loss function

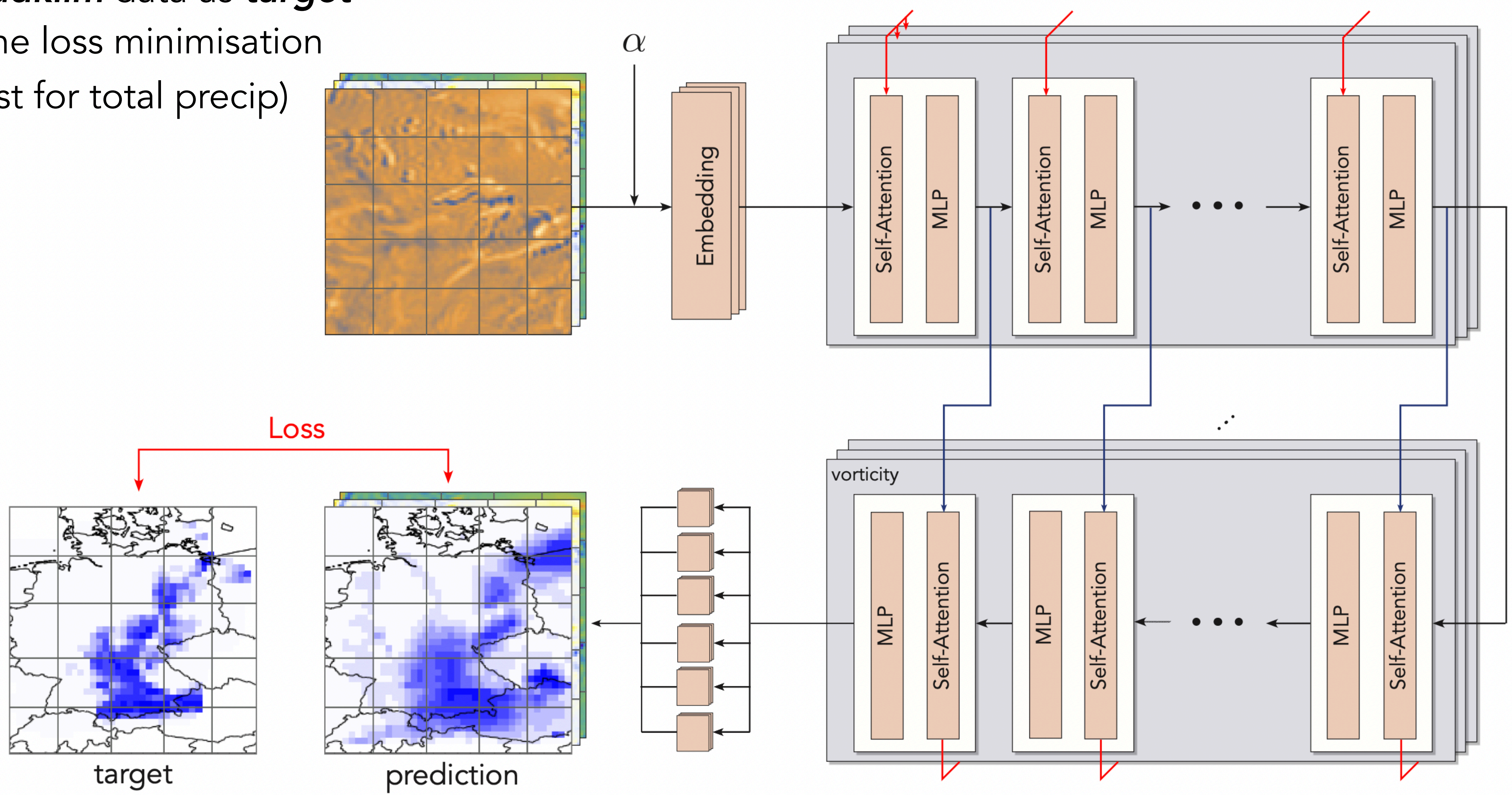


Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch

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)



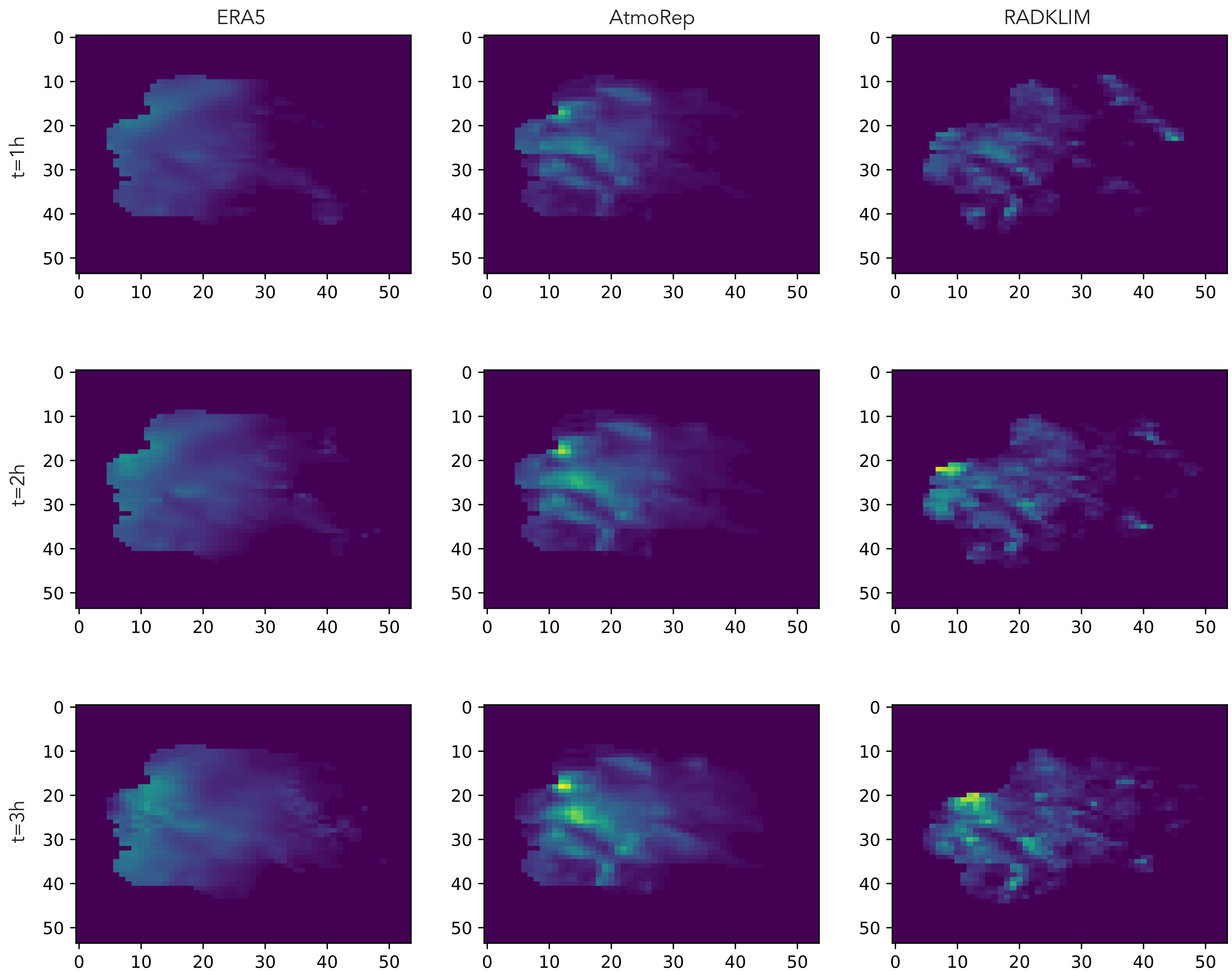
Radklim

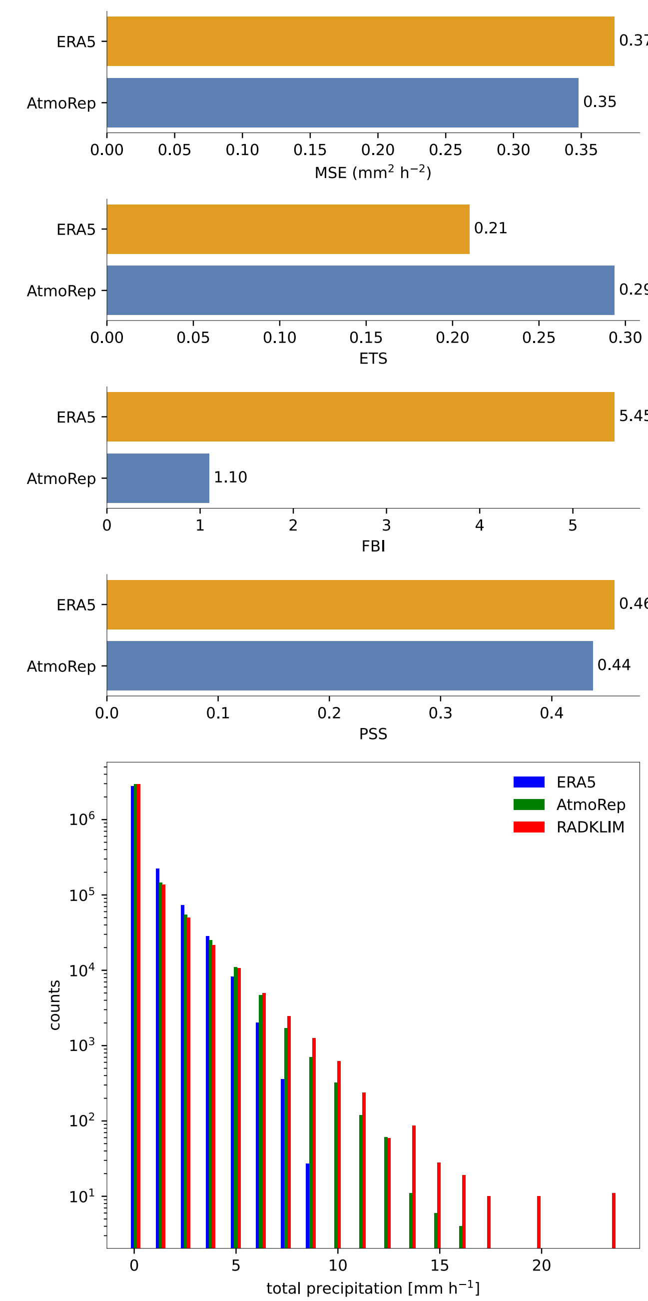
AtmoRep

Ilaria Luise, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@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, Sofia Vallecorsa CERN - ilaria.luise@cern.ch I sofia.vallecorsa@cern.ch