Emerging directions in limited-area AI-Driven Weather Forecasting

Leif Denby (lcd@dmi.dk), Weather modelling, Weather Research - with contributions from

the DMI team: Emy Alerskans, Eleni Briola, Simon Christiansen, Martin Frølund, Kasper Hintz, Ole Lindberg, Hauke Schulz, Mathias Schreiner

Joel Oskarsson (Linköping University), Simon Adamov (ETH Zurich/MeteoSwiss) and the rest of the MLLAM community

Danish Meteorological Institute

Mathematics of Data Science Seminar Series, 16/6/2025

Overview of ML activities in Weather Models

- **Neural-LAM:**
  - Graph Neural-Network using Encode-Process-Decode paradigm to emulate atmospheric flow model (traditional Numerical Weather Prediction used for operational forecasts at DMI)
- **LDCast**
  - Latent Diffusion based precipitation nowcasting from using radar observations
- **LeeWaveNet**
  - UNet (with transfer learning for predicting synthesized wavepackets) to detect and characterise lee waves over Greenland to warn air traffic
- **Applications of self-supervised learning**
  - Denoising of LIDAR-based lower-atmosphere water-vapour observations
  - Mesoscale cloud organisation in the tropics

Other activities (not covered today)

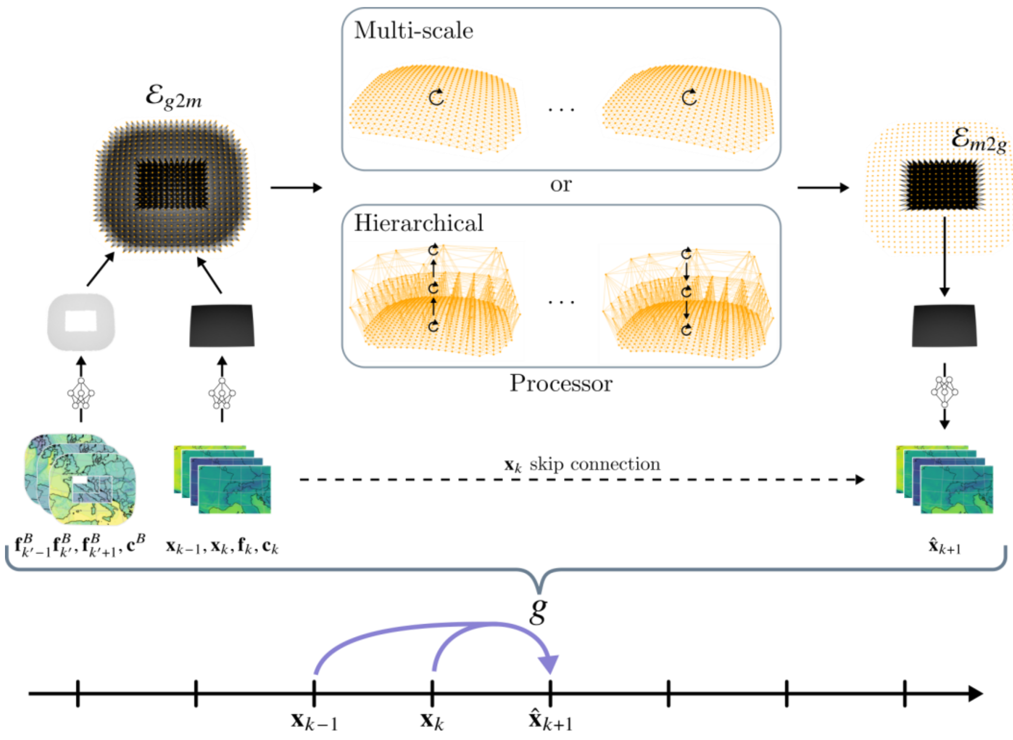
Quality Control of crowdsourced data

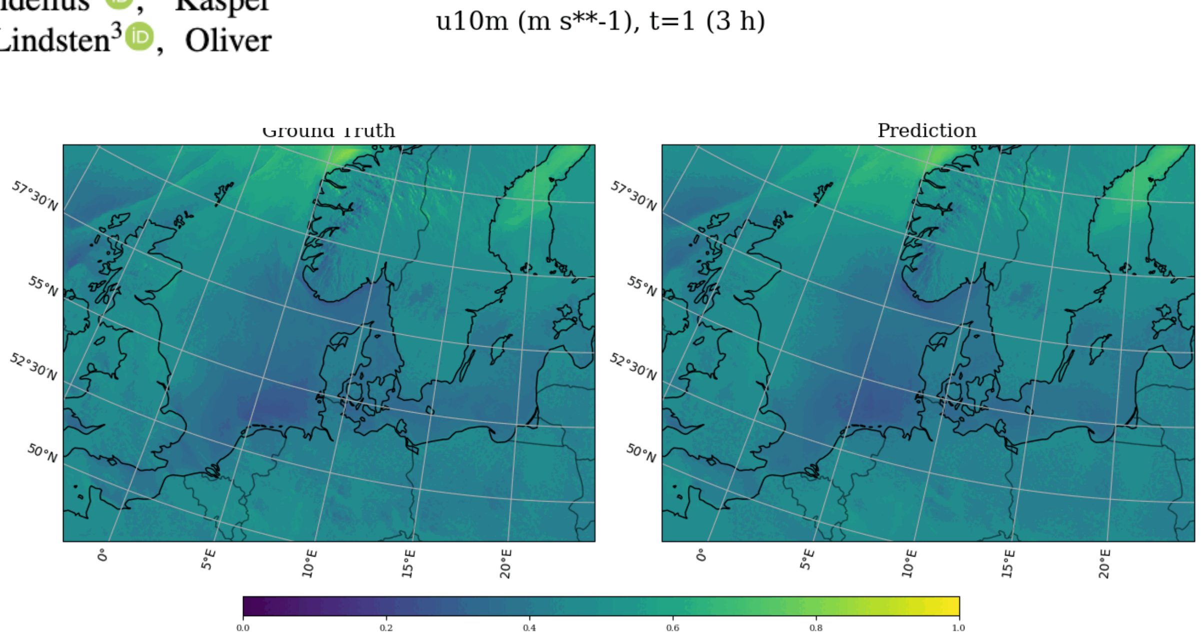
GNN Data-driven Atmospheric Dispersion

SciML Bayesian Differential equations for Road Weather Conditions

Building Neural Limited Area Models: Kilometer-Scale Weather Forecasting in Realistic Settings

Simon Adamov[†,1,2], Joel Oskarsson[†,3*], Leif Denby[4], Tomas Landelius[5], Kasper Hintz[4], Simon Christiansen[4], Irene Schicker[6], Carlos Osuna[1], Fredrik Lindsten[3], Oliver Fuhrer[1] and Sebastian Schemm[2,7]





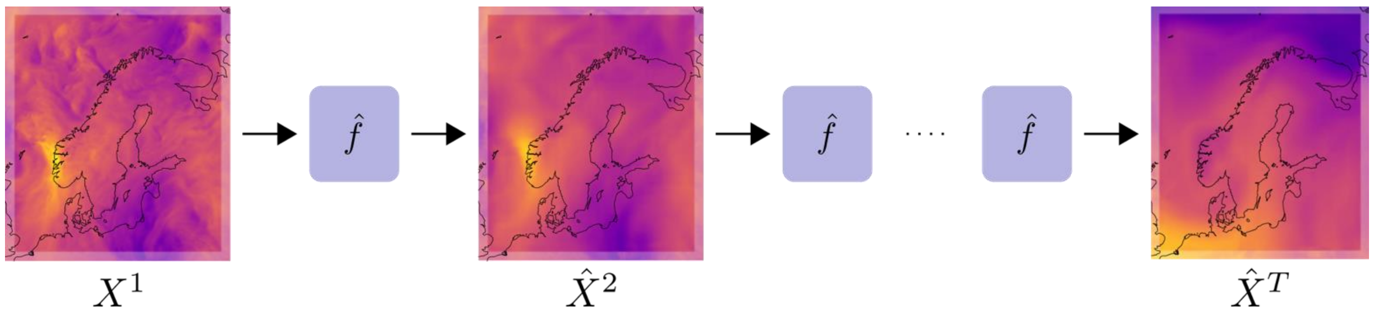
Work lead by PhD students Simon Adamov (MeteoSwiss) and Joel Oskarsson (SMHI) on training and skill of LAM models. Preprint out soon (weeks). Highlights:

- Training takes order 2K GPU hours
- Comparable, and on some metrics better, that operational NWP forecast model (Harmonie-AROME)

preprint on https://arxiv.org/abs/2504.09340

How do these models work?

- Weather state X^t
- Dynamics model $X^t = f(X^{t-1}, \ldots, X^{t-p})$
- Approximate with machine learning model $\hat{f} \approx f$



$$X^1, X^2, \ldots, X^T.$$

- Train on dataset of trajectories
  - Forecast data: Fast surrogate model
  - Reanalysis data: Surpass existing NWP

Danish Meteorological Institute

Slide from Joel Oskarsson

How do this GNN-based forecasting models work?

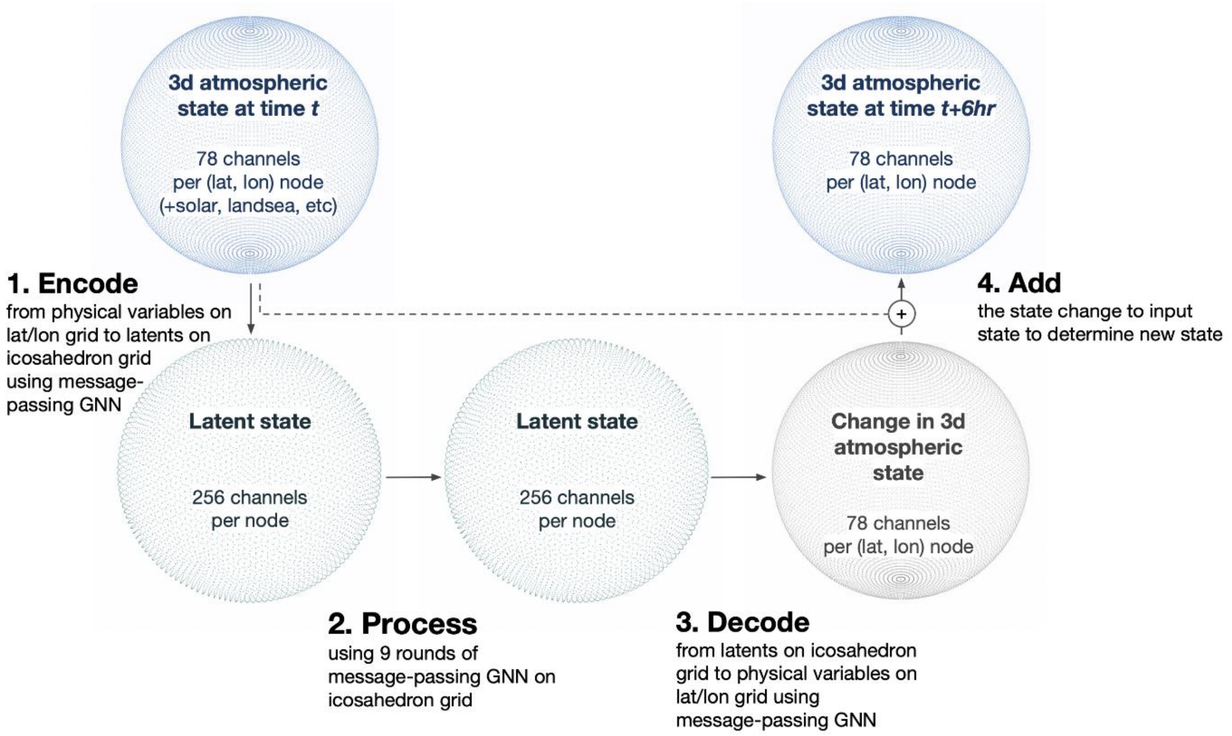


Figure 1: Using the current atmospheric state, the model evolves the state forward by 6 hours. The 3D atmospheric state is defined on a uniform latitude/longitude grid, with 78 channels per pixel (6 physical variables × 13 pressure levels = 78 channels). An Encoder GNN encodes onto latent features defined on a icosahedron grid, a Processor GNN performs additional processing of the latents, and a Decoder GNN maps back to the atmospheric state on a latitude/longitude grid.

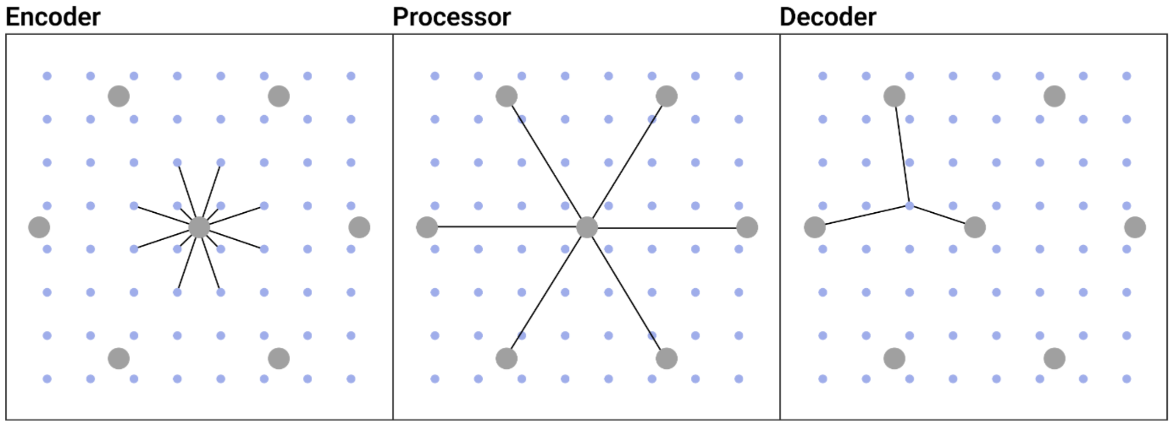
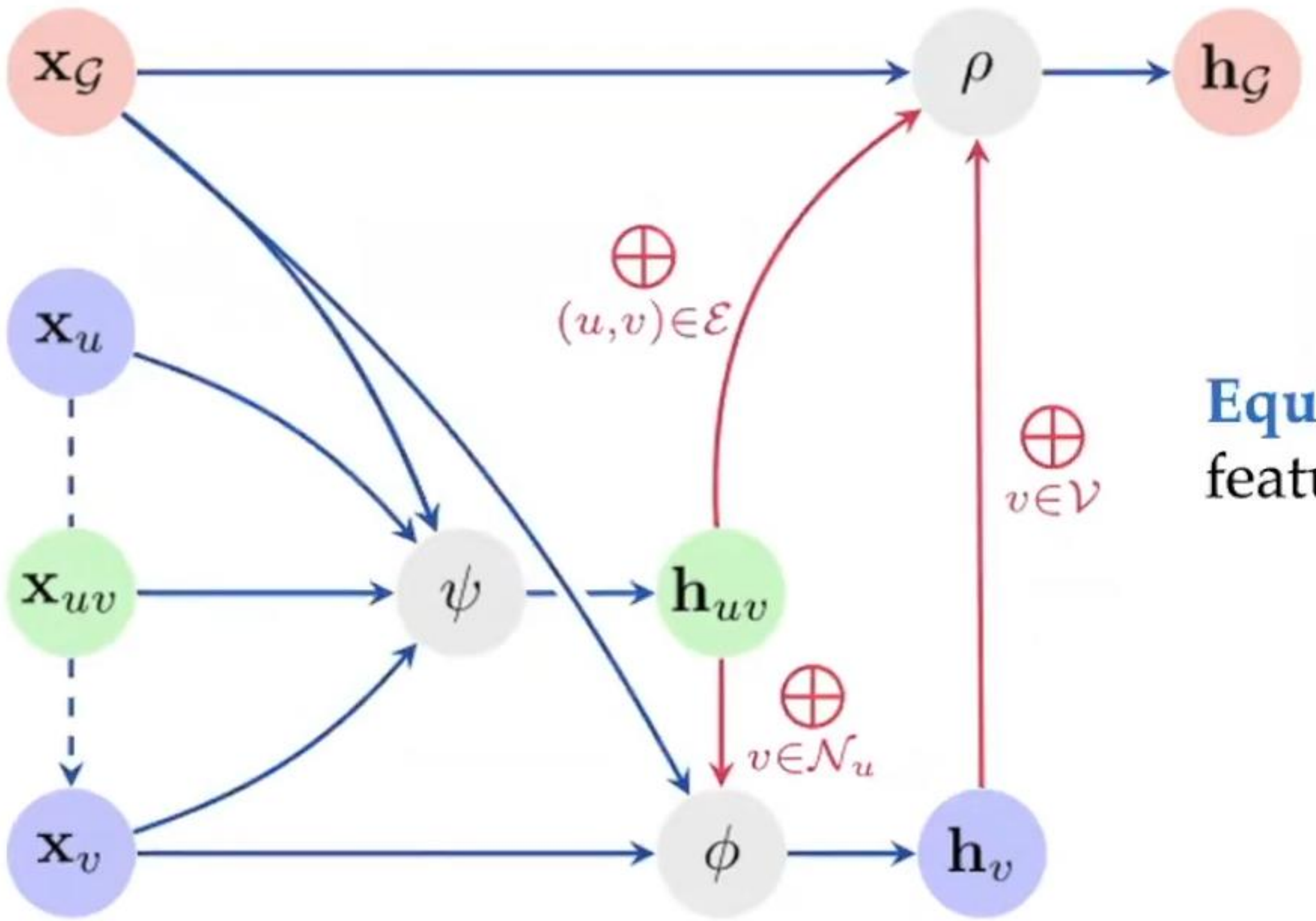


Figure 2: A schematic view of the local graph connectivity in the Encoder, Processor, and Decoder. Left: local spatial and channel information is encoded into an icosahedron node using data from nearby nodes on the input latitude/longitude grid. Center: data on the icosahedron node is further processed using data from nearby icosahedron nodes (including itself, which is not explicitly shown). Right: the output latitude/longitude data is created by decoding data from nearby icosahedron nodes.

Danish Meteorological Institute

Ryan Keisler, 2022

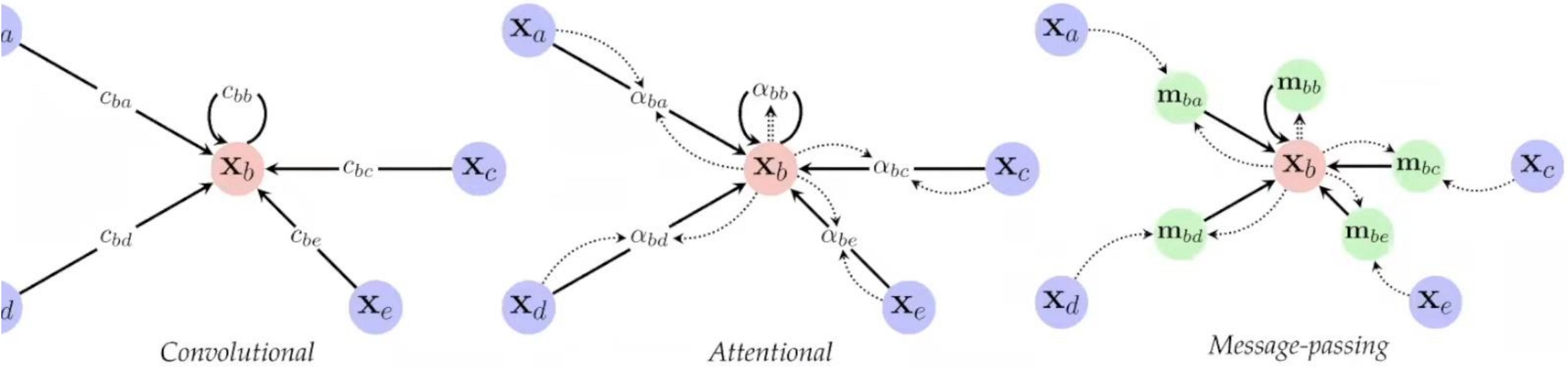
Ok, but what are GNNs (Graph Neural Networks)?



Equivariant and **invariant** layers feature extensively in GNs

Petar Veličković Graph Neural Networks: Geometric, Structural and Algorithmic Perspectives Part 2
Cambridge ELLIS Machine Learning Summer School 2022

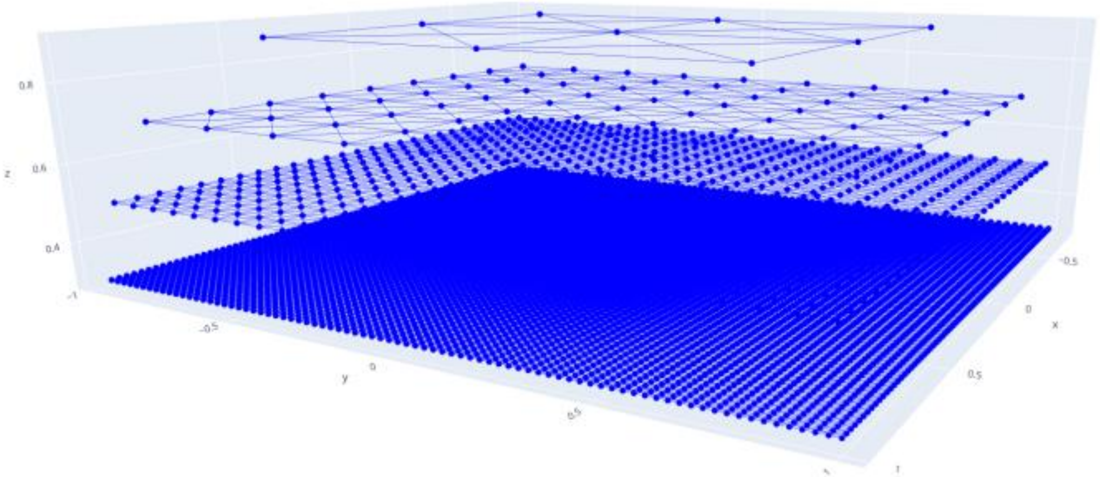
The three "flavours" of GNN layers

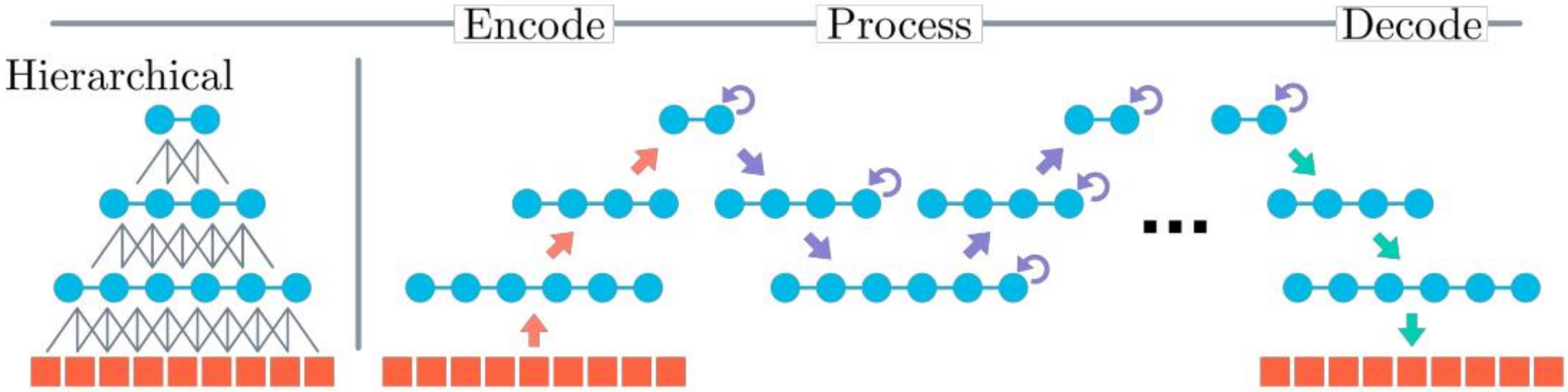


$$\mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j\in\mathcal{N}_i} c_{ij}\psi(\mathbf{x}_j)\right) \qquad \mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j\in\mathcal{N}_i} a(\mathbf{x}_i,\mathbf{x}_j)\psi(\mathbf{x}_j)\right) \qquad \mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j\in\mathcal{N}_i} \psi(\mathbf{x}_i,\mathbf{x}_j)\right)$$

Hi-LAM: Hierarchical multi-scale graph

- 4 levels of nodes in mesh graph
  - Intra-level edges
  - Inter-level edges between adjacent levels
- Sequential GNN message passing up and down the hierarchy

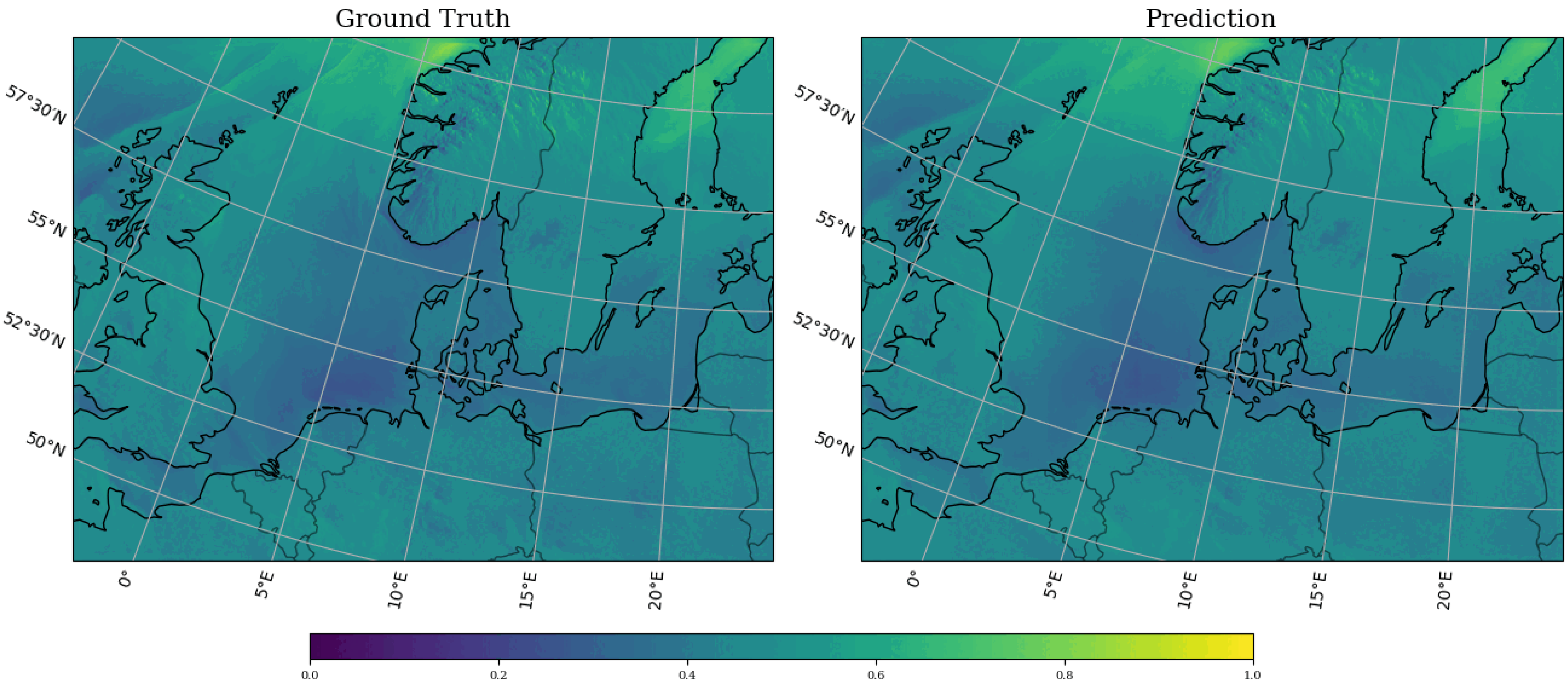




Danish Meteorological Institute

Slide from Joel Oskarsson

u10m (m s**-1), t=1 (3 h)



Danish Meteorological Institute

Skill compared to reanalysis – graph design choices

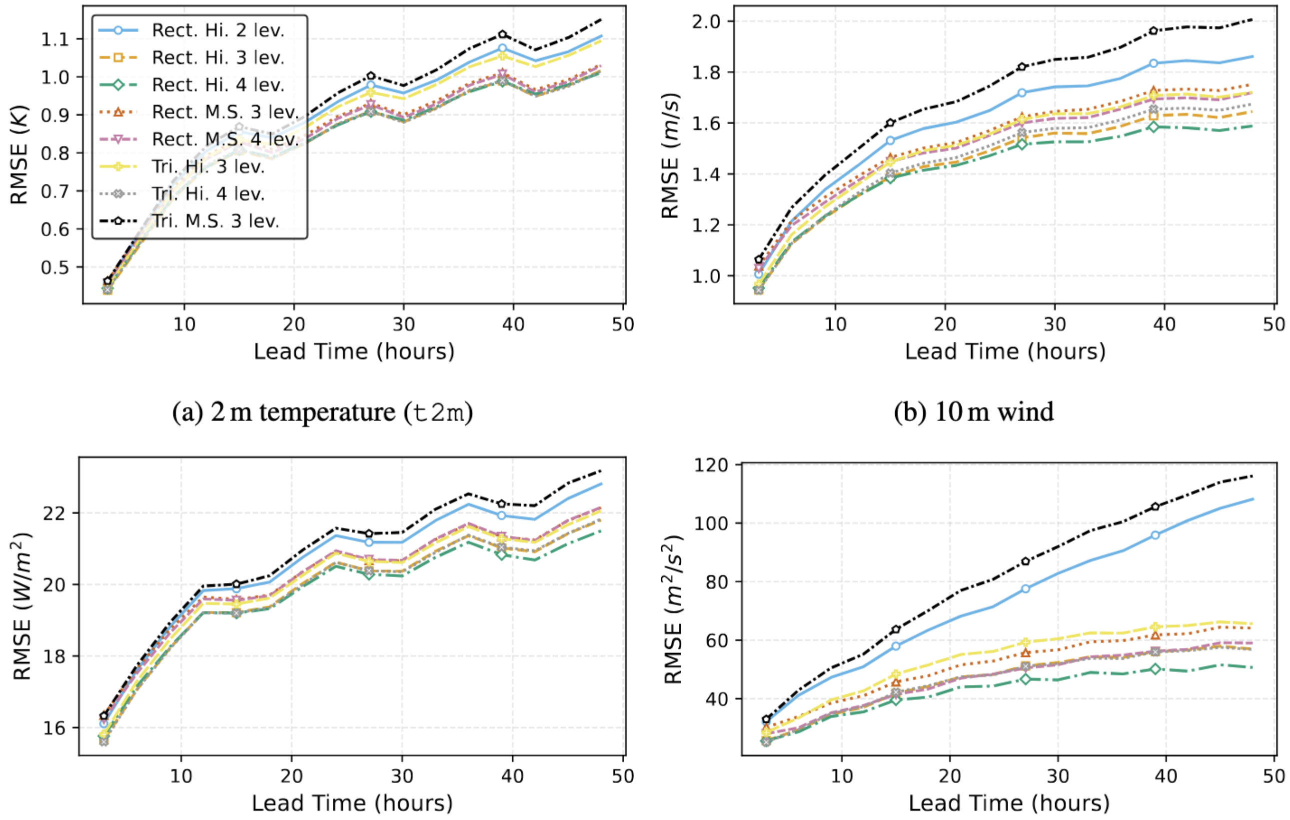
On flat vs hierarchical graph architecture:

- Difference between flat (M.S.) and hierarchical (Hi) in general small, key is to include long-range connections
- 10m wind in particular does show clear improvement with hierarchical vs flat mesh

On rectangular vs triangular mesh:

- triangular meshes show less improvement with higher levels, but maybe due to edge-length growing slower compared to rectangular mesh

Error calculated against reanalysis dataset



(a) 2 m temperature (t2m)

(b) 10 m wind

(c) Surface net longwave radiation (athb_s)

(d) Geopotential at 600 hPa (z600)

Figure 10. *RMSE on validation set for DANRA models trained using different graph configurations. We consider Rectangular (Rect.) and Triangular (Tri.) graphs, both in Hierarchical (Hi.) and Multi-Scale (M.S) setups with different number of levels (lev.). Recall that in multi-scale graphs all the levels are collapsed into one final mesh graph.*

Danish Meteorological Institute

Skill compared to station observations

On forecast skill of ML model vs operational NWP:

- Data-driven model in general better for first 9 hrs, and comparable with NWP model at least out to 18 hrs (we don't have forecast archive for DANRA beyond this)

Error calculated against DMI synop station observations for one year of forecasts with ML model

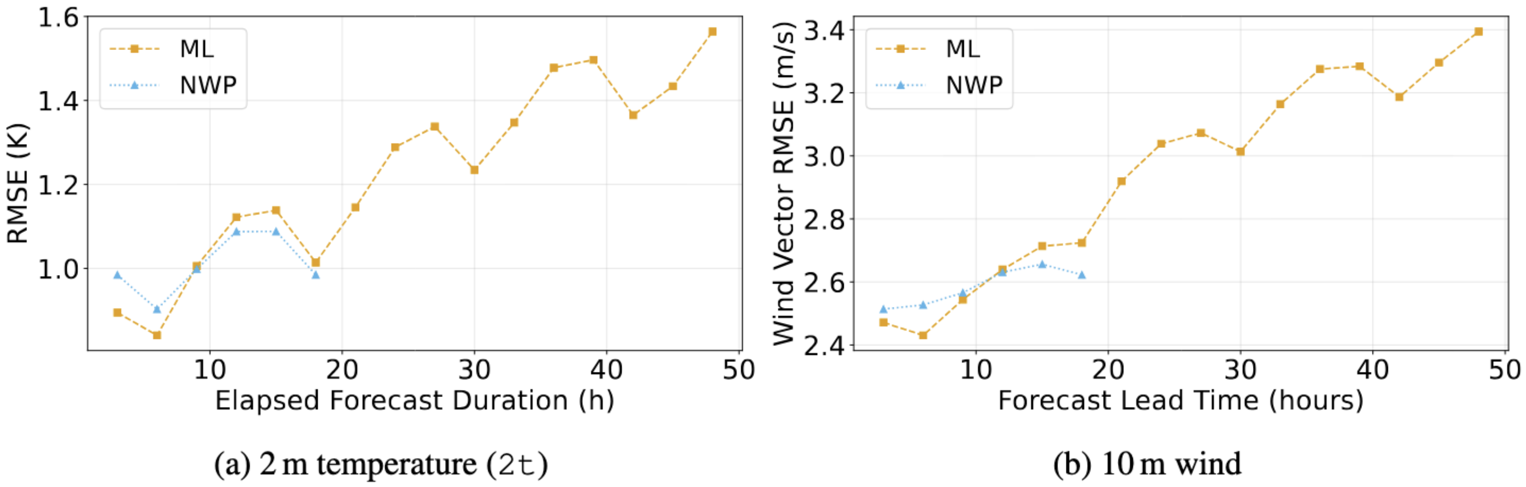
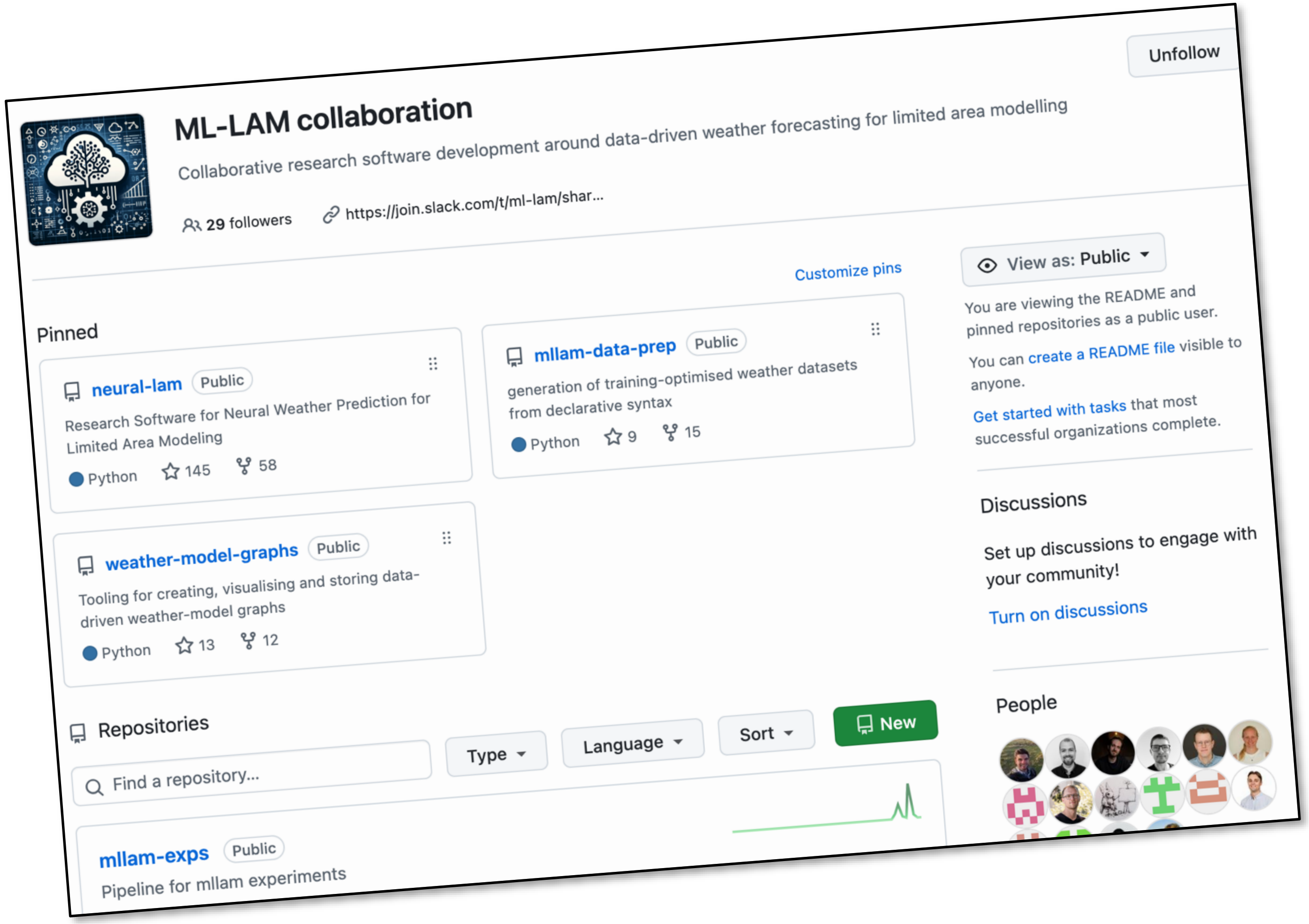


Figure 22. *RMSE along elapsed forecast duration for the DANRA models compared to station observations.*

Danish Meteorological Institute

MLLAM

Collaborative development of data-driven weather forecasting for limited area modelling



ML LAM development plan

People

- LCD: Leif Denby, lcd@dmi.dk, DMI (google: leifdenby@gmail.com)
- TL: Landelius Tomas, Tomas.Landelius@smhi.se, SMHI
- JO: Joel Oskarsson, joel.oskarsson@liu.se, https://joeloskarsson.github.io/, LiU
- TR: Thomas Rieutord, thomas.rieutord@met.ie, Met Éireann
- IrS: Irene Schicker, irene.schicker@geosphere.at, GeoSphere Austria (formerly known as ZAMG)
- FB: Fabrizio Baordo, fab@dmi.dk, DMI
- MVG: Michiel Van Ginderachter, michiel.vanginderachter@meteo.be (RMI-B)
- DY: Daniel Yazgi, daniel.yazgi@smhi.se, SMHI
- LH: Leila Hieta, leila.hieta@fmi.fi, FMI
- SK: Santeri Karppinen, santeri.karppinen@fmi.fi, FMI
- MJC: Mat Chantry matthew.chantry@ecmwf.int ECMWF
- ADS: Simon Adamov, simon.adamov@meteoswiss.ch, ETH/MeteoSwiss
- KAH: Kasper Hintz, kah@dmi.dk, DMI
- VC : Vincent Chabot vincent.chabot@meteo.fr, MF
- CB : Clément Brochet clement.brochet@meteo.fr, MF
- MP: Mikko Partio mikko.partio@fmi.fi FMI
- JLC: Jose-Luis Casado jcasador@aemet.es AEMET
- TG: Teresa García mtgarciag@aemet.es AEMET
- OF: Oliver Fuhrer oliver.fuhrer@meteoswiss.ch MeteoSwiss
- LR : Laure Raynaud laure.raynaud@meteo.fr, MF
- SA : Sara Akodad sara.akodad@meteo.fr, MF
- MP : Matthieu Plu matthieu.plu@meteo.fr, MF
- PAA: Andreas Pauling andreas.pauling@meteoswiss.ch MeteoSwiss
- TW: Tobias Wicky tobias.wicky@meteoswiss.ch MeteoSwiss
- SKC: Simon Kamuk Christiansen skc@dmi.dk, DMI
- FG: Frank Guibert frank.guibert@meteo.fr, MF
- RP: Roland Potthast Roland.Potthast@dwd.de, DWD
- ABR: Basavaraj Rajkumar Amogi basavaraj.amogi@wsu.edu WSU
- TB: Tom Beucler, tom.beucler@unil.ch, UNIL

github organisation: https://github.com/mllam/

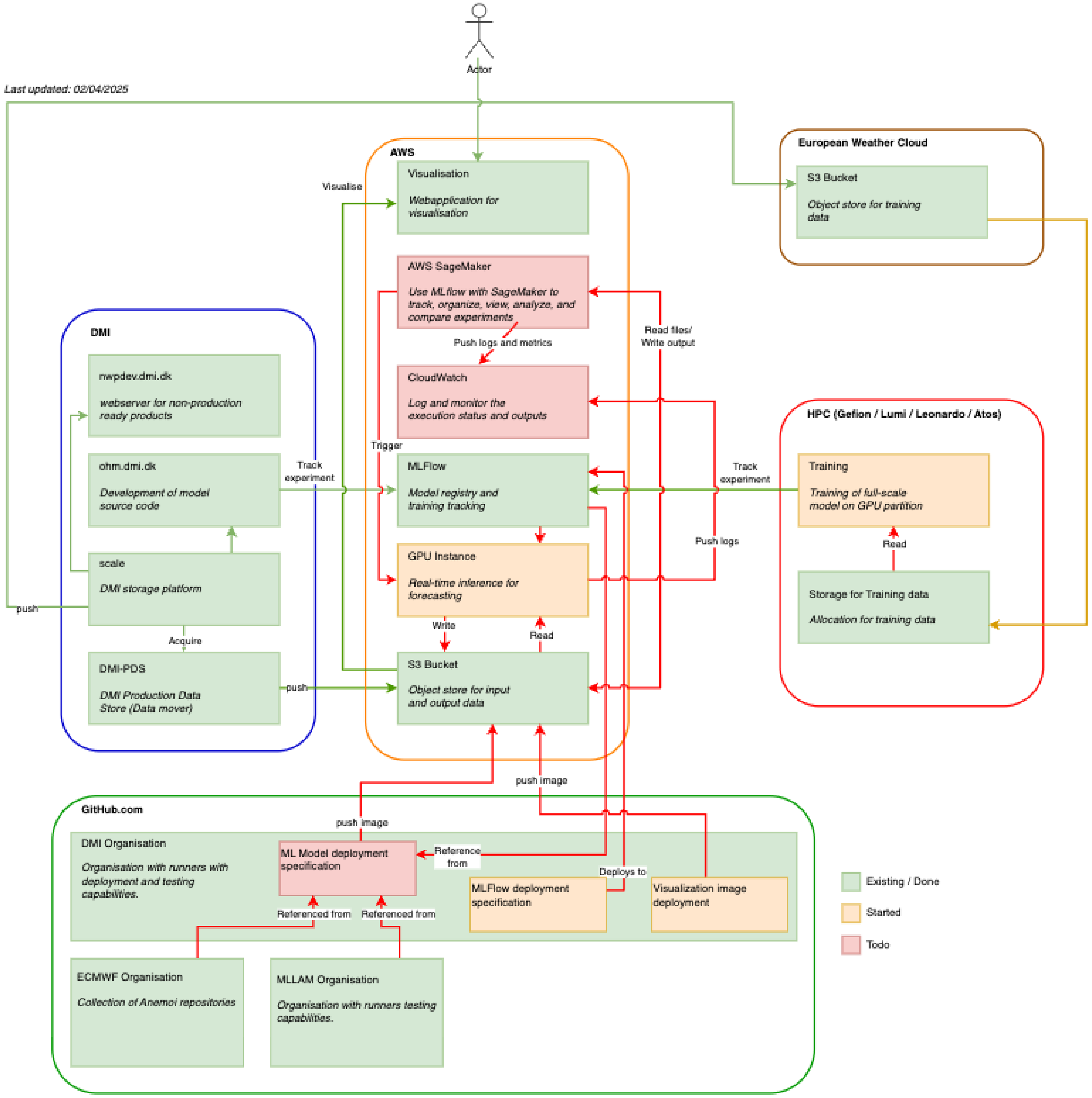
Slack space: https://join.slack.com/t/ml-lam/shared_invite/zt-2t112zvm8-Vt6aBvhX7nYa6Kbj_LkCBQ

development doc: https://bit.ly/mllam-plan

ML Pilot-project infrastructure

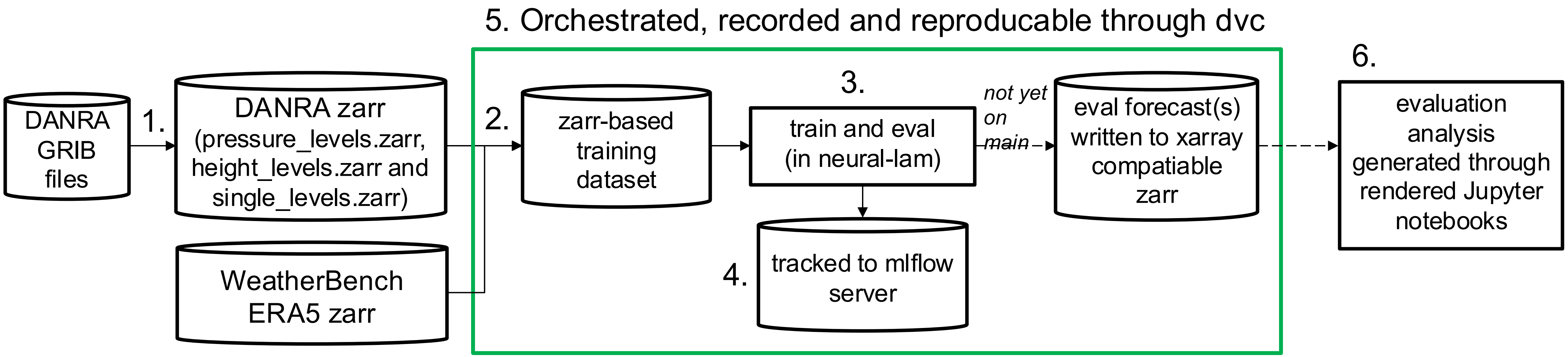
When the ML-pilot started we needed a place to:

- Track our experiments (MLFlow)
- Visualise results (Webserver)
- Store training data for Gefion (+ EuroHPCs)
- Convert input data from GRIB to Zarr.
- Store models (model-registry)
- Run inference



Danish
Meteorological
Institute

ML Development Pipeline

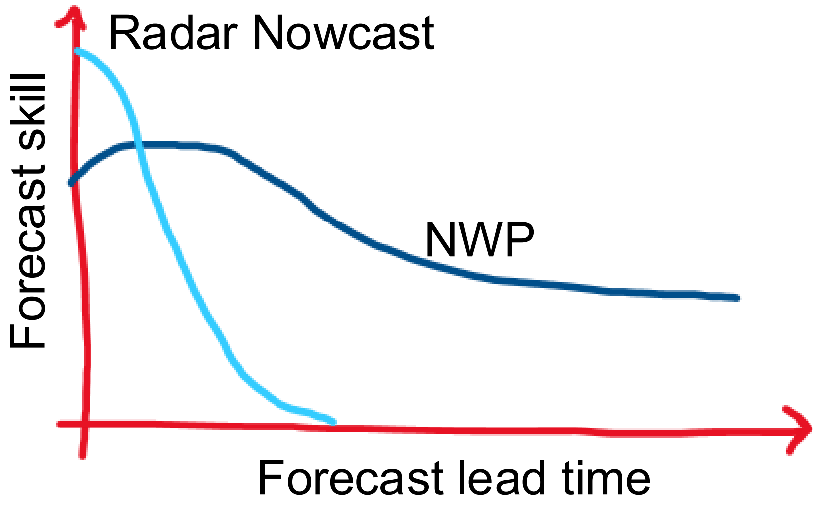


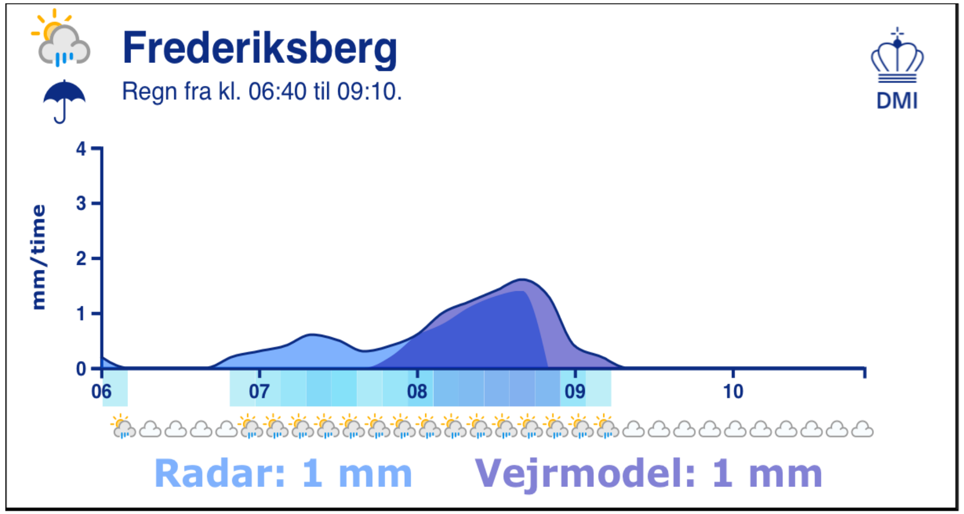
1. https://github.com/leifdenby/dmi-danra-to-zarr/, built on gribscan
2. https://github.com/mllam/mllam-data-prep, built on xarray
3. https://github.com/mllam/neural-lam
4. mlflow server hosted on (air-gapped) Gefion HPC and AWS
5. https://github.com/mllam/mllam-exps built on DVC
6. https://github.com/dmidk/dmi-mllam-verification-notebooks render to S3 bucket

LDCast

ML model for precipitation nowcasting

Precipitation Nowcasting

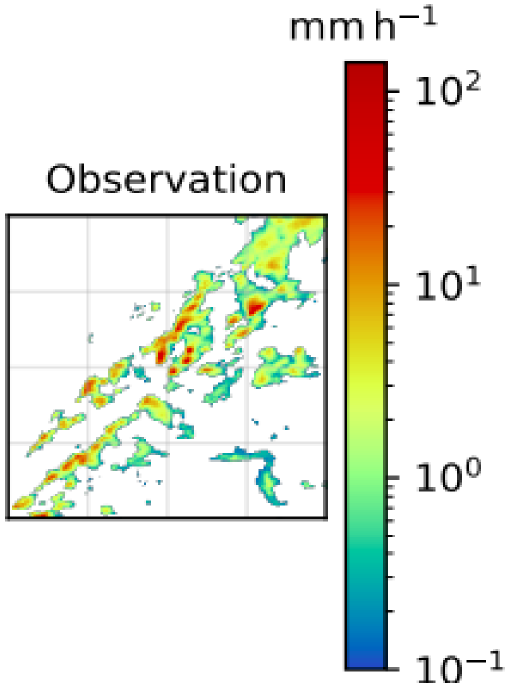


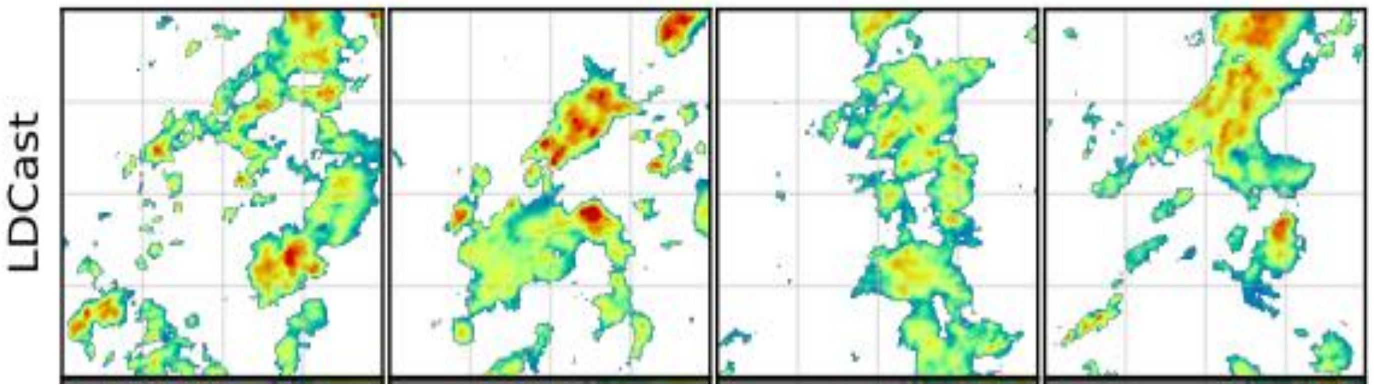


Courtesy **Thomas Bøvith**

Train latent diffusion model to predict precipitation field in 5min (next frame) from last 20 minutes

- data: radar observations - 5min, 2km resolution, ~3 years of historical data)
- diffusion model enables learning of uncertainty in prediction while retaining high spatial fidelity

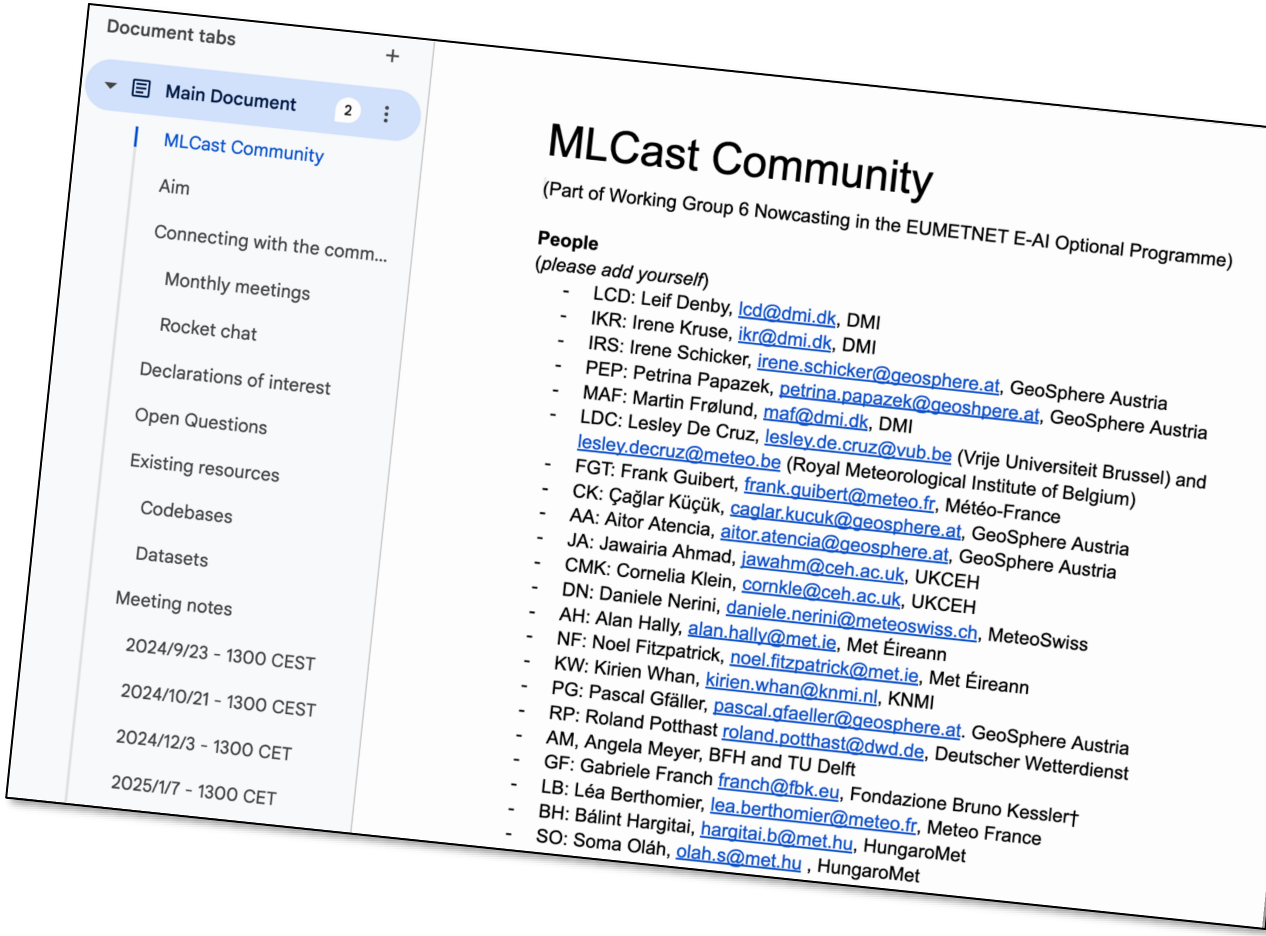




Nowcasting at DMI – joining forces in MLCast Community

Aim at DMI: train Deep Learning raster-based architectures to predict **surface irradiance** and **precipitation** based on Danish observations

- Many architectures published in litterature, no common software framework
- Formed MLCast community (E-AI WG 6) to build forum for building shared datasets, codebase and doing joint application for GPU resources



Document tabs

Main Document 2

MLCast Community

Aim

Connecting with the comm...

Monthly meetings

Rocket chat

Declarations of interest

Open Questions

Existing resources

Codebases

Datasets

Meeting notes

2024/9/23 - 1300 CEST

2024/10/21 - 1300 CEST

2024/12/3 - 1300 CET

2025/1/7 - 1300 CET

MLCast Community

(Part of Working Group 6 Nowcasting in the EUMETNET E-AI Optional Programme)

People

(please add yourself)

- LCD: Leif Denby, lcd@dmi.dk, DMI
- IKR: Irene Kruse, ikr@dmi.dk, DMI
- IRS: Irene Schicker, irene.schicker@geosphere.at, GeoSphere Austria
- PEP: Petrina Papazek, petrina.papazek@geoshpere.at, GeoSphere Austria
- MAF: Martin Frølund, maf@dmi.dk, DMI
- LDC: Lesley De Cruz, lesley.de.cruz@vub.be (Vrije Universiteit Brussel) and lesley.decruz@meteo.be (Royal Meteorological Institute of Belgium)
- FGT: Frank Guibert, frank.guibert@meteo.fr, Météo-France
- CK: Çağlar Küçük, caglar.kucuk@geosphere.at, GeoSphere Austria
- AA: Aitor Atencia, aitor.atencia@geosphere.at, GeoSphere Austria
- JA: Jawairia Ahmad, jawahm@ceh.ac.uk, UKCEH
- CMK: Cornelia Klein, cornkle@ceh.ac.uk, UKCEH
- DN: Daniele Nerini, daniele.nerini@meteoswiss.ch, MeteoSwiss
- AH: Alan Hally, alan.hally@met.ie, Met Éireann
- NF: Noel Fitzpatrick, noel.fitzpatrick@met.ie, Met Éireann
- KW: Kirien Whan, kirien.whan@knmi.nl, KNMI
- PG: Pascal Gfäller, pascal.gfaeller@geosphere.at, GeoSphere Austria
- RP: Roland Potthast roland.potthast@dwd.de, Deutscher Wetterdienst
- AM, Angela Meyer, BFH and TU Delft
- GF: Gabriele Franch franch@fbk.eu, Fondazione Bruno Kessler†
- LB: Léa Berthomier, lea.berthomier@meteo.fr, Meteo France
- BH: Bálint Hargitai, hargitai.b@met.hu, HungaroMet
- SO: Soma Oláh, olah.s@met.hu , HungaroMet

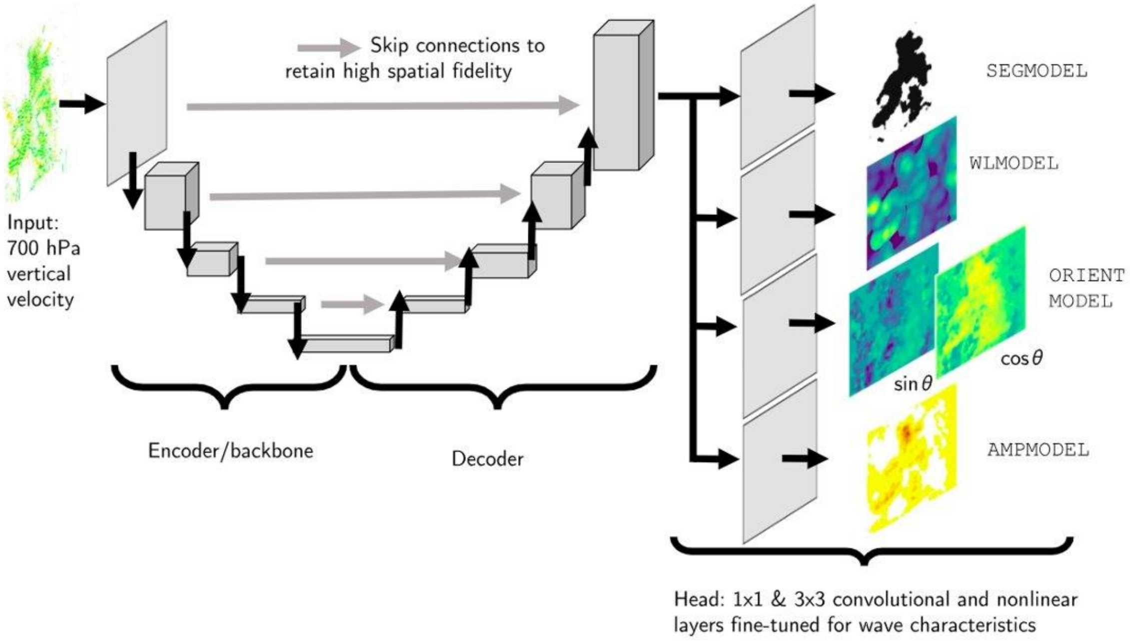
Danish Meteorological Institute

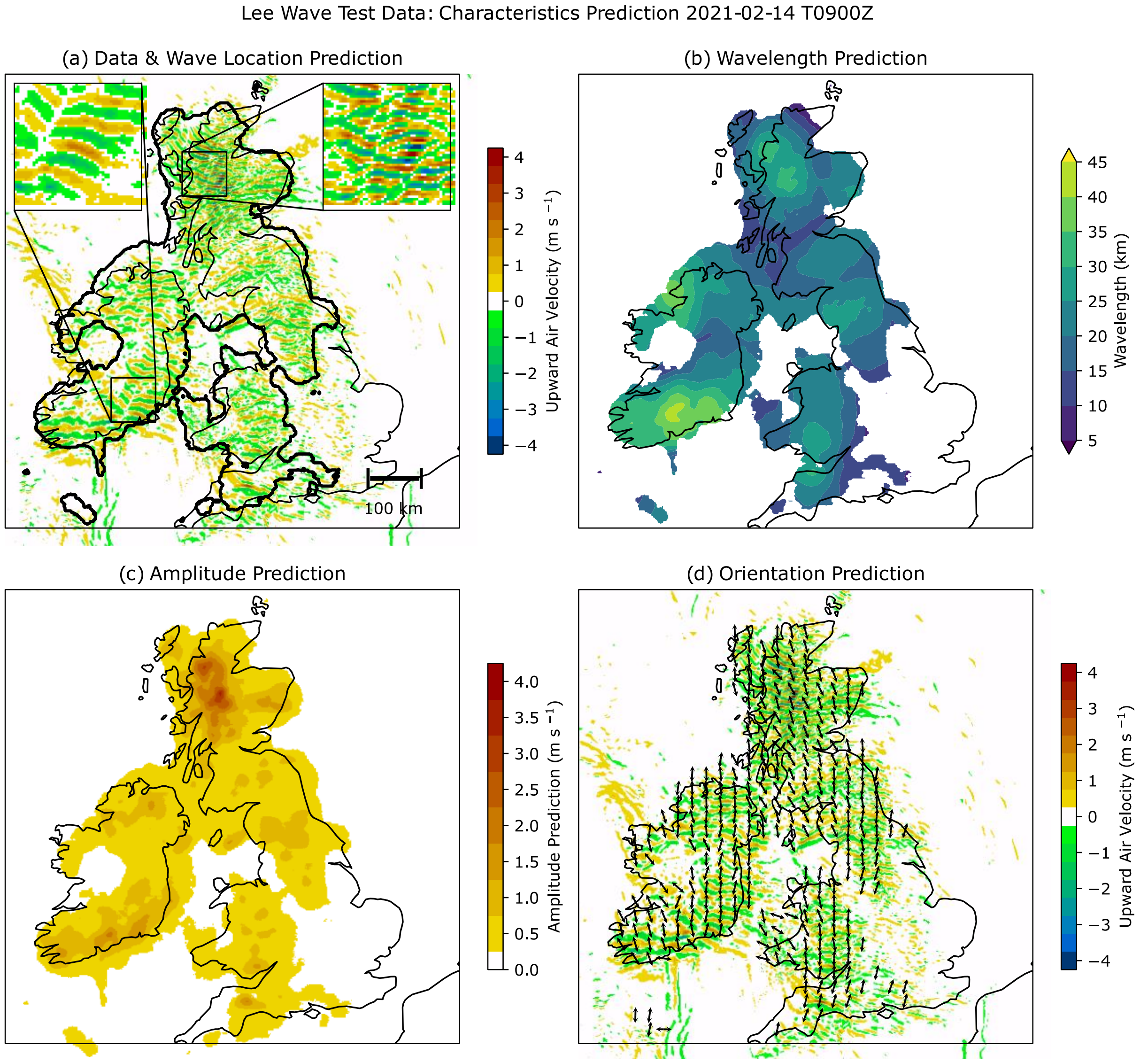
see https://bit.ly/mlcast for more details

LeeWaveNet

Detect trapped lee waves to warn aviation authorities

LeeWaveNet – detection and characterisation of atmospheric gravity waves





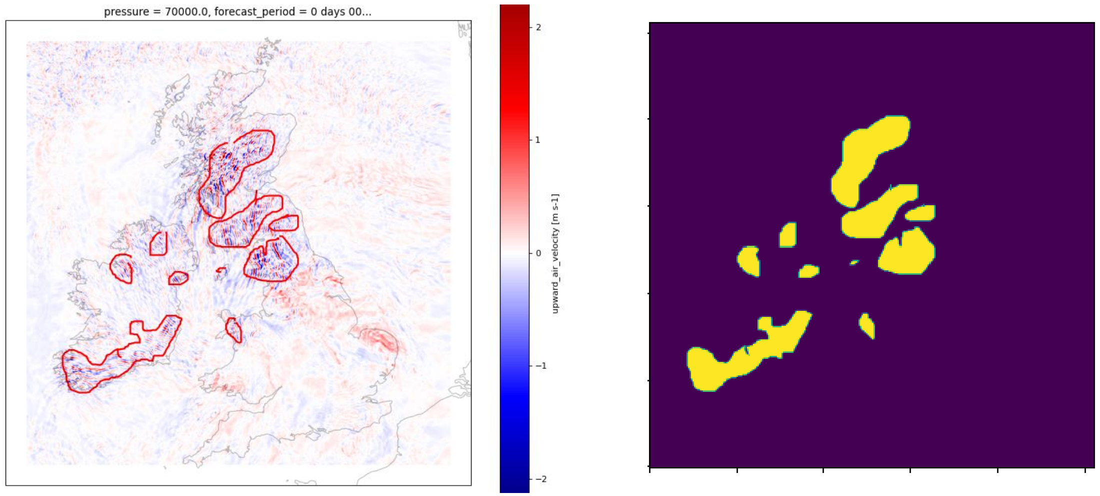
UNet-based architecture trained to detect (segment) and characterise (scalar values measuring characteristics) of trapped lee waves

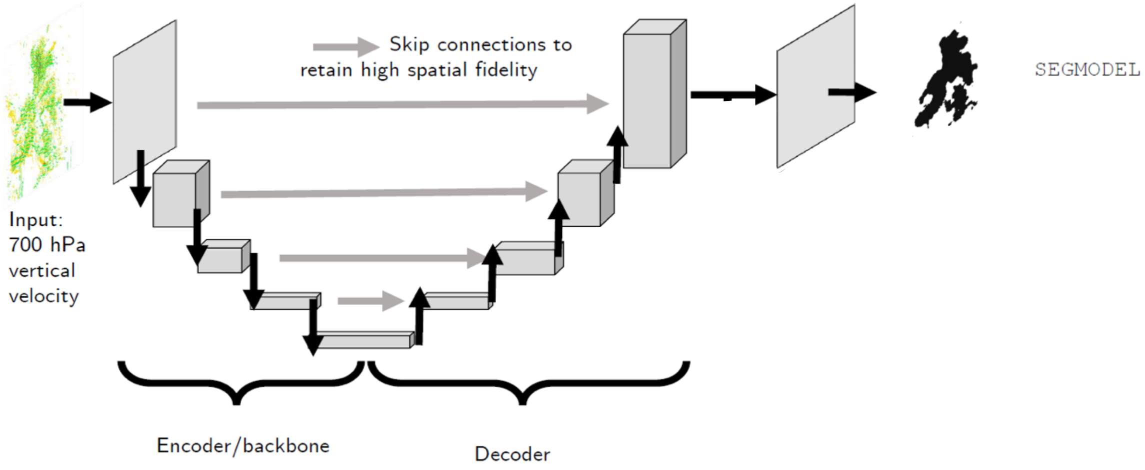
Danish Meteorological Institute

PhD project of Jonathen Coney (Univ Leeds), paper: https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/qj.4592

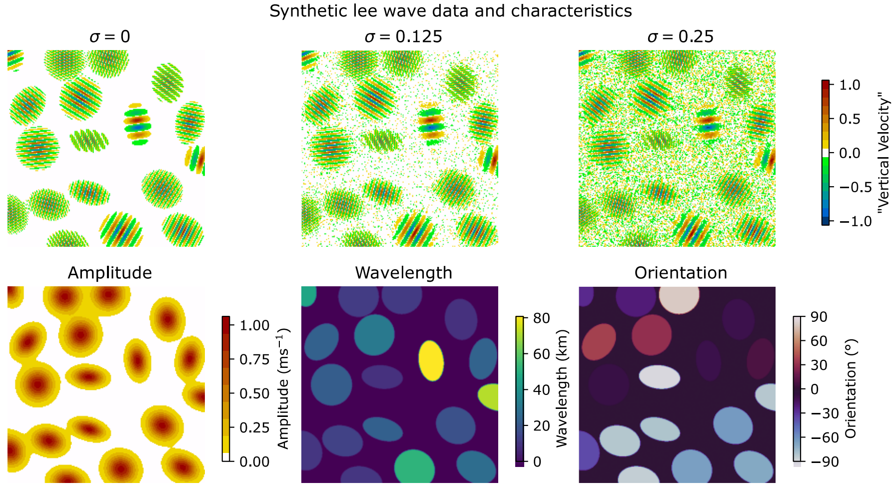
How was LeeWaveNet trained?

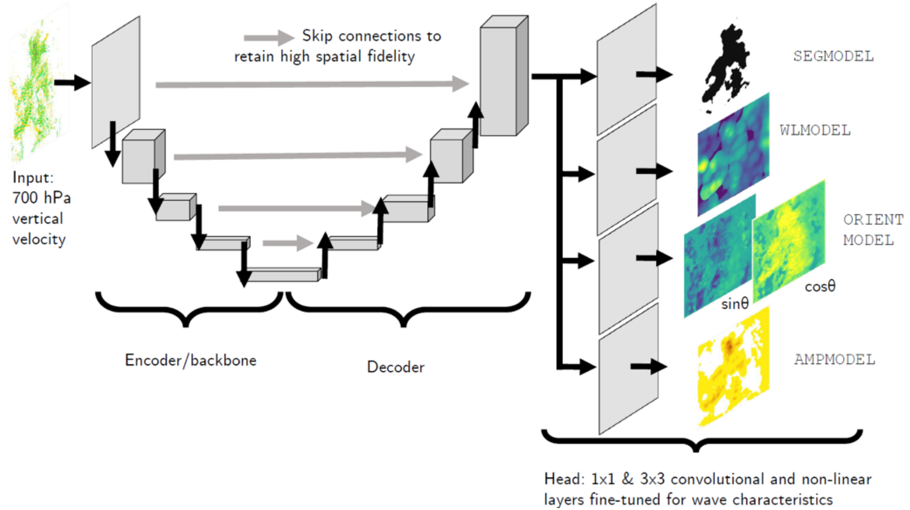
First, trained UNet to predict hand-drawn segmentation mask

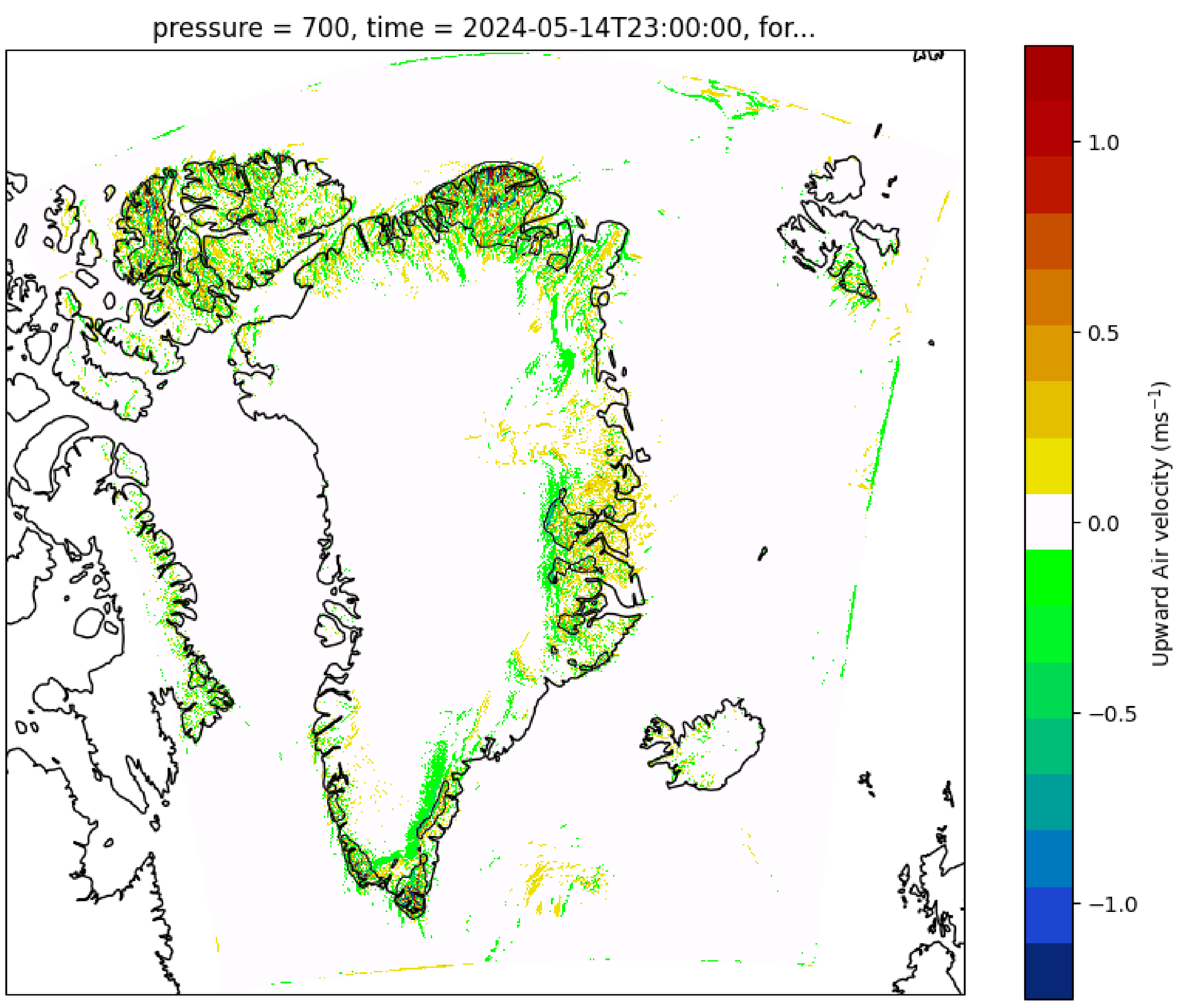


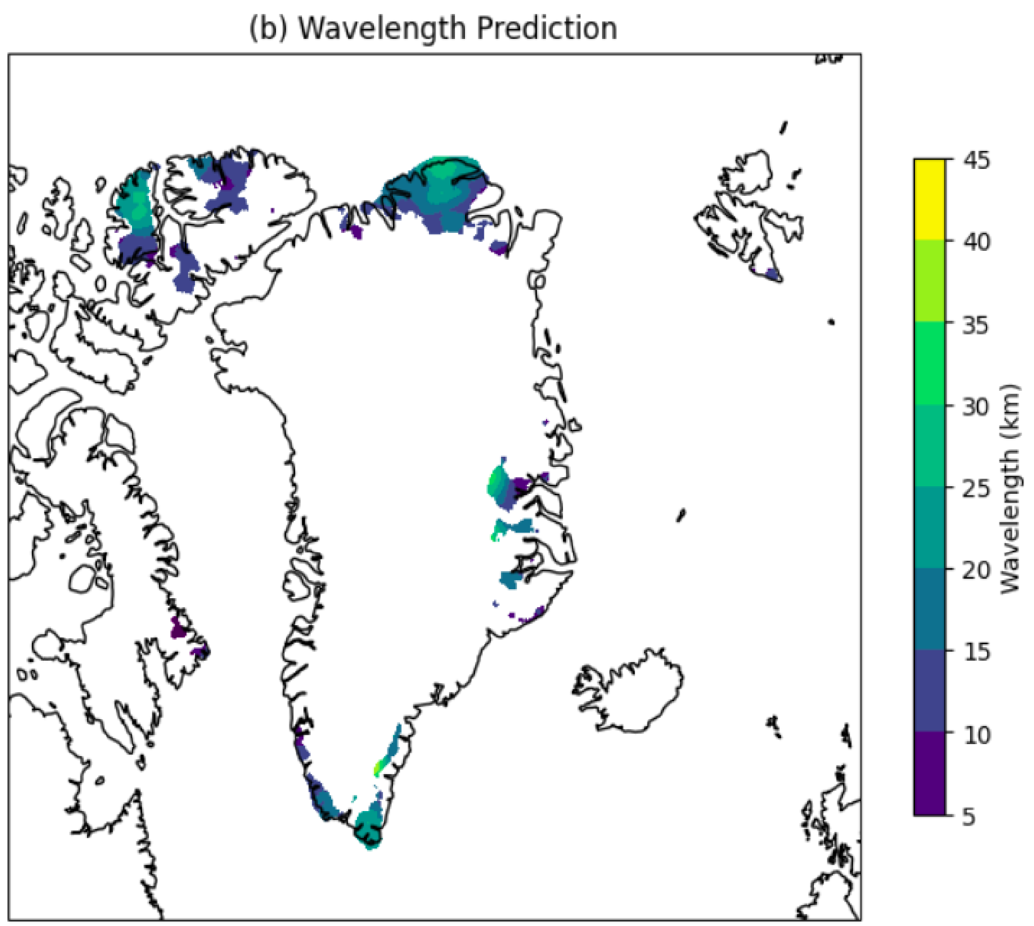


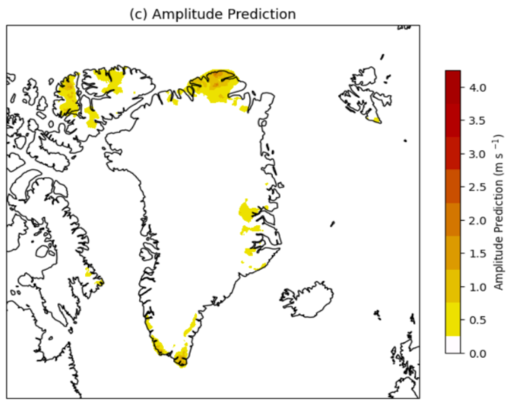
Second, UNet encode+decoder frozen, but added 1x1 convolutions to predict synthetic gravity waves

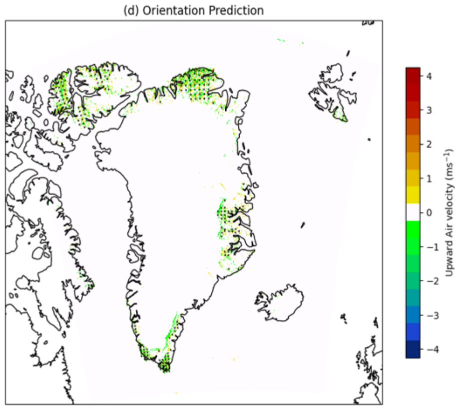








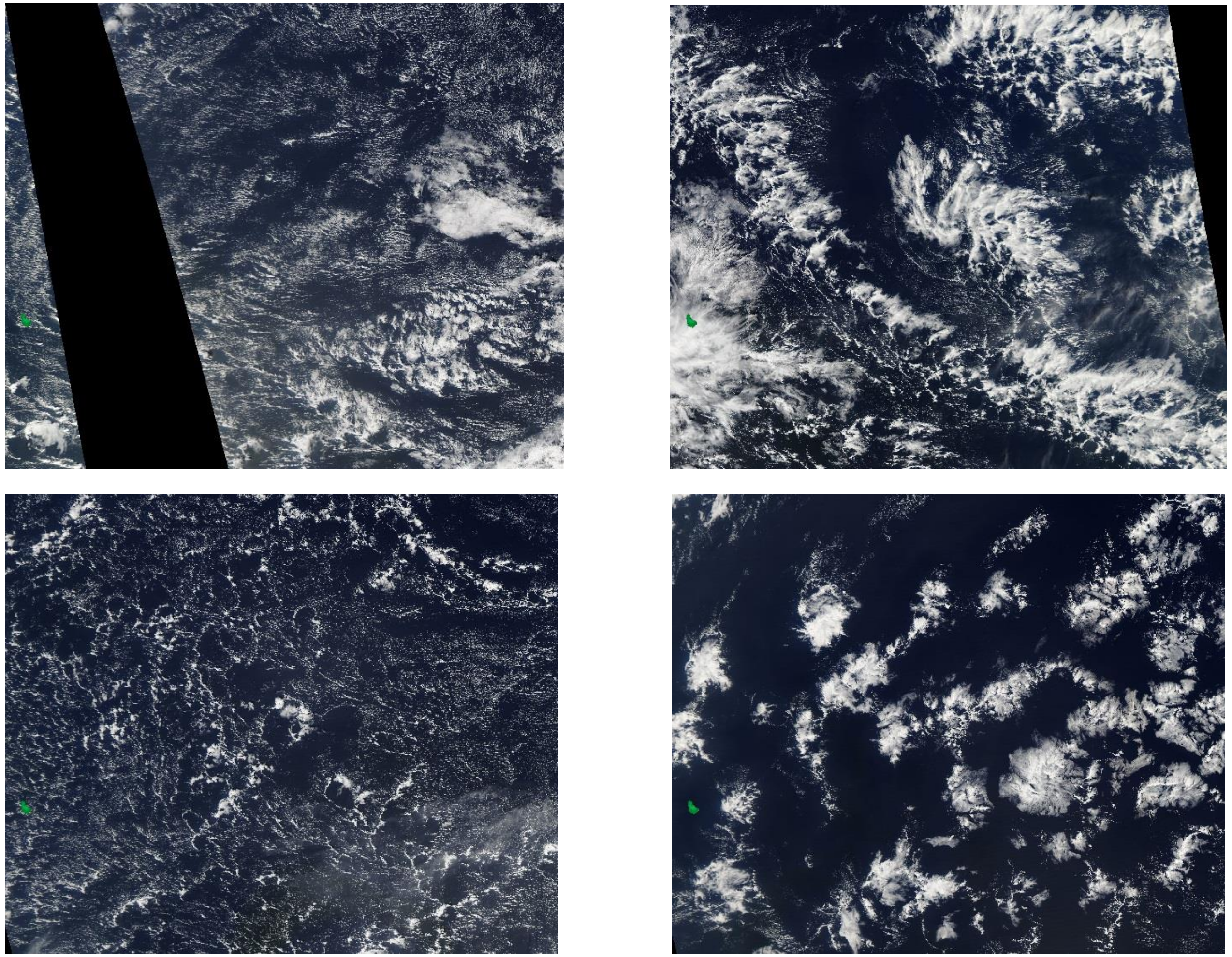




Courtesy **Eleni Briola**

Self-supervised cloud-organisation

Identifying climate-feedback effects in cloud dynamics



"sugar"

"fish"

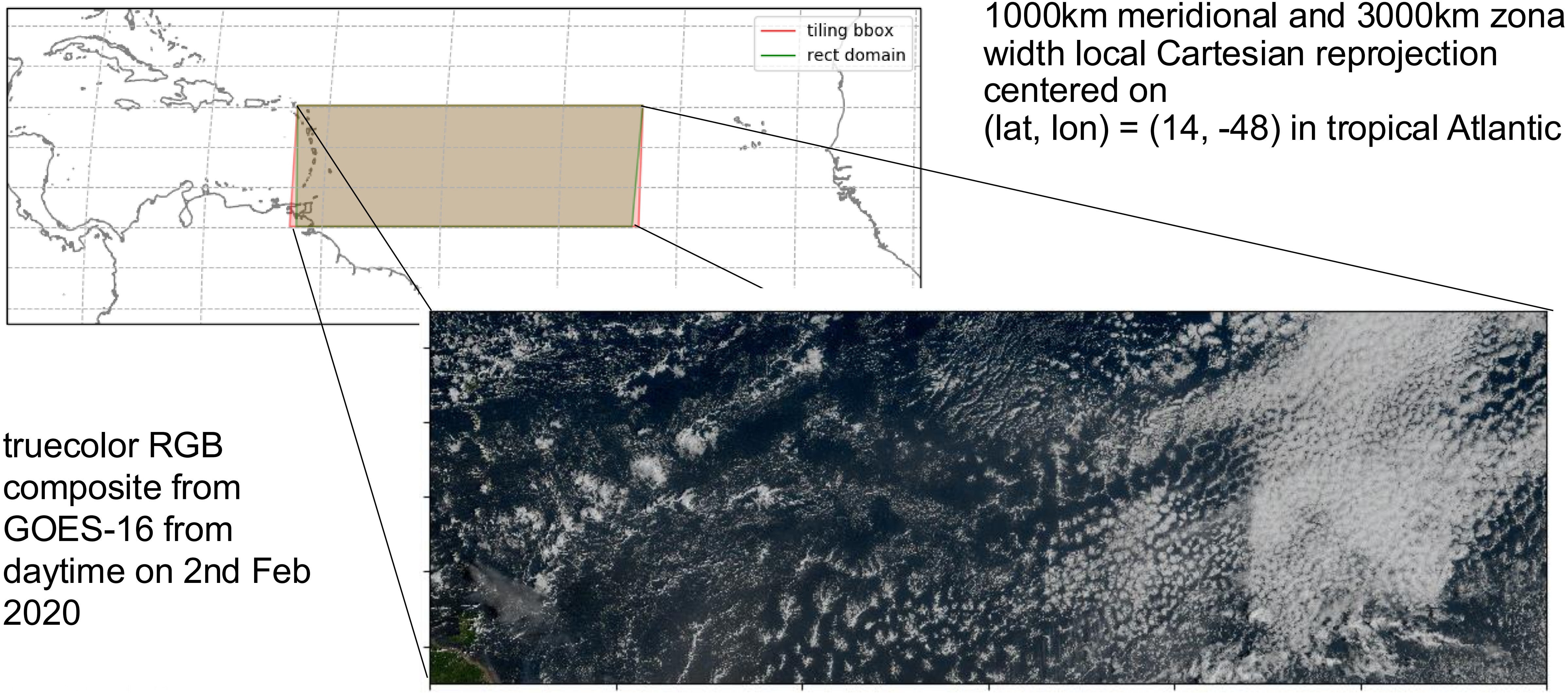
"gravel"

"flower"

Stevens et al 2020, QJRMS

What happens between the “archetypes”?

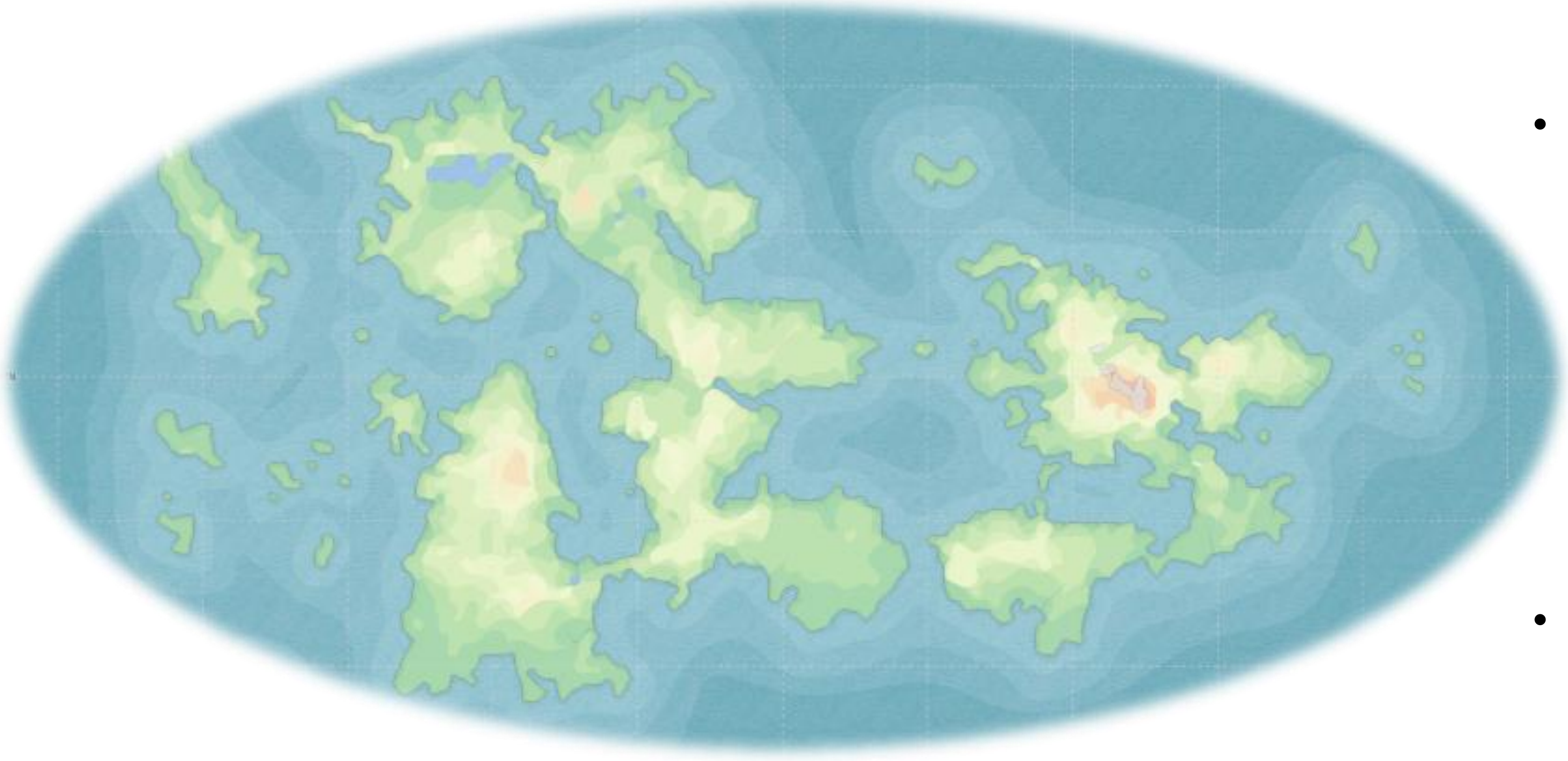
Are they all that exist?



1000km meridional and 3000km zonal width local Cartesian reprojection centered on
(lat, lon) = (14, -48) in tropical Atlantic

truecolor RGB composite from GOES-16 from daytime on 2nd Feb 2020

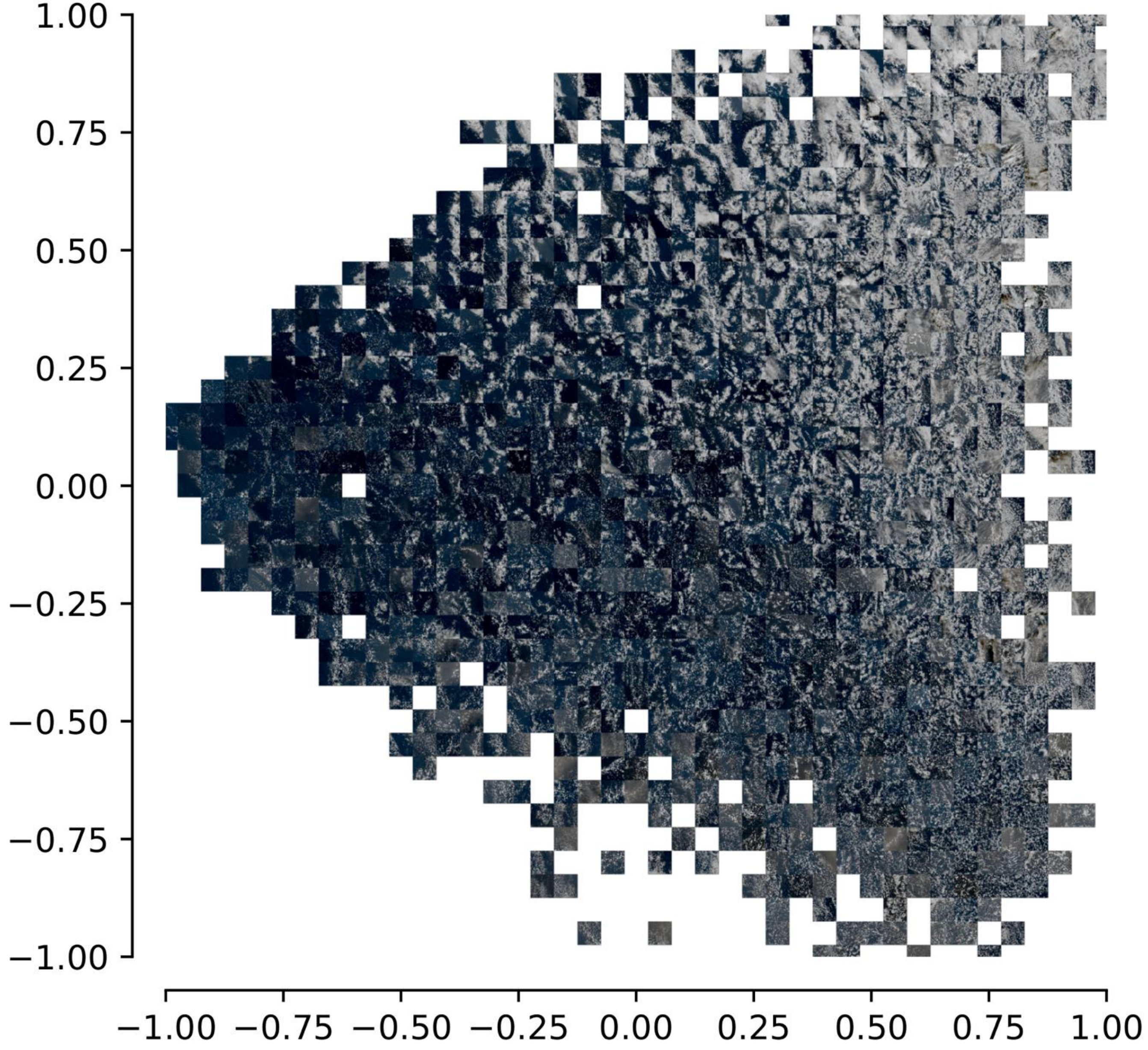
Extracting the embedding manifold



- Idea: maybe all the tile embeddings lie on some manifold in the embedding space
- Use Isomap method , to extract manifold in high-dimensional embedding space and map to 2D
  - *“Isomap seeks a lower-dimensional embedding which maintains geodesic distances between all points”*
- With this I now have a “map” of all possible types of organisation

What does the world of cloud organisation look like?

Extracting the embedding manifold



- Idea: maybe all the tile embeddings lie on some manifold in the embedding space
- Use Isomap method (Tenenbaum et al 2000) to extract manifold in high-dimensional embedding space and map to 2D
  - *"Isomap seeks a lower-dimensional embedding which maintains geodesic distances between all points"*
- With this I now have a "map" of all possible types of organisation

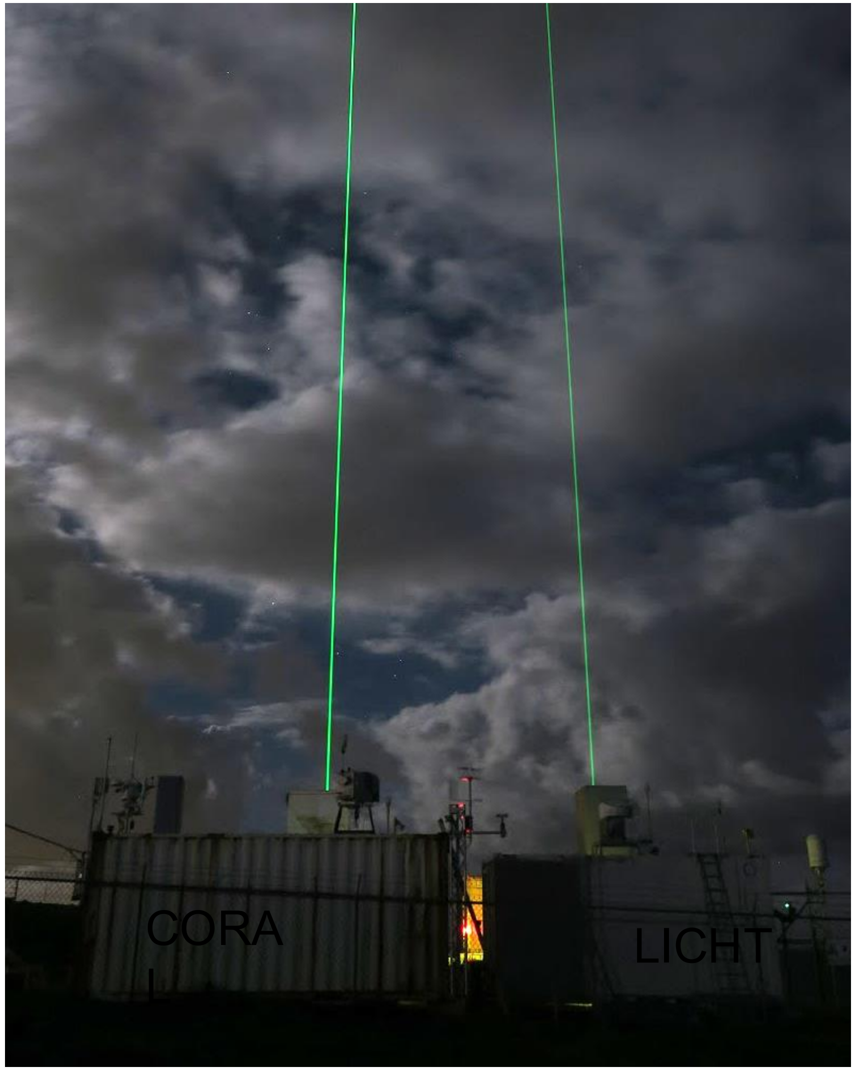
Self-supervised denoising

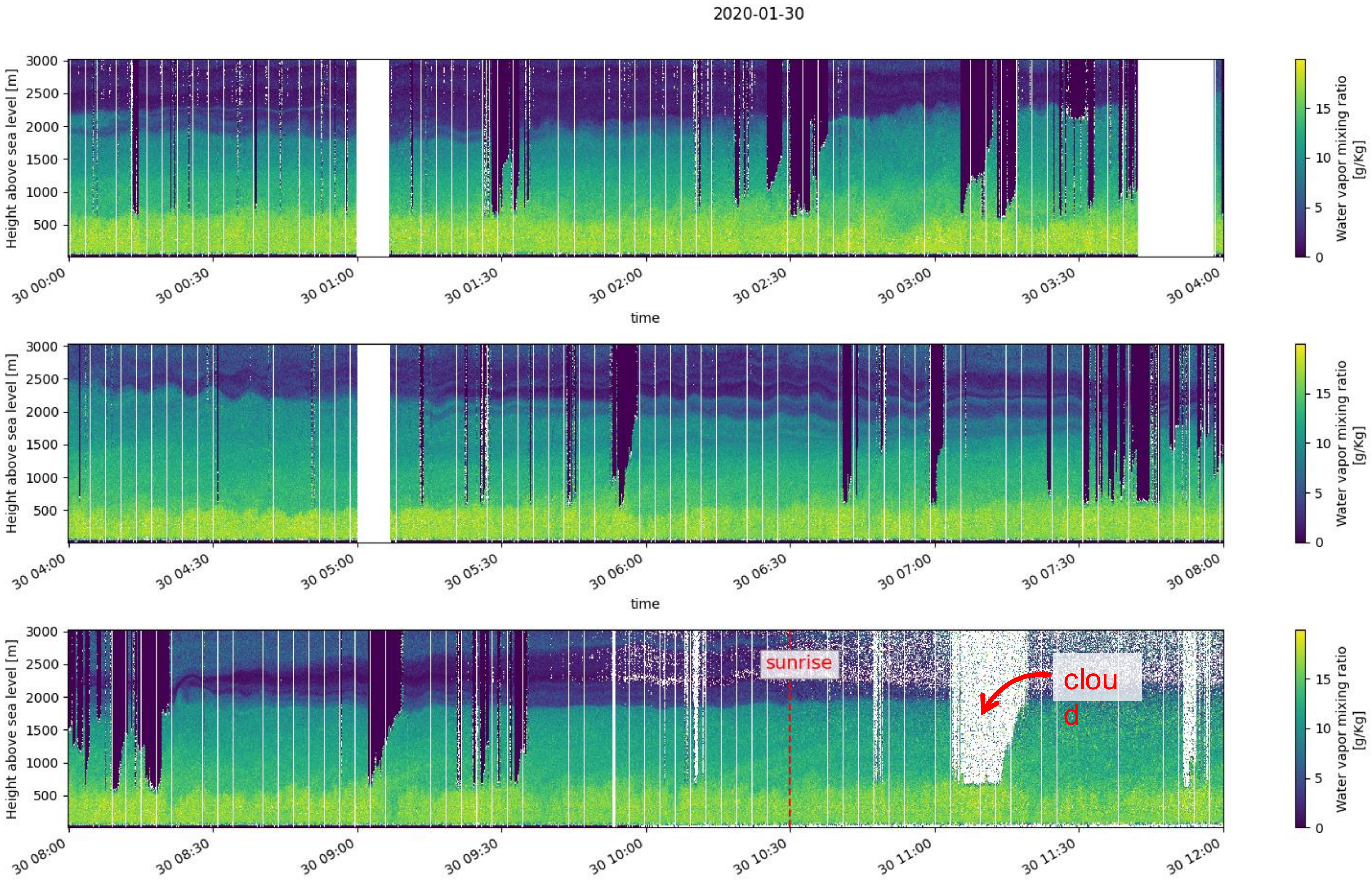
Finding cloud-triggering atmospheric structures

How do I “see” these structures?

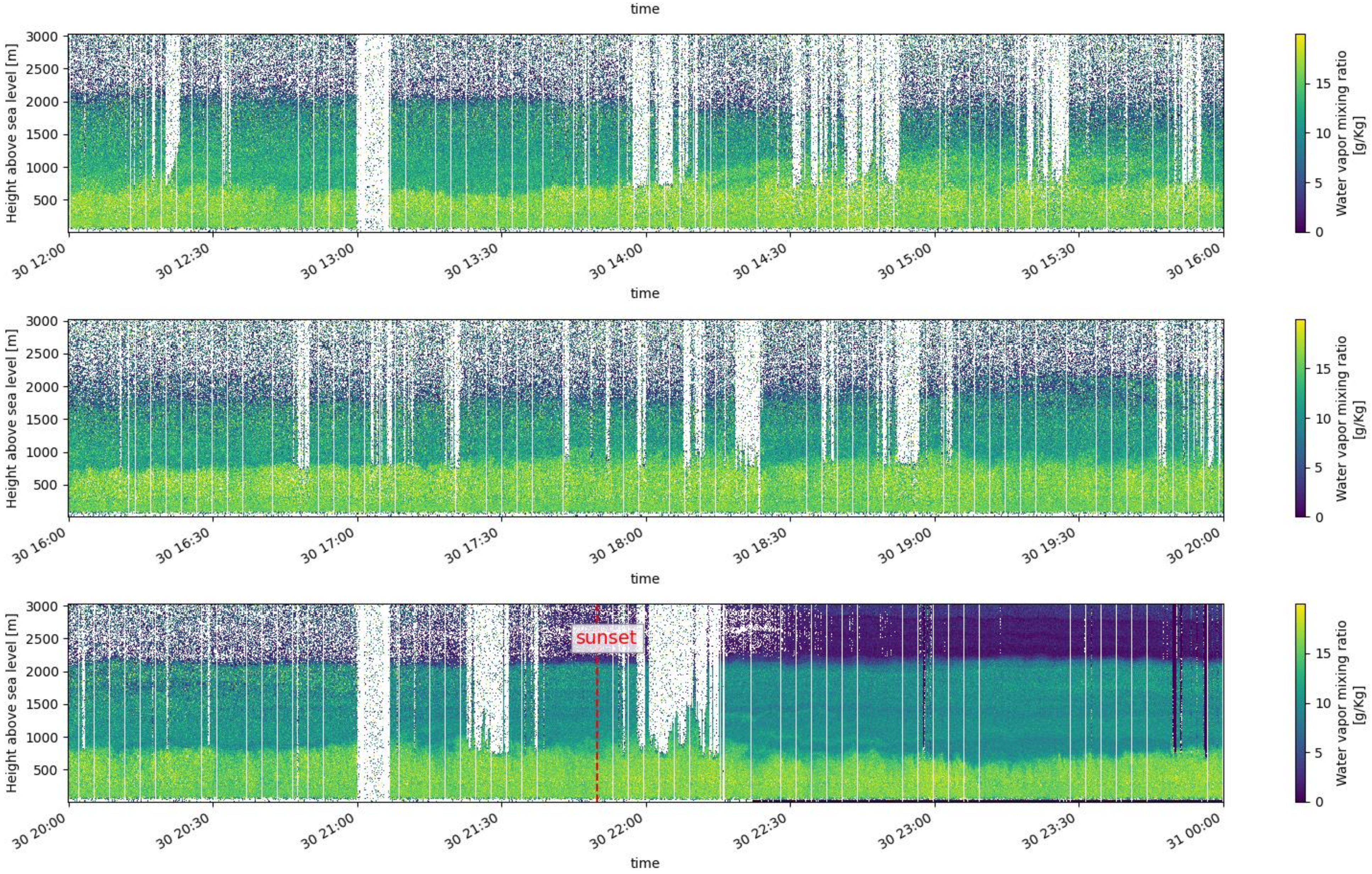
The Barbados Cloud Observatory CORAL Raman LIDAR

- Measure <u>water-vapour profiles</u> (below cloud), air temperature, aerosols and cloud properties.
- resolution:
  - horizontal wind: v ~ 5m/s
  - temporal resolution: $\Delta t = 4s$
  - => horizontal res: $\Delta x \sim 20m$
  - vertical res: $\Delta z \sim 15m$
- Developed and run by Ilya Serikov (MPI-Meteorlogy, Hamburg)

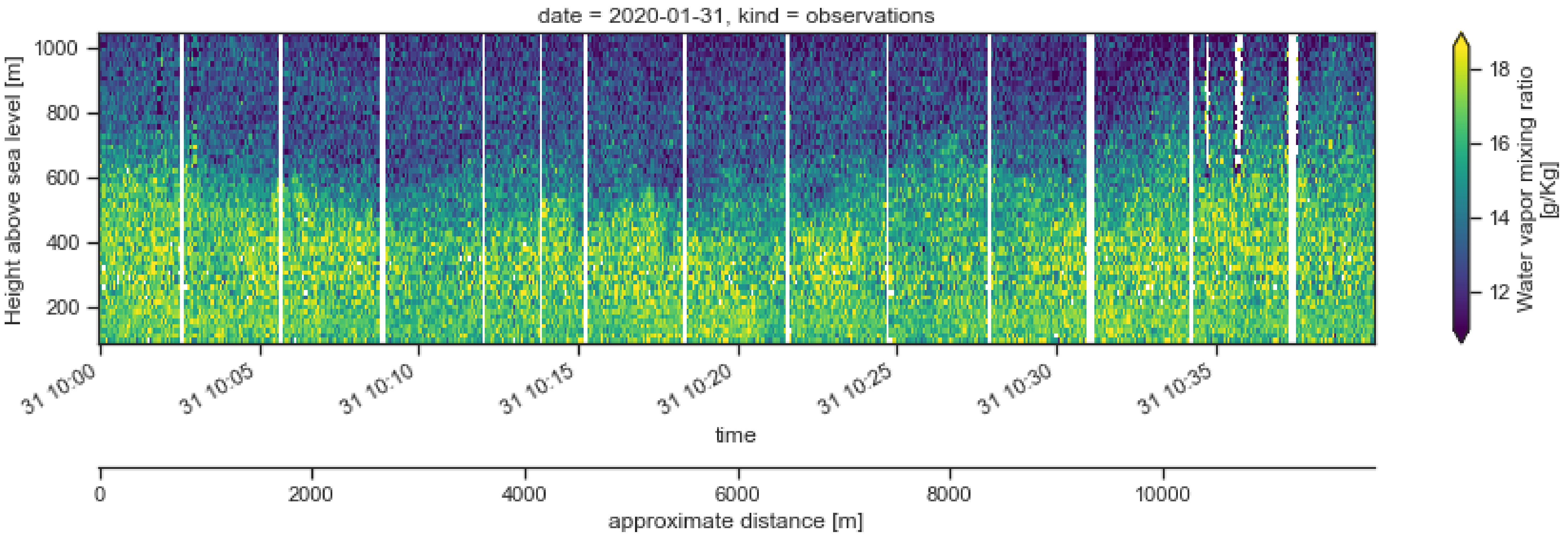




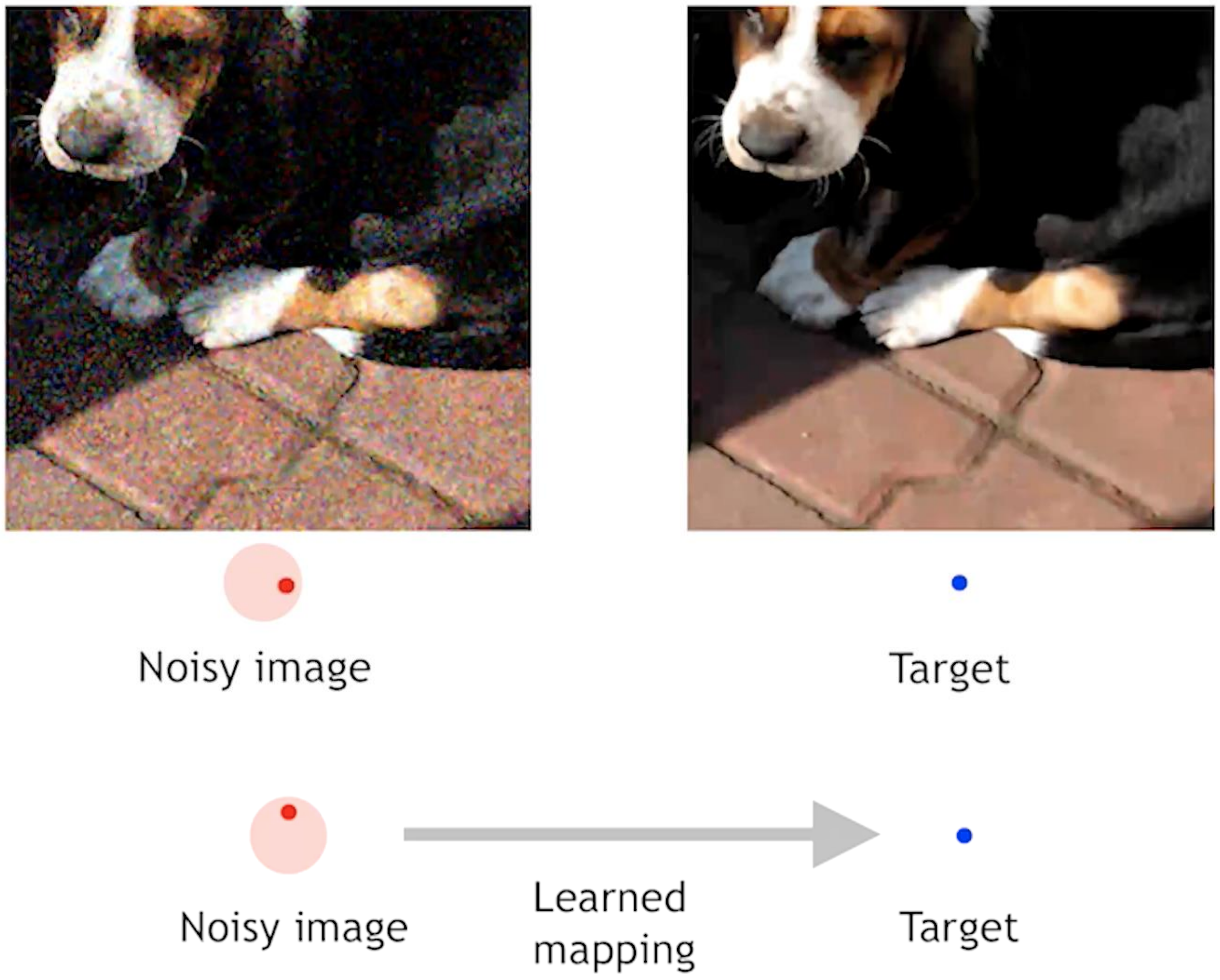
- Depth of mixed boundary layer clearly seen (~600m)
- Clouds block LIDAR, cloud-base at ~600m altitude
- More noise during daylight hours



- Depth of mixed boundary layer clearly seen (~600m)
- Clouds block LIDAR, cloud-base at ~600m altitude
- More noise during daylight hours


date = 2020-01-31, kind = observations
Height above sea level [m]
1000
800
600
400
200
31 10:00
31 10:05
31 10:10
31 10:15
31 10:20
31 10:25
31 10:30
31 10:35
time
0
2000
4000
6000
8000
10000
approximate distance [m]
Water vapor mixing ratio [g/Kg]
18
16
14
12

- For supervised learning we need pairs of noisy input and clean target data, but for real-life observations we may not have clean data
  - Could *synthesize* training data using an assumed noise distribution applied to synthetic data - need simulated data and noise model
- Can I do something with just the noisy observations?



(Krull et al 2019)

- Assume noise at any two points in input is uncorrelated
- Exploit that image contains a high degree of structure
- Learn correction to point value from looking only at neighbouring pixels. Network forced to ignore central pixel by overwriting with random pixel in neighbourhood during training
  - If central pixel is included network simply learns identity
- Idea: if noise is uncorrelated then the only thing the network can learn from the context (surrounding) pixels is the true denoised value of a pixel

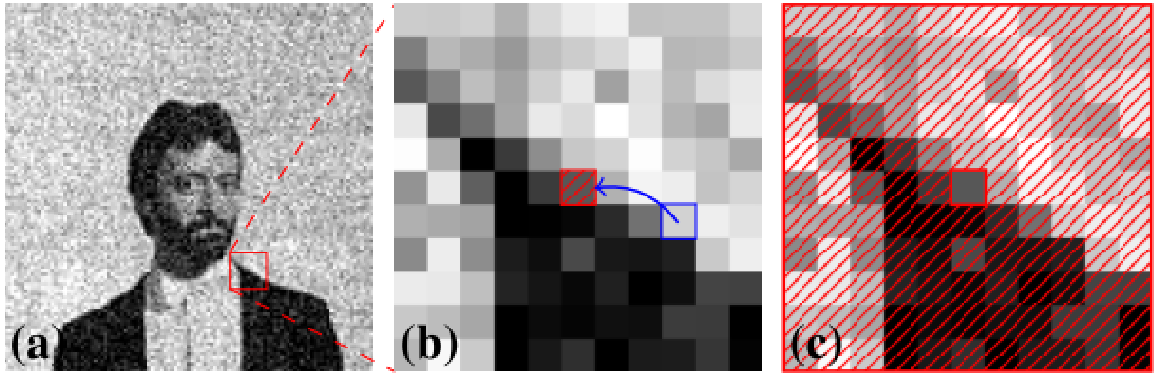
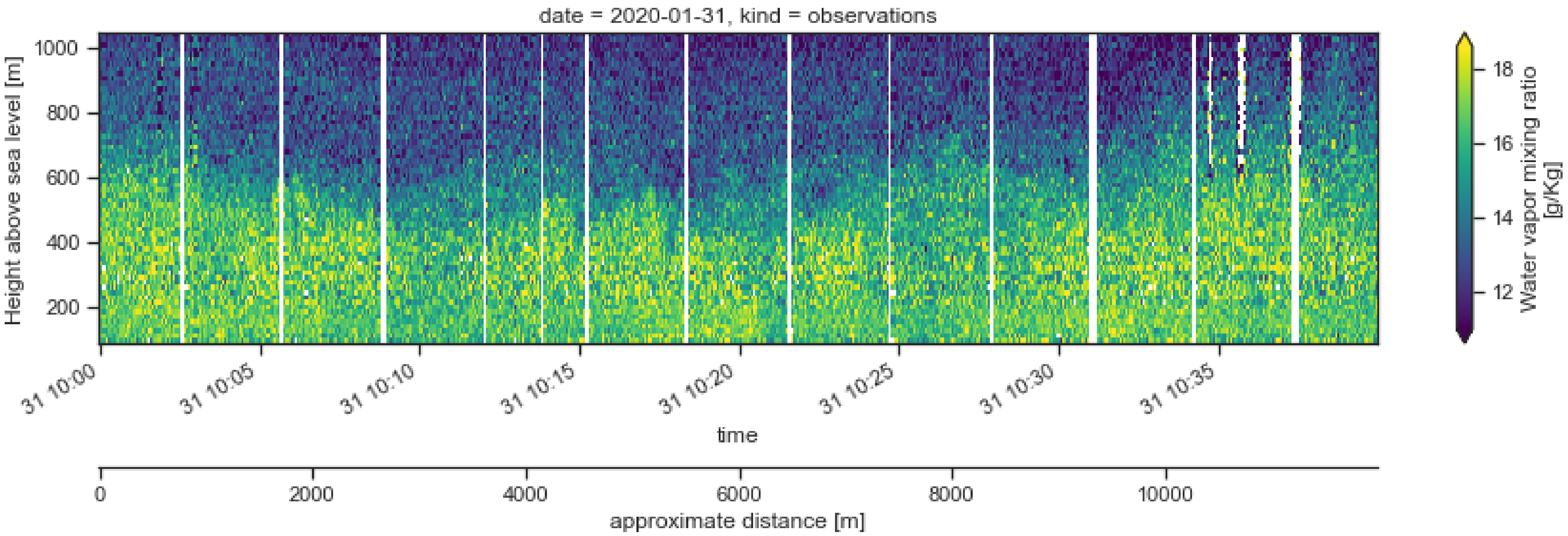
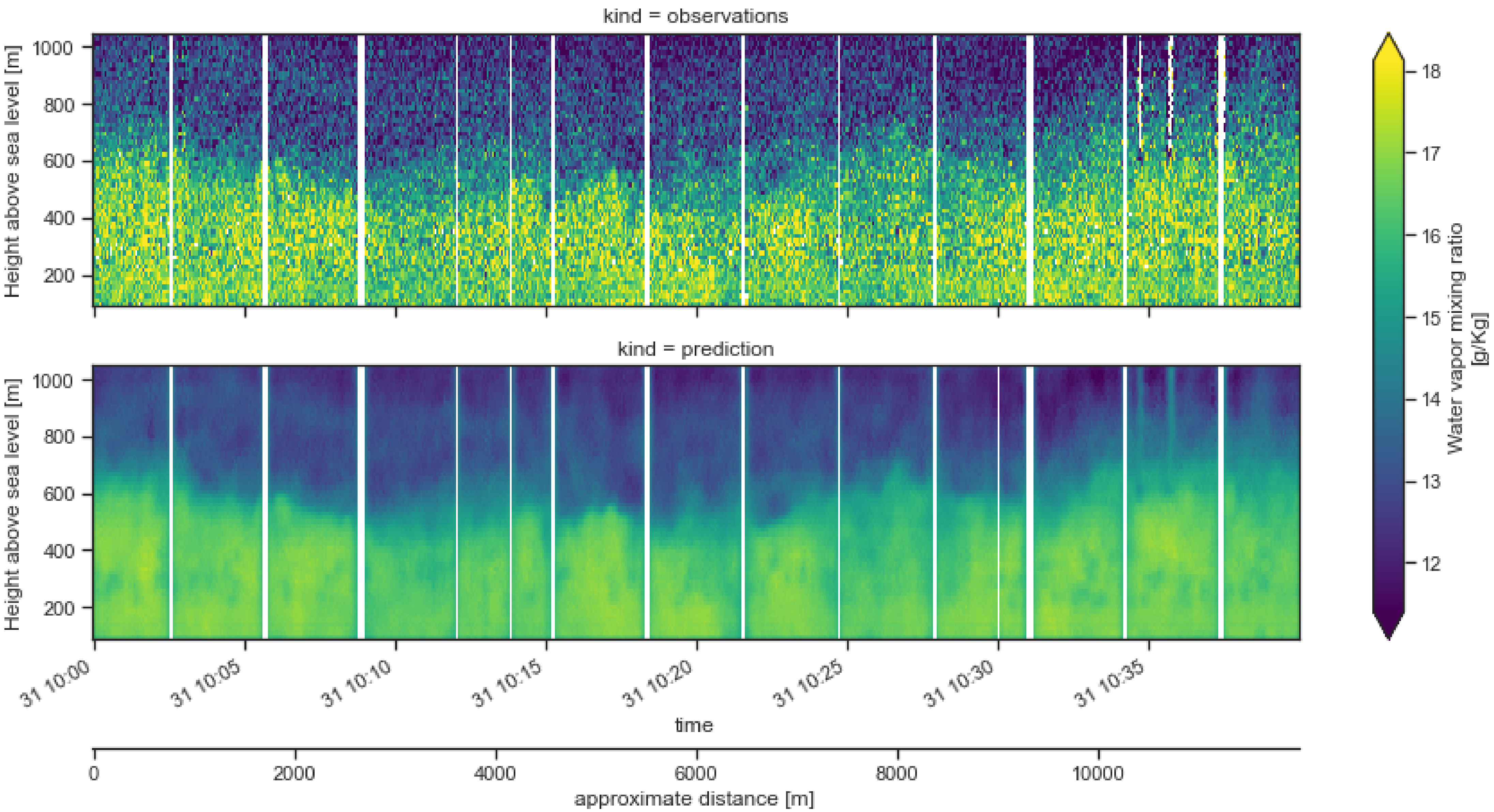
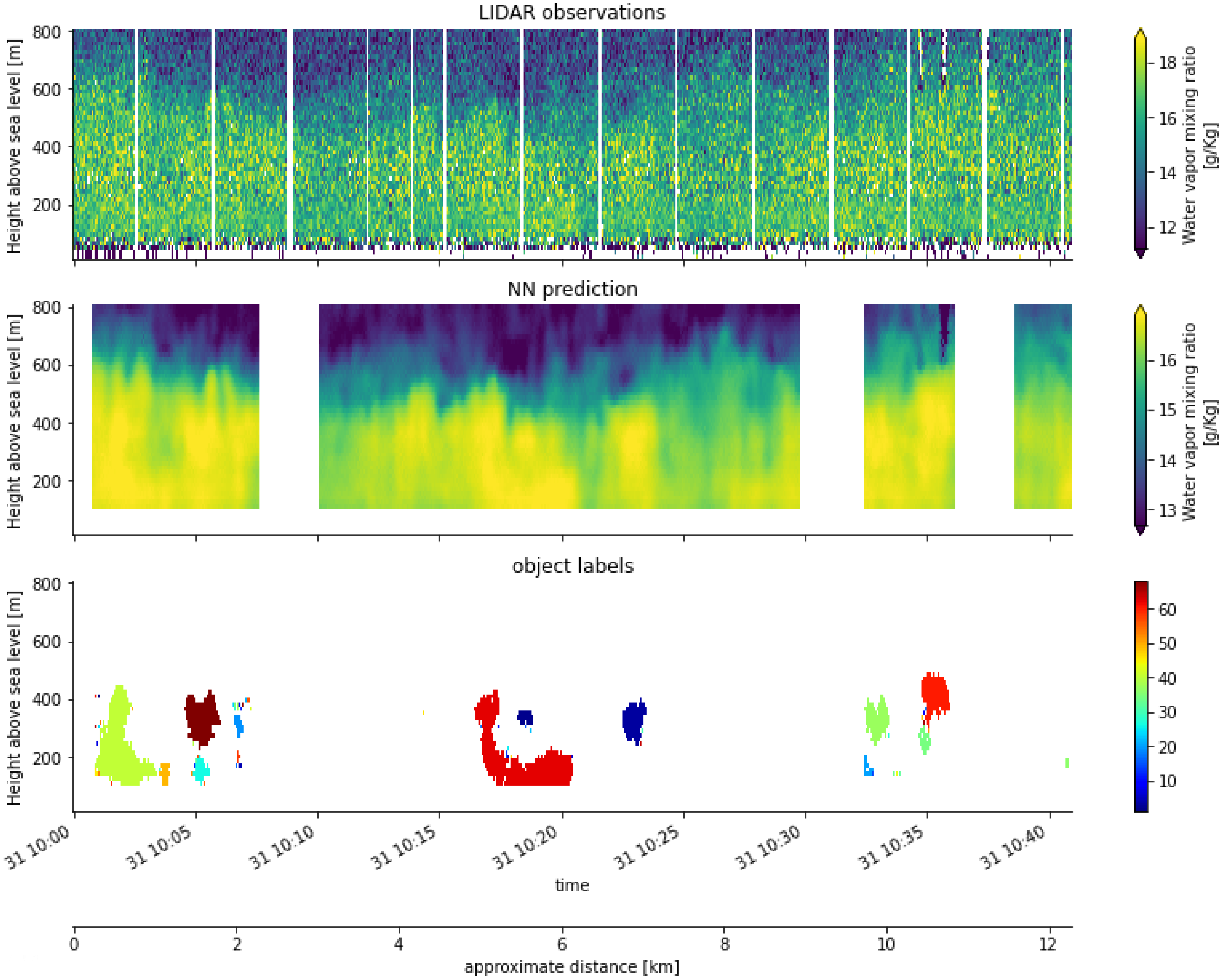


Figure 3: Blind-spot masking scheme used during NOISE2VOID training. **(a)** A noisy training image. **(b)** A magnified image patch from (a). During N2V training, a randomly selected pixel is chosen (blue rectangle) and its intensity copied over to create a blind-spot (red and striped square). This modified image is then used as input image during training. **(c)** The target patch corresponding to (b). We use the original input with unmodified values also as target. The loss is only calculated for the blind-spot pixels we masked in (b).


date = 2020-01-31, kind = observations
Height above sea level [m]
1000
800
600
400
200
31 10:00
31 10:05
31 10:10
31 10:15
31 10:20
31 10:25
31 10:30
31 10:35
time
0
2000
4000
6000
8000
10000
approximate distance [m]
Water vapor mixing ratio [g/Kg]
18
16
14
12



kind = observations
kind = prediction
Height above sea level [m]
200
400
600
800
1000
31 10:00
31 10:05
31 10:10
31 10:15
31 10:20
31 10:25
31 10:30
31 10:35
time
0
2000
4000
6000
8000
10000
approximate distance [m]
Water vapor mixing ratio [g/Kg]
12
13
14
15
16
17
18

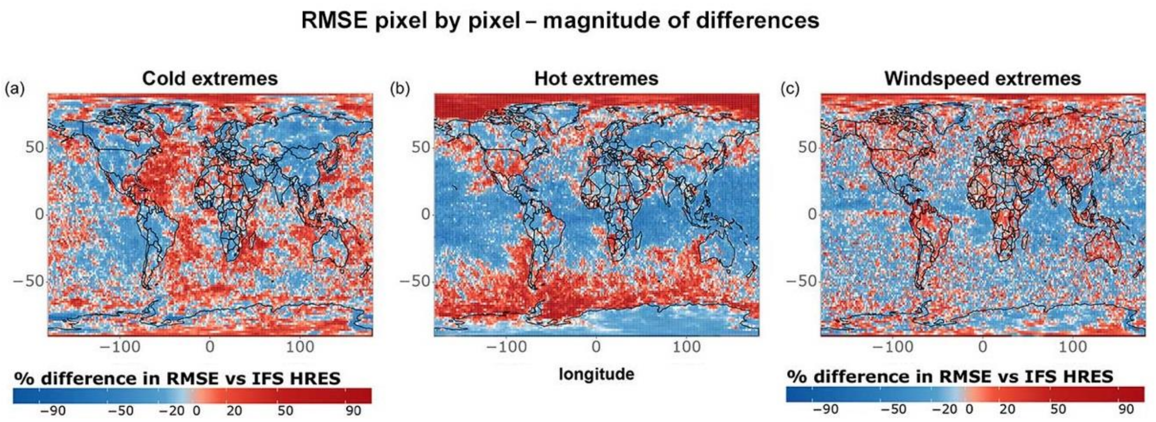

LIDAR observations
NN prediction
object labels
Height above sea level [m]
Water vapor mixing ratio [g/Kg]
time
approximate distance [km]
31 10:00
31 10:05
31 10:10
31 10:15
31 10:20
31 10:25
31 10:30
31 10:35
31 10:40

Future directions: outstanding issues with data-driven models

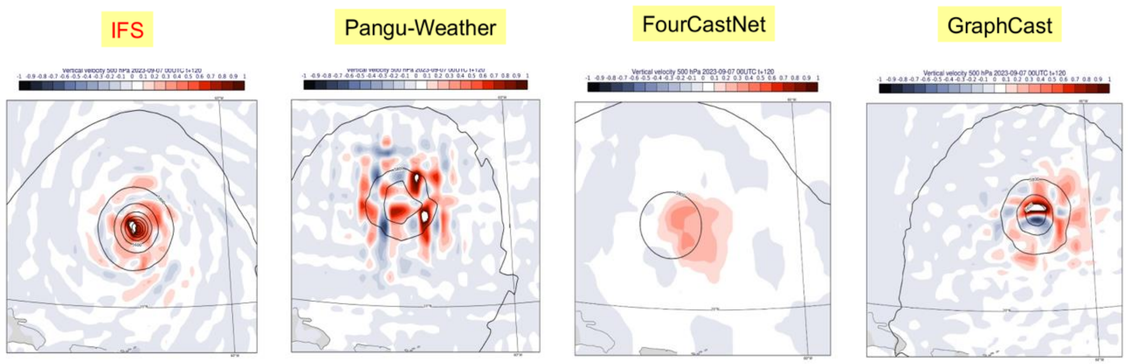
Frontiers of ML-based purely data-driven weather forecasting, are (lack of) representation of:

- extremes, e.g. temperature and particularly rainfall
- highly non-linear physics, e.g. cloud formation
- physical consistency, e.g. geostrophic balance

Analysis so far has only been on global (Δx ~ 20km) forecasts, the issues are likely to be more acute a km-scale (Δx ~ 2km).



Olivetti & Messori 2024



Bonavita 2023

Building collaborations with DTU

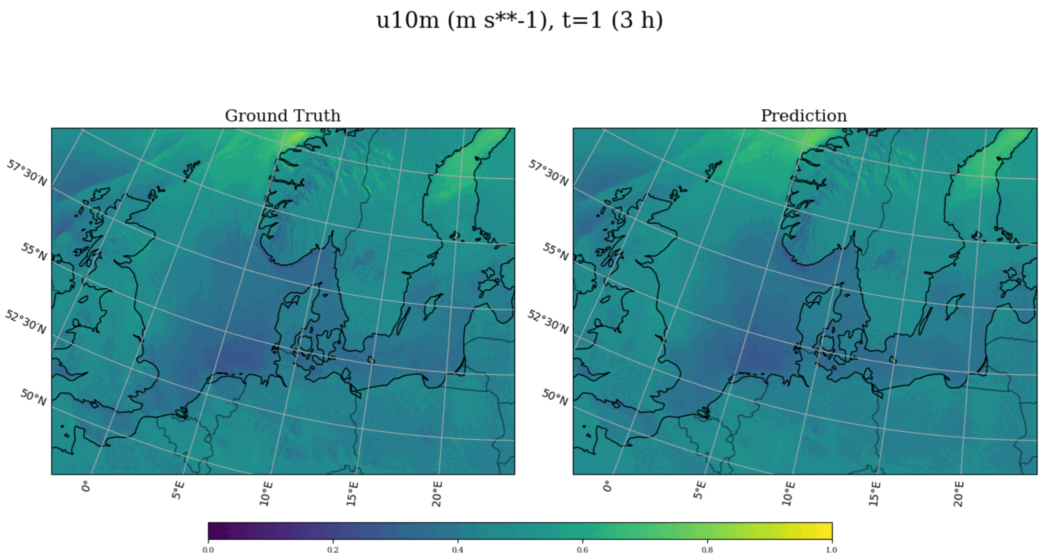
Why work with us? :)

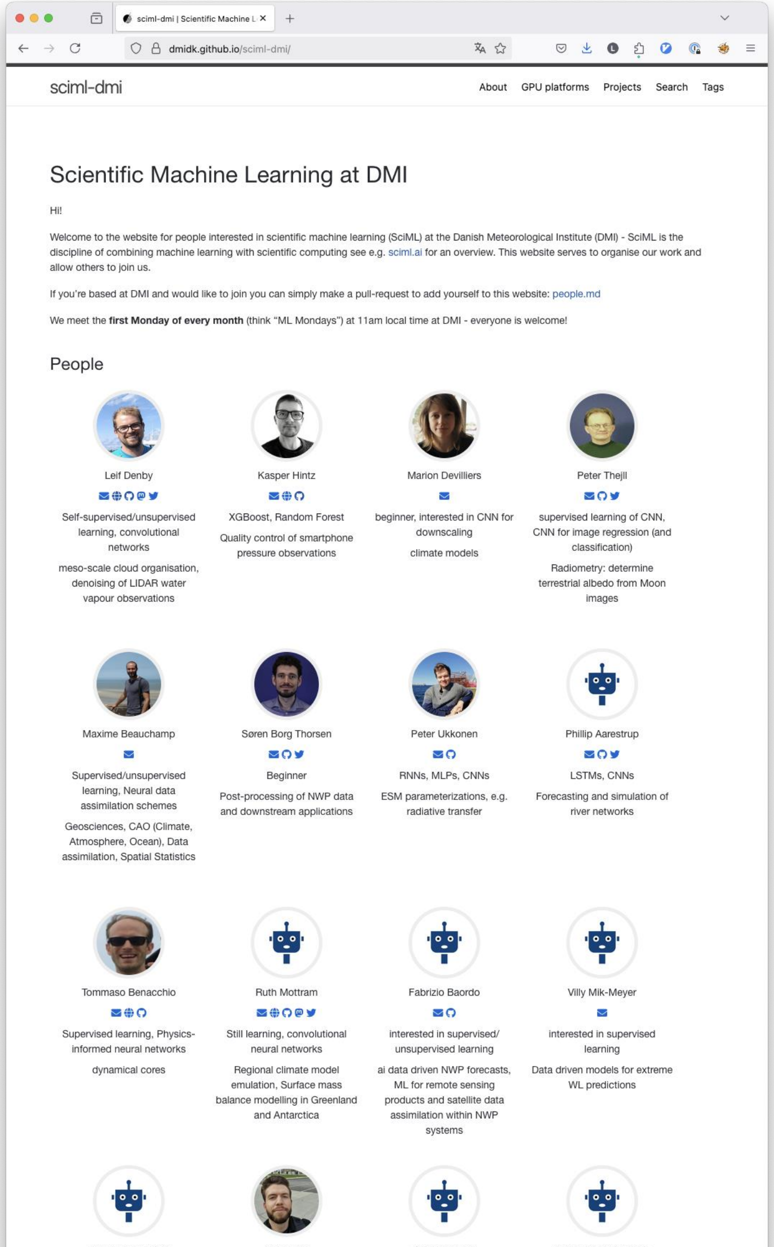
- Interesting datasets (wealth of observations, reanalysis, etc)
- Interesting applications (weather preparedness, disaster response, etc)
- Operationalisation – see your research used
- Expertise in atmospheric science, scientific computing and increasingly SciML

How to get in touch?

https://dmidk.github.io/sciml-dmi/

- Join or present at our SciML monthly meetings - contact Simon (skc@dmi.dk) for details
- Email anyone on our team
  - Me, Leif Denby (lcd@dmi.dk) – I'm going on leave till September though
- Talk to me after this :)





Danish Meteorological Institute

Dec 2023: Ensemble data-driven model (GenCast, Google)

“Producing a single 15-day trajectory with GenCast takes around a minute on a Cloud TPU v4, and so N ensemble members can also be generated in around a minute with N TPUs, enabling the use of orders of magnitude larger ensembles in the future”

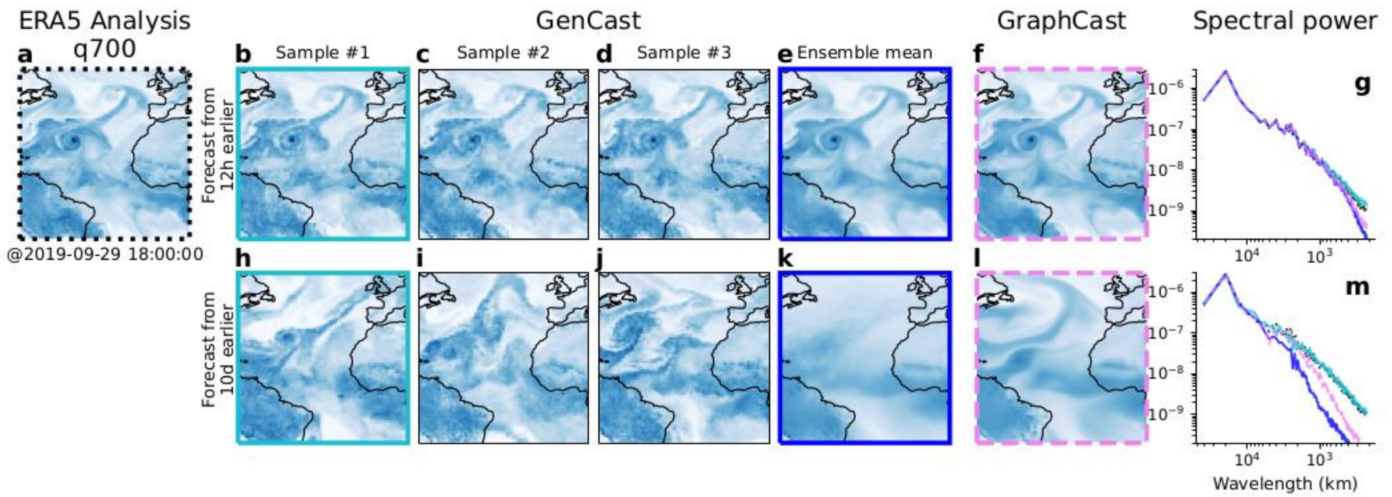
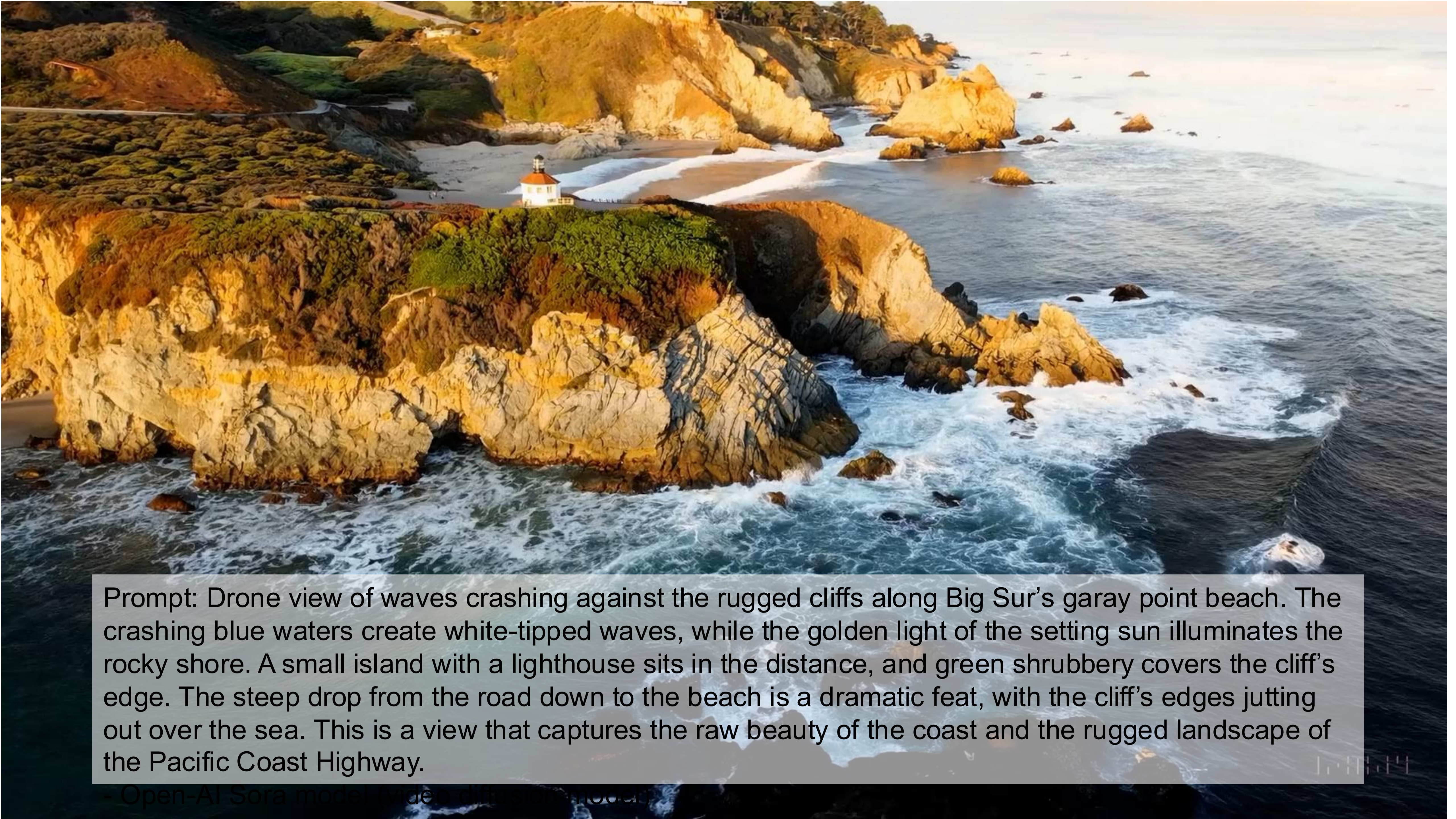


Figure 3 | Visualization of representative states produced by GenCast compared to GraphCast. (a) ERA5 analysis state for specific humidity at 700hPa at 6pm on the 29th of September of 2019. (b-d) 3 sample forecasts of this state by GenCast from 12 hours earlier. (e) Ensemble average obtained by taking the mean of 50 sample forecasts by GenCast from 12 hours earlier. (f) Forecast by the GraphCast (model which is deterministic), made 12 hours earlier. (g) Spectrum of the fields shown in panels (a-f), with colors matching the frames of the panels. (h-m) Same as (b-g), but for forecasts made 10 days earlier. Unlike deterministic GraphCast, which expresses uncertainty as blurring which increases with lead time (f, l), we observe how the sample forecasts produced by GenCast are sharp (g, m), regardless of whether the forcasts are for 12 hours ahead (g, b-d) (where the three samples are very similar) or 10 days ahead (m, h-j) (where the three samples differ more). The samples can still be averaged to produced a blurry mean state (e, k). Additional visualizations and an explanation of how this date/time was selected for visualisation are available in Appendix A.8.



Prompt: Drone view of waves crashing against the rugged cliffs along Big Sur’s garay point beach. The crashing blue waters create white-tipped waves, while the golden light of the setting sun illuminates the rocky shore. A small island with a lighthouse sits in the distance, and green shrubbery covers the cliff’s edge. The steep drop from the road down to the beach is a dramatic feat, with the cliff’s edges jutting out over the sea. This is a view that captures the raw beauty of the coast and the rugged landscape of the Pacific Coast Highway.

- Open-AI Sora model (video diffusion model)