



Koninklijk Meteorologisch Instituut

Institut Royal Météorologique

Königliches Meteorologisches
Institut

Royal Meteorological Institute

Comparing stretched-grid and limited-area modelling for data-driven regional weather forecasting

D. Van den Bleeken, M. Van Ginderachter and P. Termonia (RMI)
J. Wijnands, B. François and S. Buurman (KNMI)

Machine Learning Pilot Project

EWGLAM-SRNWP 2025
Norrköping, Sweden: 22 – 25 September 2025



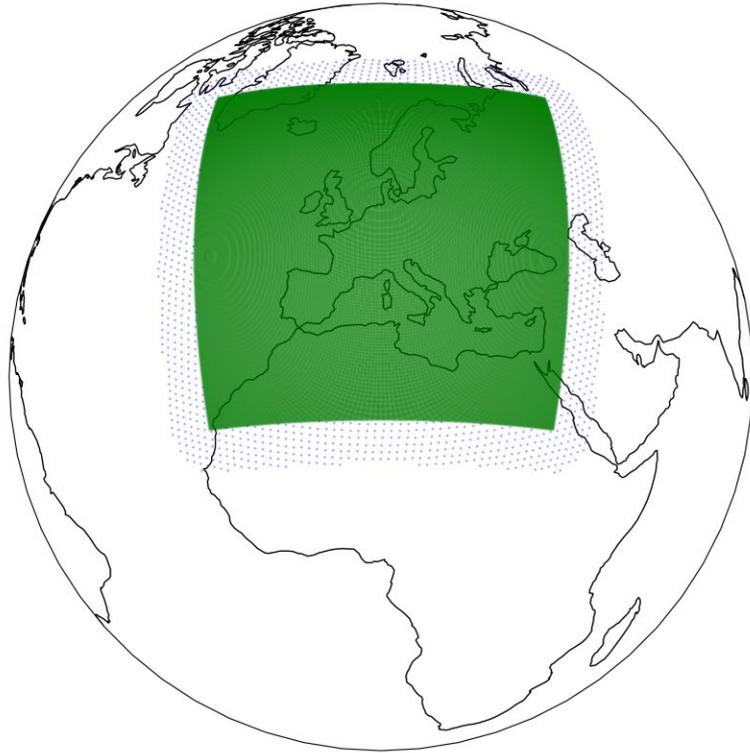
arXiv 2507.18378

Comparison of two types of regional MLWP models

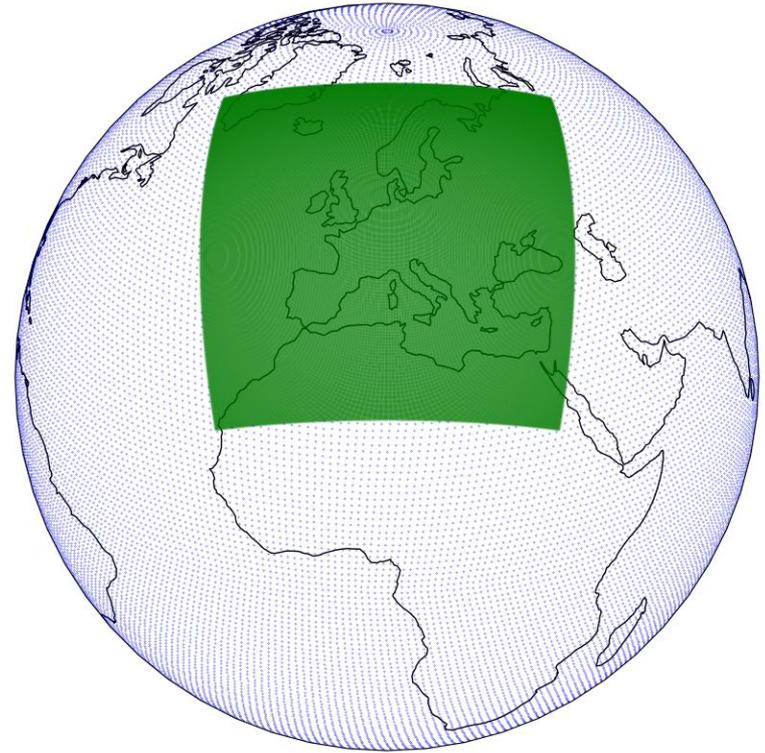
- Developed and trained in identical setting (anemoi)
- Common regional dataset: CERRA reanalysis (5.5km resolution)

OUTLINE

1. Introduction of the models
2. Comparative evaluation
 - a. Overall performance
 - b. Per variable differences
 - c. Effect of initialization and forcings
 - d. Differences in generalizability
 - e. Impact of model size
3. Summary & outlook



Limited area model (LAM)



Stretched grid model (SGM)

GNN's with Encoder - Processor - Decoder architecture (à la Graphcast)

Based on two spatial grids:

- data grid (physical grid)
- hidden grid (lower resolution grid for processing)

Graph-transformers:

- Encoder: data → hidden edges based on proximity
- Processor: hidden → hidden edges based on scale
- Decoder: hidden → data edges based on proximity

Implemented through the anemoi software

Data grid: LAM \subset SGM

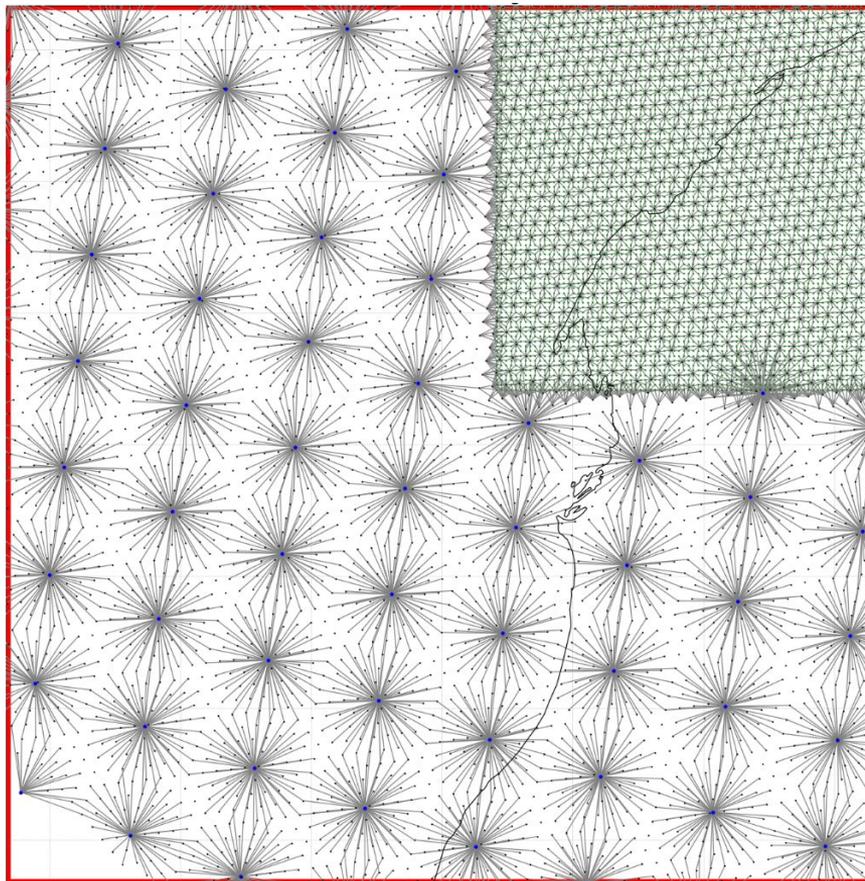
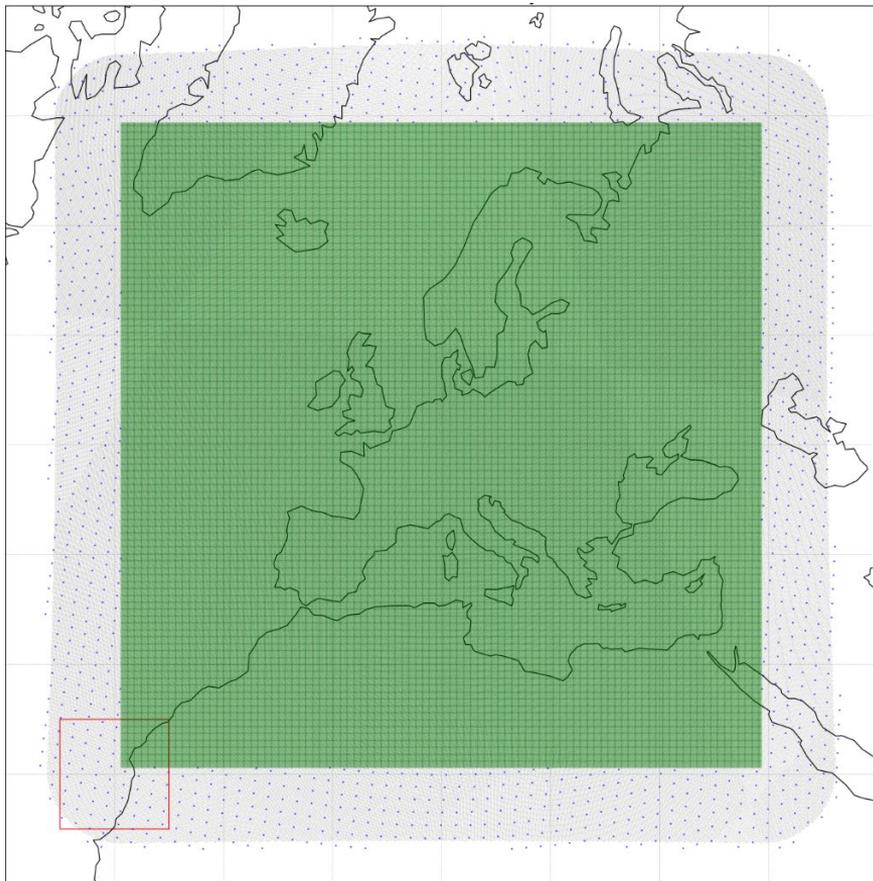
- Regional domain: 5.5 km (native) resolution CERRA grid
- Outside: \approx 100 km (o96) resolution ERA5 grid

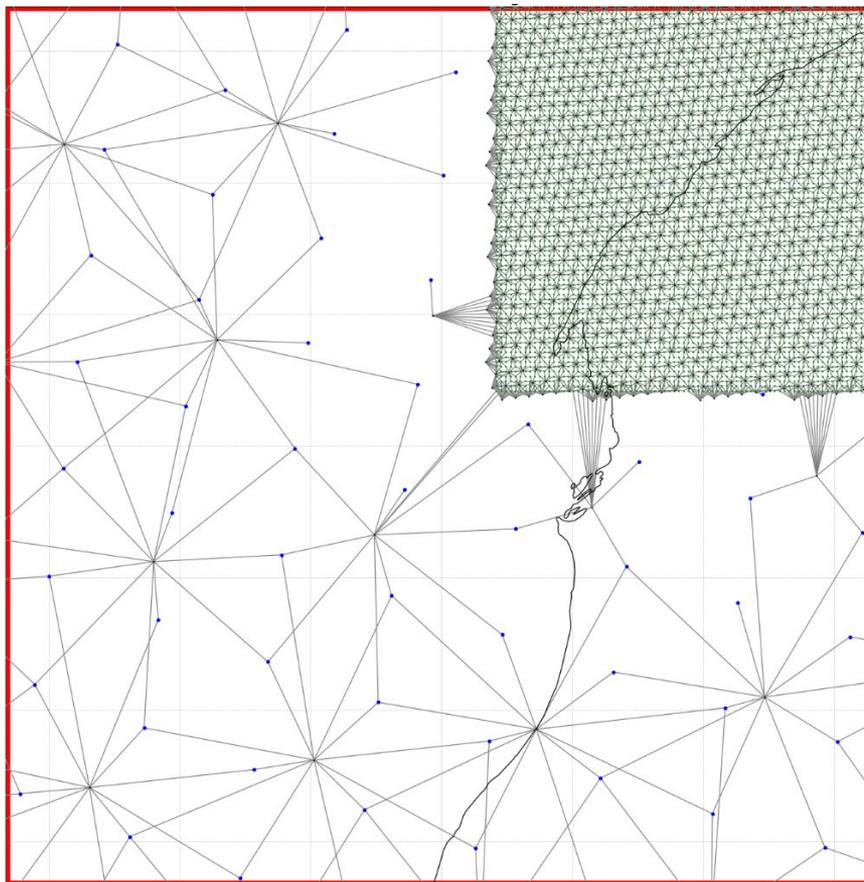
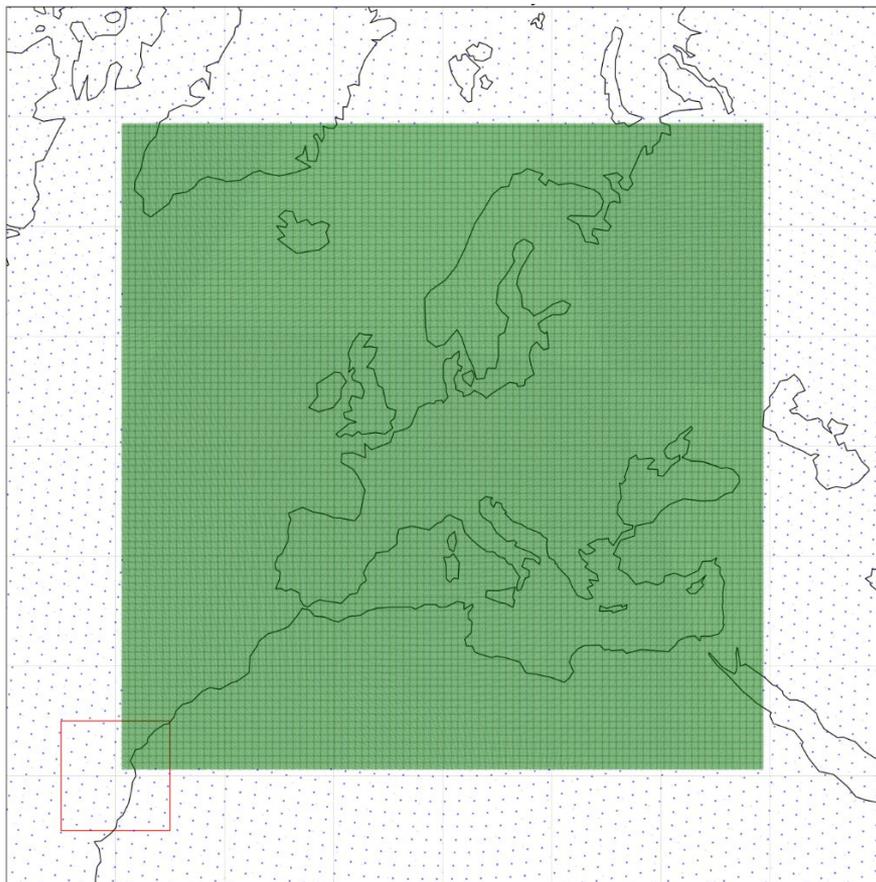
Hidden grid: triangular refined icosahedral

- LAM: \approx 15 km resolution (refinement 9) regional domain + 600 km
- SGM: \approx 15 km resolution (refinement 9) regional domain + 11 km
 \approx 240 km resolution (refinement 5) remainder of globe

Edges: (data to hidden)

- SGM: hidden nodes connected to 12 nearest neighbours (nn) in data grid
- LAM: hidden nodes in regional domain connected to 12 nn in data grid
 hidden nodes outside regional domain connected to all data points within 66km radius \Rightarrow effective boundary size in data grid: 666 km





- Dataset: CERRA + ERA5
1984 - 2018: Training | 2019: Validation | 2020: Test
- Two stages
 - a. Main
 - 260k training steps
 - Single 6h timestep
 - b. Rollout
 - 5k training steps
 - Multiple 6h timesteps: incremental up to 12 steps (72h)

	SGM	LAM
HPC hardware	Leonardo/Atos	LUMI
normalization	ERA5 statistics	CERRA statistics
Training loss weighting	Global domain: 75% Regional domain: 25%	Boundary domain: 0% Regional domain: 100%
Validation metrics	In- and/or outside regional domain	Inside regional domain
Rollout	Regional domain: autoregressive Global domain: autoregressive	Regional domain: autoregressive Boundary domain: external forcing

Main evaluation settings:

- Period: **2020/01/02 - 2020/12/25**
- Over regional domain, vs CERRA reanalysis/SYNOP stations

Various models and inference setups were considered

`[structure]-[size]-[rollout]-[setup]`

E.g. LAM-1024-R12-ideal **or** SGM-512-R01-operational

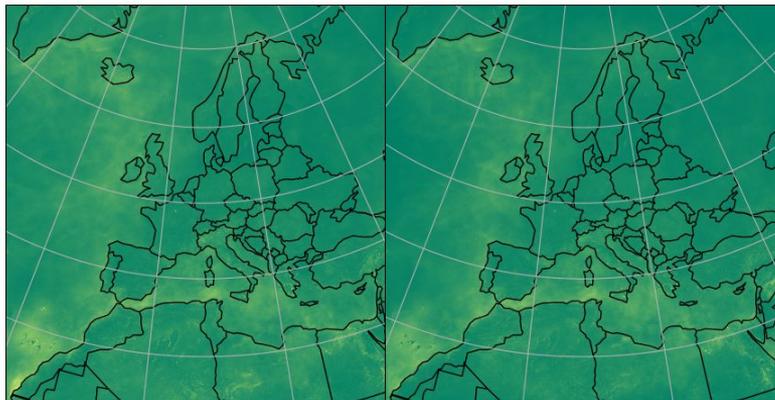
Selection of experiments in this presentation

(more in the paper, even more investigated)

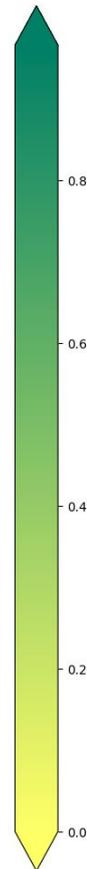
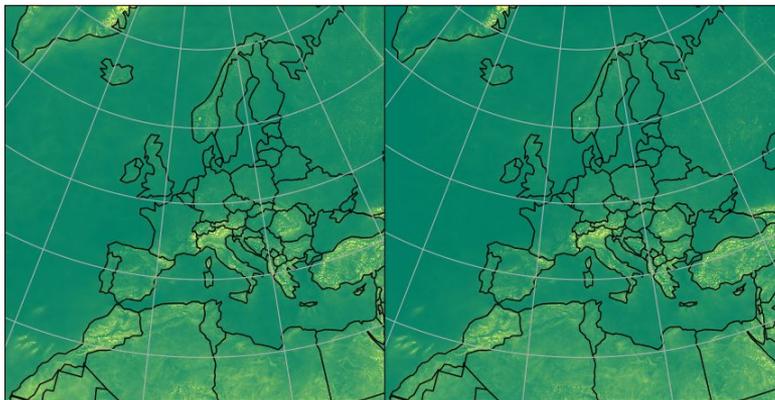
SGM

LAM

2m temperature



10m wind speed

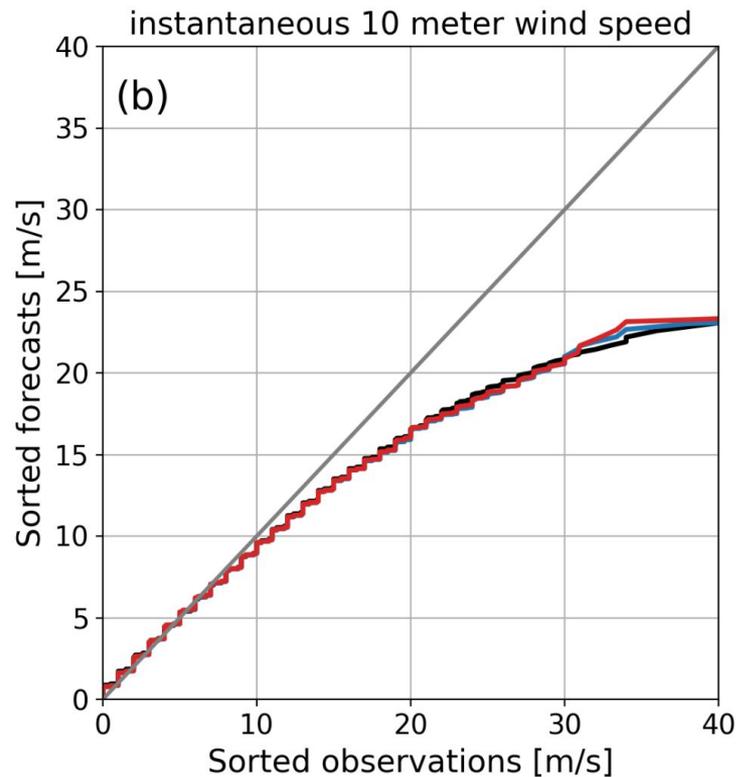
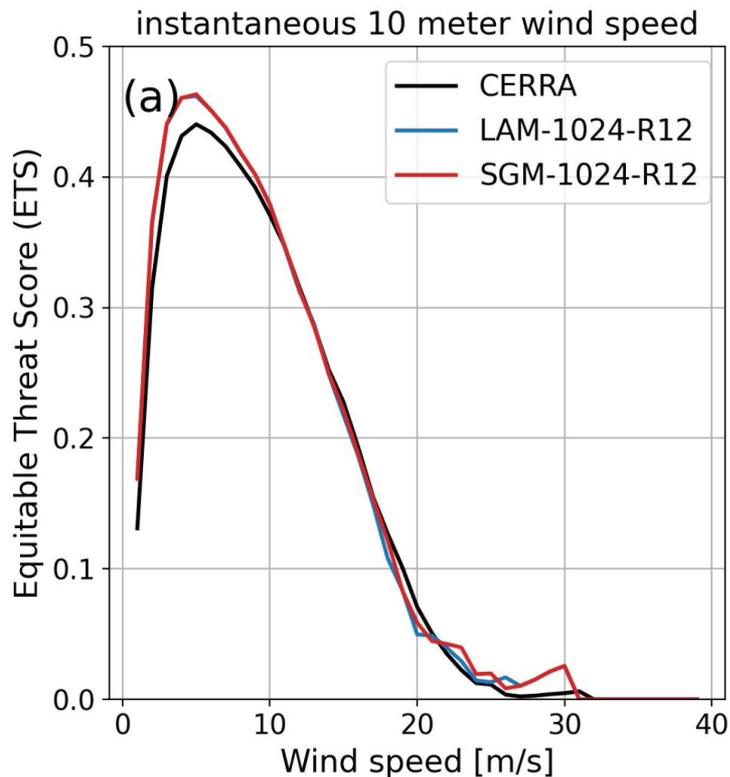


1024-R12-ideal
00+72 UTC

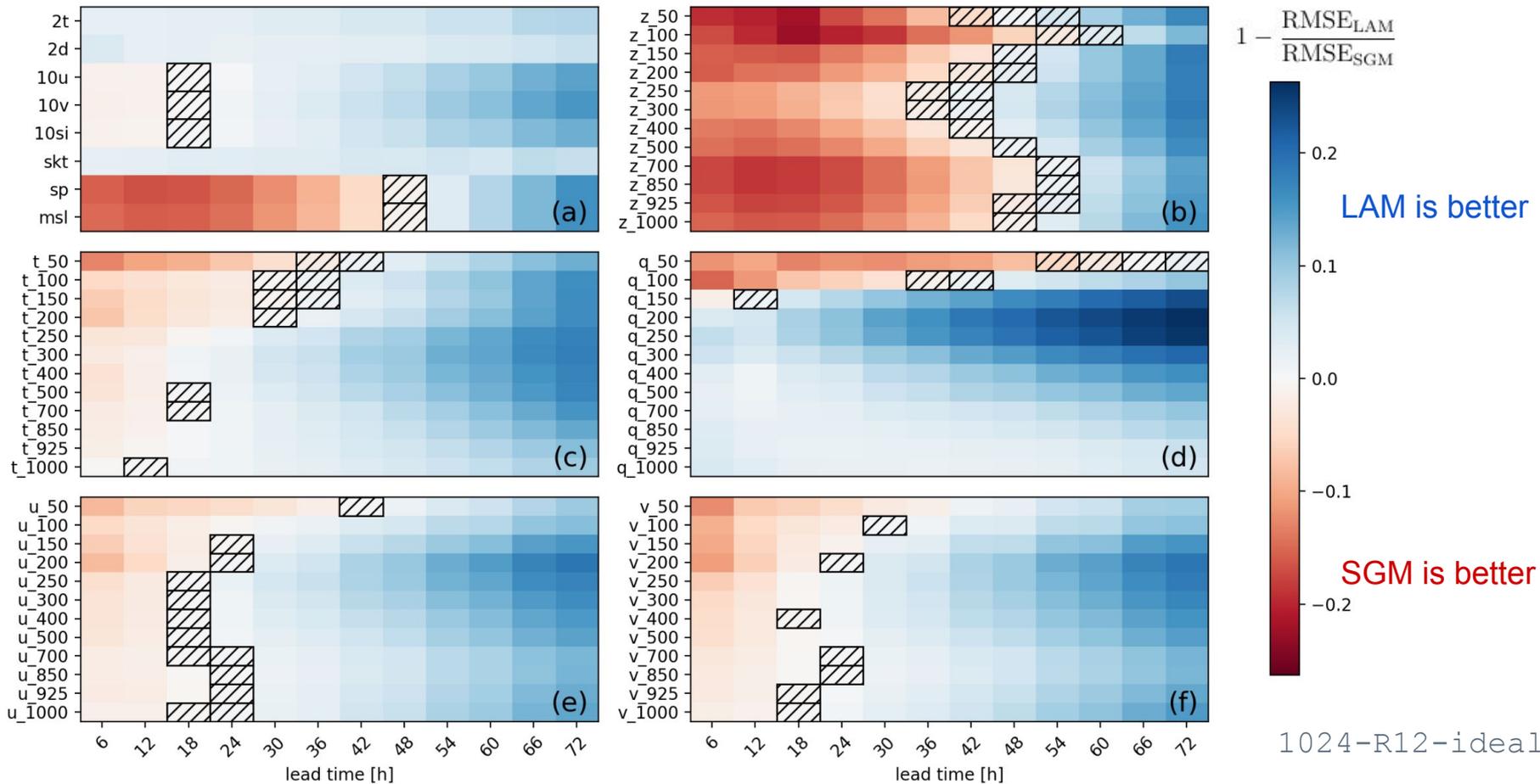
Skill vs climatology $1 - \frac{MSE_{\text{model}}}{MSE_{\text{clim}}}$

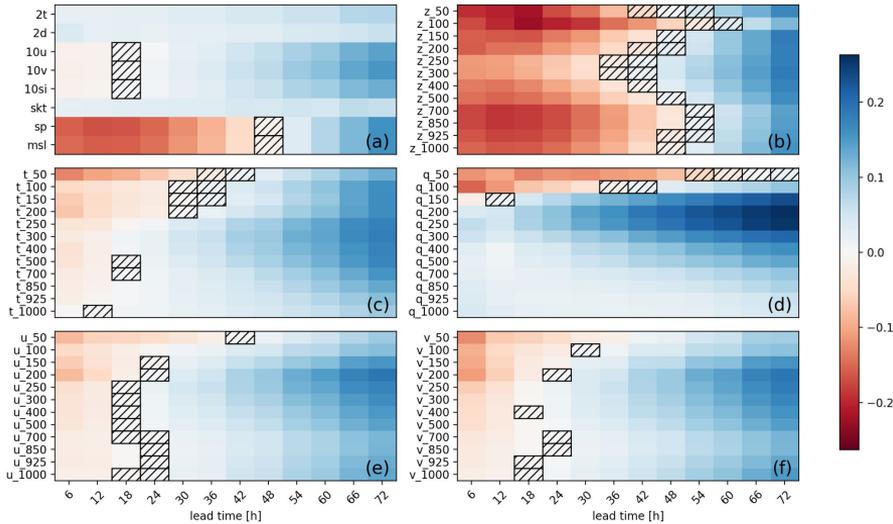
Both models show similar forecast performance:

- Good skill throughout domain
- Weak points coincide
→ physical
- E.g. both models struggle in mountainous areas with 10m wind

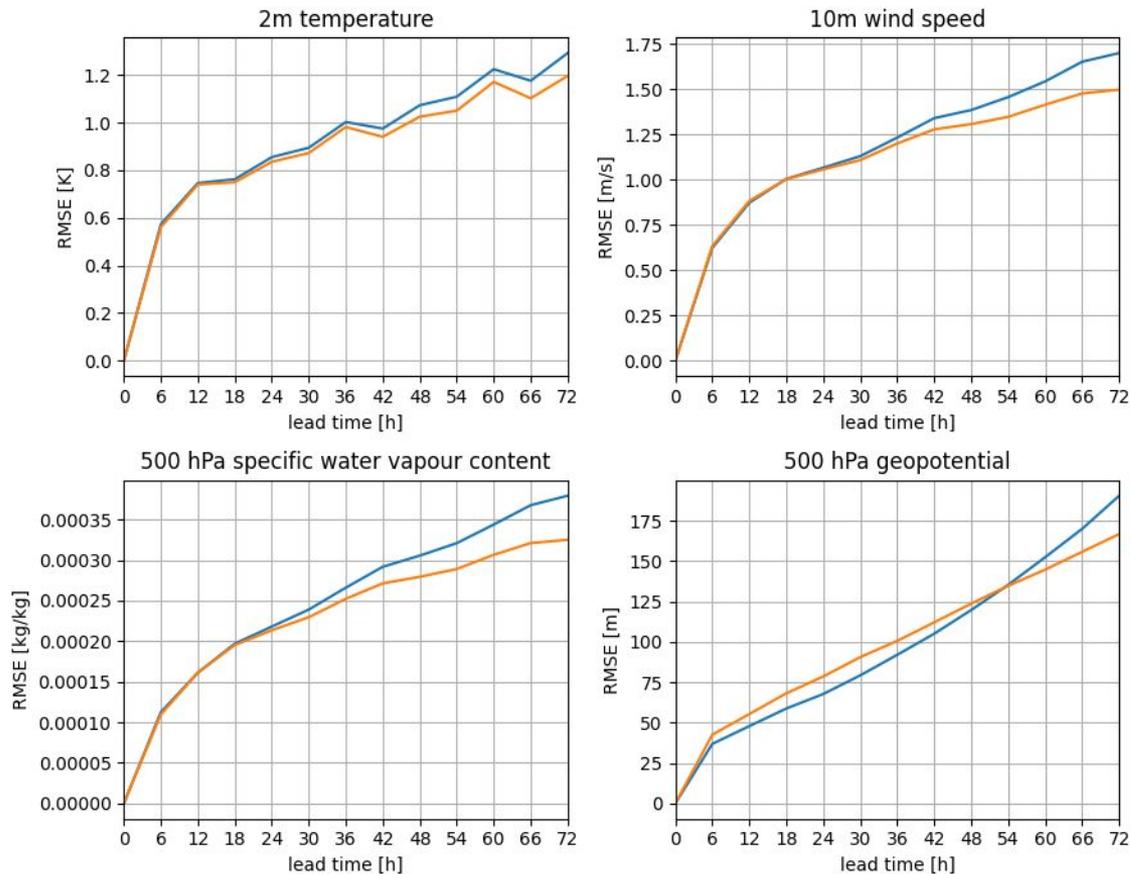


Models underestimate extremes, but so does the CERRA reanalysis.





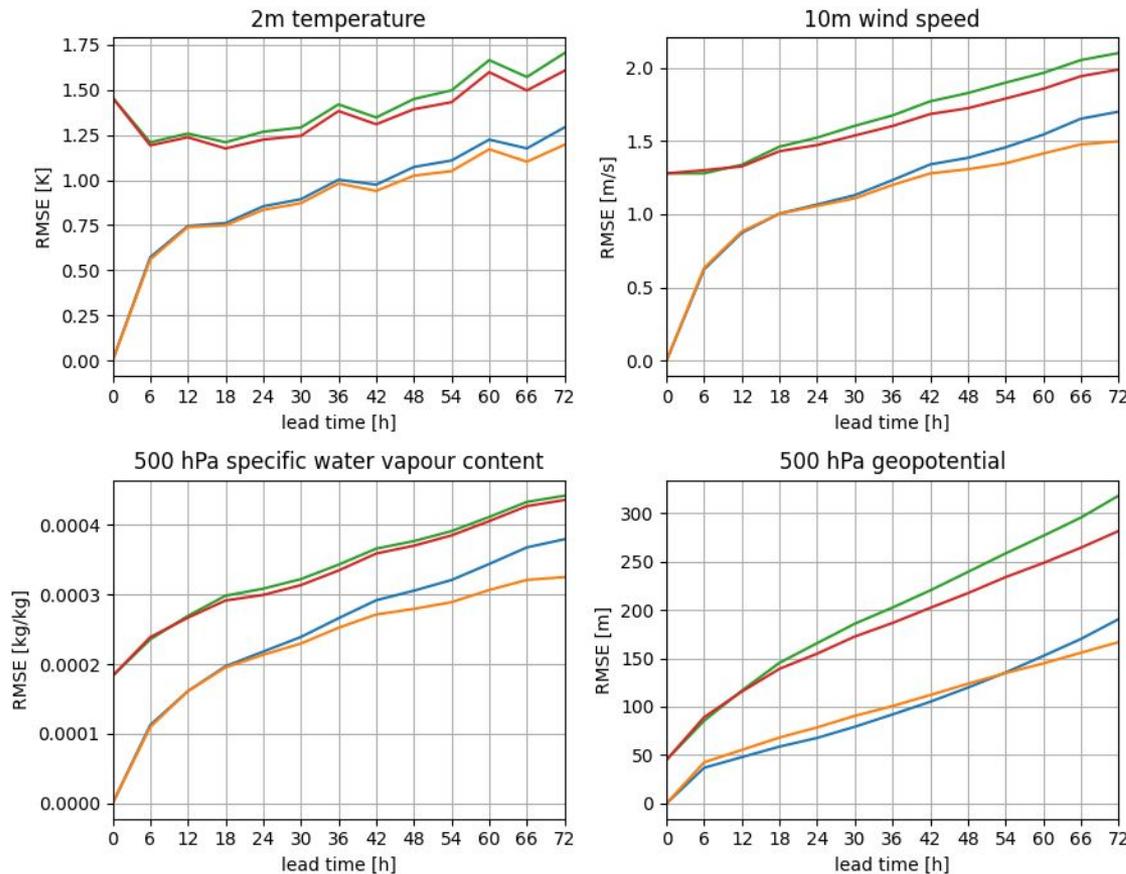
- At +6h SGM outperforms LAM for most variables, strongly so for the synoptic ones. LAM is marginally better for near surface temperatures and humidity.
- Skill of LAM wrt SGM decreases as one moves up in the atmosphere
- LAM takes over at later lead times (see next slides)



— SGM ideal
— LAM ideal

Ideal initialization:
Both models initialized on
CERRA + ERA5 reanalysis

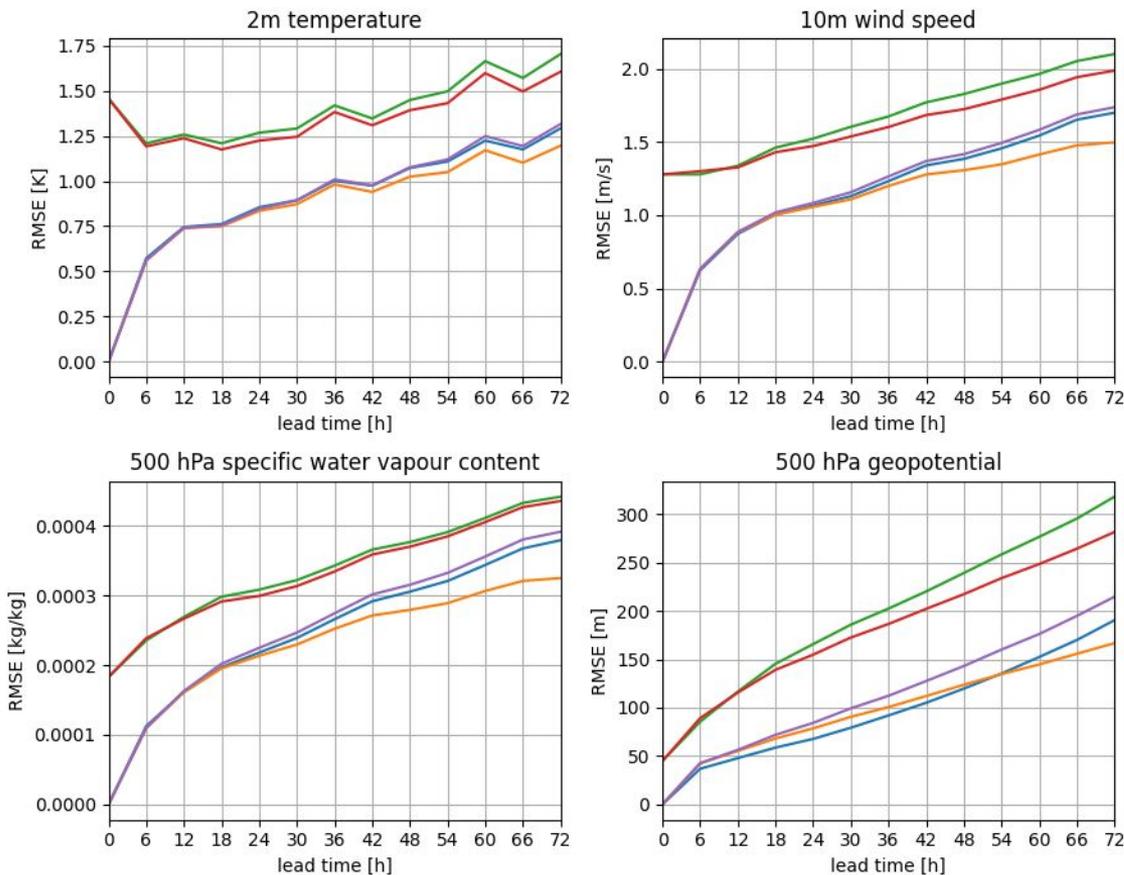
Ideal forcing:
LAM takes ERA5 boundary
forcings



- SGM ideal
- LAM ideal
- SGM oper
- LAM oper

‘operational’ initialization:
Both models initialized on
(regridded) IFS analysis

‘operational’ forcing:
LAM takes (regridded) IFS
forecasts as boundary
forcings

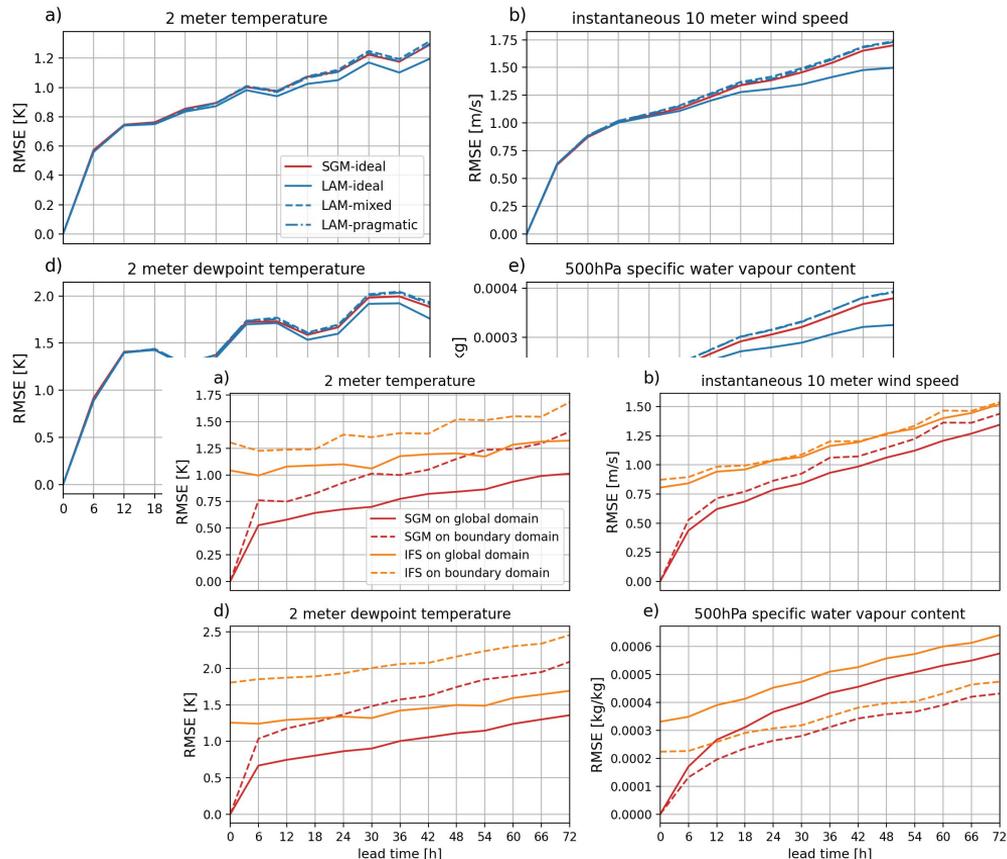


- SGM ideal
- LAM ideal
- SGM oper
- LAM oper
- LAM sgf

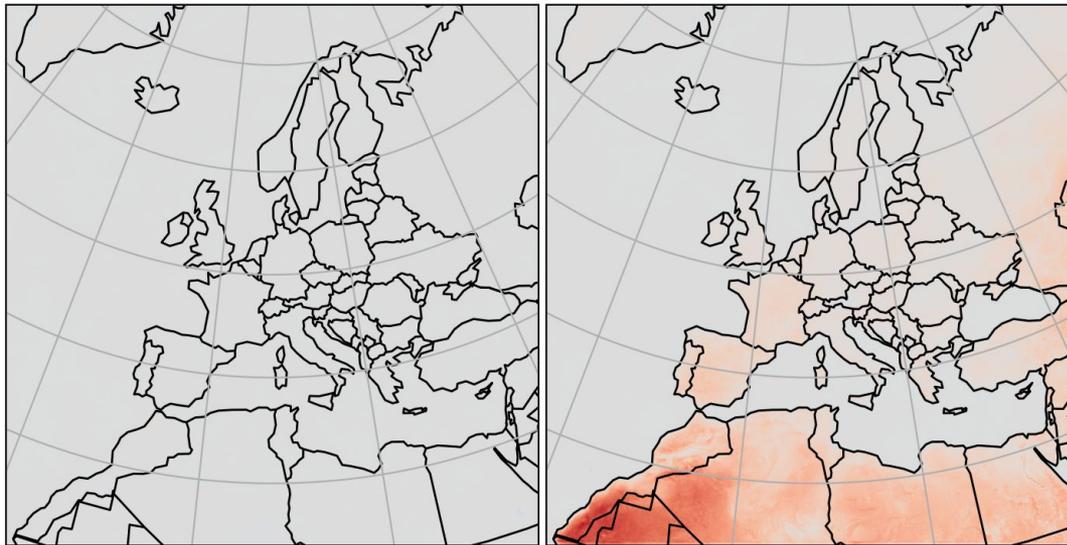
Neutral forcing:
LAM takes SGM forecasts as
boundary forcings

- Well adaptive to change in initialization/forcing
- Lam provides flexibility in choice of boundary forcing
- External forcing impacts error growth and can be stabilizing factor.

All of this was without fine-tuning!



2m temperature: difference in MSE [K^2] +6h



LAM is better

SGM is better

Known

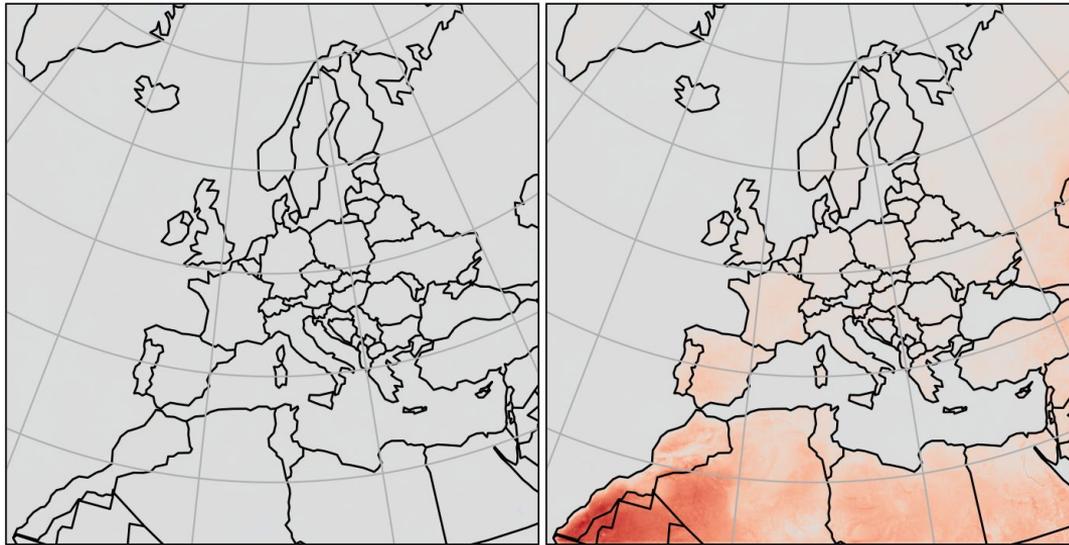
00, 06, 12, 18 UTC

Unknown

03, 09, 15, 21 UTC

At unknown times of day SGM significantly outperforms LAM over Northern Africa (NA)

2m temperature: difference in MSE [K²] +6h



LAM is better

SGM is better

Known

00, 06, 12, 18 UTC

Unknown

03, 09, 15, 21 UTC

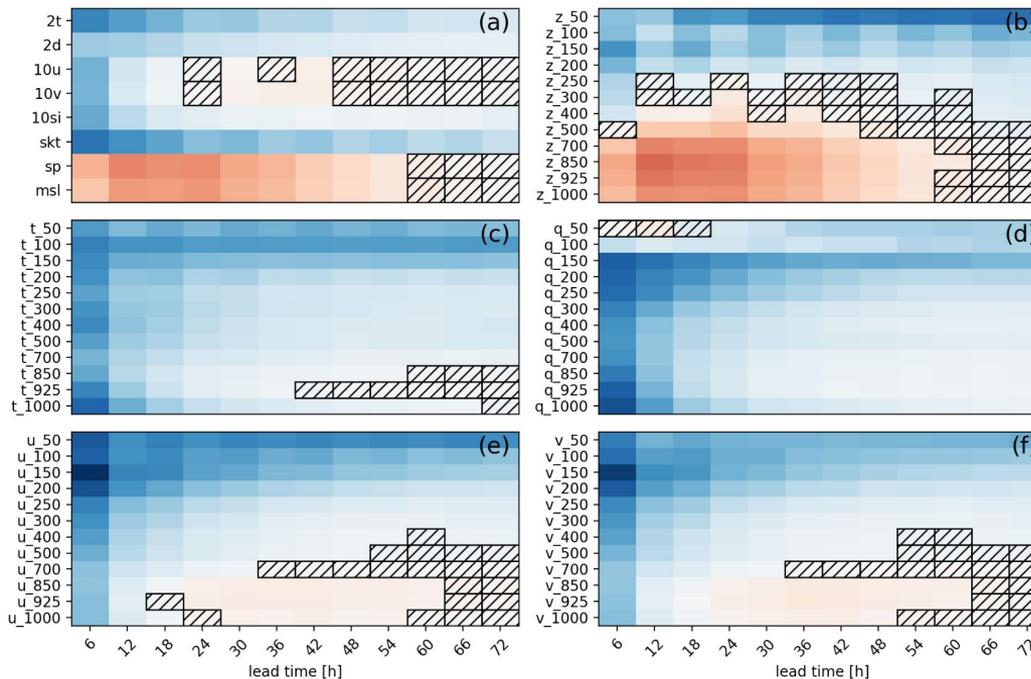
SGM has access to all time zones, LAM only to the CERRA time zones
 SGM seems capable of transporting physics of arid areas similar to NA learned in other geographic location/time zone to this one.

Small (512 channels)
vs
Large (1024 channels)

For some variables, like lower level geopotential height, smaller is better.

Model realizes it cannot learn the more complicated higher variance fields and goes for an easy win by focusing on larger scale variables?

SGM-R01-ideal



Larger is better

Smaller is better

SGM and LAM both have comparable and good forecast skill

Exhibit usual drawbacks of MLWP: underestimate extremes and activity.

- SGM benefits from global part in training data:
 - Better on synoptic scale variables
 - Better temporal generalizability
- LAM only needs to learn regional weather, allows it to have marginally better performance on some small scale variables like 2m temperature
- Forecast quality at later lead times depends strongly on quality of global forecast, LAM can leverage this.

Among others:

- Transfer learning to higher resolution
- Fine-tuning on operational data
- Higher forecast frequency
- Detailed comparison to regional NWP
- Probabilistic models
- Alternative approaches:
 - mixed resolution
 - future boundaries
 - downscaling

THANK

The Royal Meteorological
Institute

Het Koninklijk
Meteorologisch Instituut

L'Institut Royal
Météorologique

Das Königliche
Meteorologische Institut



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.