Koninklijk Meteorologisch Instituut

Institut Royal Météorologique

Königliches Meteorologisches Institut

Royal Meteorological Institute

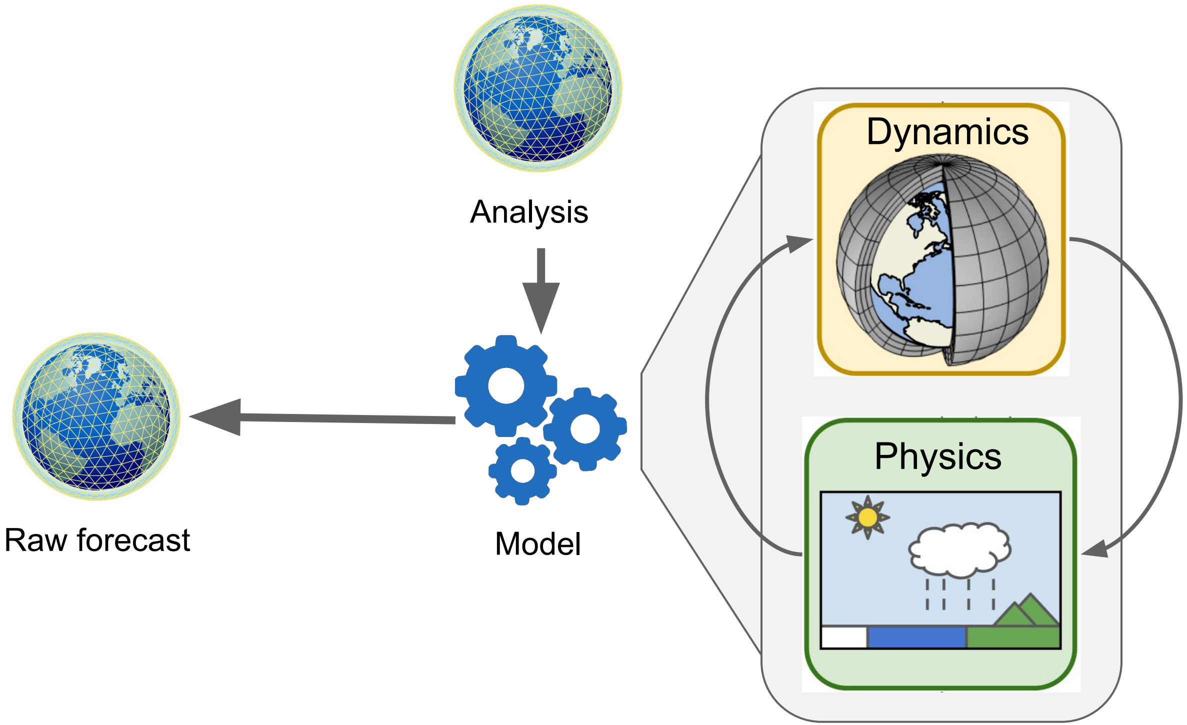
Catching the Wind:
Machine Learning Weather Prediction with anemoi

Michiel Van Ginderachter, Dieter Van Den Bleeken, Piet Termonia, Jef Philippé

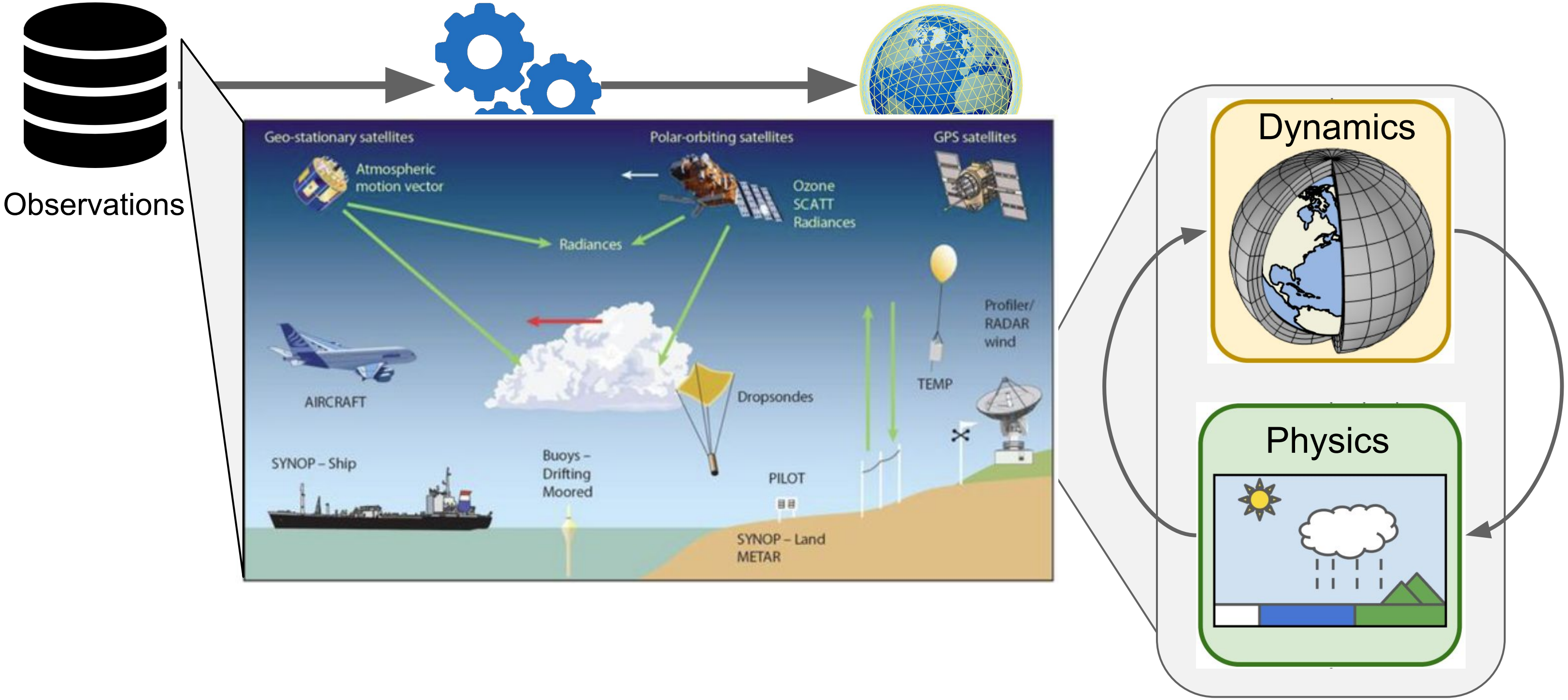
17.12.2025 | LUMI-BE User Day

Background

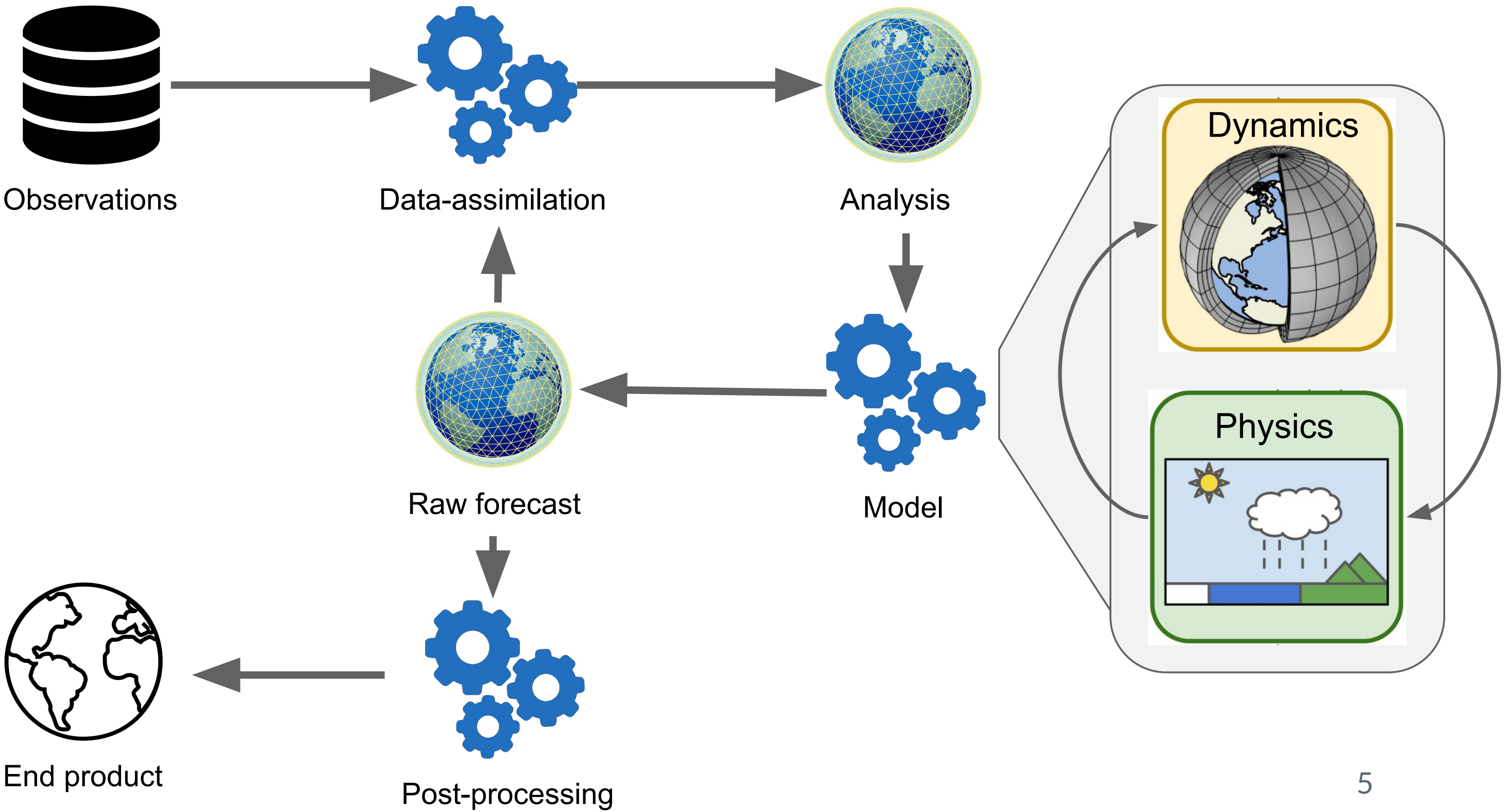
Numerical Weather Prediction Pipeline



Numerical Weather Prediction Pipeline



Numerical Weather Prediction Pipeline



70 years of Numerical Weather Prediction

1970 - Now: Slow but steady progress

REVIEW

doi:10.1038/nature14956

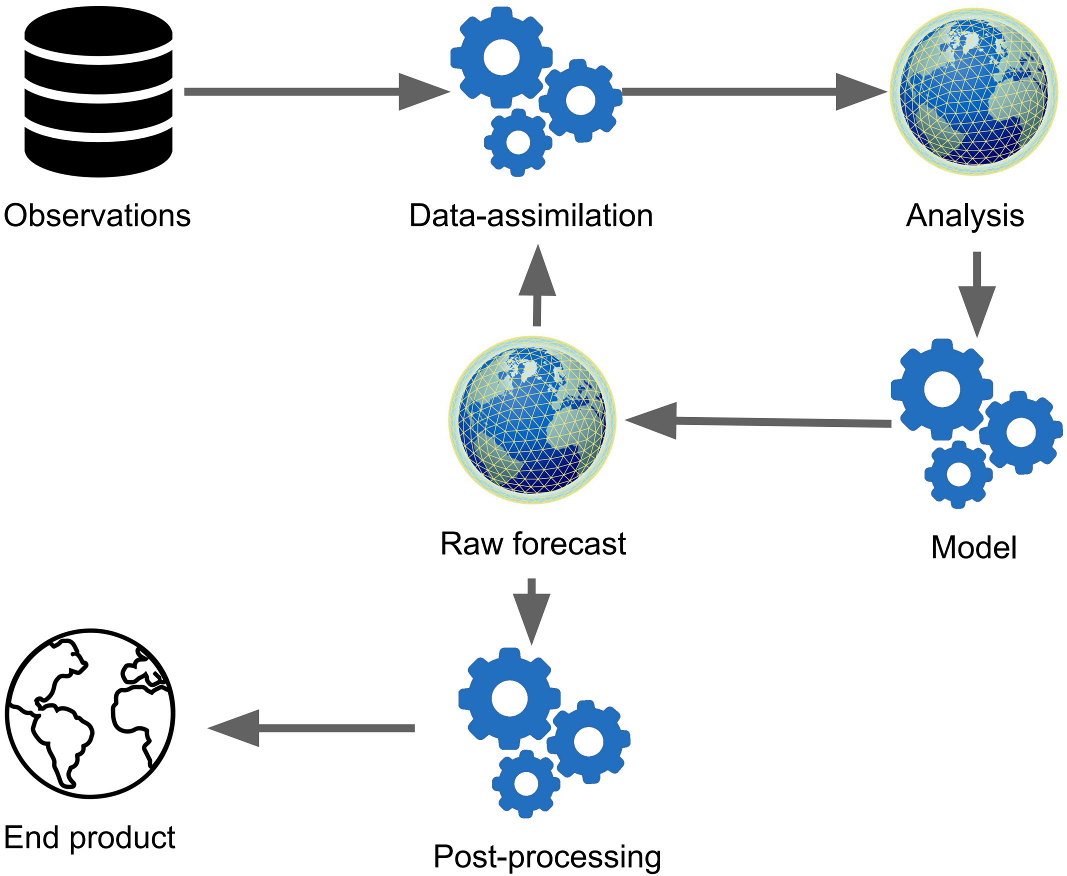
The quiet revolution of numerical weather prediction

Peter Bauer[1], Alan Thorpe[1] & Gilbert Brunet[2]

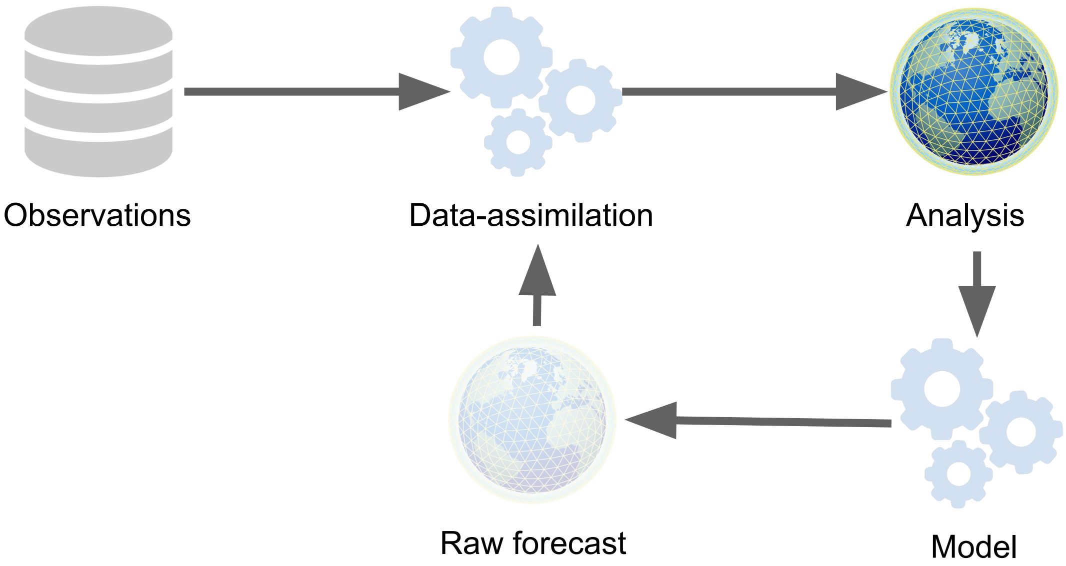
Advances in numerical weather prediction represent a quiet revolution because they have resulted from a steady accumulation of scientific knowledge and technological advances over many years that, with only a few exceptions, have not been associated with the aura of fundamental physics breakthroughs. Nonetheless, the impact of numerical weather prediction is among the greatest of any area of physical science. As a computational problem, global weather prediction is comparable to the simulation of the human brain and of the evolution of the early Universe, and it is performed every day at major operational centres across the world.

ECMWF

Reanalysis: Predicting the past



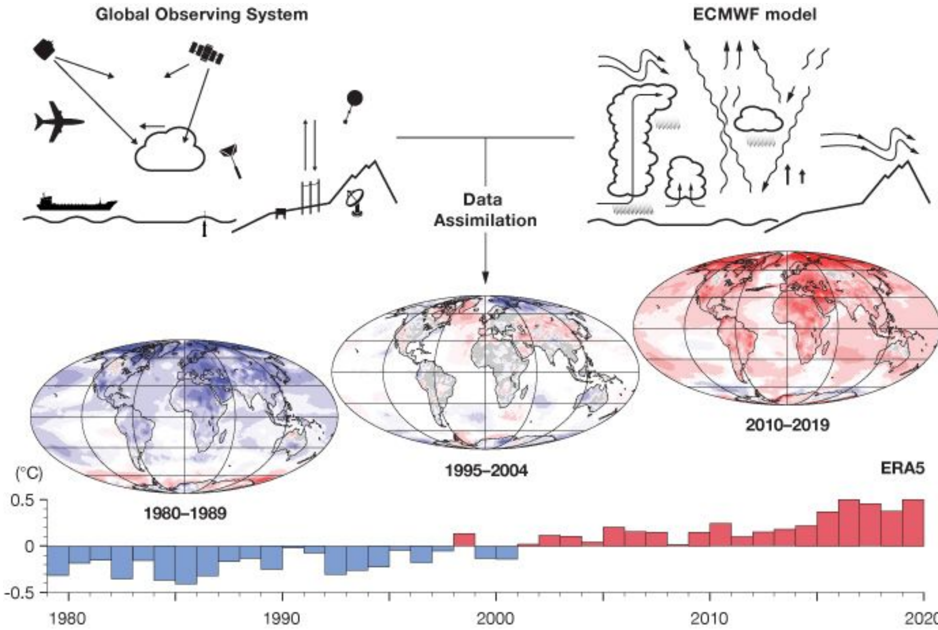
Reanalysis: Predicting the past



ECMWF ReAnalysis v5 (ERA-5)

Provides the most complete global picture currently possible of past weather and climate.

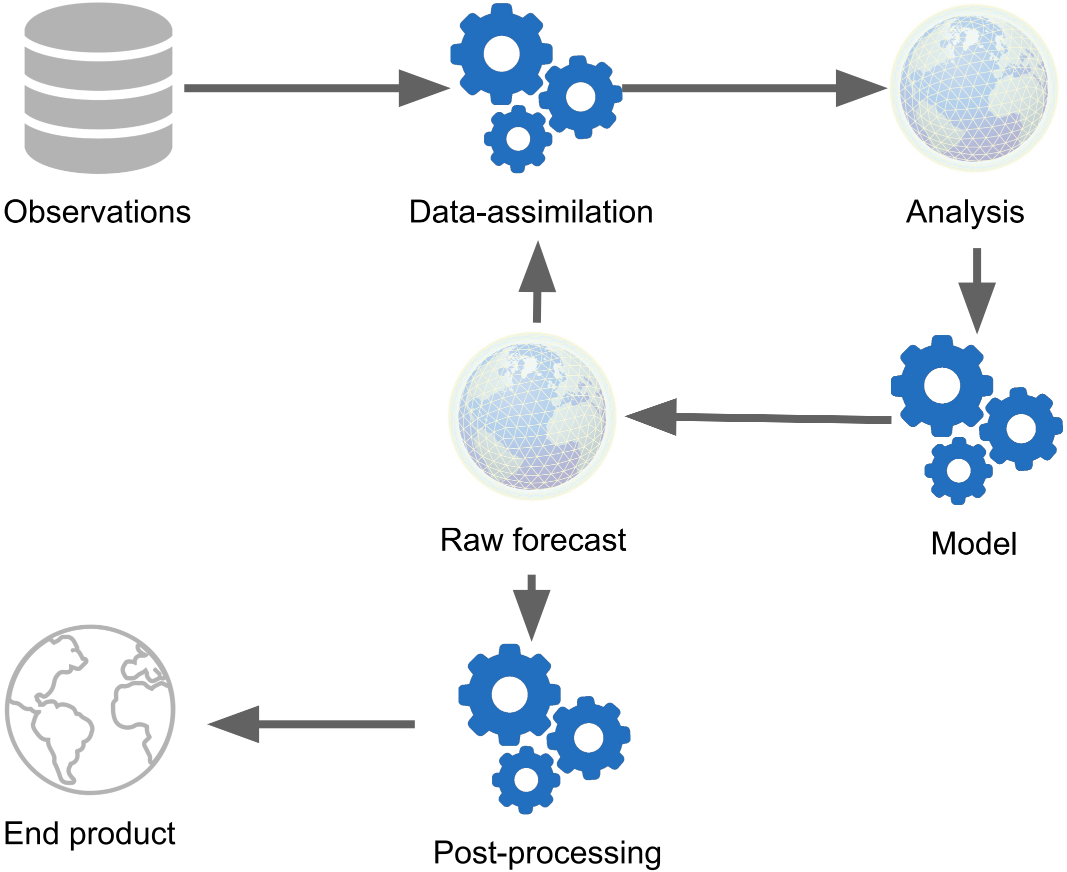
- hourly full 3D status of atmosphere
- 1940 - Now
- ~ 30 km resolution
- crucial for climate research
- 5 PB of data



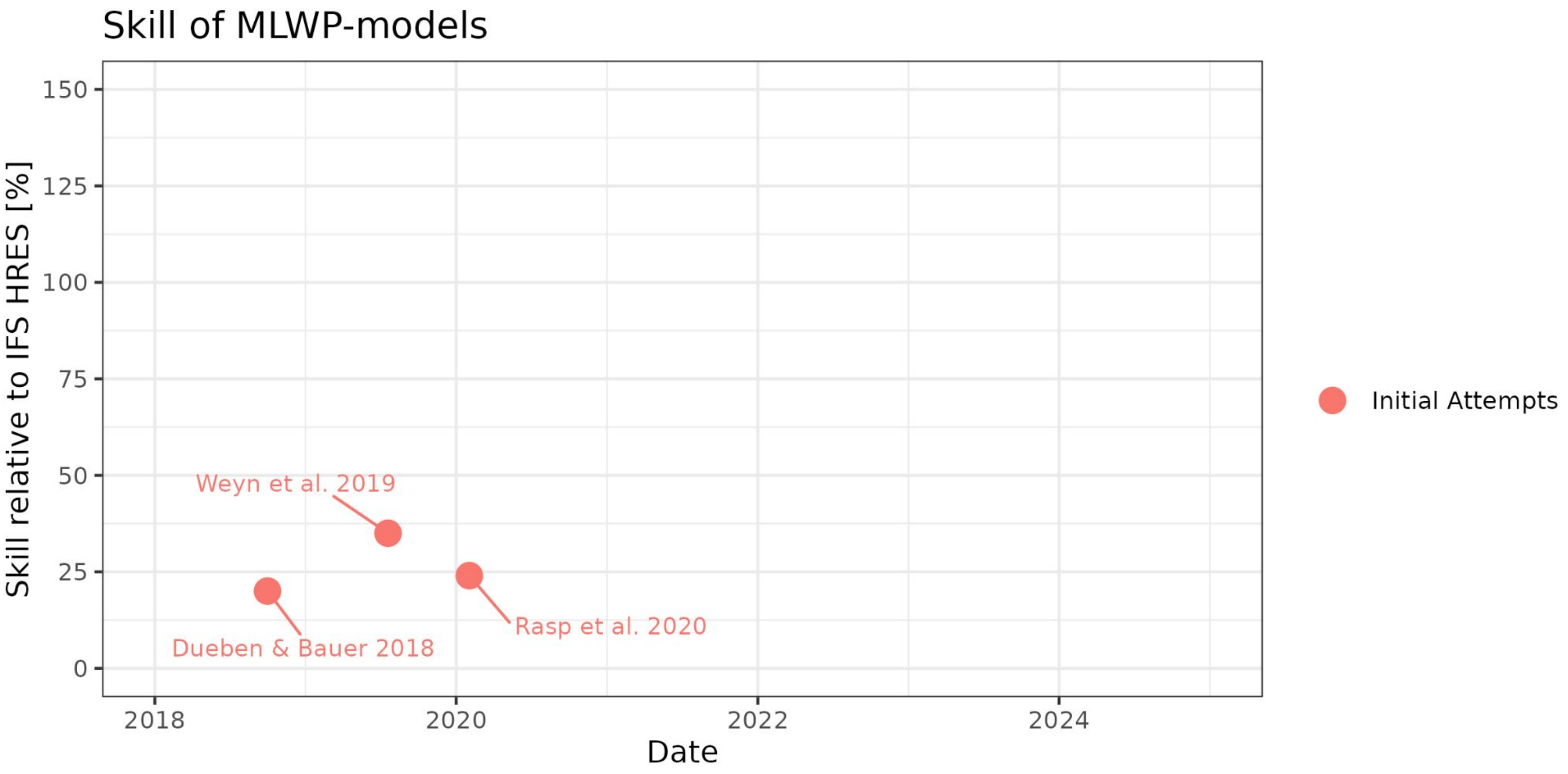
ECMWF

Machine Learning

Machine Learning the Weather Forecast

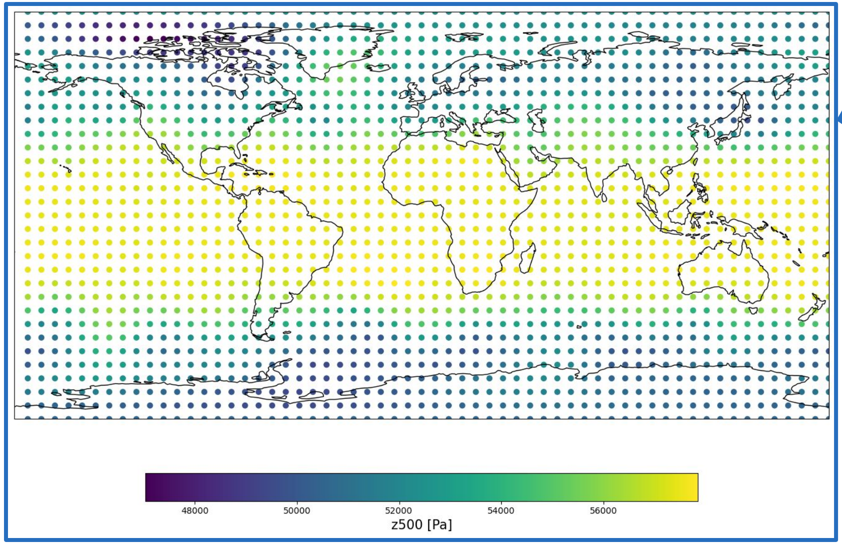


Machine Learning Weather Prediction: A Second Revolution



Initial Attempt

Z500 at time 00:00 h



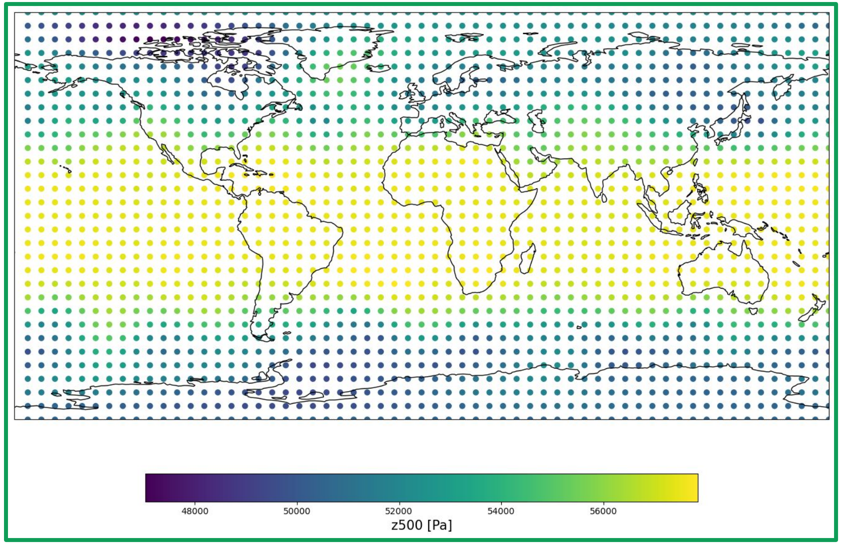
INPUT

LEARN

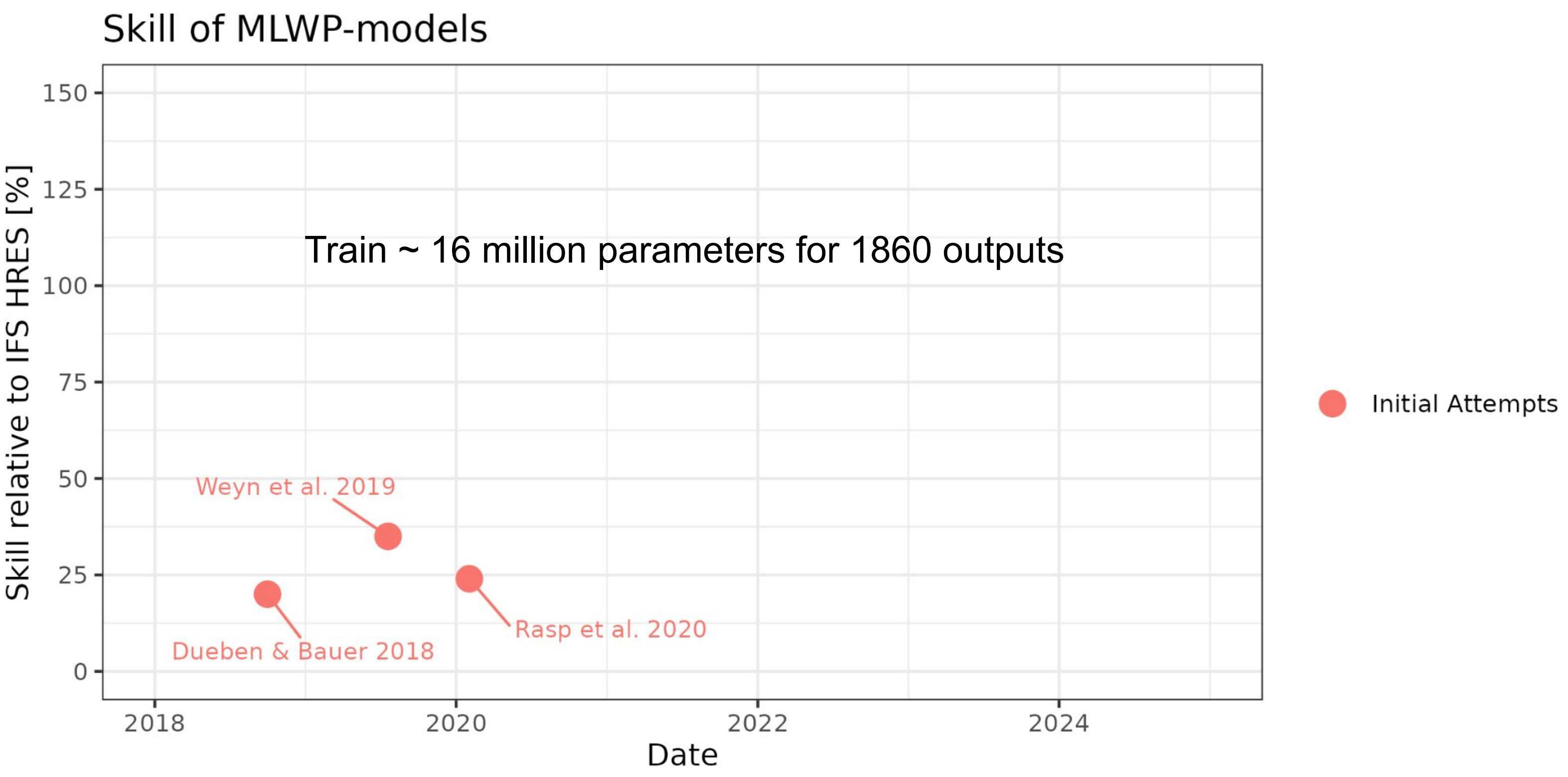
$$\mathbf{y} = \mathbf{F}(\mathbf{x};\, \mathbf{p})$$

OUTPUT

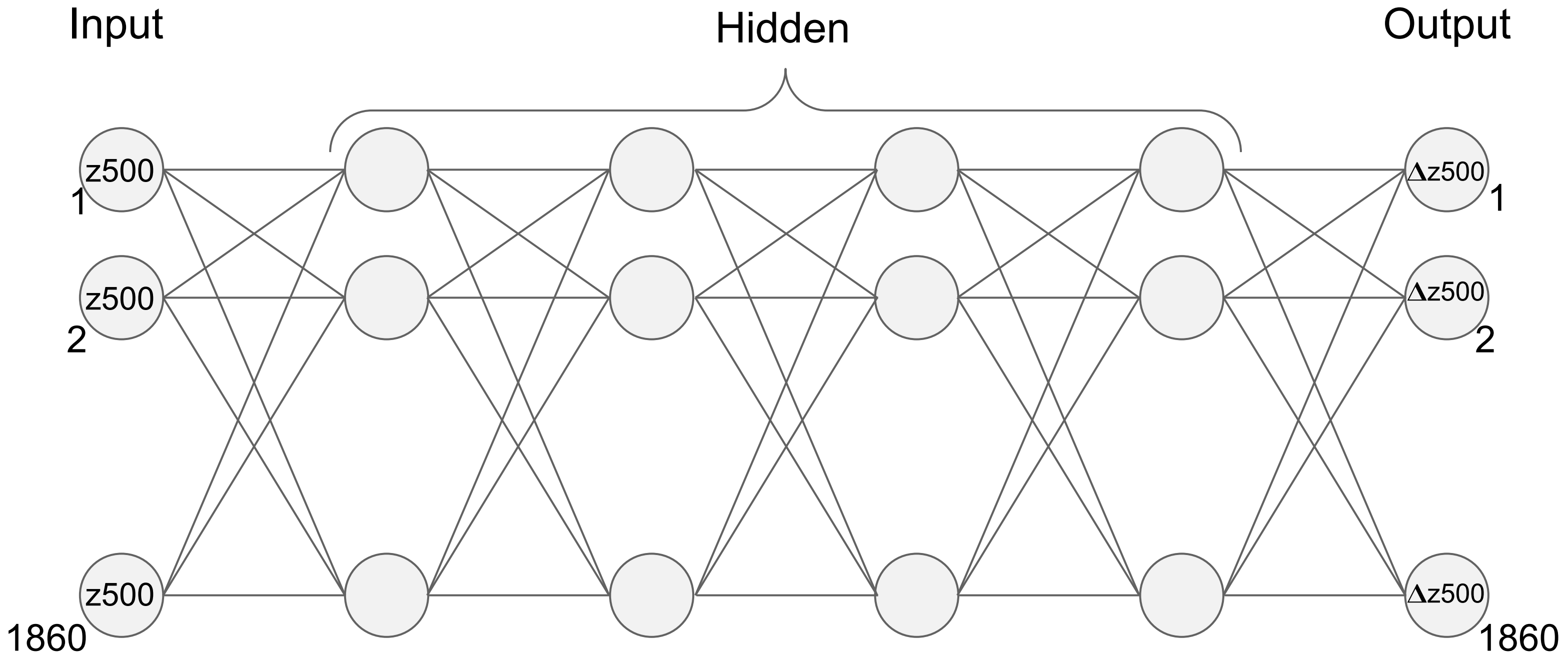
Z500 at time 01:00 h



Machine Learning Weather Prediction: A Second Revolution



Dueben & Bauer 2018

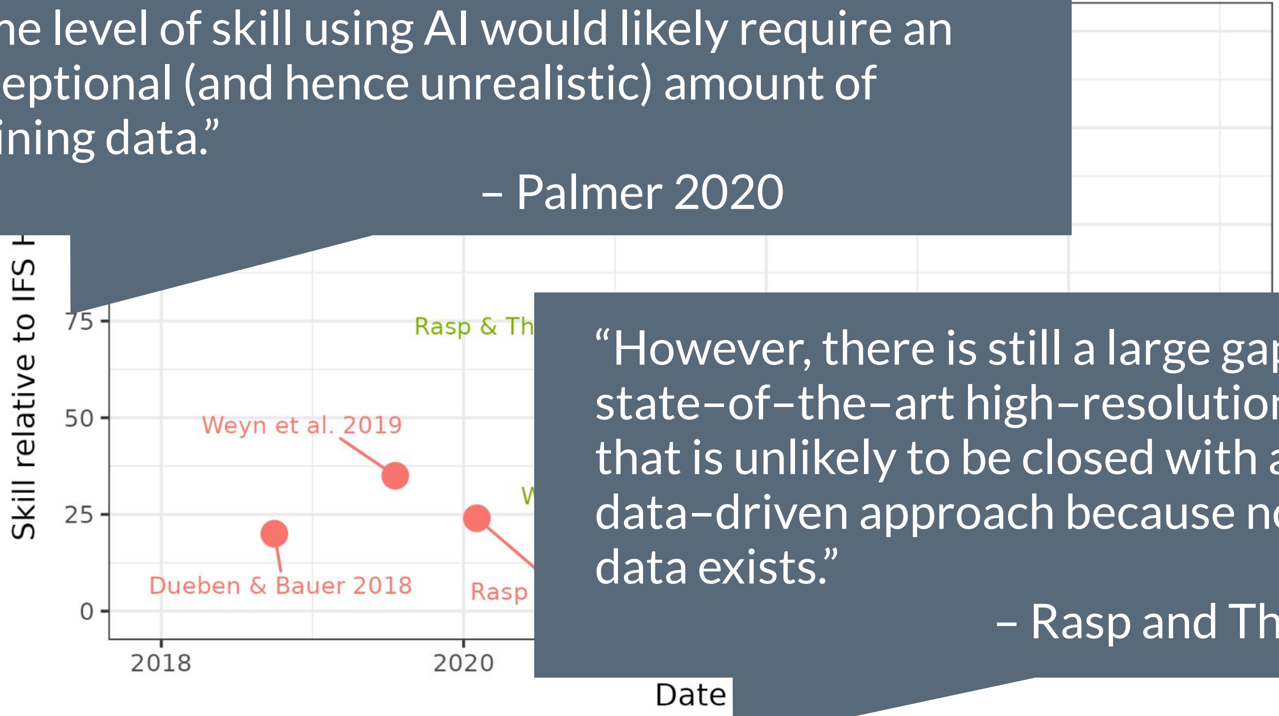


z500(t+Δt) = z500(t) + Δz500

Train ~ 16 million parameters for 1860 outputs

“An ability to forecast reliably the probability of highly nonlinear phenomena in the medium range using NWP, requires high-quality models run from high-quality initial conditions. To do this with the same level of skill using AI would likely require an exceptional (and hence unrealistic) amount of training data.”

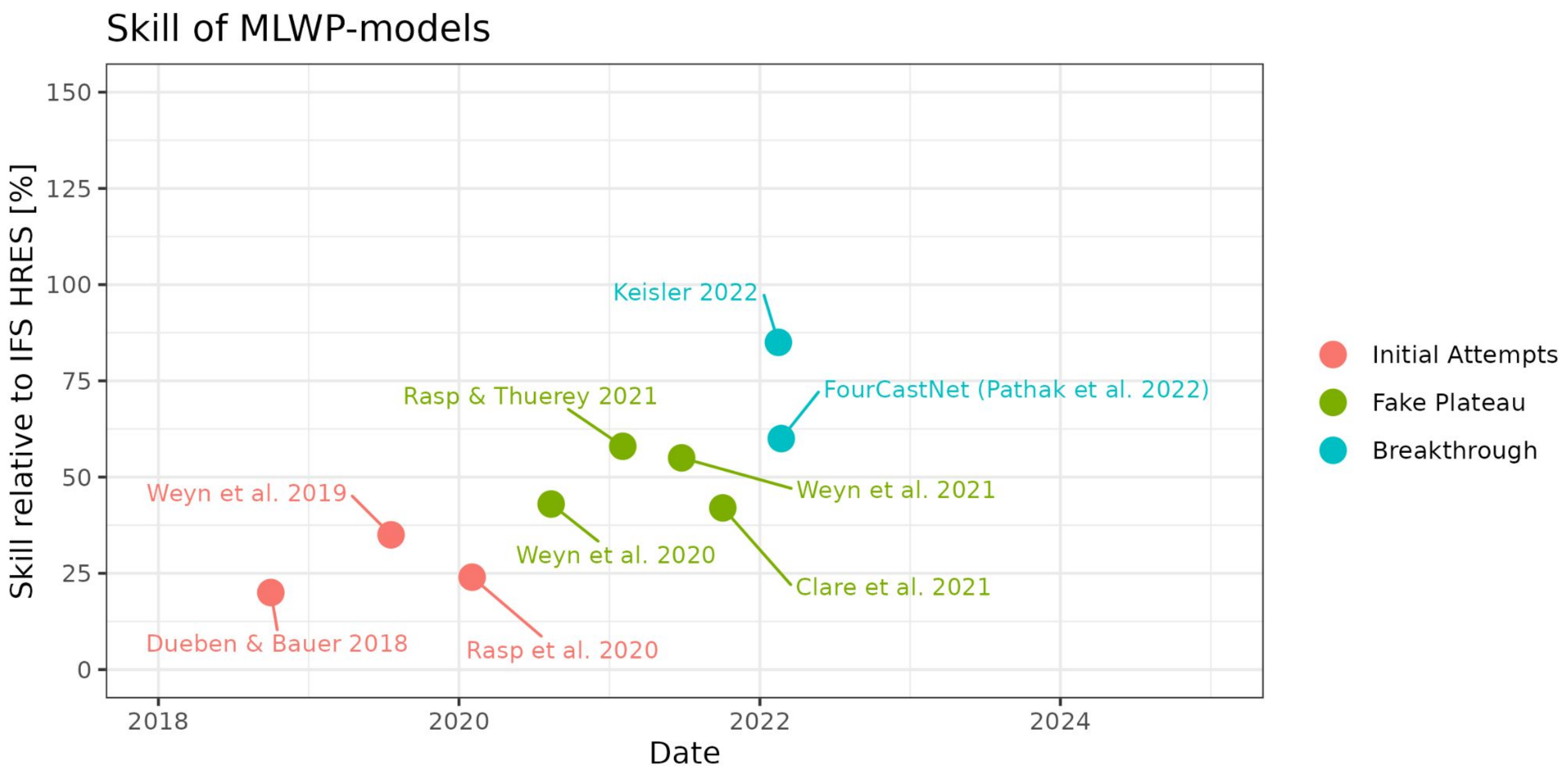
– Palmer 2020



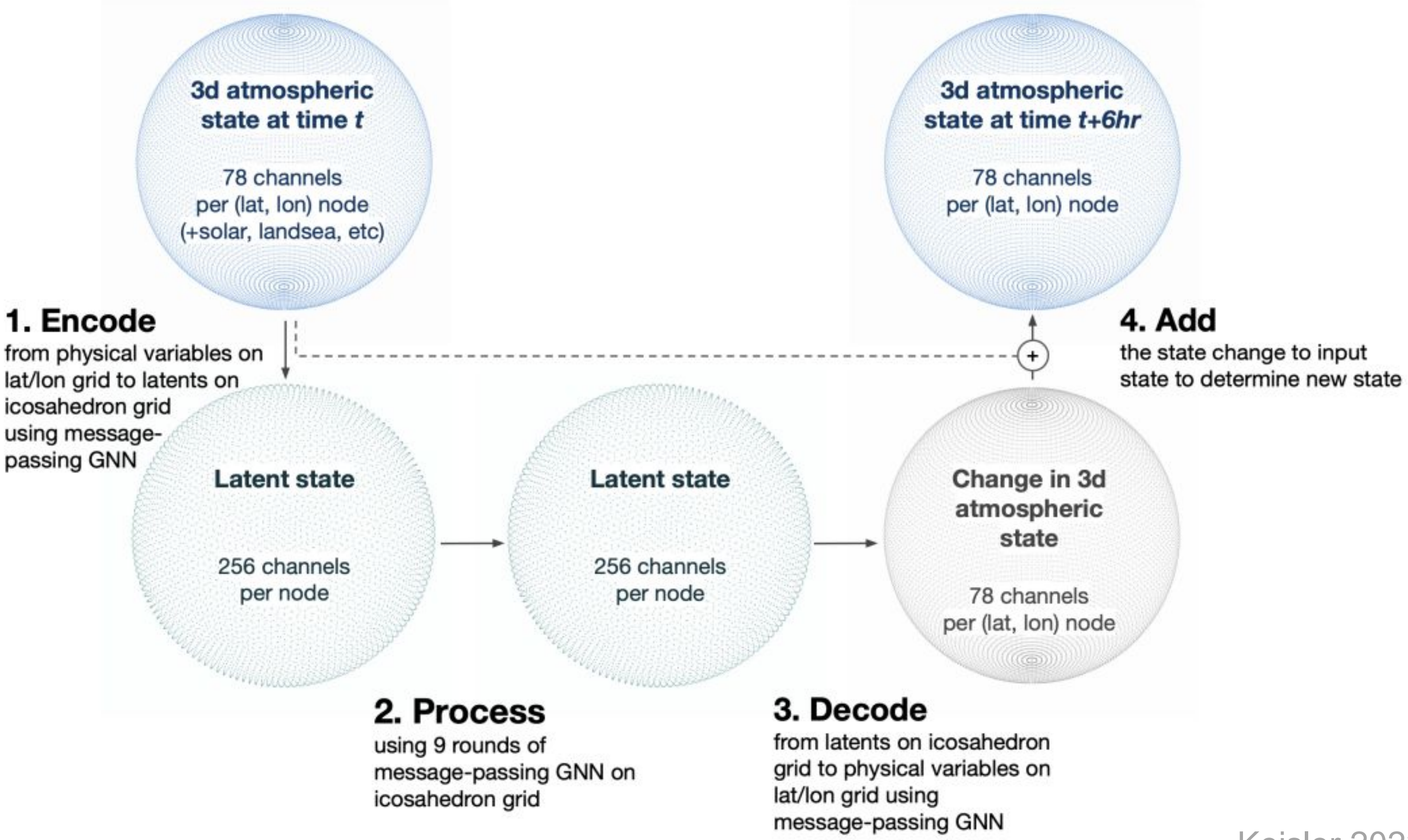
“However, there is still a large gap to current state–of–the–art high–resolution weather models that is unlikely to be closed with a purely data–driven approach because not enough training data exists.”

– Rasp and Thuerey 2021

MLWP: A Second Revolution

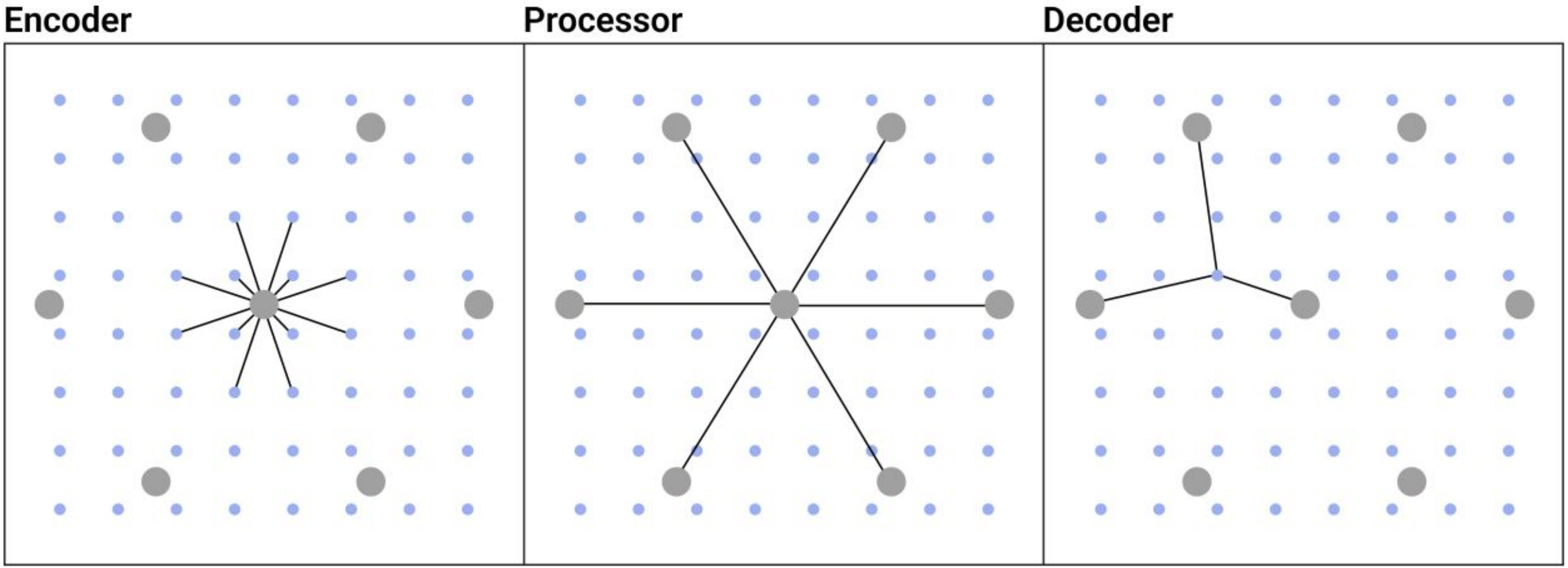


Keisler 2022: Graph Neural Networks



Keisler 2022

Keisler 2022: Graph Neural Networks

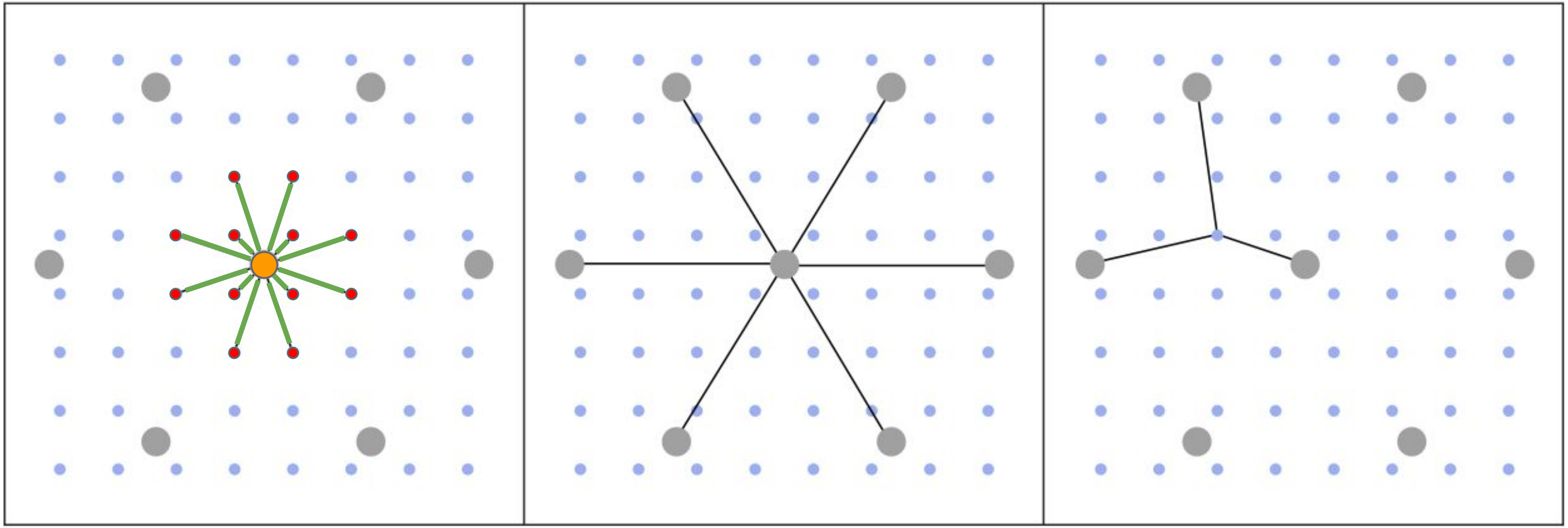


Keisler 2022

Train ~6.7 million parameters for ~5 million outputs

Keisler 2022: Graph Neural Networks

Encoder **Processor** **Decoder**

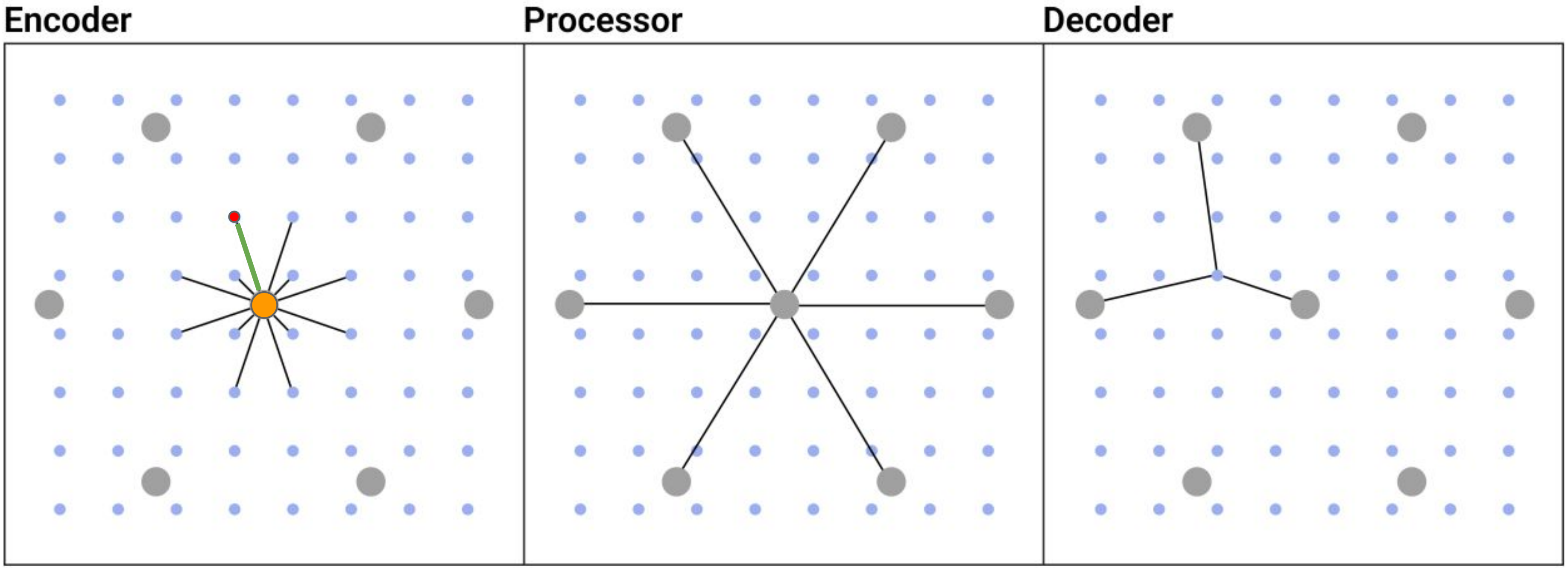


Mesh node $\mathbf{v}_r^M$

$\mathbf{e}^{G2M}$ Grid2Mesh edge

Grid node $\mathbf{v}_s^G$

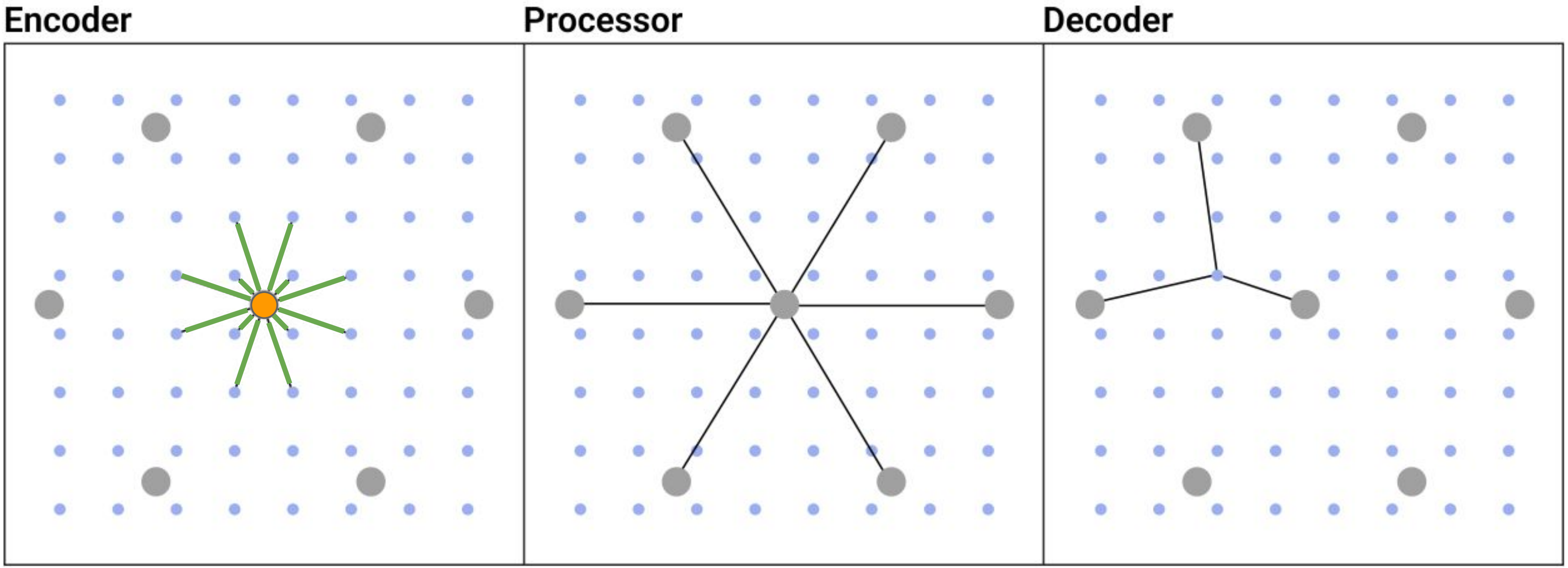
Keisler 2022: Graph Neural Networks



1. Update Grid2Mesh edges

$$\mathbf{e}'^{G2M} = f(\mathbf{e}^{G2M}, \mathbf{v}_s^G, \mathbf{v}_r^M)$$

Keisler 2022: Graph Neural Networks

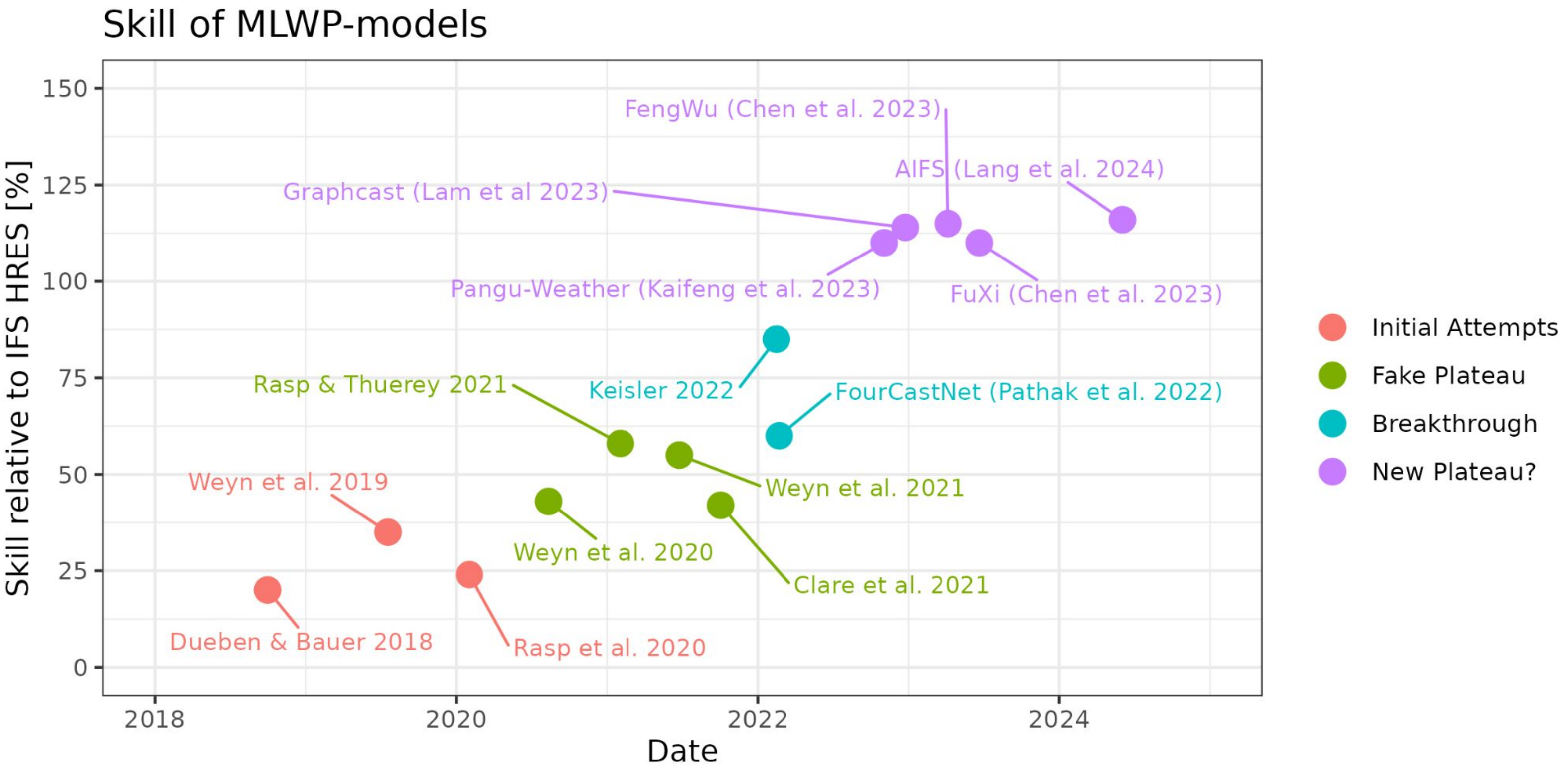


2. Update mesh nodes

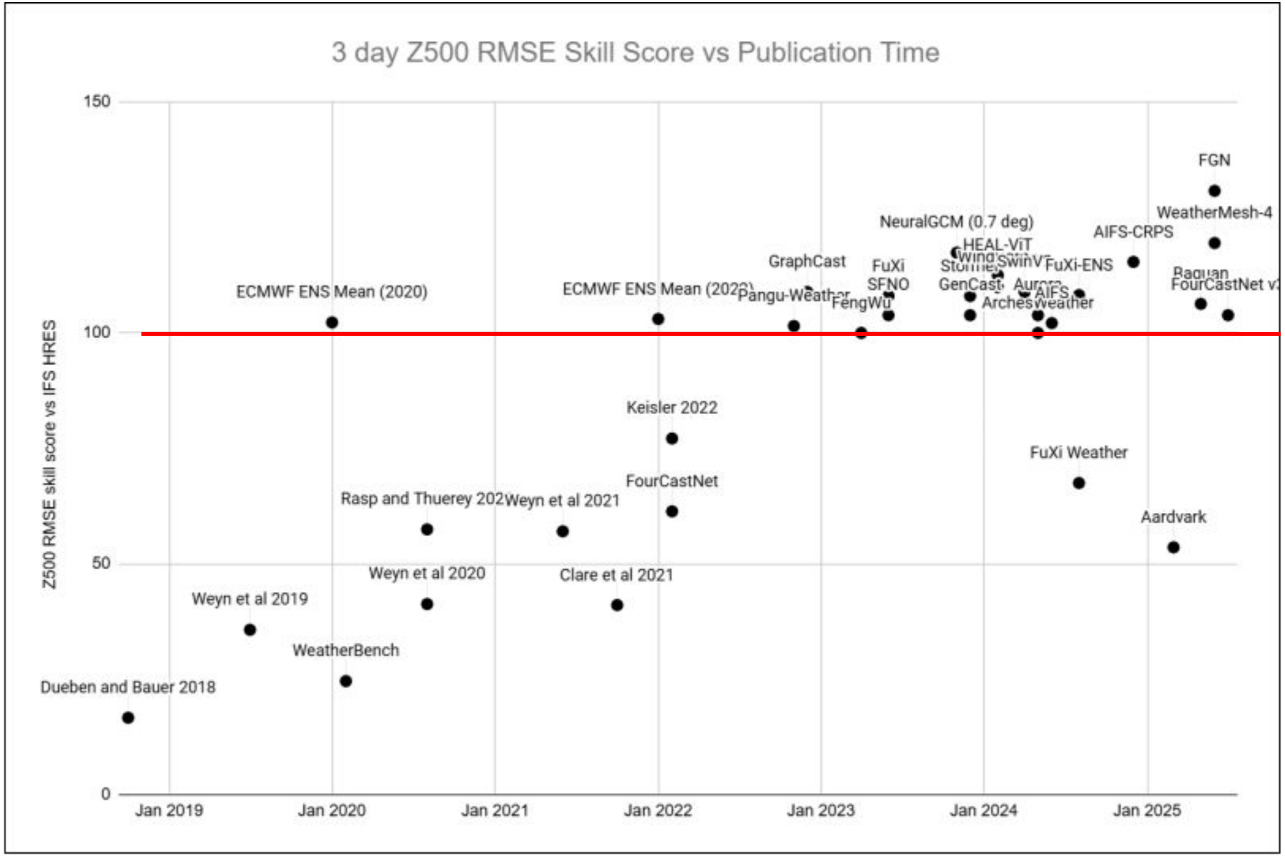
$$\mathbf{v}'^{M}_{i} = f\left(\mathbf{v}^{M}_{i}, \sum \mathbf{e}'^{G2M}\right)$$

$f =$ 

MLWP: A Second Revolution

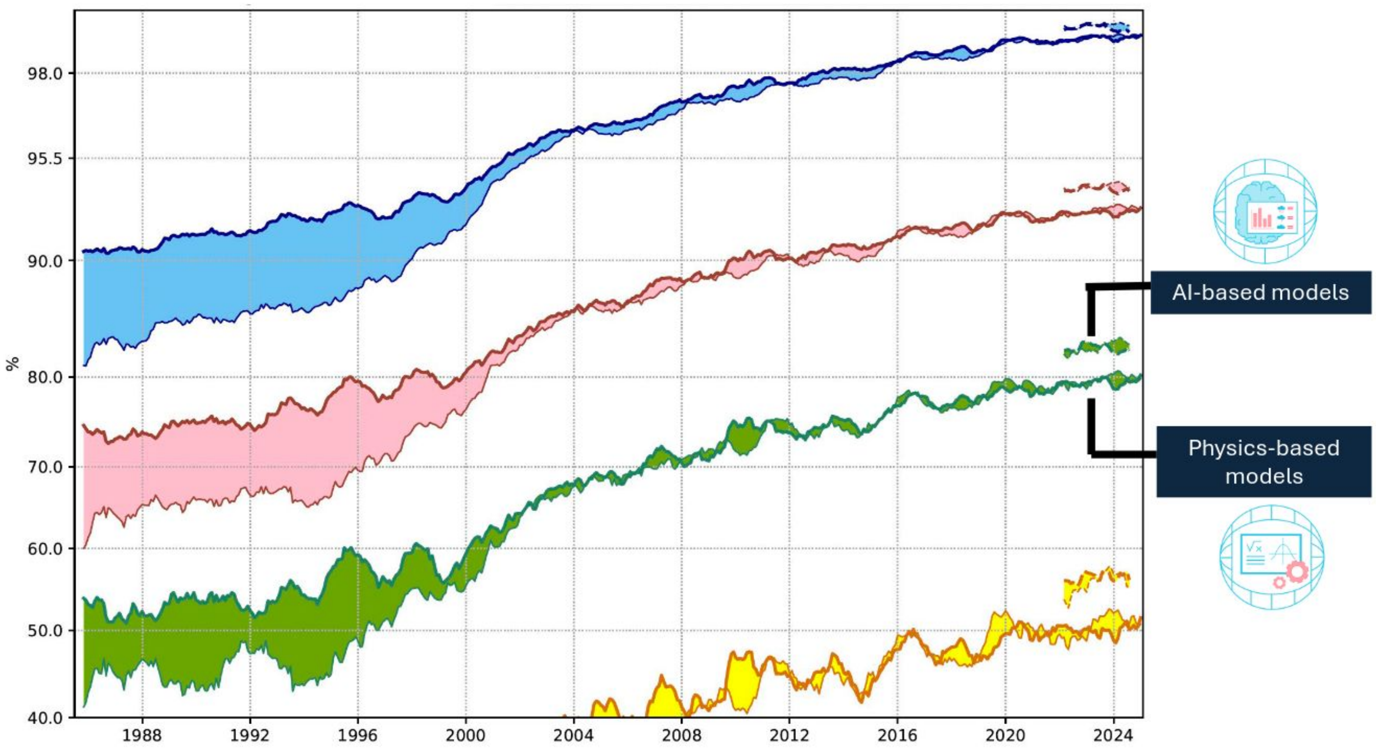


More than hype



Rasp, Stephan (2024). AI-Weather SotA vs Time. figshare. Dataset. https://doi.org/10.6084/m9.figshare.28083515.v1

More than hype



Courtesy of M. Chantry

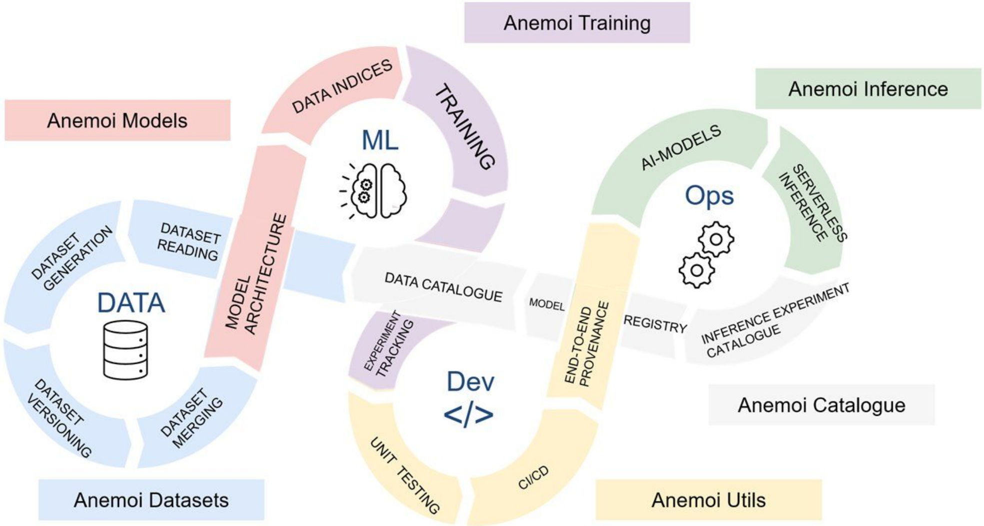
anemoi

anemoi

Anemoi Framework

- Provide **complete toolkit** to develop **data-driven meteorological forecasting** covering the whole ML lifecycle

- Anemoi is developed through a **collaborative European initiative**



07/05/2025

EMS Technology Achievement Award 2025 for ANEMOI

The EMS Technology Achievement Award for significant technology achievements and innovations in the field of meteorology and earth observation –

The European Meteorological Society is awarding the EMS Technology Achievement Award 2025 to the Anemoi Framework:

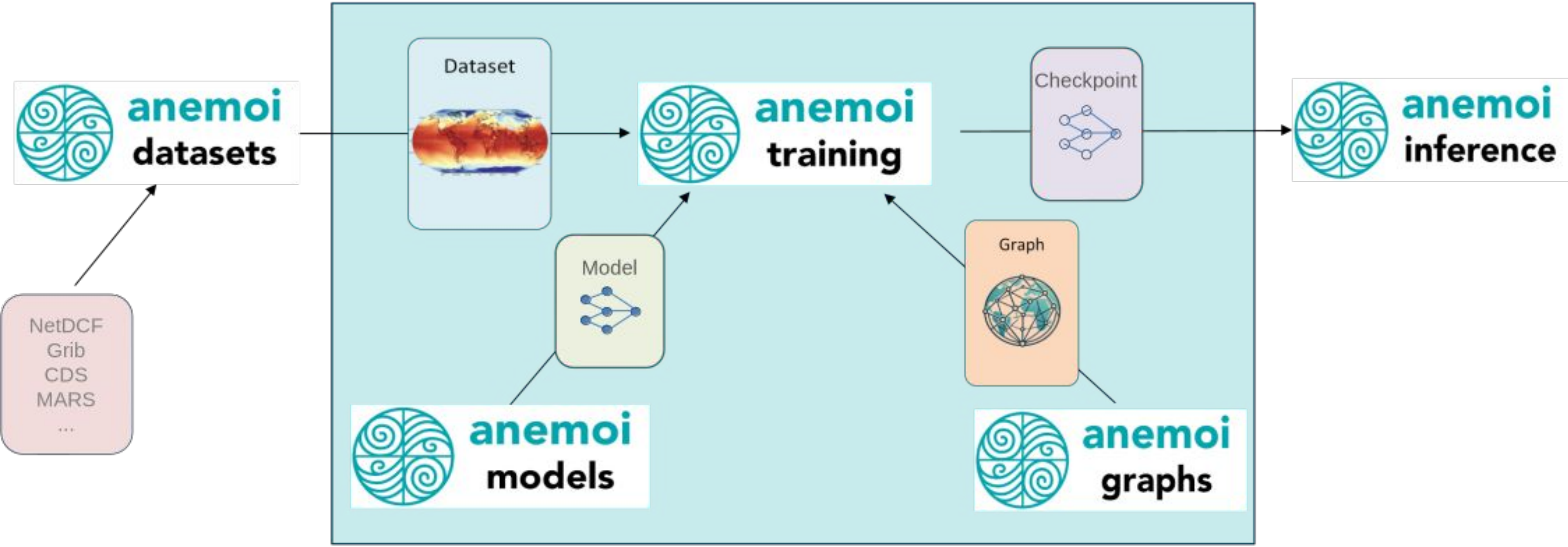
> *The Anemoi Framework is an excellent example of a European collaborative effort, which offers enhanced forecast accuracy through the use of advanced machine-learning methodologies. The flexible open-source approach enables various European stakeholders to further integrate AI in operational forecasting.*

https://www.emetsoc.org/ems-technology-achievement-award-2025-for-anemoi/

Anemoi in a nutshell

https://github.com/ecmwf/anemoi

https://anemoi.readthedocs.io



"anemoi-core"

Built on shoulder of giants

- Don’t reinvent the wheel: rely on existing tools
- Main dependencies

PyTorch PyTorch Lightning PyG

earthkit Zarr xarray

Hydra Pydantic mlflow

Design choices

- Open-Source
  - Collaboration with member states
- Focus on best use of resources (File systems, GPUs, ...)
  - Do not starve the GPUs during training, due to slow I/Os
- Makes Research-to-operations as simple as possible
  - Inference and training are independent
  - Each component collects metadata that can be used by the others

High resolution data-driven weather forecasting

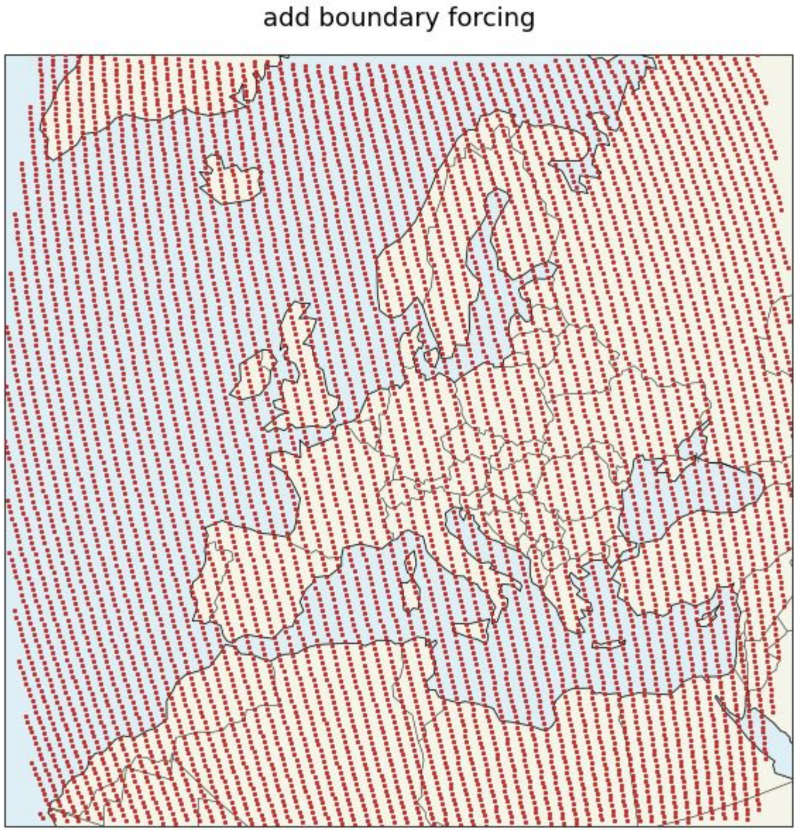
Leverage high-resolution datasets

- CERRA - dataset
  - 5.5 km
  - 36 years

regional high-resolution reanalysis

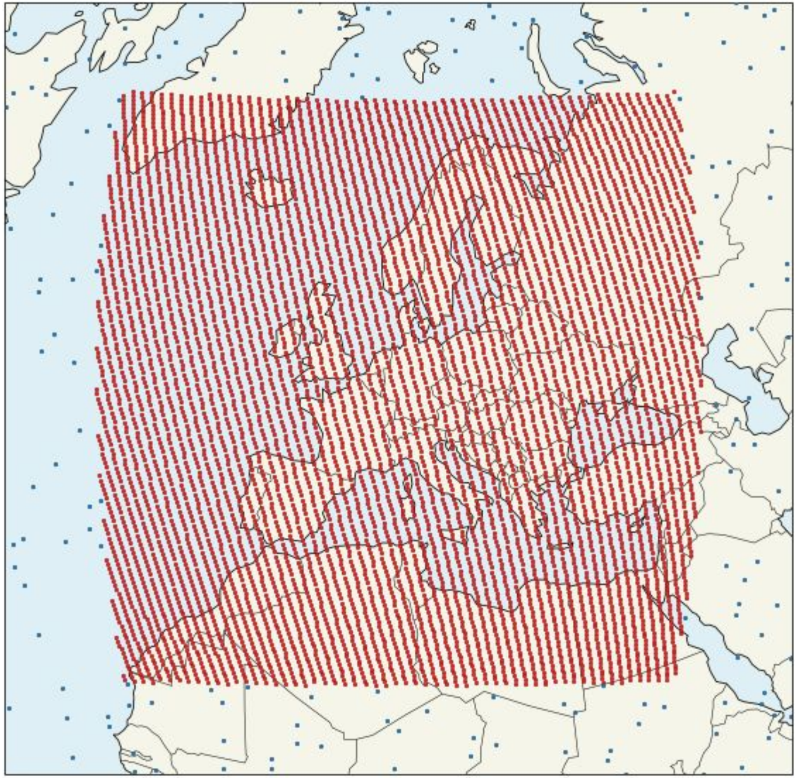
Leverage high-resolution datasets

- CERRA - dataset
  - 5.5 km
  - 36 years
- Limited Area Model
  - High resolution
  - Lower resolution boundary



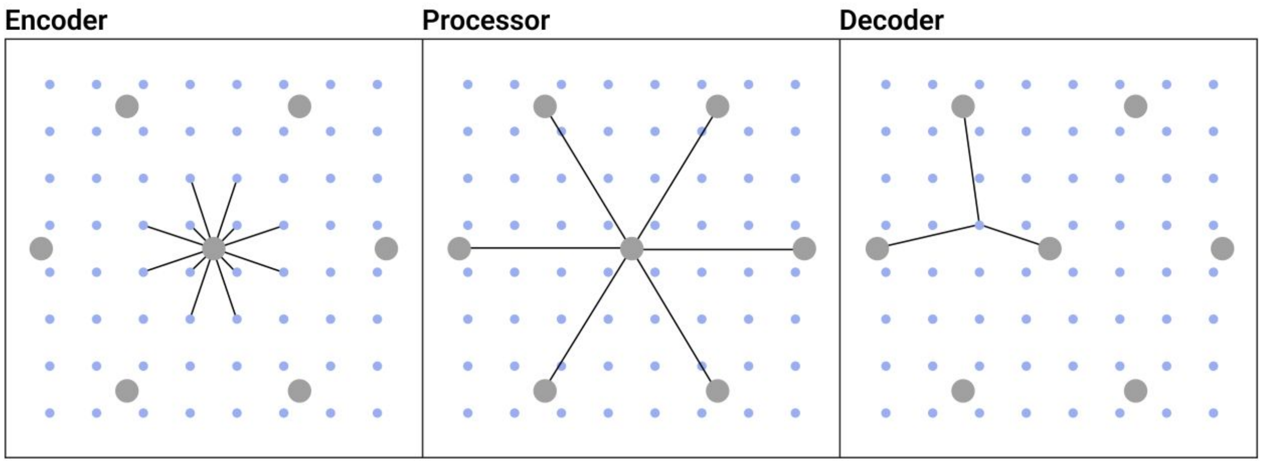
Leverage high-resolution datasets

- CERRA - dataset
  - 5.5 km
  - 36 years
- Limited Area Model
  - High resolution
  - Lower resolution boundary
- Learn only on regional domain



Model architecture

- Encoder - Processor - Decoder
- 6h timestep
- Graph-Transformer (1024 channels) → 246 million trainable parameters



ML-training workflow on LUMI

Training setup

- Model no longer fits on 1 GPU
  - **Model-parallel**: model is *sharded* over multiple GPUs
  - 8 GPUs (= 1 node)

- **Data-parallel**: multiple samples divided over multiple GPUs
  - batch size 16
  - 16 nodes

⇒ 128 GPUs x 19 days

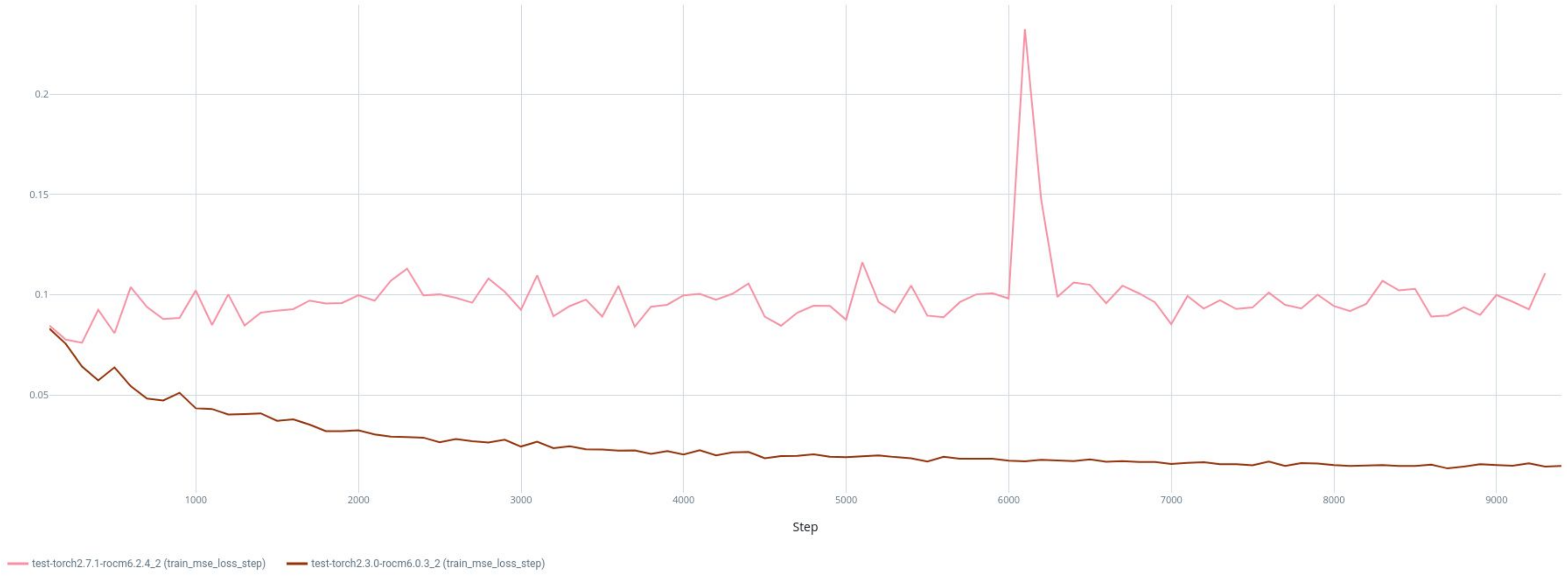
⇒ ~30.000 GPUh

LUMI environment

- Use Cotainr to build container from base container
  - ROCm 6.0.3 + torch2.3 + anemoi
- But also developments needed
  - Container: ROCm 6.0.3 + torch2.3 + parts of anemoi
  - Virtual environment: anemoi-graphs | anemoi-training
- Some issues with NCCL timeout

```
Epoch 2:   0%|          | 977/253732 [1:24:43<365:16:44,  0.19it/s, v_num=b133, train_mse_loss_step=0.0198, val_mse_loss_step=0.0218,
val_mse_loss_epoch=0.0211, train_mse_loss_epoch=0.0239]
Epoch 2:   0%|          | 977/253732 [1:24:43<365:16:51,  0.19it/s, v_num=b133, train_mse_loss[rank111]:[E ProcessGroupNCCL.cpp:563] [Rank 7]
Watchdog caught collective operation timeout: WorkNCCL(SeqNum=1970765, OpType=ALLGATHER, NumelIn=18446744073709551615,
NumelOut=18446744073709551615, Timeout(ms)=600000) ran for 600026 milliseconds before timing out.
[rank109]:[E ProcessGroupNCCL.cpp:563] [Rank 5] Watchdog caught collective operation timeout: WorkNCCL(SeqNum=1970765, OpType=ALLGATHER,
NumelIn=18446744073709551615, NumelOut=18446744073709551615, Timeout(ms)=600000) ran for 600054 milliseconds before timing out.
[rank109]:[E ProcessGroupNCCL.cpp:1537] [PG 1 Rank 5] Timeout at NCCL work: 1970765, last enqueued NCCL work: 1970765, last completed NCCL work:
1970756.
[rank109]:[E ProcessGroupNCCL.cpp:577] [Rank 5] Some NCCL operations have failed or timed out. Due to the asynchronous nature of CUDA kernels,
subsequent GPU operations might run on corrupted/incomplete data.
[rank109]:[E ProcessGroupNCCL.cpp:583] [Rank 5] To avoid data inconsistency, we are taking the entire process down.
[rank109]:[E ProcessGroupNCCL.cpp:1414] [PG 1 Rank 5] Process group watchdog thread terminated with exception: [Rank 5] Watchdog caught
collective operation timeout: WorkNCCL(SeqNum=1970765, OpType=ALLGATHER, NumelIn=18446744073709551615, NumelOut=18446744073709551615,
Timeout(ms)=600000) ran for 600054 milliseconds before timing out.
Exception raised from checkTimeout at ../torch/csrc/distributed/c10d/ProcessGroupNCCL.cpp:565 (most recent call first):
```

Let’s update our software stack



Let’s update our software stack

- Move to torch 2.7 (it has some nice features)

- But it requires ROCm >= 6.2 (available on LUMI)

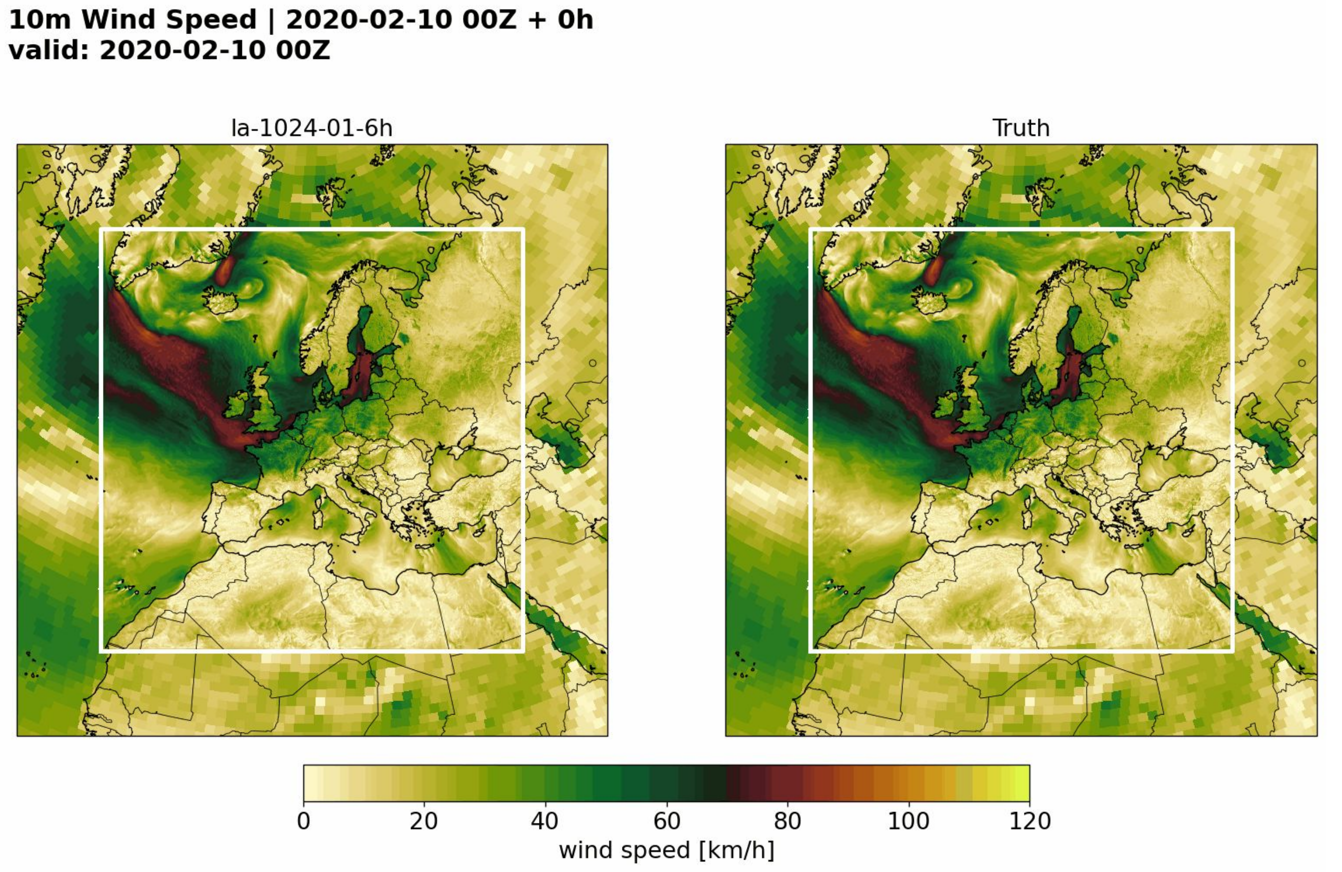
⇒ Issues with precision handling

ROCm 6.0 + torch 2.3: 16-mixed

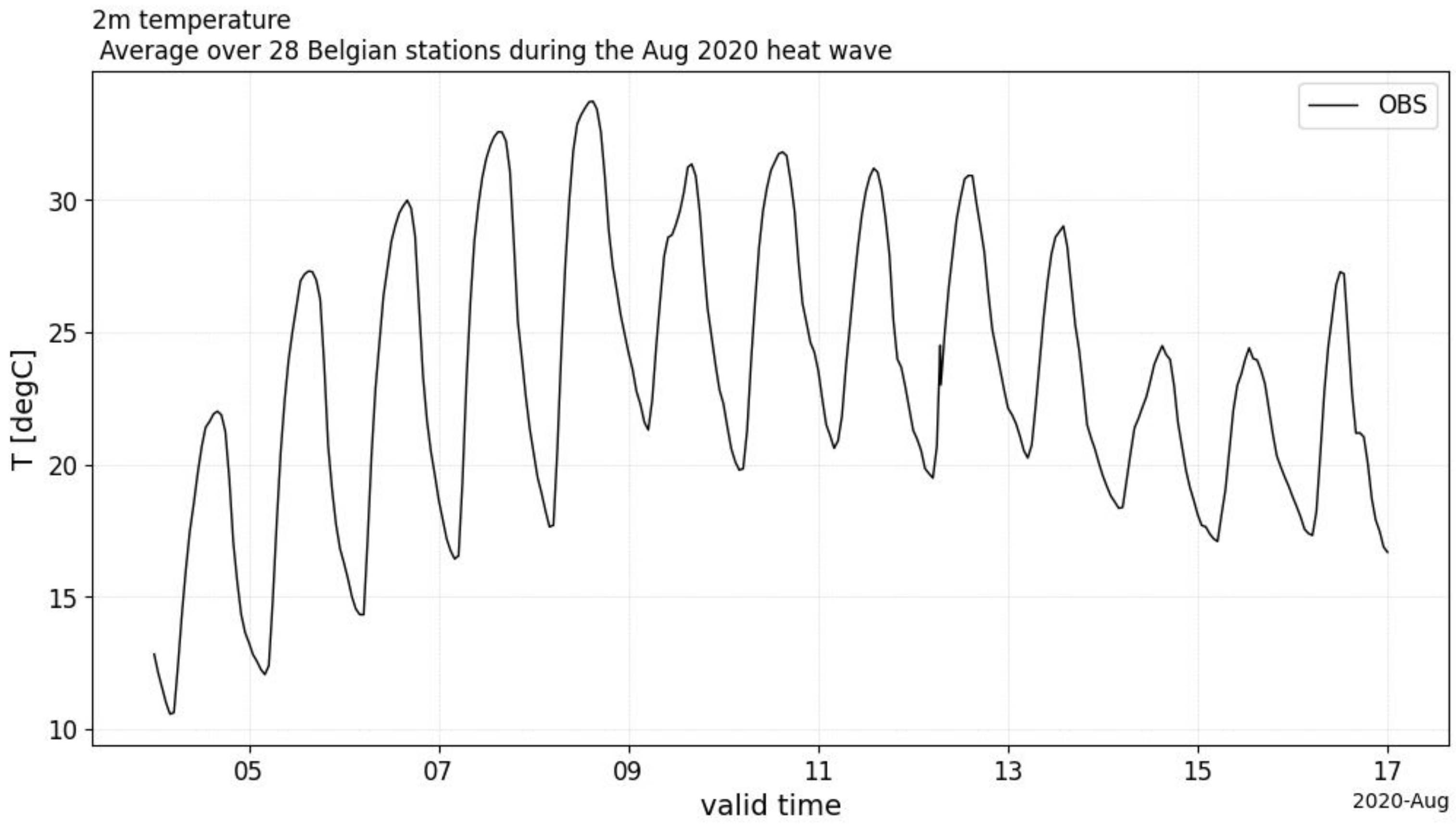
ROCm 6.3 + torch 2.7: 32-true | bf16-mixed

Some results

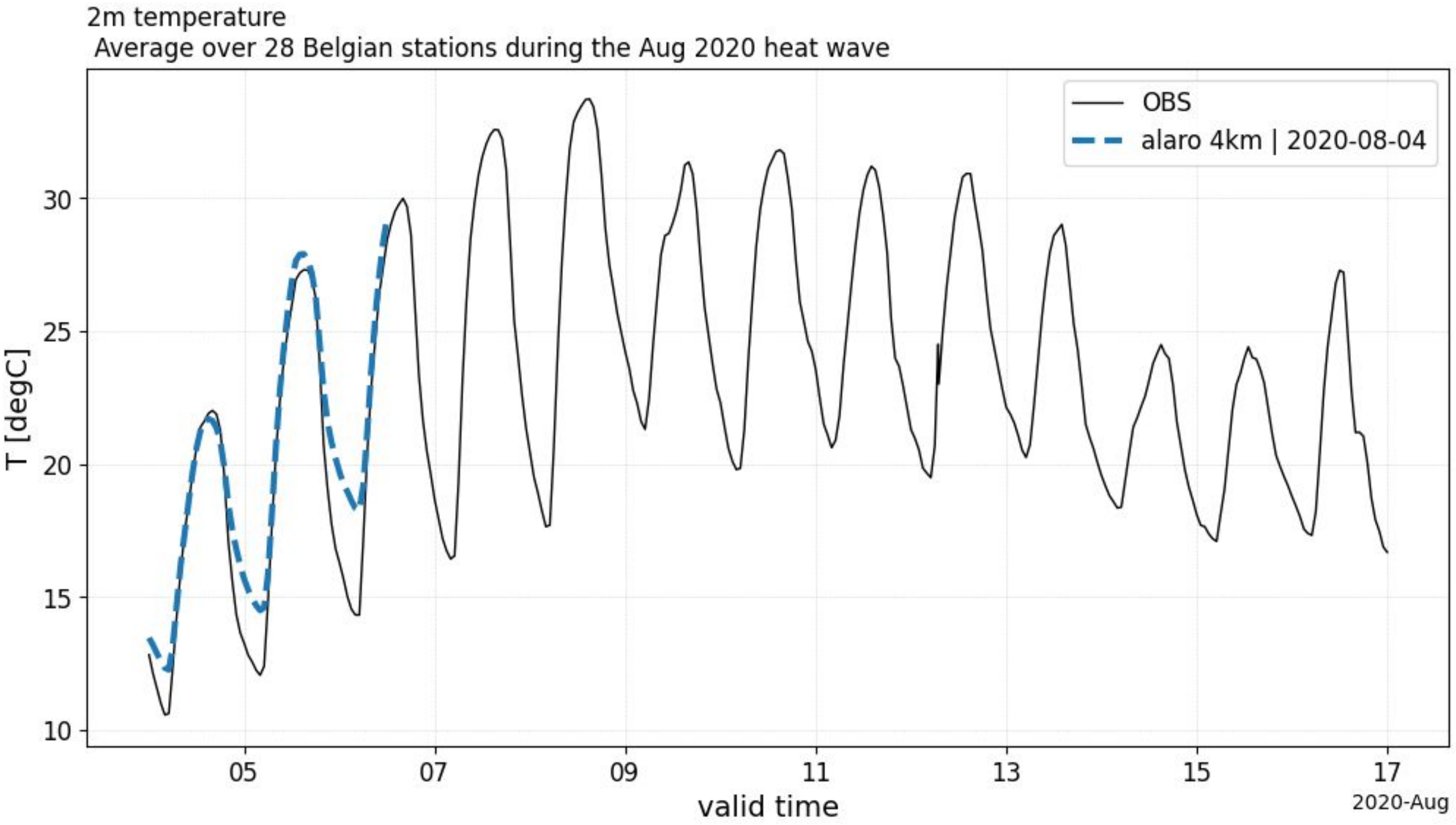
Case studies: Storm CIARA



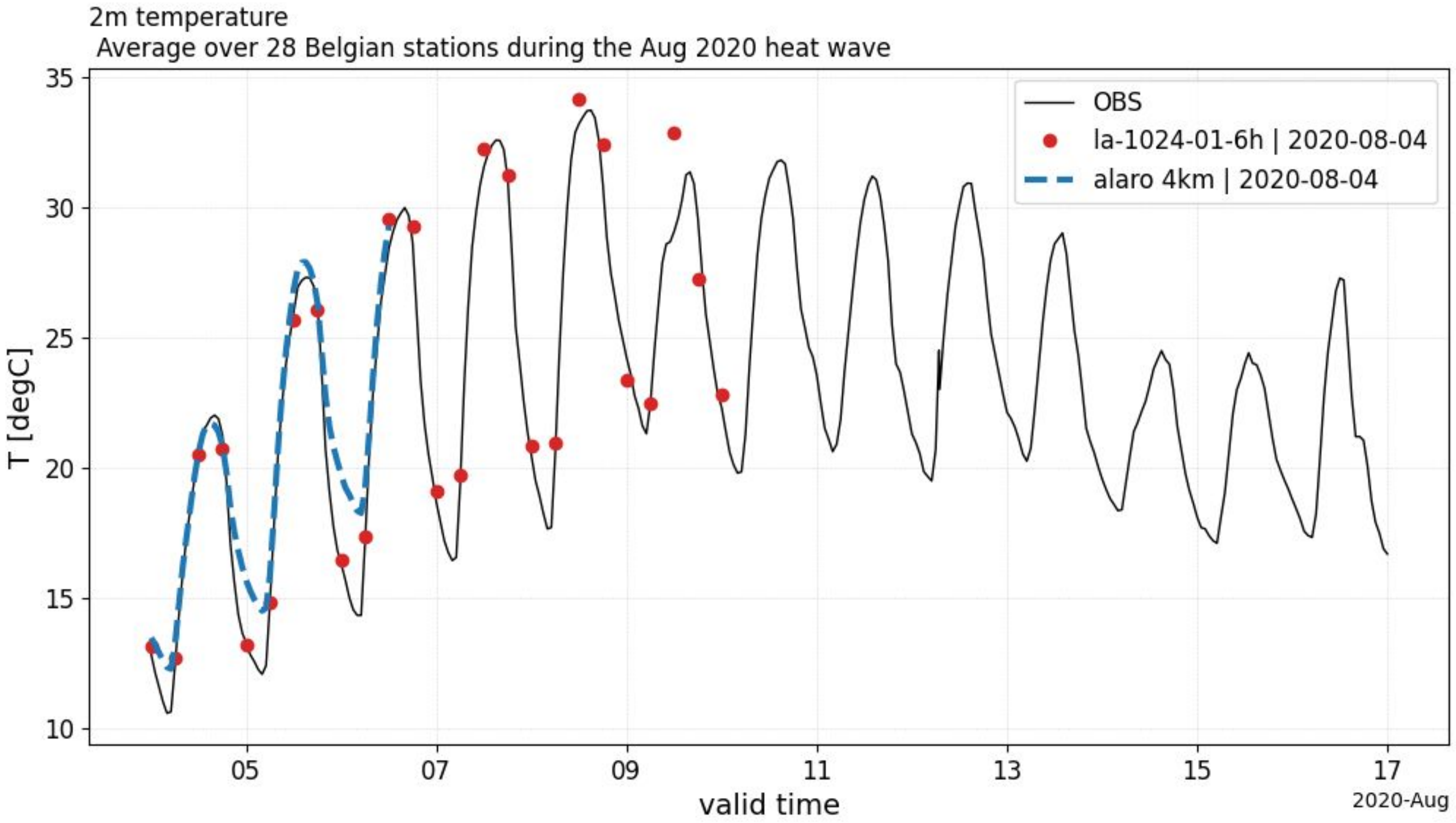
Case studies: Heatwave



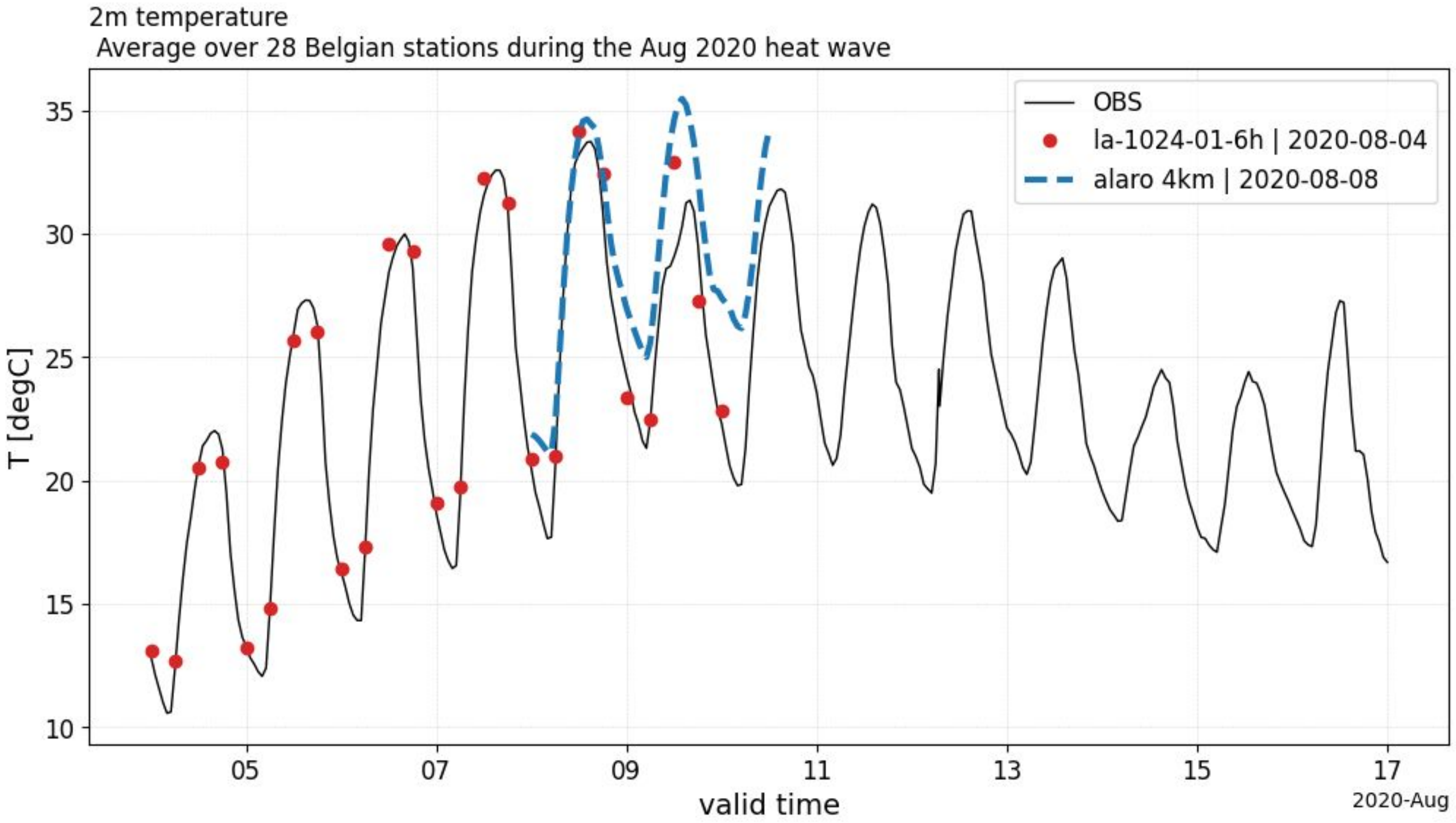
Case studies



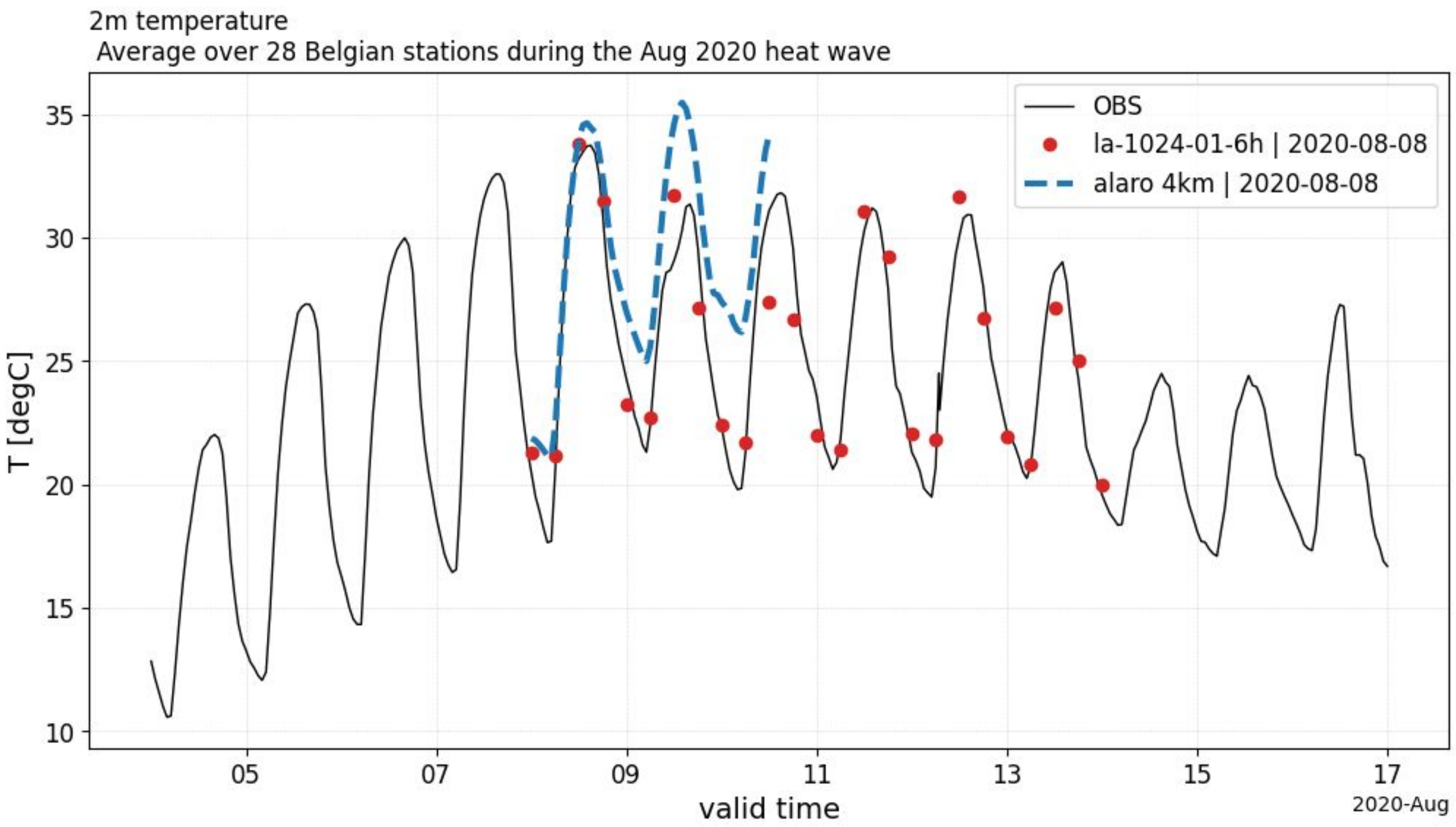
Case studies



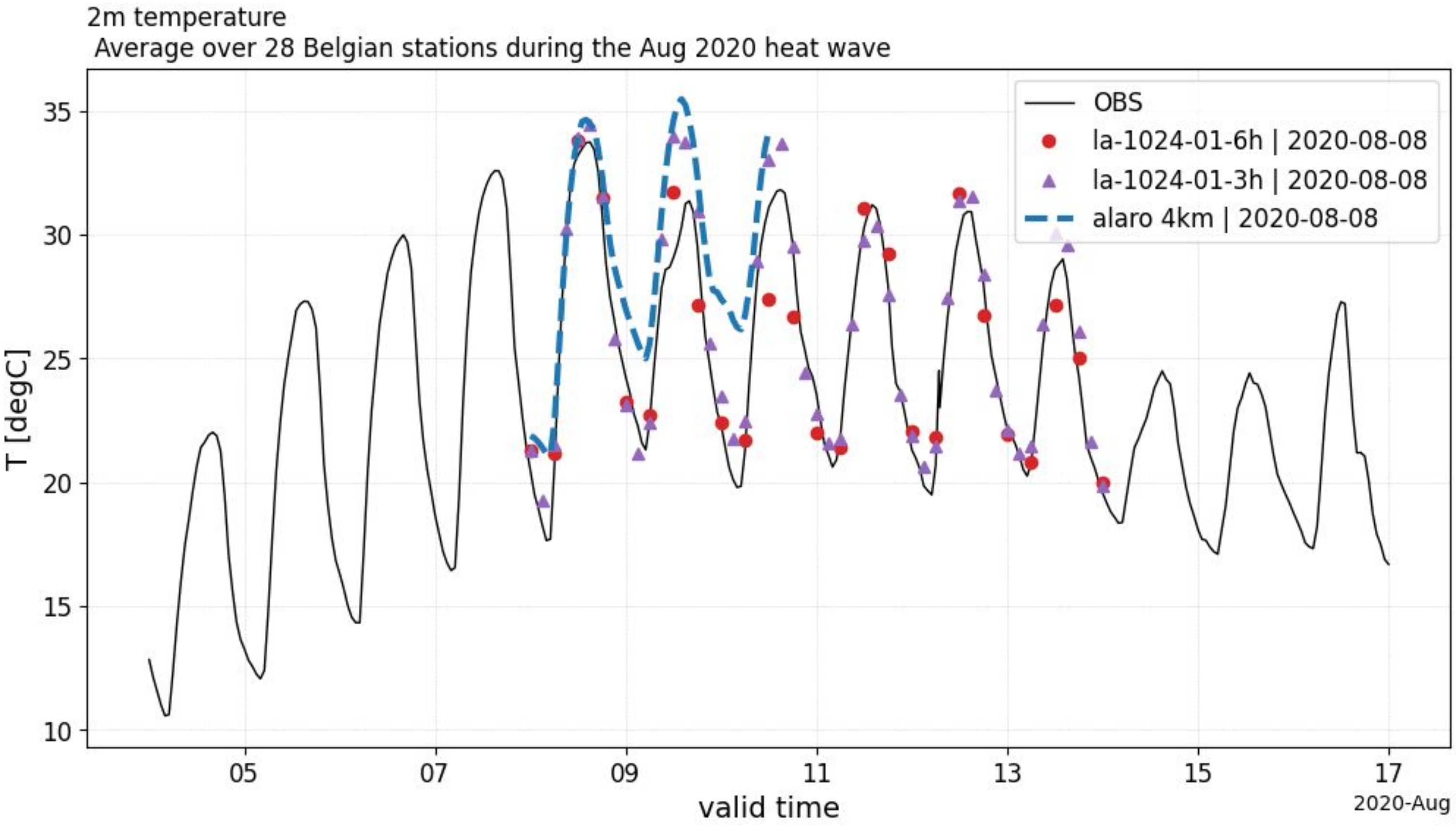
Case studies



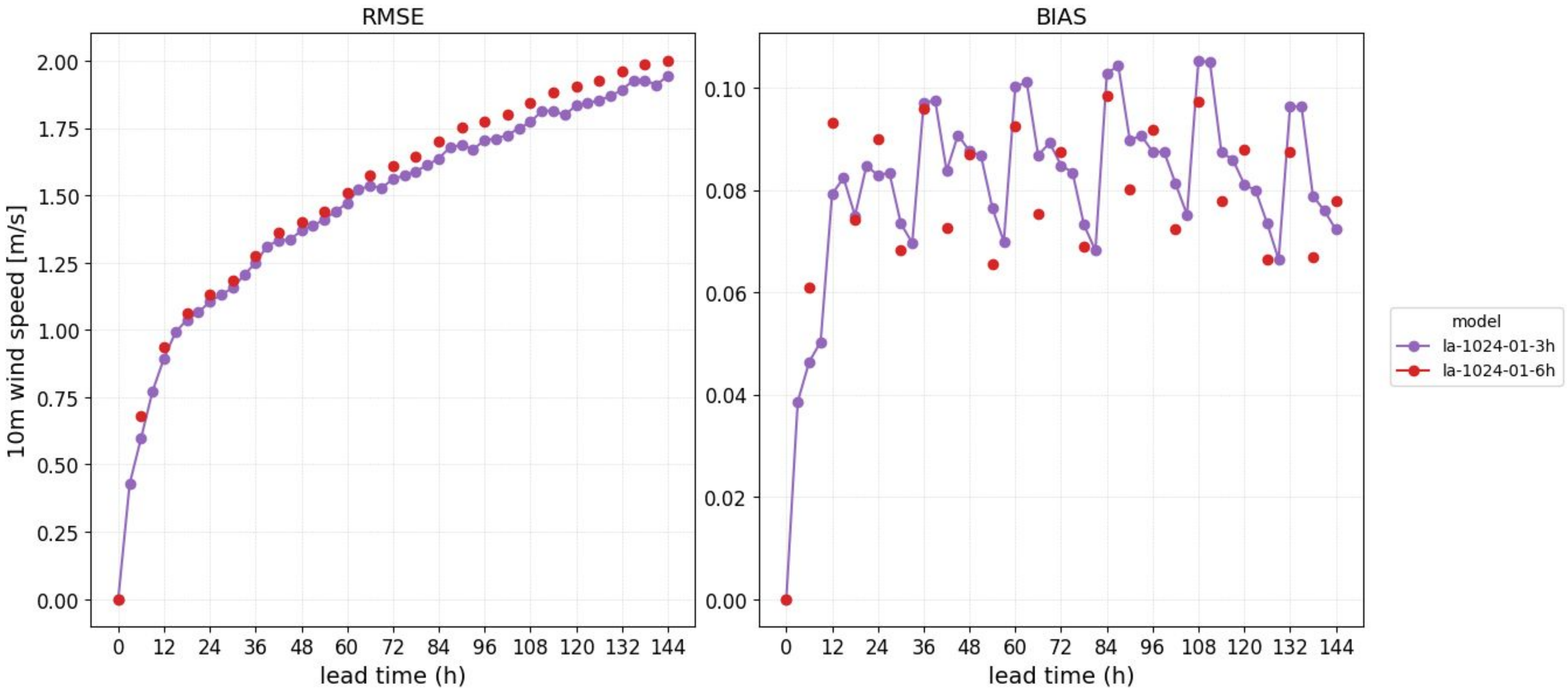
Case studies



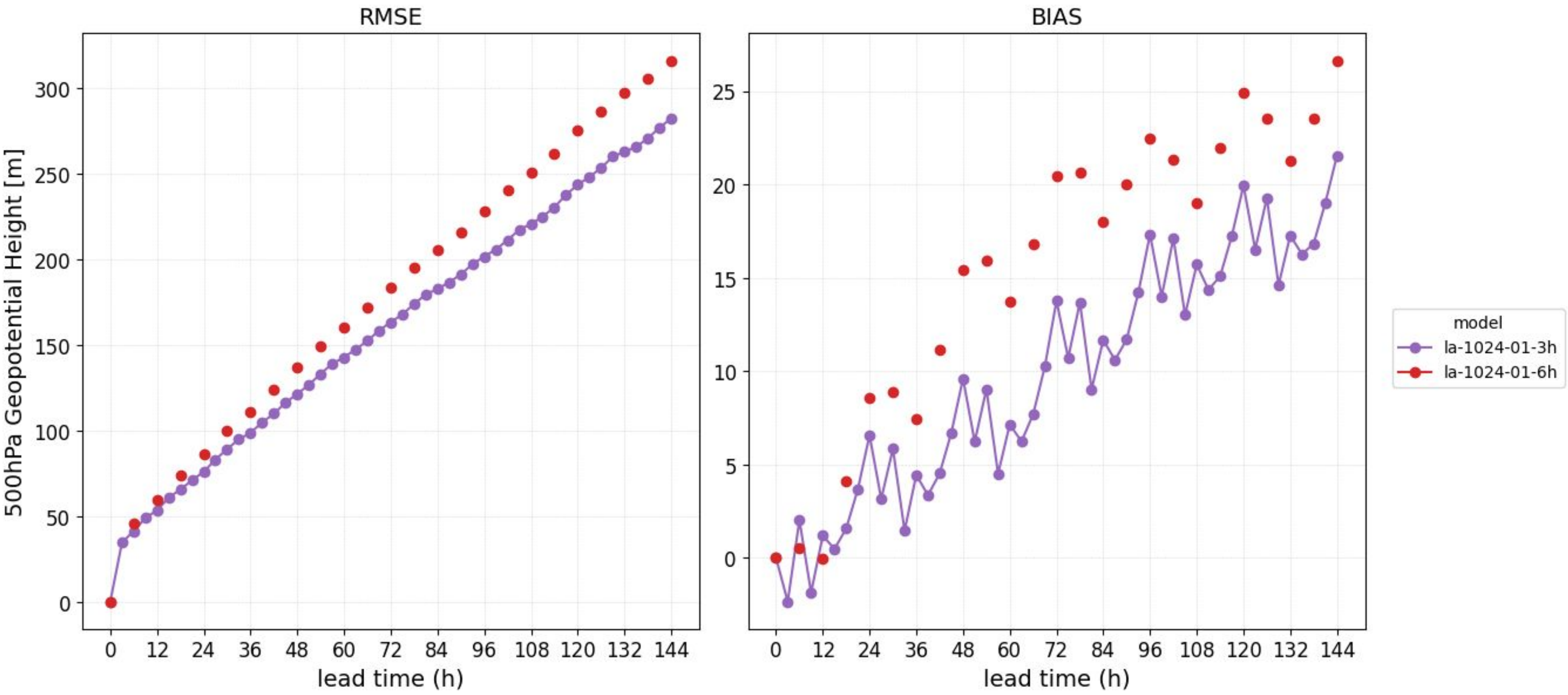
Case studies



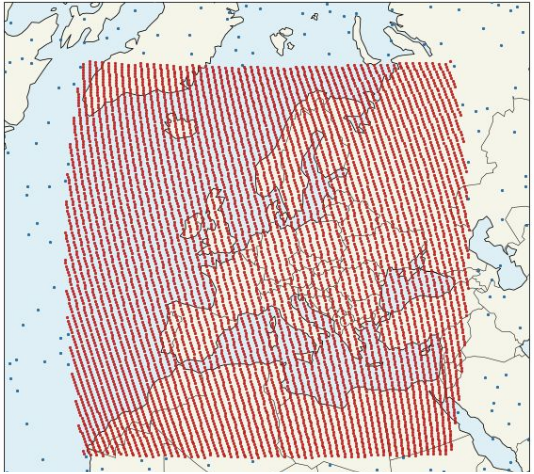
General scores

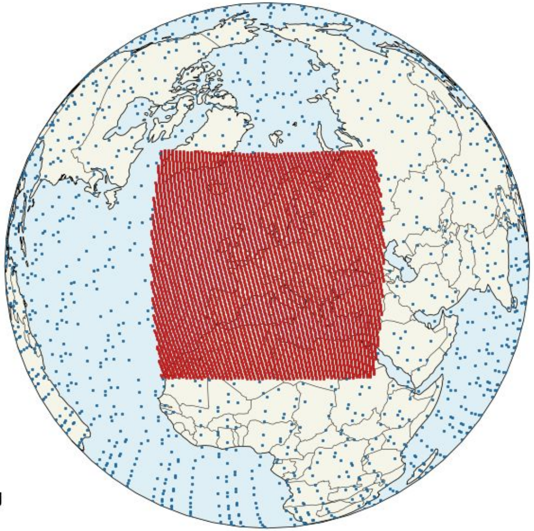


General scores



Comparison study: LAM vs Stretched-Grid





A comparison of stretched-grid and limited-area modelling for data-driven regional weather forecasting

Jasper S. Wijnands, Michiel Van Ginderachter, Bastien François, Sophie Buurman, Piet Termonia, Dieter Van den Bleeken

Regional machine learning weather prediction (MLWP) models based on graph neural networks have recently demonstrated remarkable predictive accuracy, outperforming numerical weather p
area model (LAM) and stretched-grid model (SGM) approaches have emerged for generating high-resolution regional forecasts, based on initial conditions from a regional (re)analysis. While LA
incorporates a global domain at lower resolution. This study aims to understand how the differences in model design impact relative performance and potential applications. Specifically, the stre
generating deterministic regional forecasts over Europe. Using the Anemoi framework, models of both types are built by minimally adapting a shared architecture and trained using global and re
experiments have been conducted to explore their relative performance and highlight key differences. Results show that both LAM and SGM are competitive deterministic MLWP models with g
regional domain. Various differences were identified in the performance of the models across applications. LAM is able to successfully exploit high-quality boundary forcings to make predictions
data is difficult to acquire. SGM is fully self-contained for easier operationalisation, can take advantage of more training data and significantly surpasses LAM in terms of (temporal) generalisabi
to guide their choice between LAM and SGM in developing an operational data-driven forecasting system.

Subjects: **Atmospheric and Oceanic Physics (physics.ao-ph)**; Machine Learning (cs.LG)
Cite as: arXiv:2507.18378 **[physics.ao-ph]**
(or arXiv:2507.18378v1 **[physics.ao-ph]** for this version)
https://doi.org/10.48550/arXiv.2507.18378

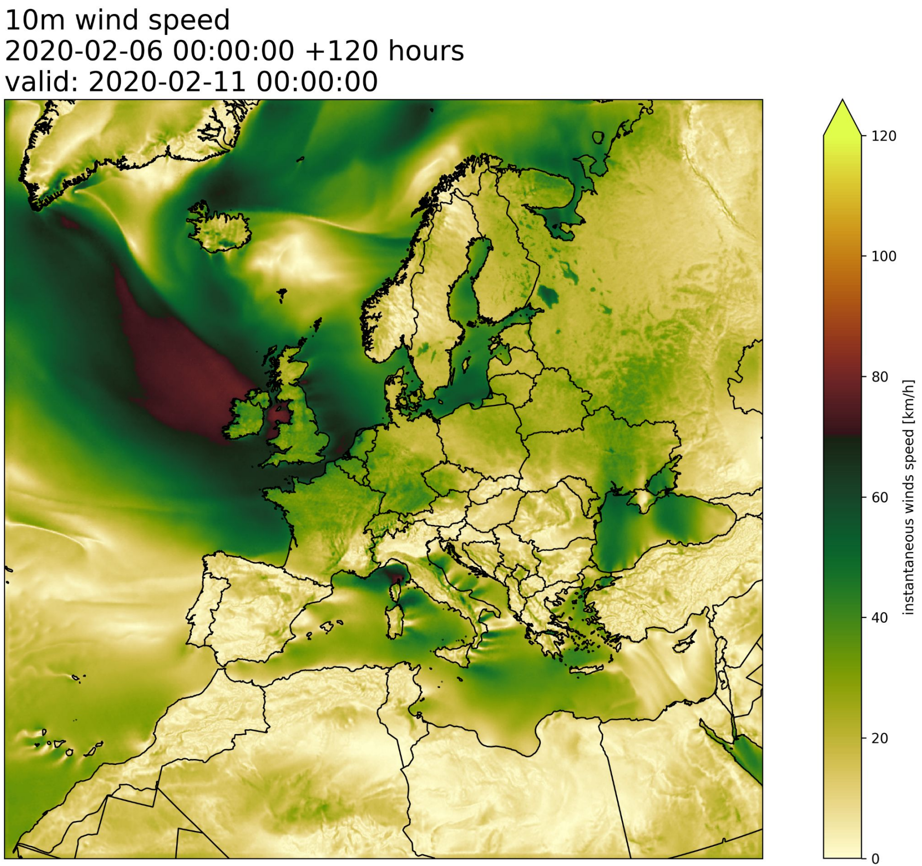
Submission history

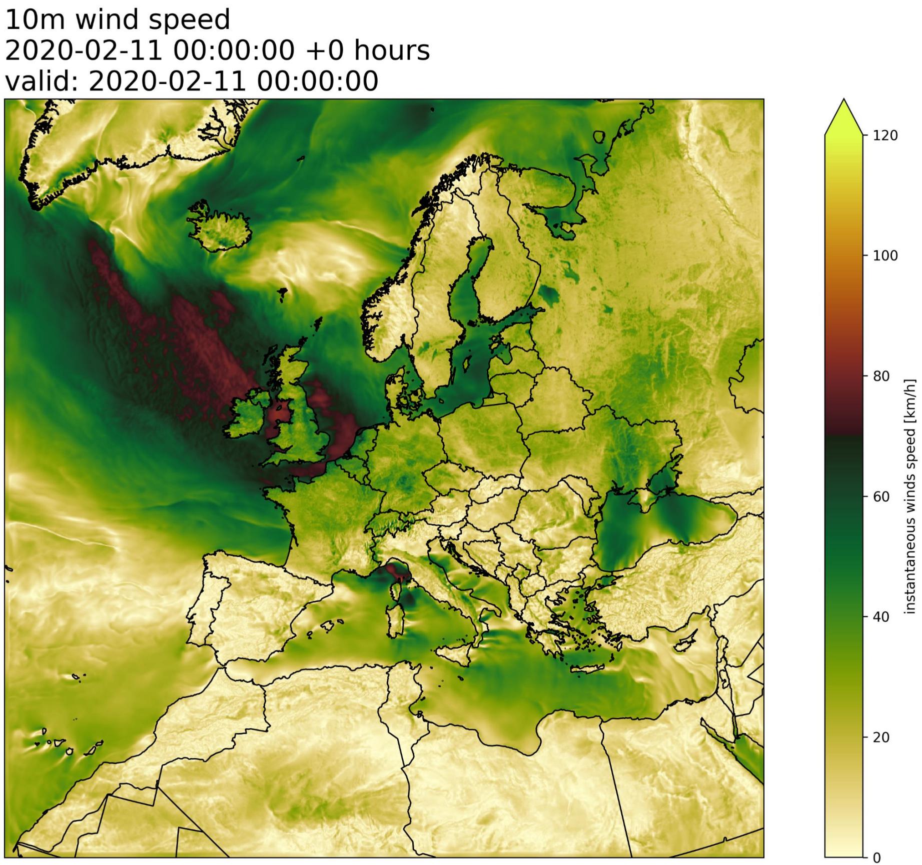
From: Dieter Van Den Bleeken [view email]
[v1] Thu, 24 Jul 2025 12:54:08 UTC (16,227 KB)

https://doi.org/10.48550/arXiv.2507.18378

Some challenges

Smoothing

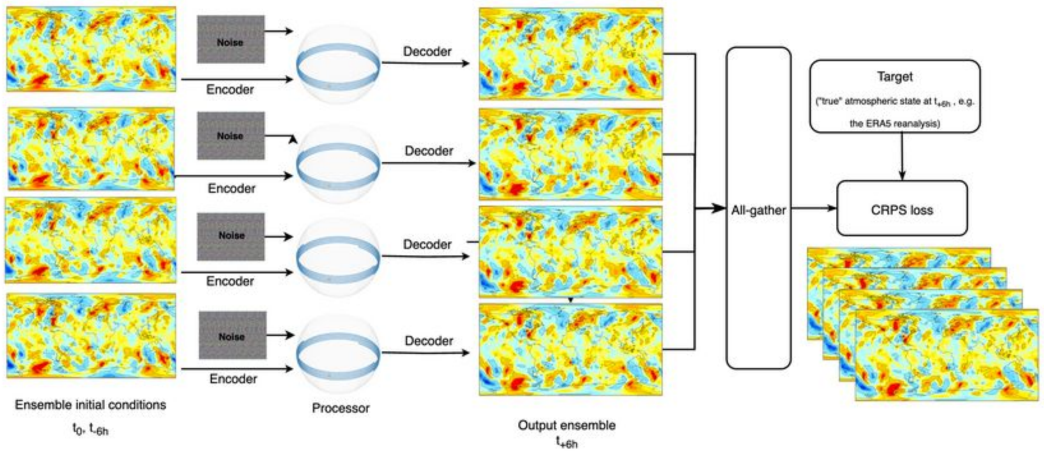

10m wind speed
2020-02-06 00:00:00 +120 hours
valid: 2020-02-11 00:00:00
0
20
40
60
80
100
120
instantaneous winds speed [km/h]

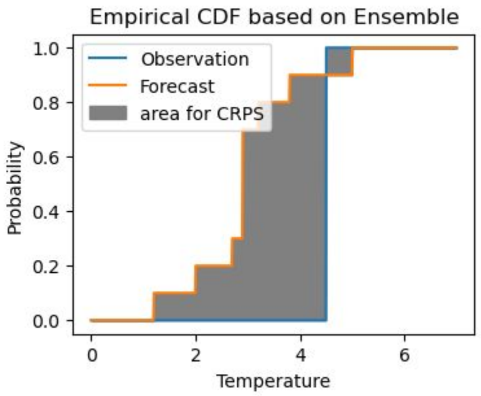

10m wind speed
2020-02-11 00:00:00 +0 hours
valid: 2020-02-11 00:00:00
0
20
40
60
80
100
120
instantaneous winds speed [km/h]

Solution: Going probabilistic

- Inject noise into processor

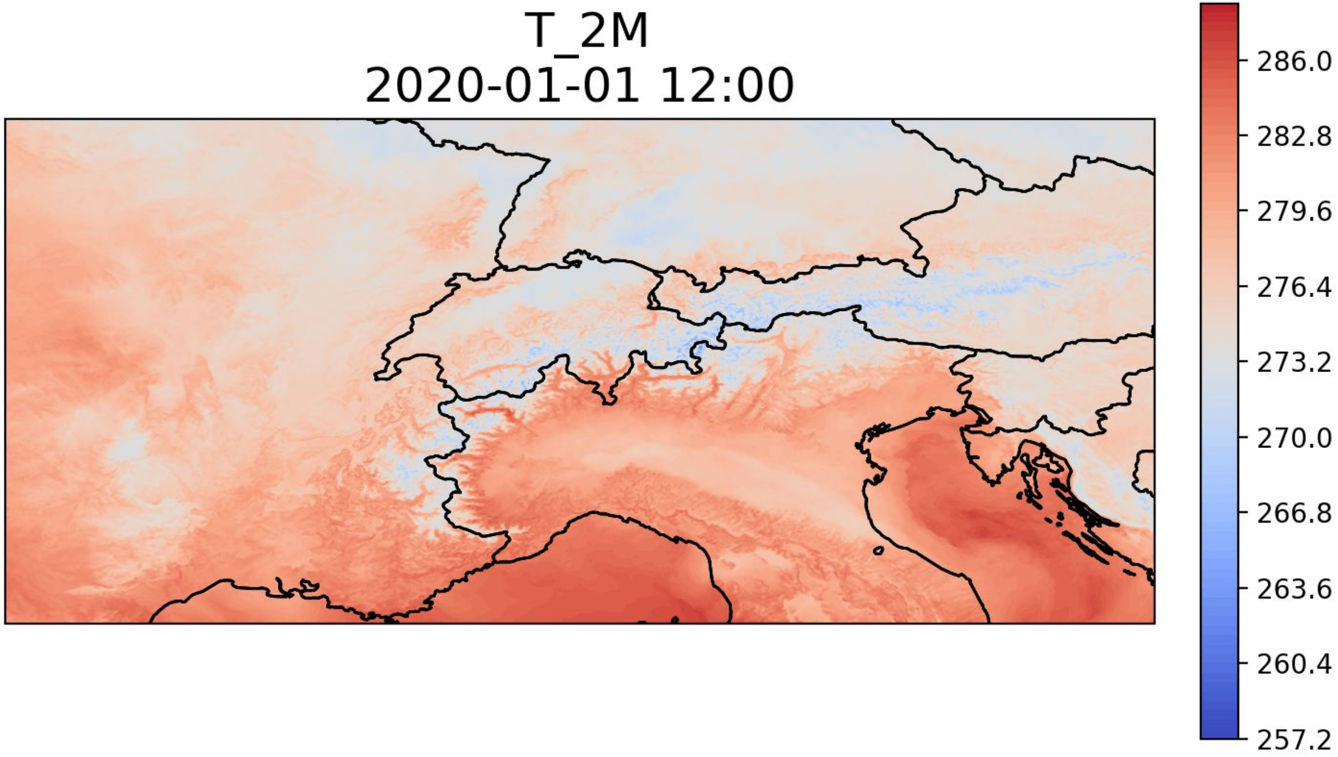
- Probabilistic loss function
    - Continuously Ranked Probability Score





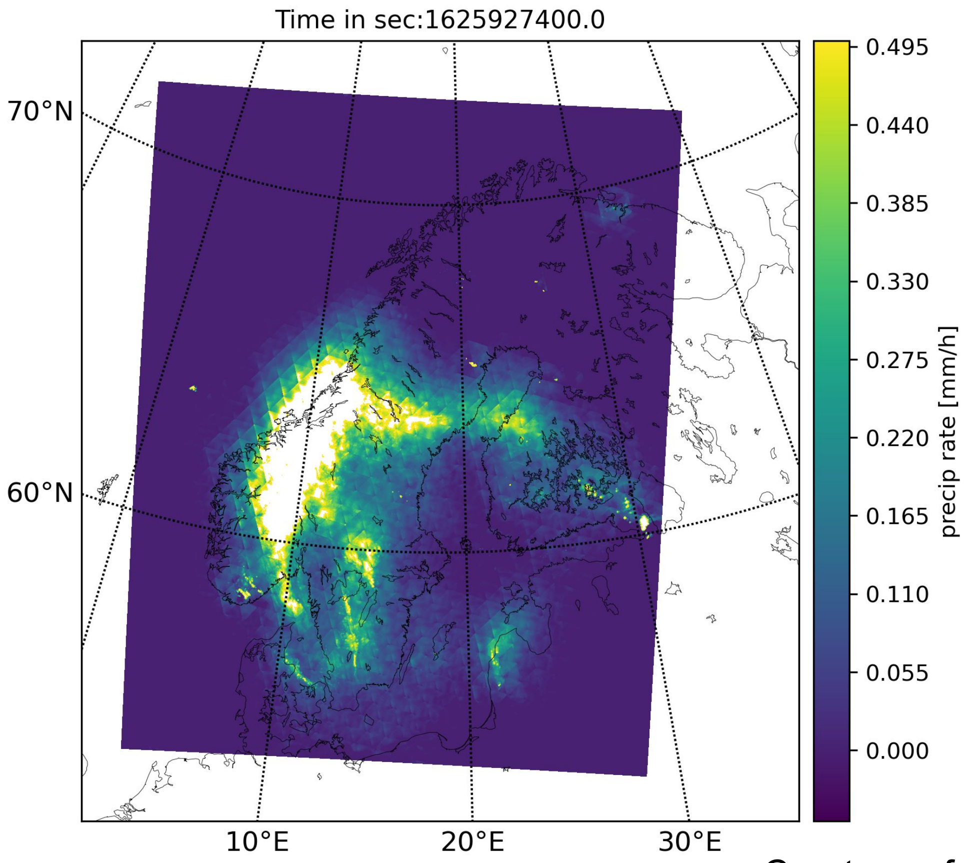
- Ready to start training run on LUMI with custom Triton Kernel (25 - 40% speedup)

Artifacts



Courtesy of MeteoSwiss

Artifacts



Courtesy of MetNo

Solution: Unknown

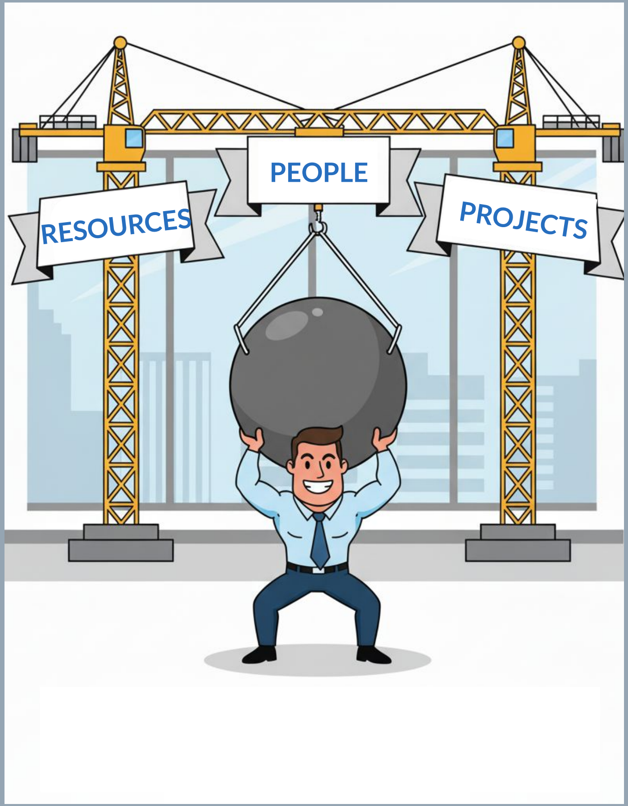
- ~~Overfitting?~~
- Graph + Transformer?
- Decoder?

Closing Remark

RMI is *punching above its AI-weight:*

- International collaboration:
  Anemoi

- Resources:
  LUMI



THANK

The Royal Meteorological Institute

Het Koninklijk Meteorologisch Instituut

L'Institut Royal Météorologique

Das Königliche Meteorologische Institut

RMI

The RMI provides reliable public service realized by empowered staff and based on research, innovation and continuity.

Het KMI verleent een betrouwbare dienstverlening aan het publiek en de overheid gebaseerd op onderzoek, innovatie en continuïteit.

L'IRM fournit un service fiable basé sur la recherche, l'innovation et la continuité au public et aux autorités.

Vertrauenswürdige Dienstleistungen für Öffentlichkeit und Behörden begründet auf Forschung, Innovation und Kontinuität.