Met Office

The Moonshot Project

Tom Dunstan[1], Tom Dodds[1], Hannah Brown[1], Aled Owen[1], Sophia Moreton[1], Karina Williams[1], Cyril Morcrette[2]

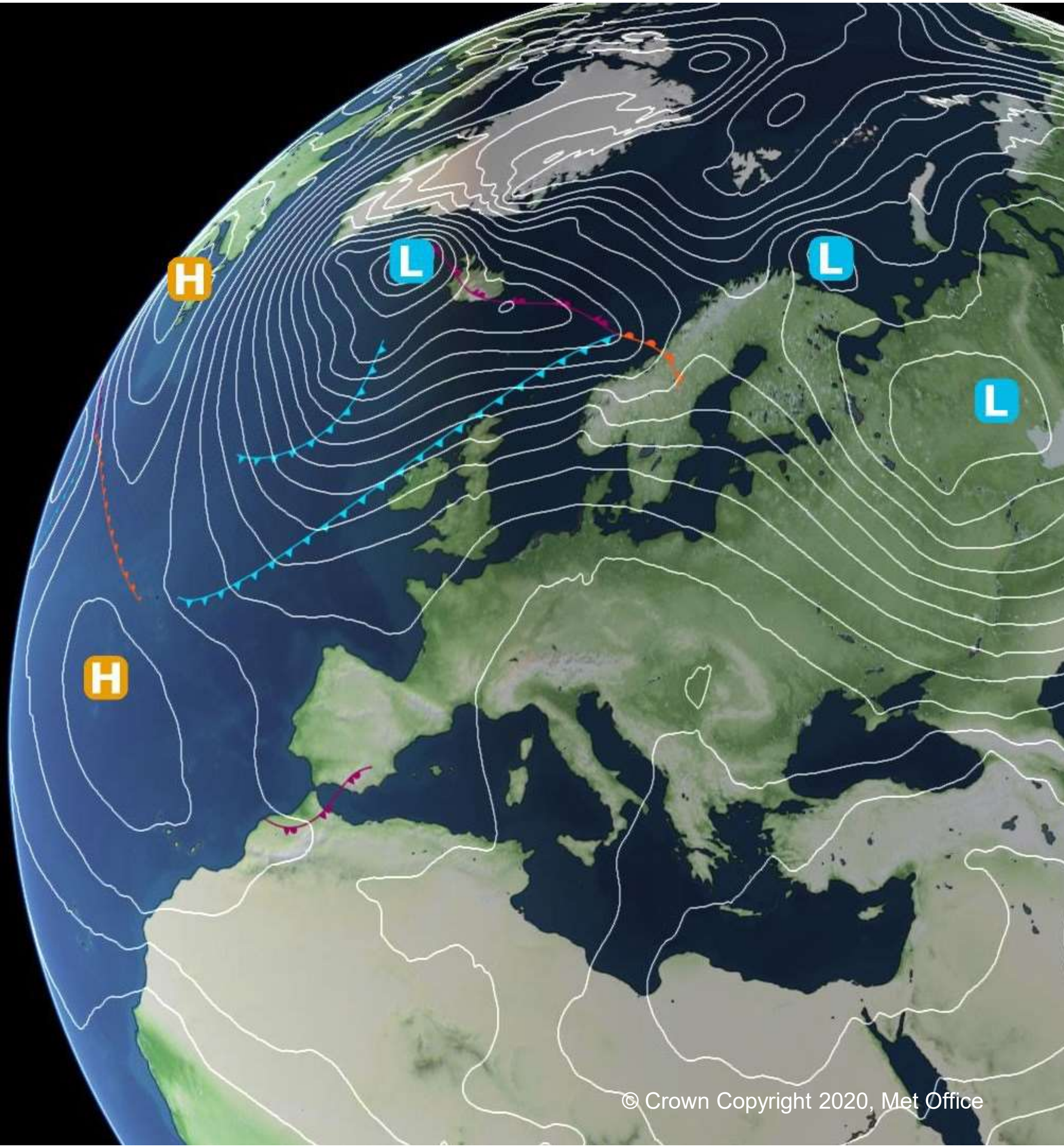
Bernat Puig-Camps[3], George Williams[3], Clodagh Lynch[3], Luke Vinton[3]

Large-scale deep learning for the earth system,
University of Bonn, 4-5 September 2023

1 Informatics Lab, Met Office

2 Atmospheric Processes and Parametrizations, Met Office

3 Microsoft Data & AI Service Line



Moonshot timeline

Rasp et al. Weatherbench, https://doi.org/10.1029/2020MS002203

Dueben and Bauer (2018), Geosci. Model Dev. https://doi.org/10.5194/gmd-11-3999-2018

'Machine learning for weather and climate modelling' *Phil Trans Royal Soc A, (2021)*

2022

Pathak et al. FourCastNet, Feb 2022

Keisler, Feb 2022

Phase 0
Project started April 2022 following a 'horizon scan' late 2021

Bi et al., Pangu Weather, Nov, 2022

Lam et al. GraphCast Dec 2022

2023

Nguyen et al. Climax, January 2023

Phase 1
Microsoft Collaboration March - July

Chen et al. FengWu April 2023

Chen et al. FuXi, June 2023

Phase 2
AI4NWP
July -

Met Office

Moonshot Vision

The key enabler for this amazing progress is ERA5, but…

- X It is expensive and slow to produce; long improvement cycle
- X Many important processes / scales are not well represented (as with any global DA)
- X Single-source training means biases are inherited

How to go beyond reanalysis data?

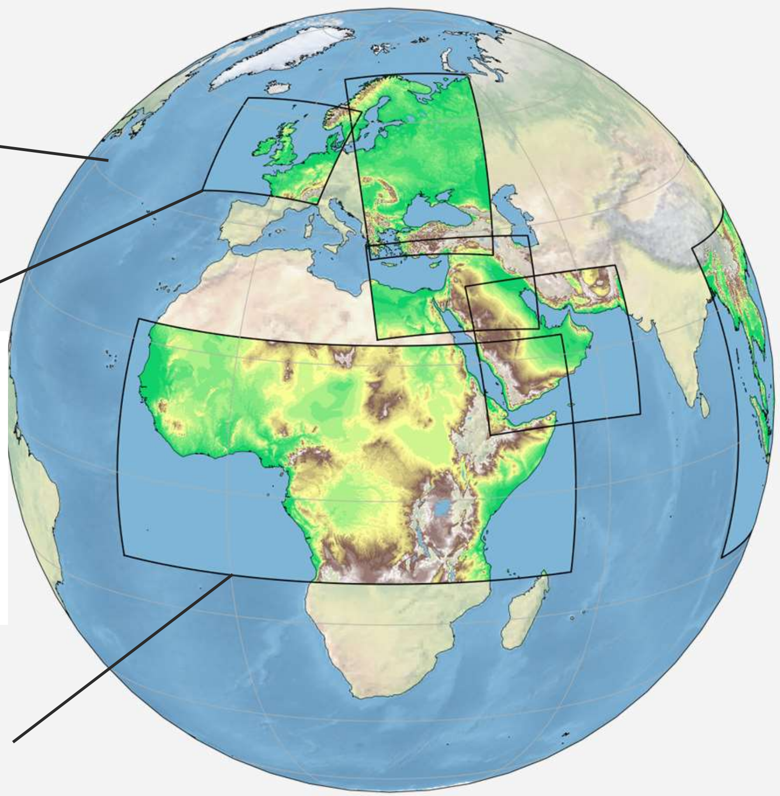
- Downscale to incorporate high-res data – CNNs, GANs, VAEs, Diffusion models all show promise.
- No upscale energy transfer. How important?
- Input diversity: will multi-model, multi-mode training data reduce bias?

Operational model archives

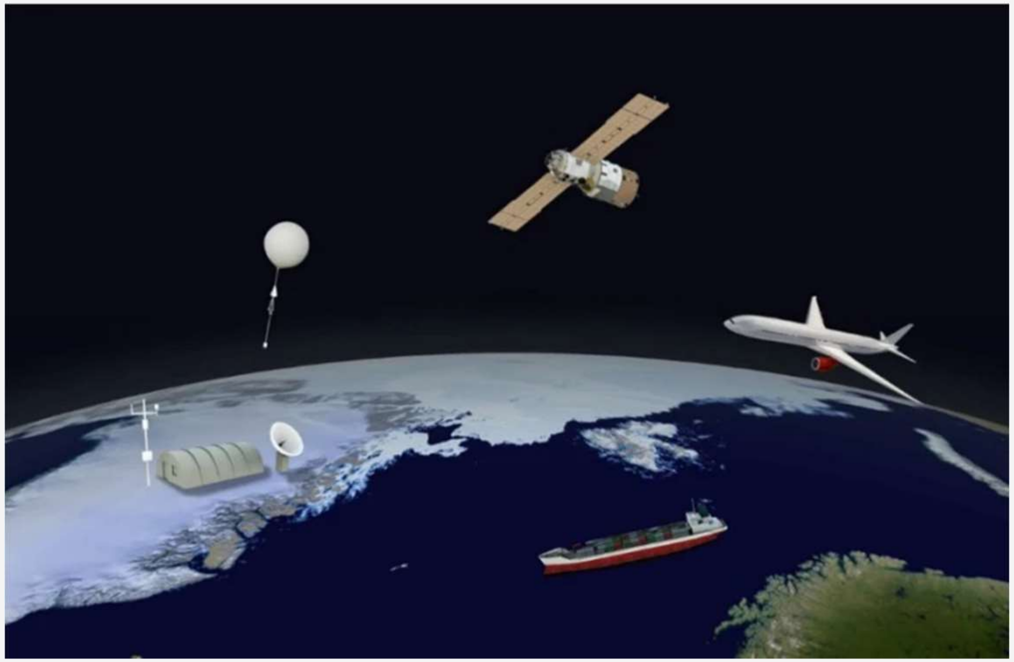
Global NWP:
- 10–20km atmosphere/land
- 70 vertical levels (80km top)
- Coupled hourly to 25km ocean/sea ice
- Hybrid 4DVar/En-4DEnVar Data Assimilation (DA)
- Forecasts currently out to ~8 days

UK NWP:
- 1.5–2.2km atmosphere/land
- 70 vertical levels (40km top)
- Sea temperature from separate 1.5km ocean forecast
- Hourly 4DVar DA
- Forecasts hourly out to 5 days

Other Models:
- Run in support of Defence/International Development



Observation archives



- Geostationary/polar orbiting satellites.
- GNSS radio occultation.
- Direct/indirect aircraft observations.
- Balloon launched radiosondes.
- Radar data.
- Ground (and sat) based LIDAR.
- Surface stations, ships, buoys etc.

Can we unlock the information in our data archives?

Shorter, possibly non-overlapping time spans compared to reanalysis

Discontinuities due to upgrades in the DA system and/or forecast models

Obs data are sparse, irregular, often missing.

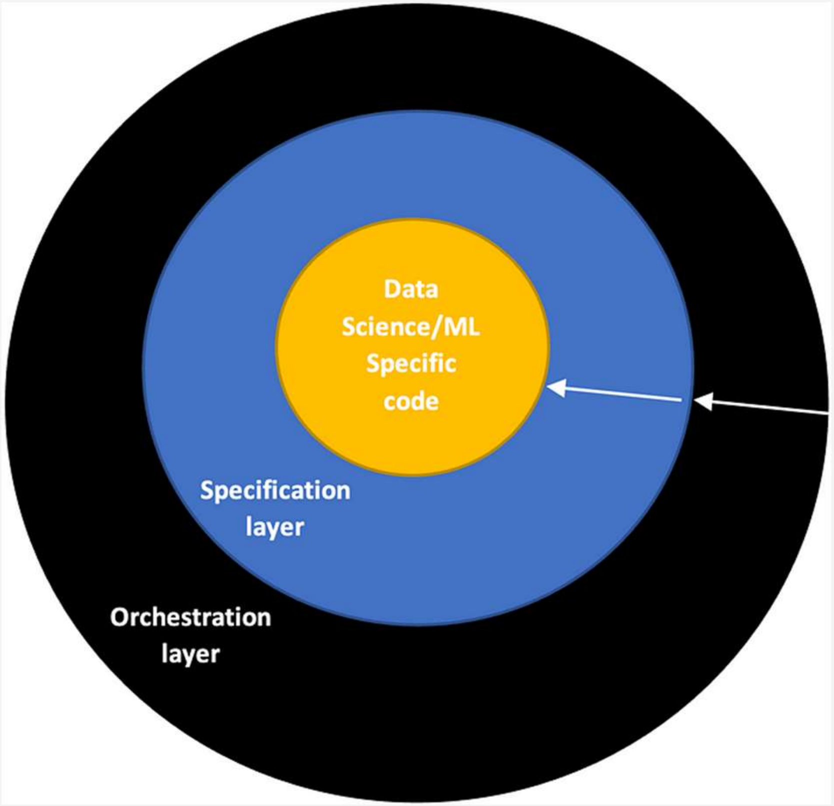
Raw data volumes are prohibitively large

The rest of this talk

- Baseline model & results
- Lines of research for Phase 2
- The AI4NWP Programme

Infrastructure

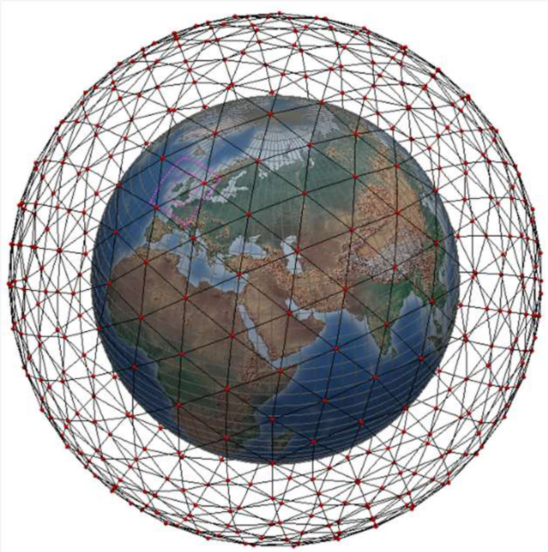
- Working with Microsoft colleagues to produce efficient ML workflows at-scale using AzureML
- ‘Scaffolding’ approach used to ensure scalable, reproducible, and traceable workflows
- Scientific progress slowed while we adopted this way of working, but we expect it to pay dividends later.

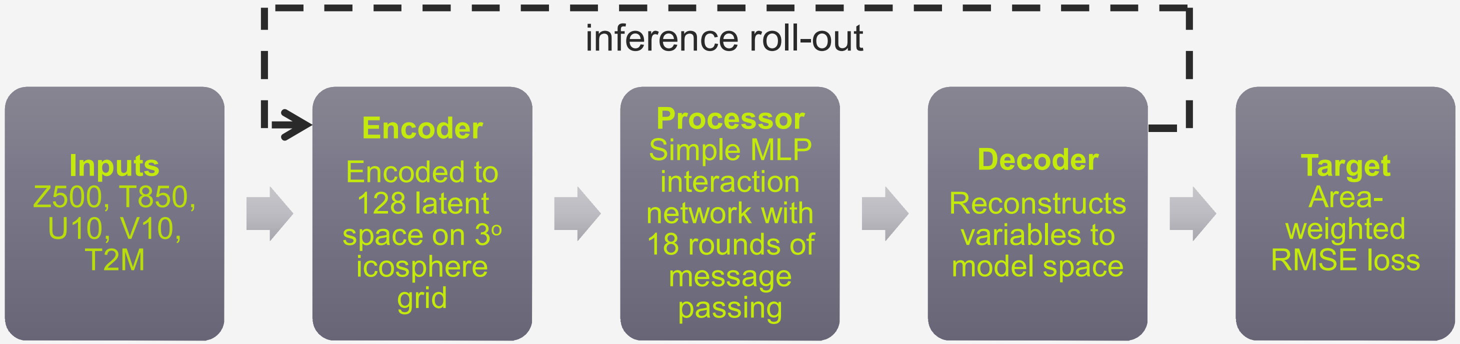


“A layered approach to MLOps”, Bernat Puig-Camps (https://medium.com/data-science-at-microsoft/a-layered-approach-to-mlops-d935beefca2e)

Baseline Model

- GNN based model (Pfaff 2020, Keisler 2022, Lam 2023)
- Minimal viable model for fast dev cycle and low costs
- Single-level 3^{o} icosphere mesh, 5 variables per lat/lon
- Trained over 50 epochs of ERA5 1979-2017 at 1^{o}, verification on 2018
- Pytorch-Accelerated with mixed-precision and distrubuted data parallelism

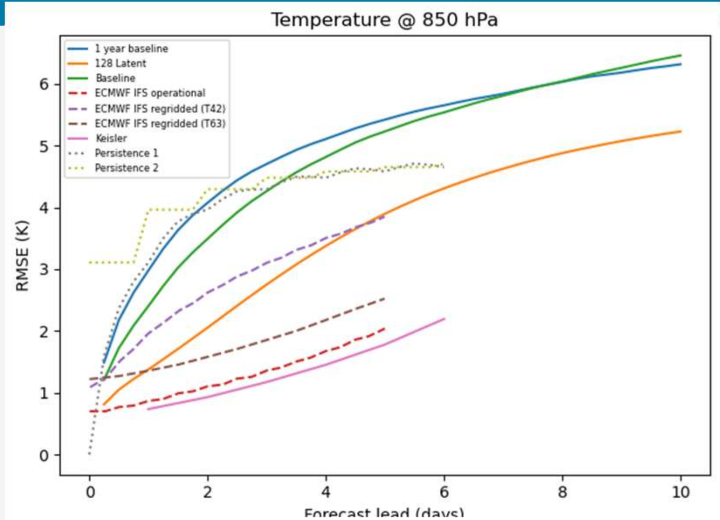


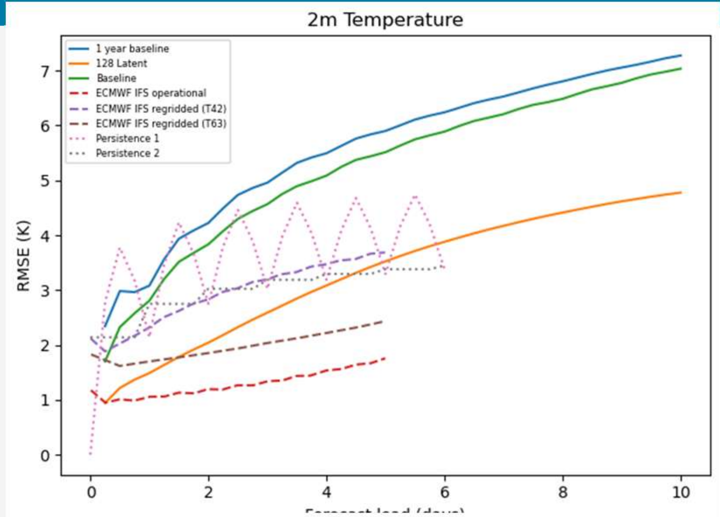


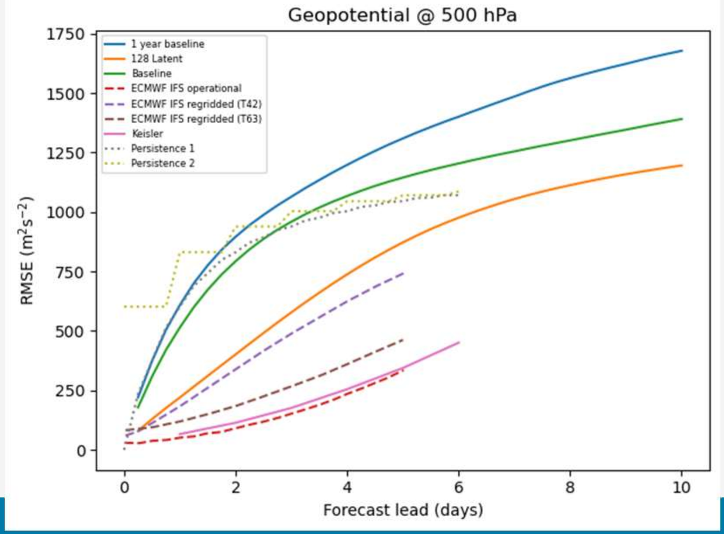
Baseline Results

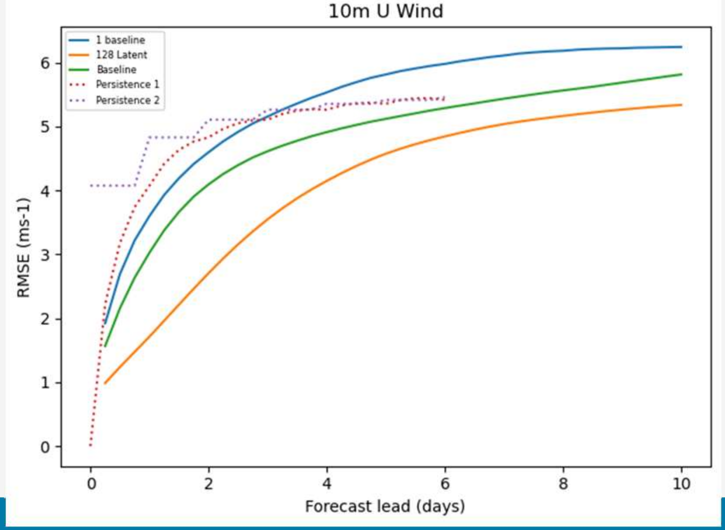
- Simple model but performing well.

- Tests with increased number of input levels and increased latent space show large improvements in performance and no sign of over-fitting.

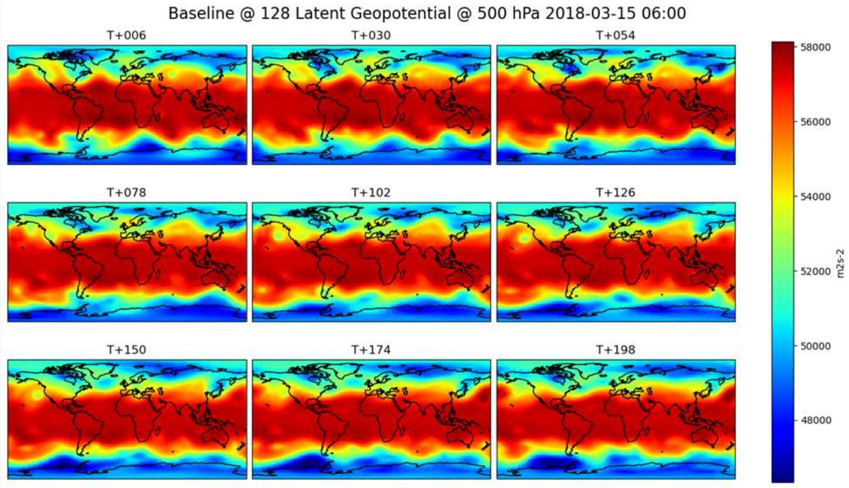


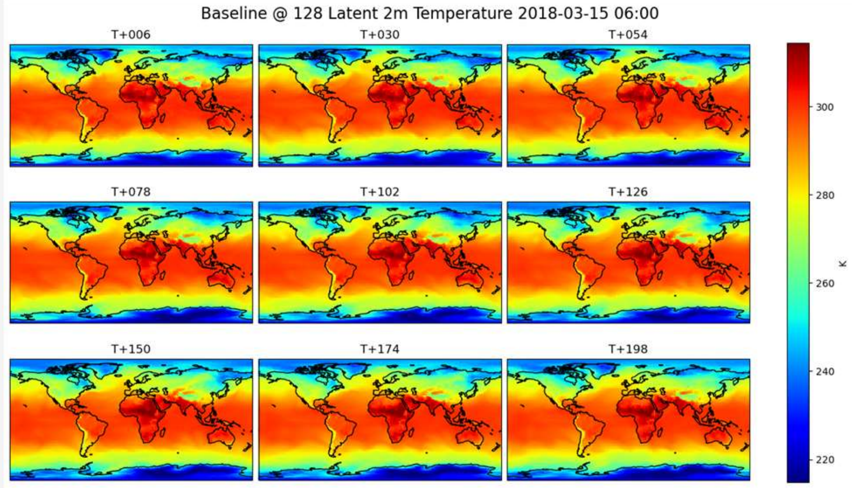


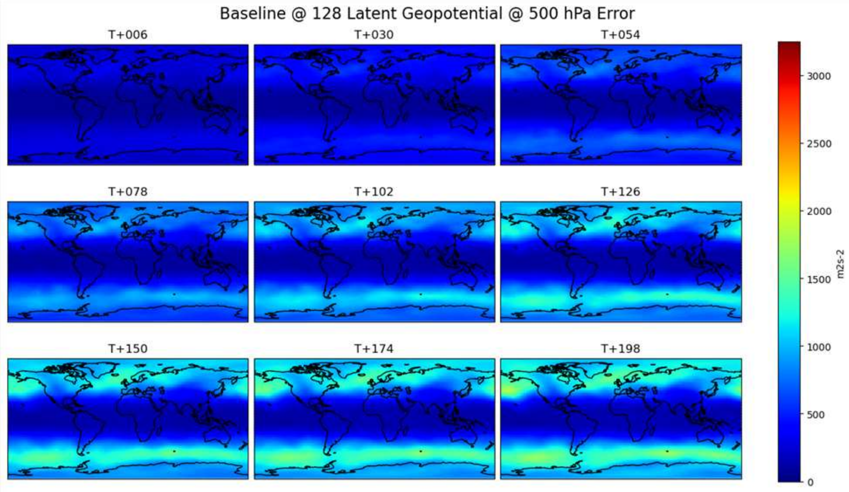


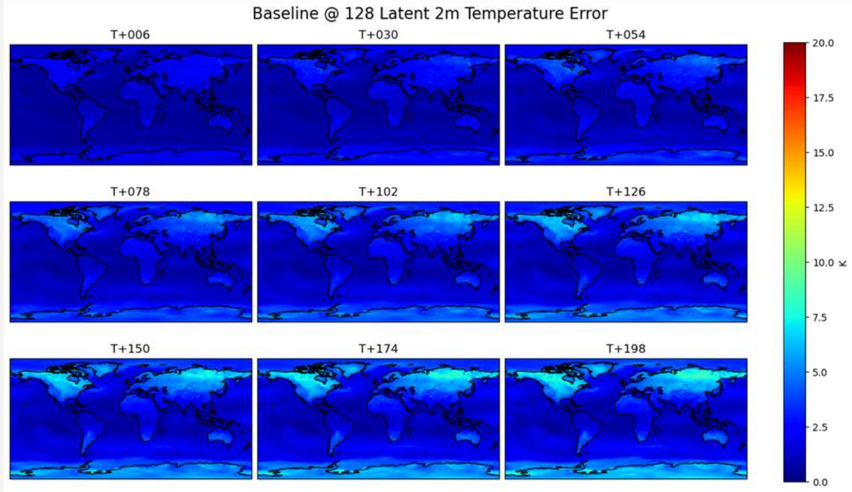


Baseline Results









What next?

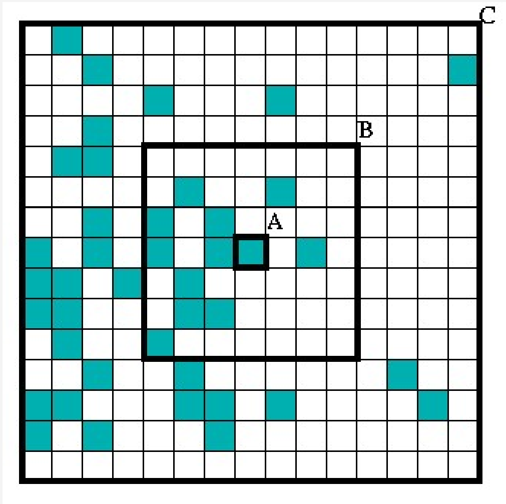
Integrating UKV data

- Δx = 1.5km, 70 levels
- Data volumes per sample are ~50% of ERA5 at 0.25°
- Only ~5 years of UKV data, so pre-train on ERA5 then fine-tune
- Loss function must be adapted for high-res / convection permitting region.

Novel loss functions

Value of high-res models comes mainly from near-surface variables, especially precip. This will not be captured by standard loss functions (RMSE)

- Blurring needs to be better understood, time-step may also be important e.g. Smith et al. https://arxiv.org/abs/2305.00100
- Scale splitting e.g Annau et al. https://arxiv.org/abs/2302.08720
  - Large scale: content loss (RMSE)
  - Small scale: probabilistic loss



- Area-based loss functions using thresholds (FSS, SEEPS) can penalise blurring.
- Area size could also be a function of forecast period /lead time to accommodate error growth

Met Office

Data reduction and feature engineering

Data reduction, i.e. encoding, compression, dimension reduction is fundamental to ML-based forecasting

Can we take some of this outside the forecast model in a pre-processing step? For large-scale multi-source training this will be essential

Potential Vorticity (PV):

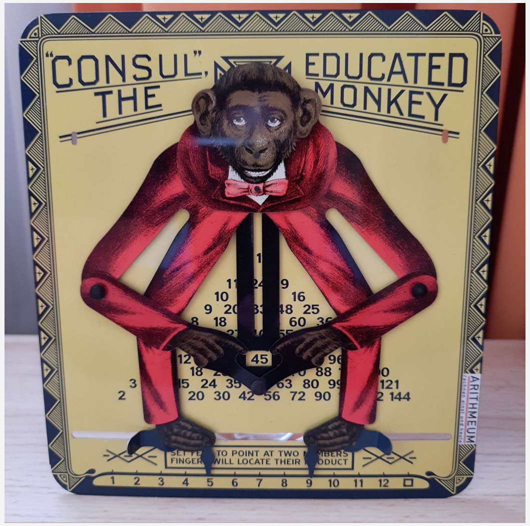
- Single scalar that encapsulates flow and thermodynamic information
- Conserved on isosurfaces of theta
- Analytically inverting PV to get state variables is challenging

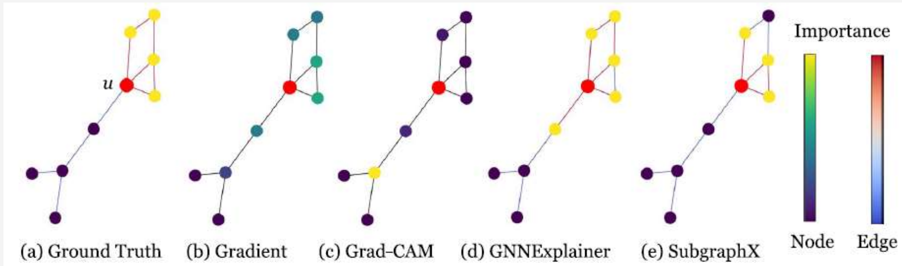
Auto-encoding:

- Single model columns or small clusters of model columns to reduce input dimensionality.
- Latent space not optimised for dynamics

Explainable AI

- Trust:
  - Building trust in AI modelling both internally and externally
- Discovery:
  - This is a completely new way of modelling dynamical systems, so what can we learn?
- Optimisation:
  - Select regions to gain the most benefit from high-res input data or obs, possibly as a function of lead time.





From Agarwal et al. 2023, Evaluating explainability for graph neural networks, Sci Data 10, 144 (2023). https://doi.org/10.1038/s41597-023-01974-x

AI for Numerical Weather Prediction (AI4NWP) Programme

An end-to-end ML system for global and regional forecasting