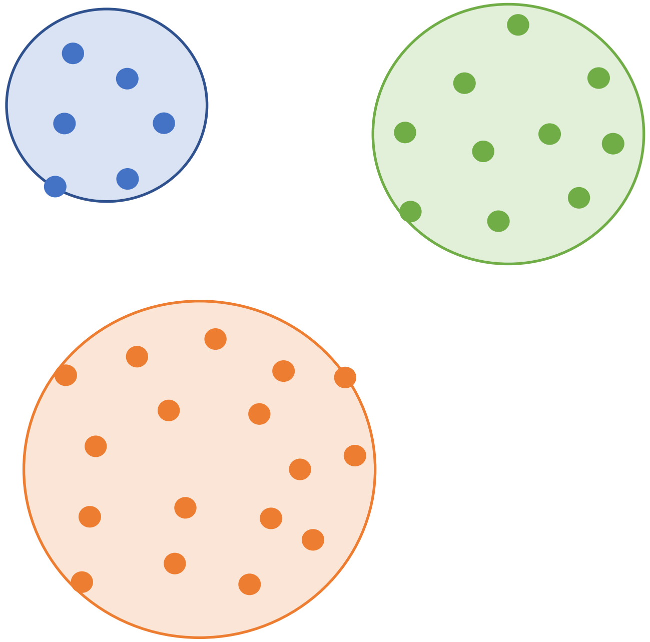
Post-processing with Machine Learning

Predictability Training Course

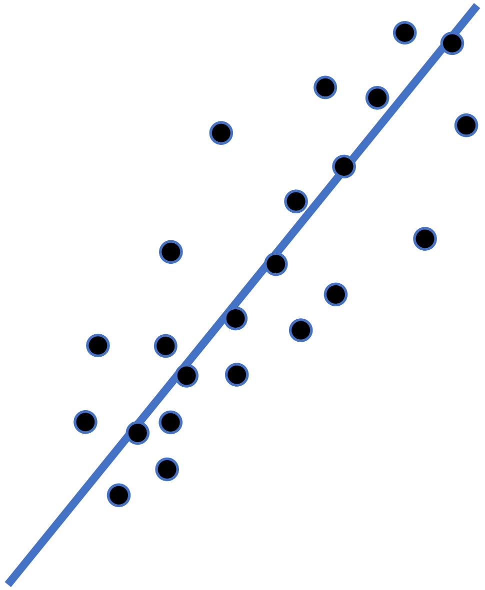
Mariana Clare

with thanks to Zied Ben Bouallègue, Matthew Chantry and Martin Leutbecher

Unsupervised Learning Methods	Supervised Learning Methods	Reinforcement Learning



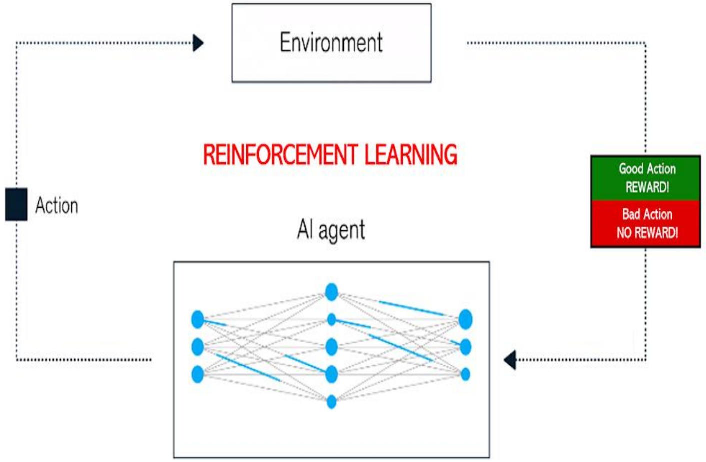
Unlabelled data
provide input data



Labelled data
provide input and target data

Training data | Test data

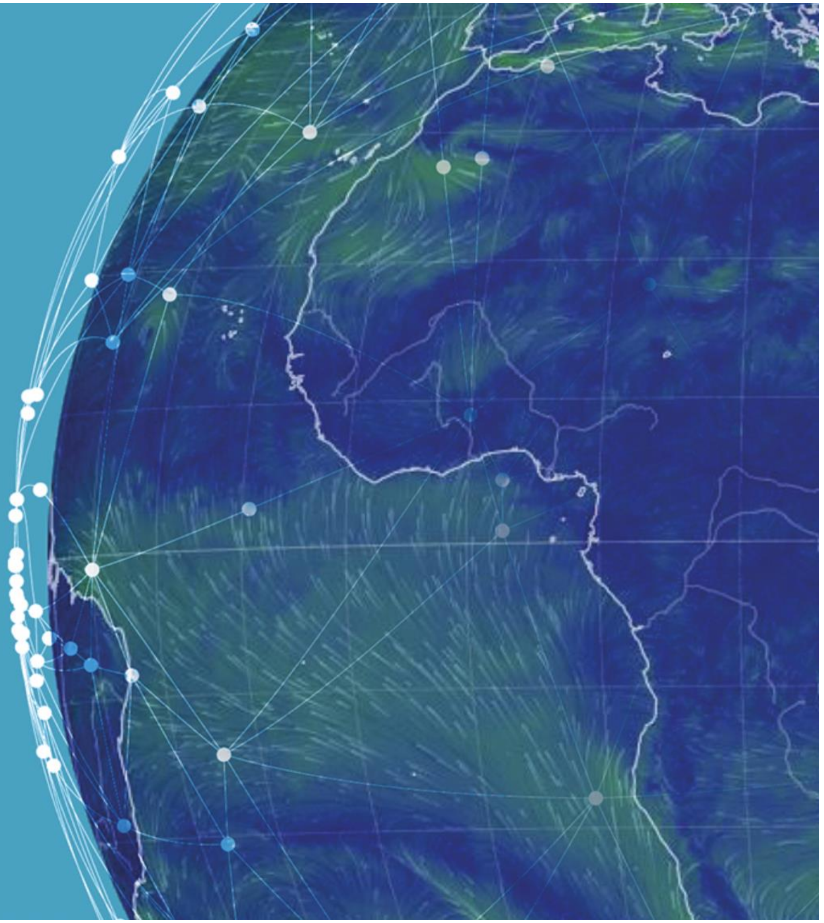
Total data



Learns from mistakes

MOOC Machine Learning in Weather & Climate

https://lms.ecmwf.int/pages/index.html



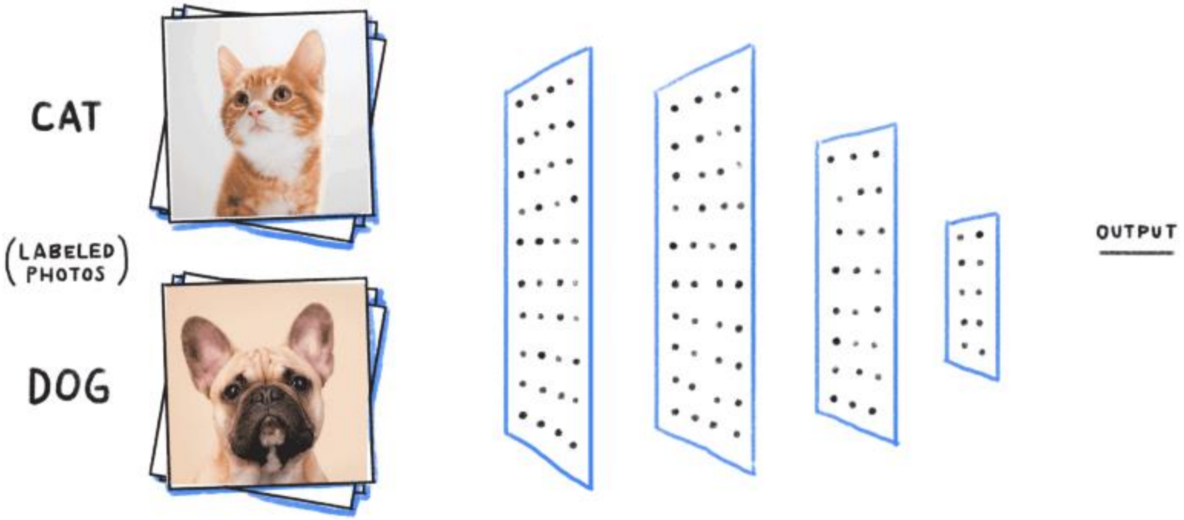
Neural Networks

Simple Neural Network

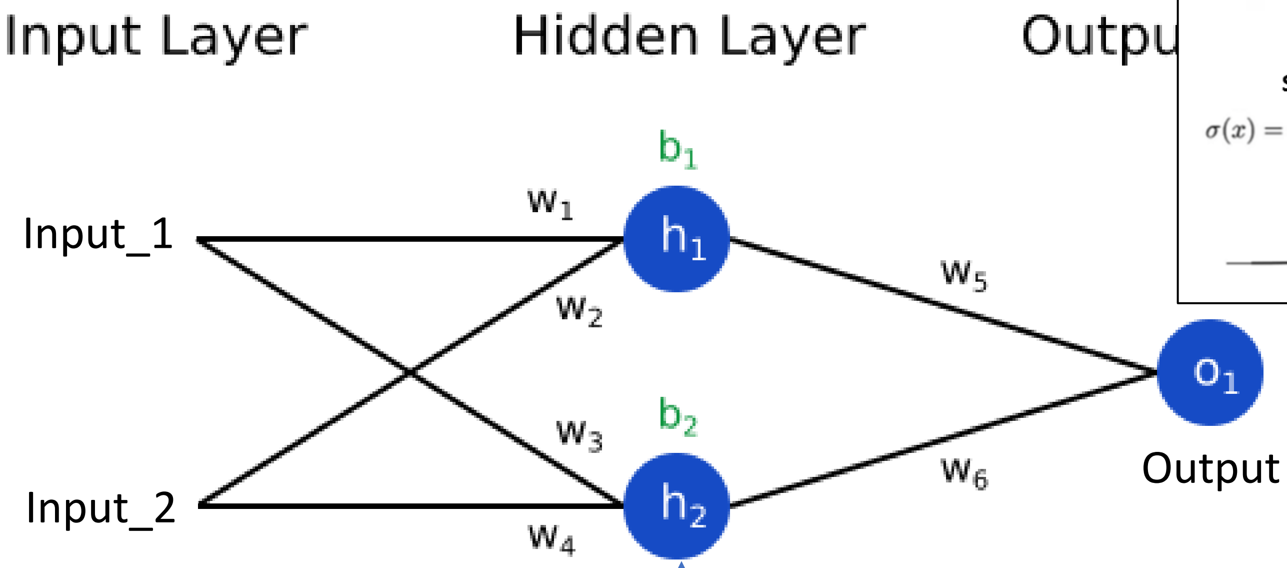
Training stage

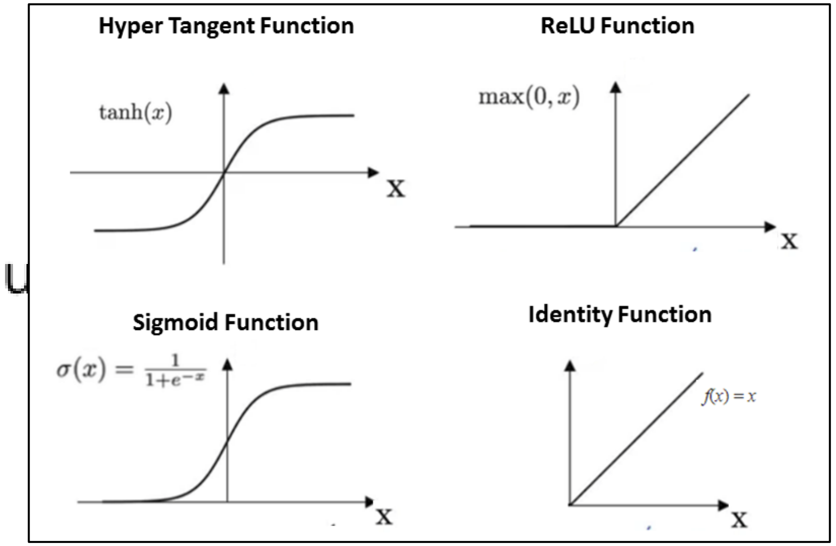
Training Data → Neural Network → Output Targets

Prediction stage



Simple Neural Network



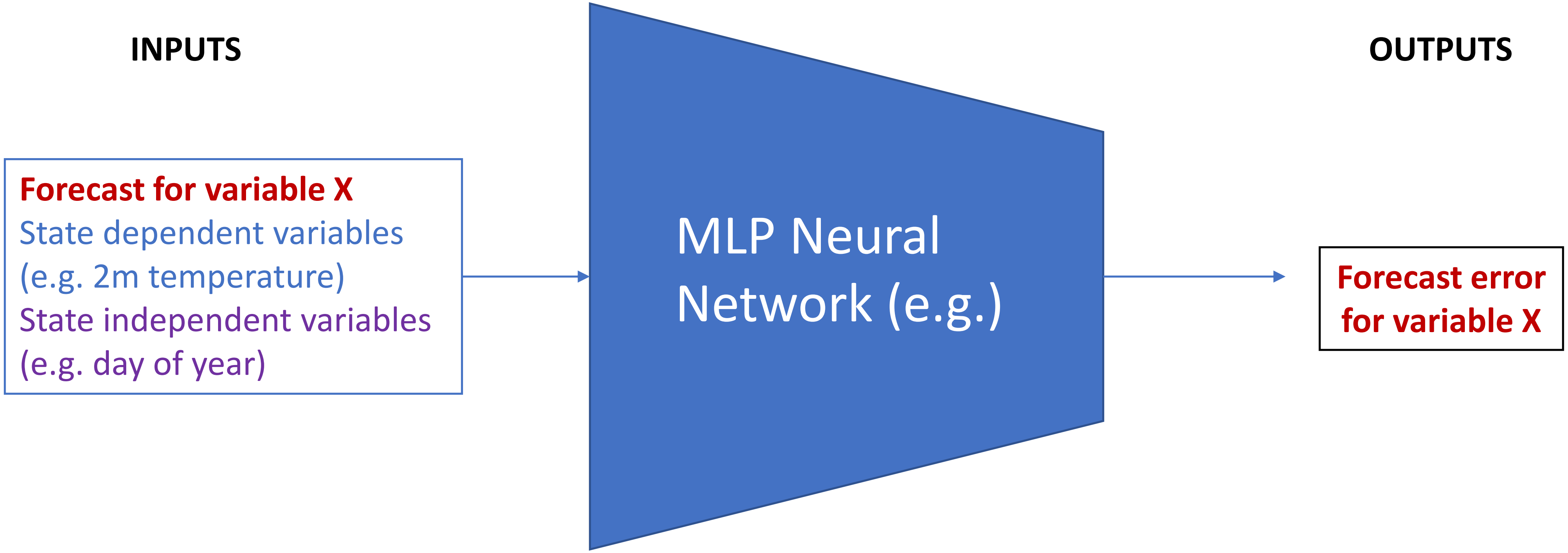


Activation function
$f(w_3x_3+w_4x_4)$

Train weights and biases to minimize a loss function like $\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_{true} - y_{pred})^2$

Methodology

Suppose post-processing variable X



Useful set-up for post-processing a deterministic forecast to a deterministic forecast

Bouallègue, Z. B., Cooper, F., Chantry, M., Düben, P., Bechtold, P., & Sandu, I. (2023). Statistical Modeling of 2-m Temperature and 10-m Wind Speed Forecast Errors. Monthly Weather Review, 151(4), 897-911.

Generative methods and downscaling

Downscaling

MSE often penalises a NN for making bold predictions, resulting in "blurry images".

This is not appropriate for downscaling and therefore different approaches must be considered.

Generative Methods take a probabilistic approach and sample small scale uncertainty

Input

Low-Resolution Model Data

e.g. ERA5 at 0.25^{o}

Generative Neural Network

Target

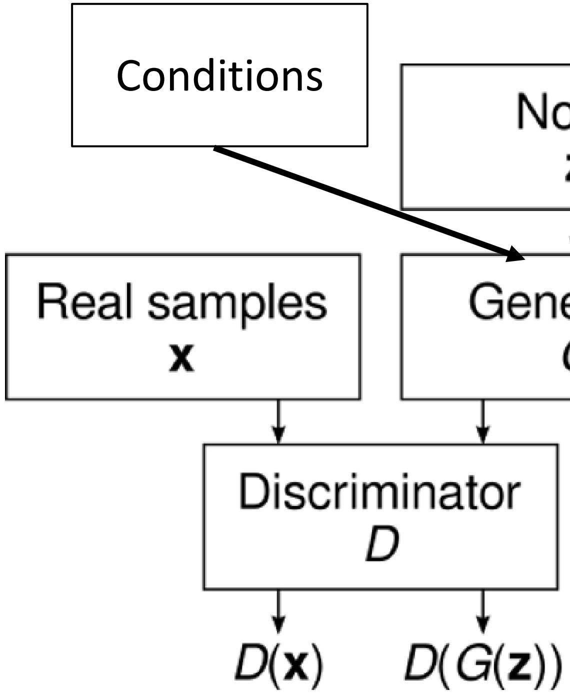
High-Resolution Model/Observation Data

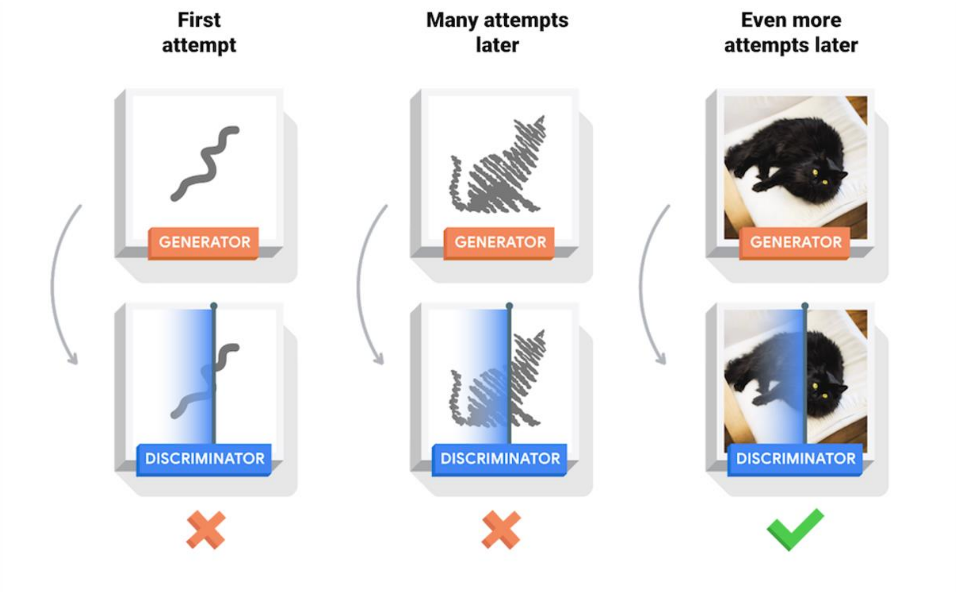
e.g. NIMROD 1km Radar Data

Generative Adversarial Networks (GANs)

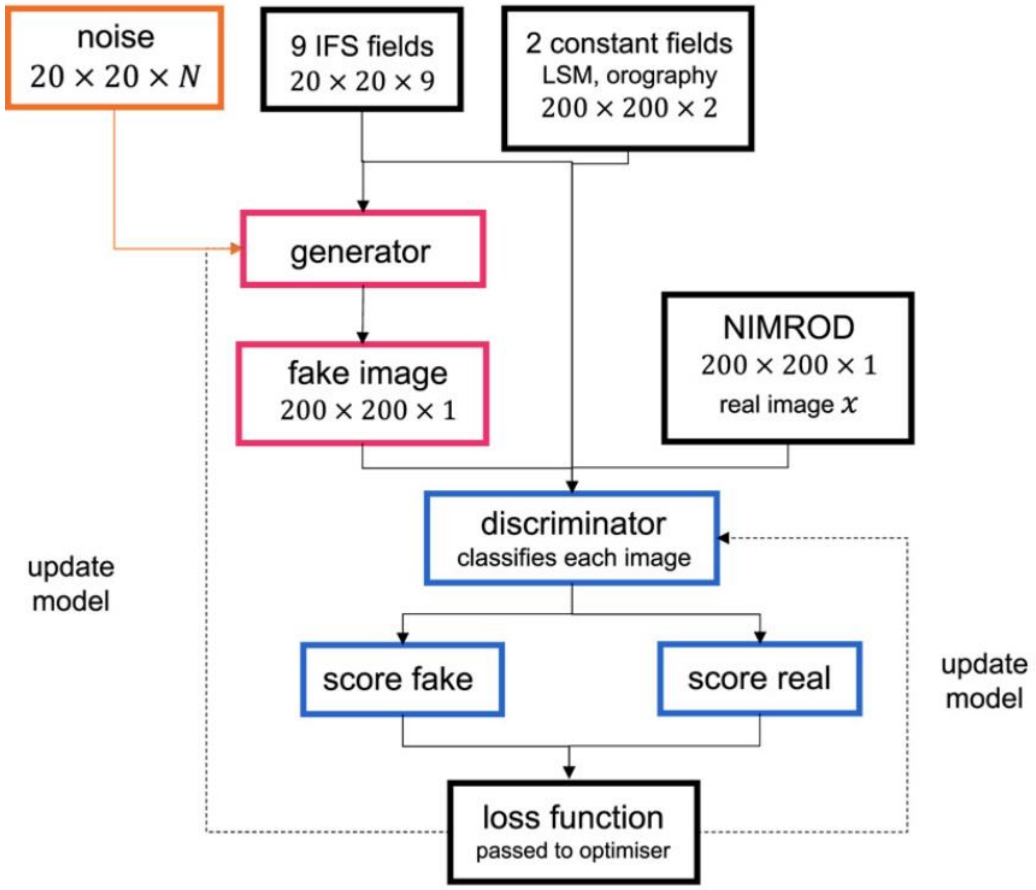
GANs consist of two neural networks competing against each other:
a Generator (learns to generates realistic data) and a Discriminator (learns to distinguish fake data from real data). Often use a conditional GAN which has inputs other than noise

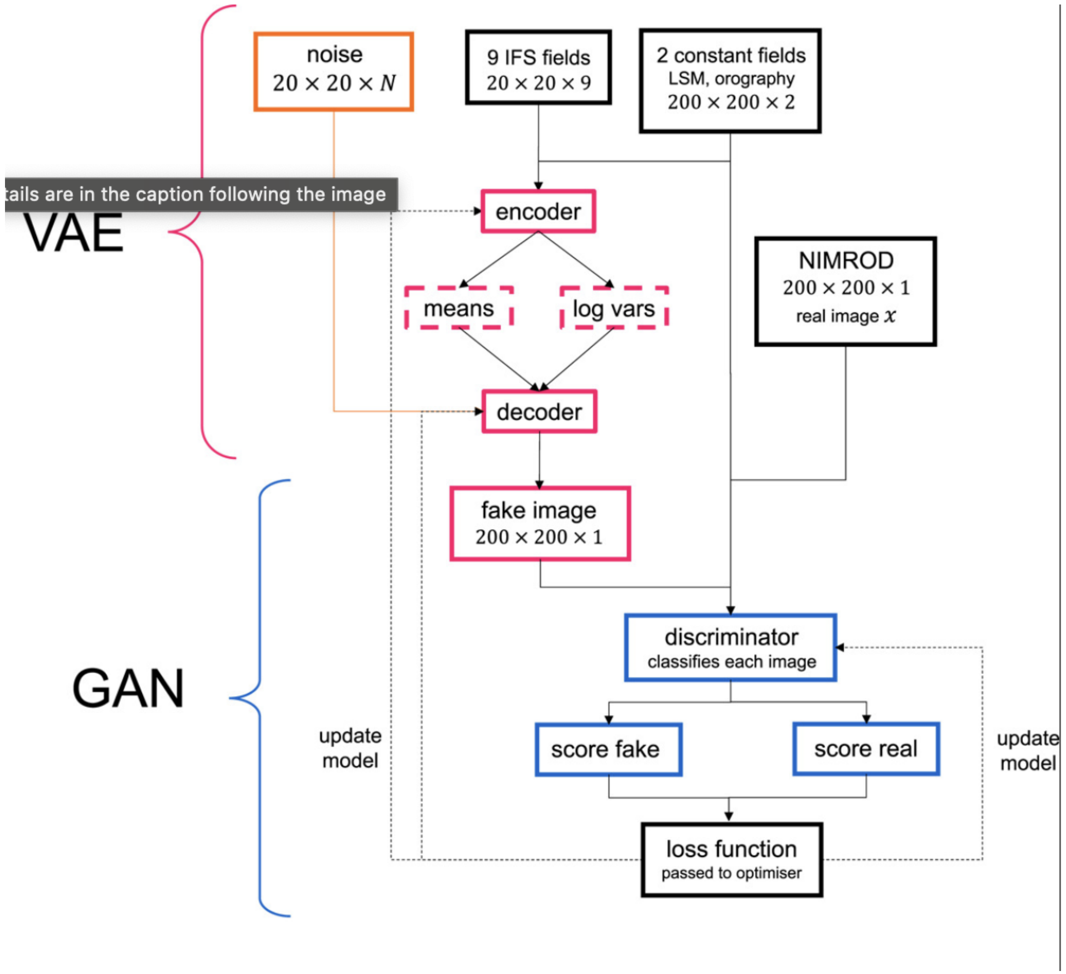
As the generator improves, the discriminator will not be able to discriminate between real and generated samples





VAE-GAN

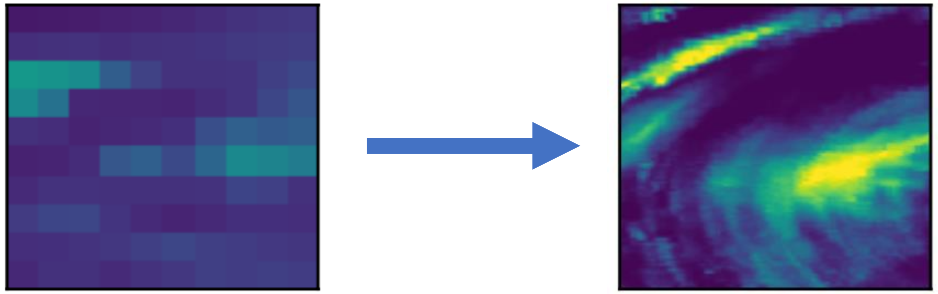




Harris, L., McRae, A. T., Chantry, M., Dueben, P. D., & Palmer, T. N. (2022). A generative deep learning approach to stochastic downscaling of precipitation forecasts. Journal of Advances in Modeling Earth Systems, 14(10), e2022MS003120.

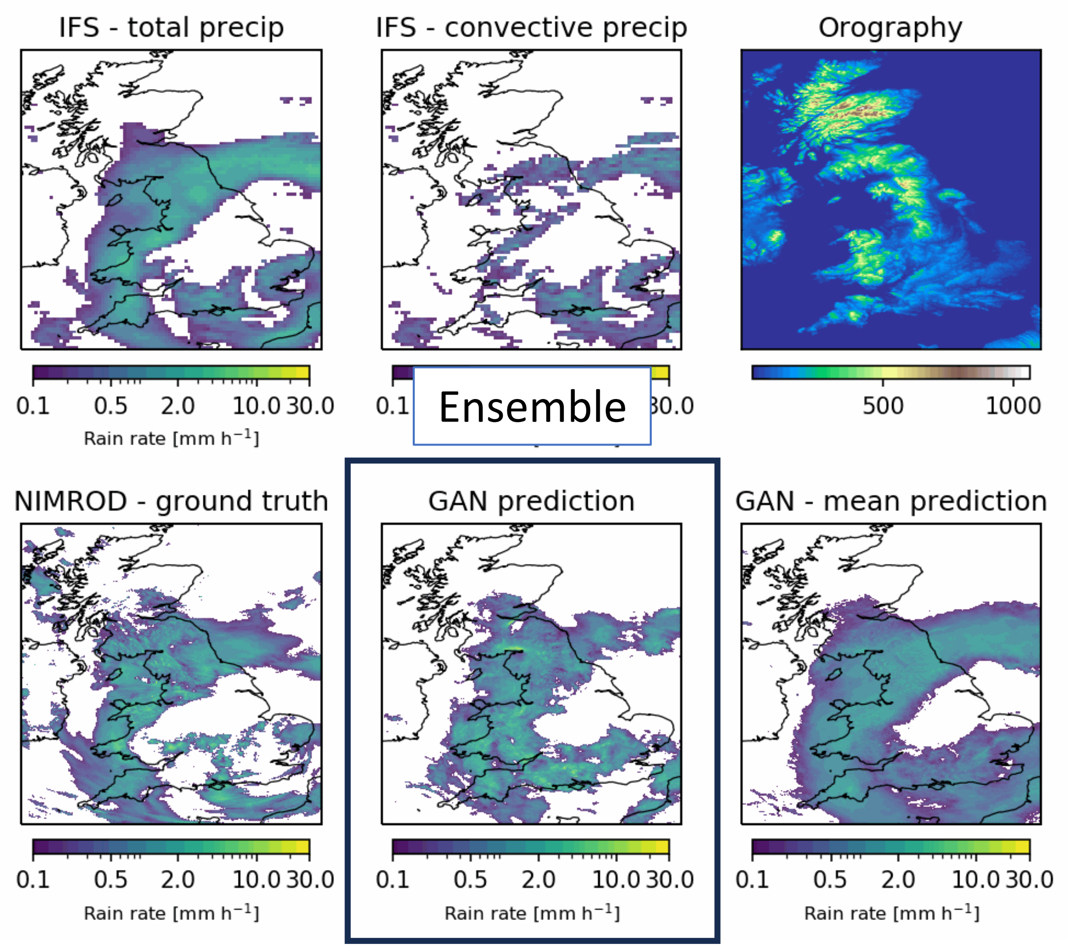
Downscaling

Downscale precipitation from IFS to higher resolutions using a GANs trained using higher resolution orographic information



10km IFS

1km radar



Training dates: 2016- 2018
Validation dates: 2019
Test dates: 2020

Hourly resolution

Approximately 70 million trainable parameters

Harris, L., McRae, A. T., Chantry, M., Dueben, P. D., & Palmer, T. N. (2022). A generative deep learning approach to stochastic downscaling of precipitation forecasts. Journal of Advances in Modeling Earth Systems, 14(10), e2022MS003120.

Case Study: PoET

Ben Bouallegue Z., Weyn J., Clare M., Dramsch J., Düben P. & Chantry M (2023). Improving medium-range ensemble weather forecasts with hierarchical ensemble transformers. arXiv preprint arXiv:2303.17195.

Reforecasts: An important dataset for post-processing

A r**eforecast dataset** is a collection of forecasts produced by the same model system with start and forecast dates from the past, usually going back for a considerable number of years.

This thus creates a consistent large dataset from which it is possible to learn errors and biases of the model

Idea of using reforecasts is an old idea and was being pushed for as far back as 15 years ago in Hagedorn (2008) for example.

Hagedorn, R. (2008). Using the ECMWF reforecast dataset to calibrate EPS forecasts. *ECMWF Newsletter*, *117*, 8-13.

Postprocessing Ensembles

Variables: Precipitation and 2m Temperature

Training Dataset

Input

Target

Ensemble Reforecast
(10 perturbed + 1 control)

generated twice a week in 2020 for a 20-year period

ERA5

Test Dataset

Operational Ensemble
(50 perturbed + 1 control)

using twice a week S2S forecast

ERA5

Postprocessing Ensembles

Neural Network

Training dataset: 2000-2016
Validation dataset: 2017-2018

1 million unknown parameters

Statistical Benchmark

Window centred around the forecast validity date over 2000-2018

Precipitation: 60-day window
2m Temperature: 30-day window

3 unknown parameters per initialisation time per lead-time per grid point

Test dataset: 2021

ML methods require more data than statistical methods (generally) and require two separate datasets in training to prevent overfitting

Member-by-Member *(statistical benchmark)*

Correction applied to each ensemble member X^m individually

$$X_C^m = \alpha + \beta\bar{X} + \gamma(X^m - \bar{X})$$

where α nudges the ensemble mean, β scales the ensemble mean and γ adjusts the ensemble spread

Constrained to preserve two different reliability conditions:

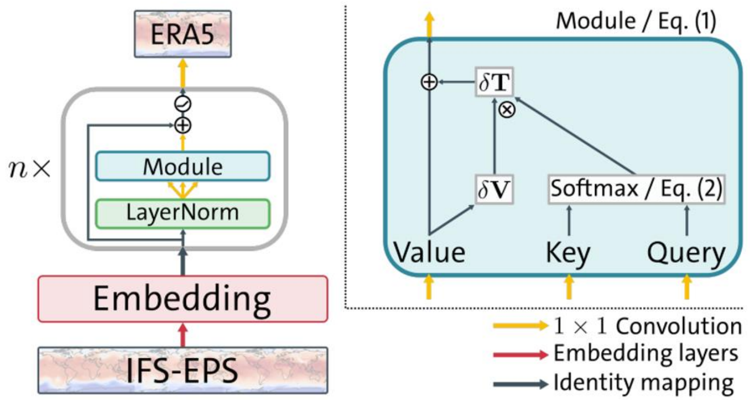
1. *Climatological reliability:* Forecast variability is equal to observation variability
2. *Weak ensemble reliability:* Average ensemble variance agrees with mean squared forecast error

Van Schaeybroeck, B., & Vannitsem, S. (2015). Ensemble post-processing using member-by-member approaches: theoretical aspects. *Quarterly Journal of the Royal Meteorological Society*, *141*(688), 807-818.

PoET – Postprocessing Ensembles with Transformers

Transformers compute relevance of each sequence element to every other element.

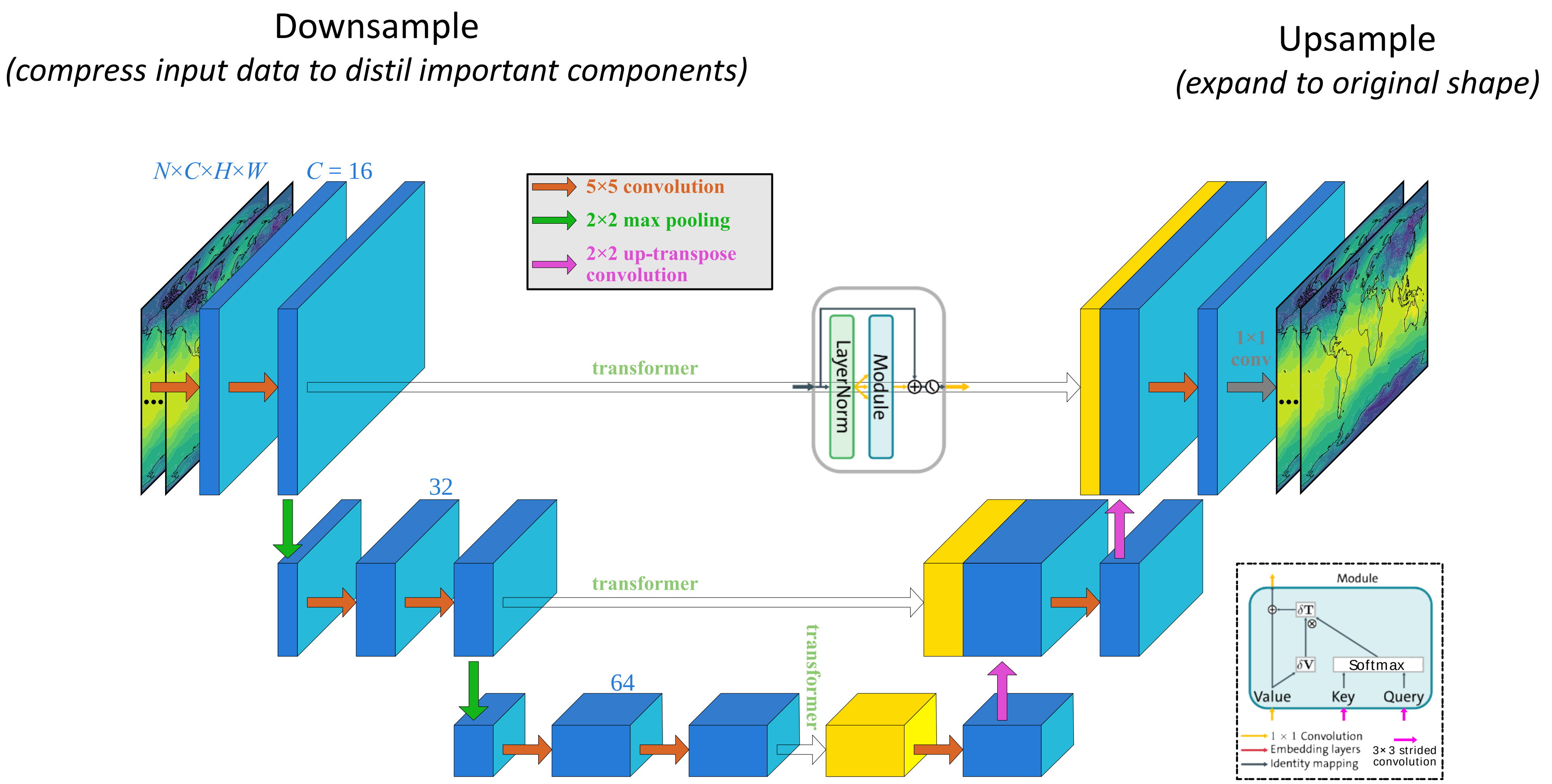
Very relevant technique for ensemble post-processing because applying along the ensemble member dimension, transformer learns similarities between members and aggregates information across the ensemble



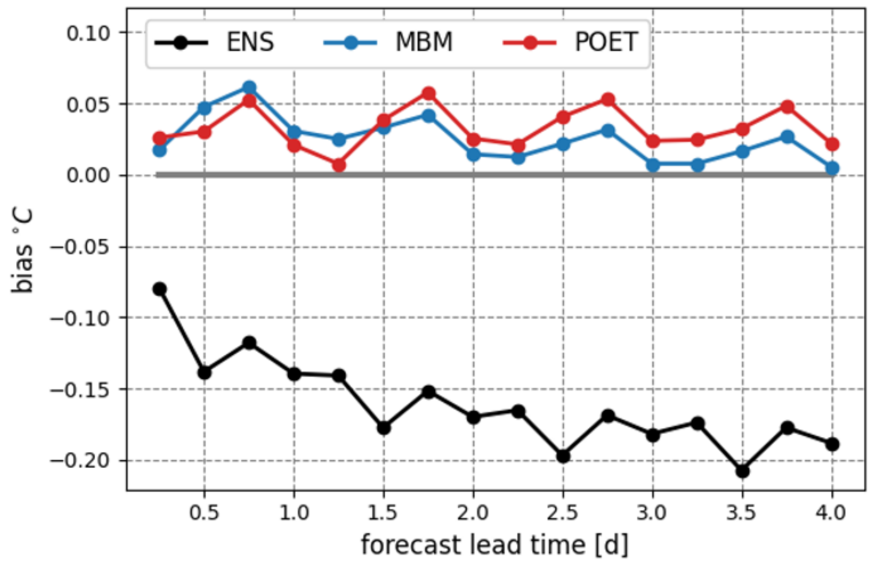
Values represent embedding of original ensemble member, keys and queries compute similarity between ensemble members on global scale

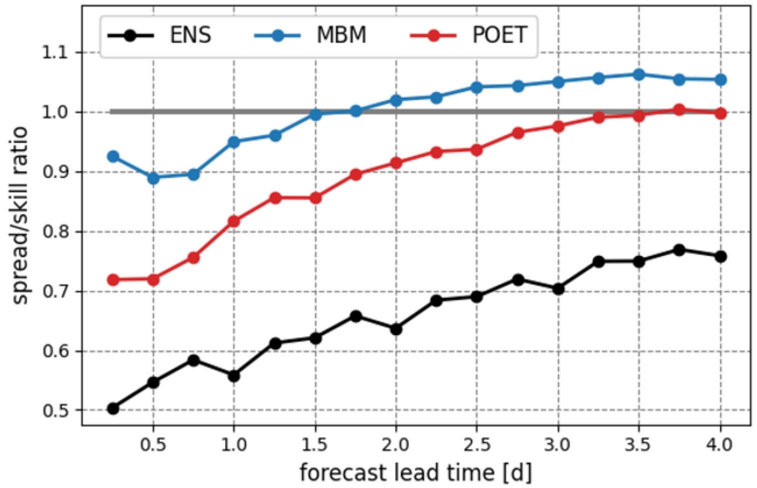
PoET – Postprocessing Ensembles with Transformers

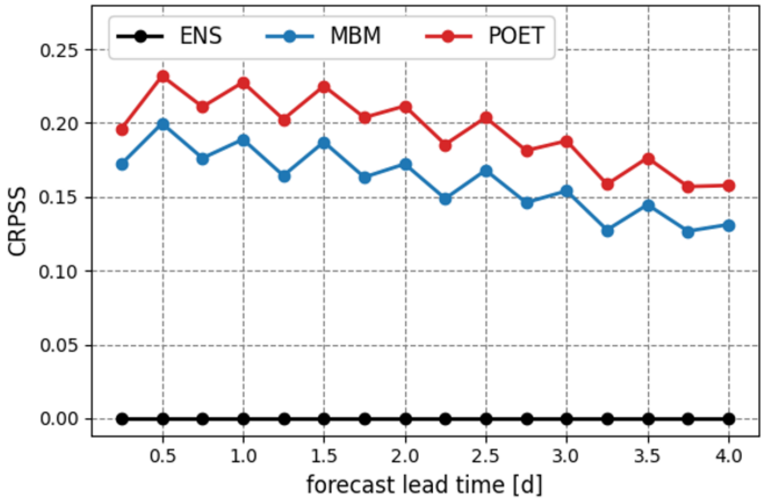
PoET combines a Transformer with a U-Net (learning information at different scales) to optimise CRPS



Results: 2m Temperature





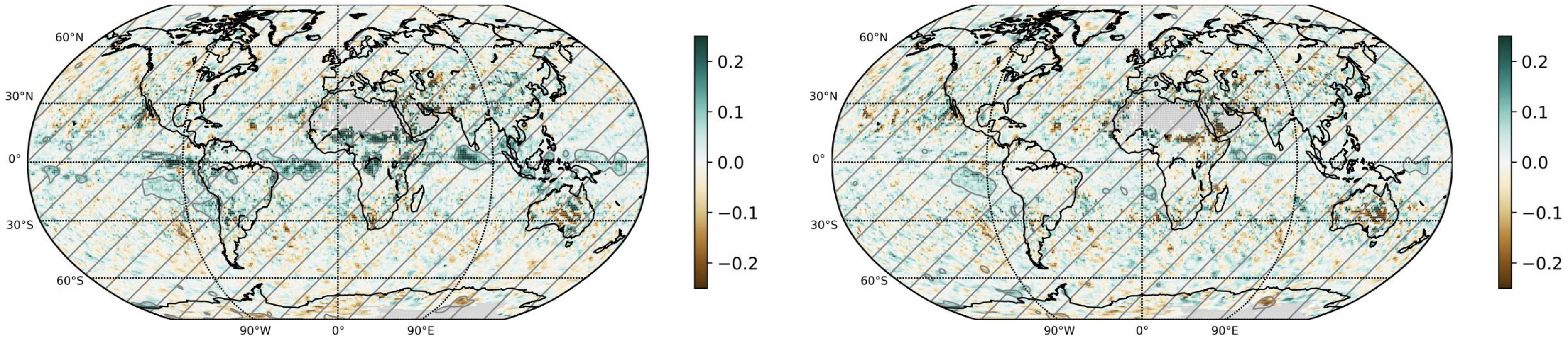


Both methods reduce bias and improve calibration

MBM closer to ideal spread/skill relationship

PoET improves headline score CRPS

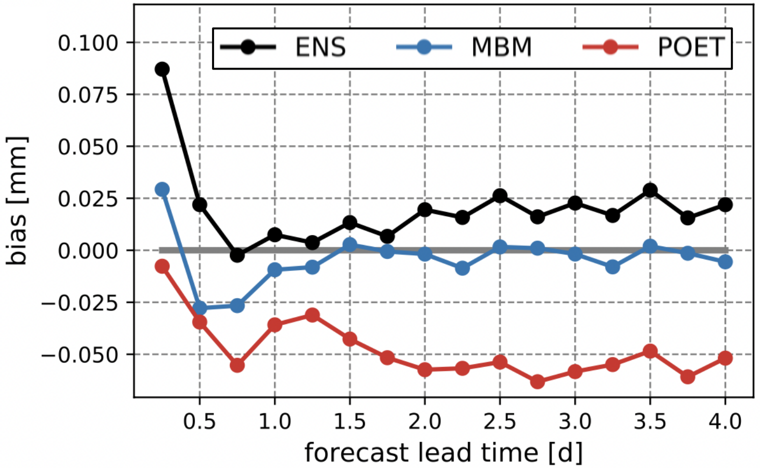
Results: Precipitation



CRPSS of precipitation of PoET Left: with respect to raw ensemble; Right: with respect to MBM

Both methods reduce bias and improve calibration

PoET improves headline score CRPS

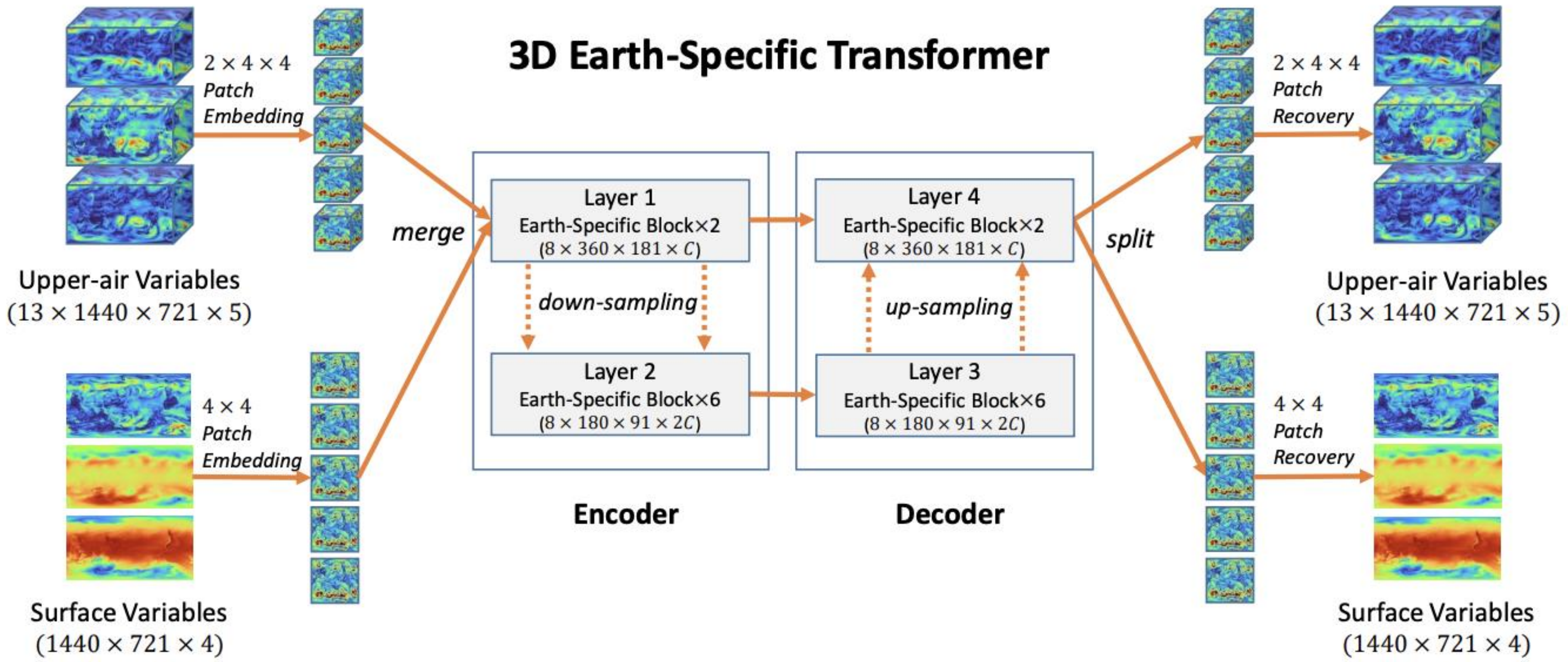


Post-processing data-driven forecasts

Is there still a role for using machine learning for post-processing with data-driven forecasts given they are trained towards the truth?

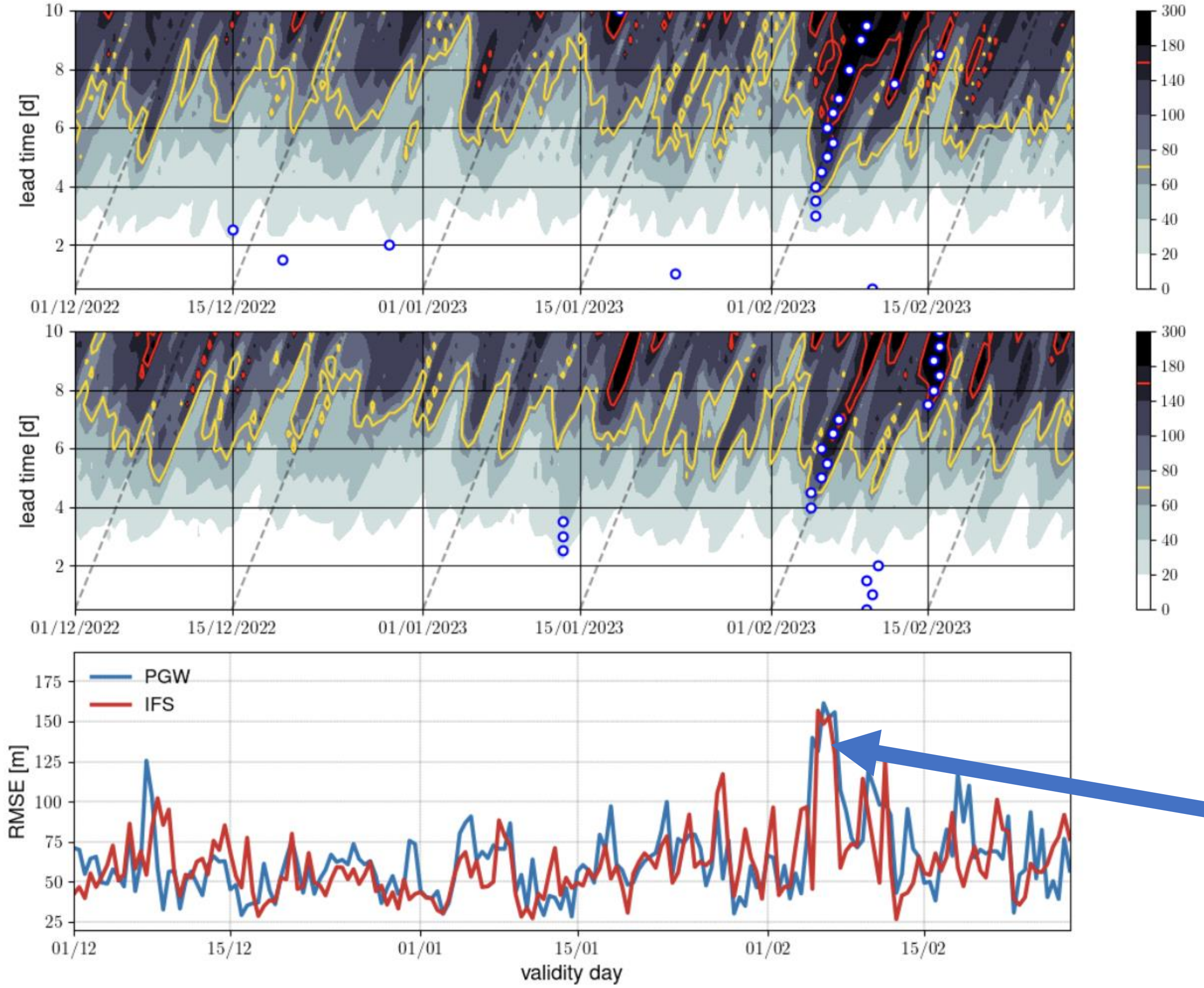
Or is it just trying to solve the same problem twice? In other words a "hat on a hat"?

Pangu Weather Model



Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. Nature, 1-6.

Predictability barrier plots for Z500

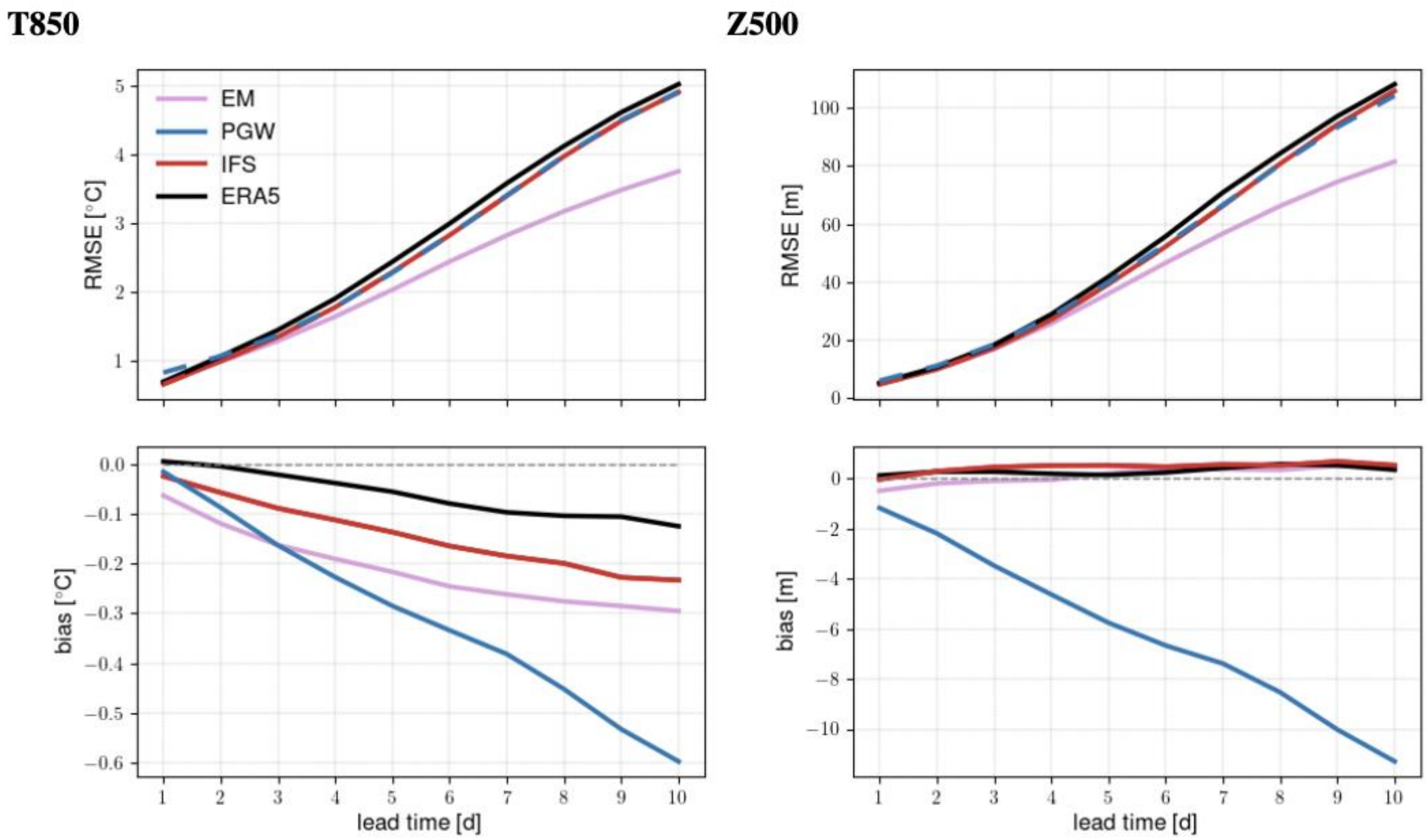


Transverse structure indicates rapid error growth leading to a poor forecast at all lead times (likely forecast initialisation issue)

Vertical structure indicates a weather situation difficult to predict for consecutive runs (likely due to predictability barriers).

Ben-Bouallegue, Z., Clare, M. C., Magnusson, L., Gascon, E., Maier-Gerber, M., Janousek, M., ... & Pappenberger, F. (2023). The rise of data-driven weather forecasting. *arXiv preprint arXiv:2307.10128*.

Verification



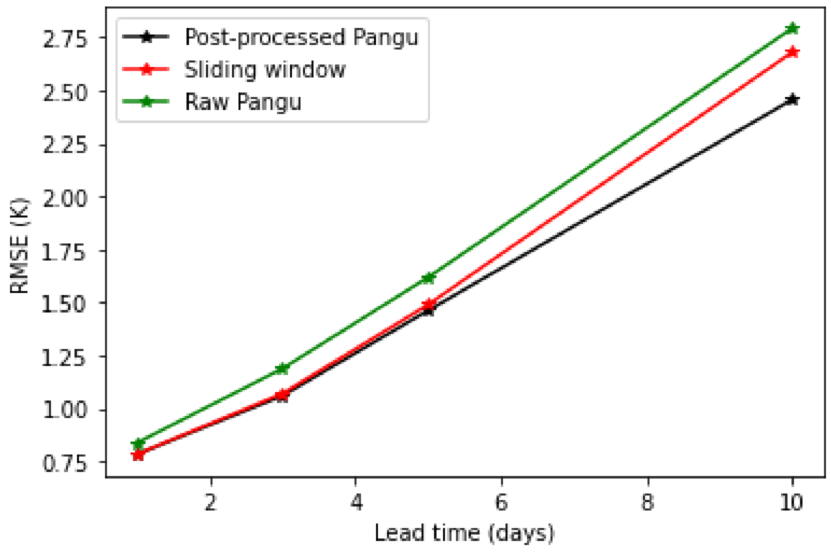
$$\sqrt{\overline{(f-o)^2}}$$

$$\overline{f-o}$$

Ben-Bouallegue, Z., Clare, M. C., Magnusson, L., Gascon, E., Maier-Gerber, M., Janousek, M., ... & Pappenberger, F. (2023). The rise of data-driven weather forecasting. *arXiv preprint arXiv:2307.10128*.

Post-processing Pangu

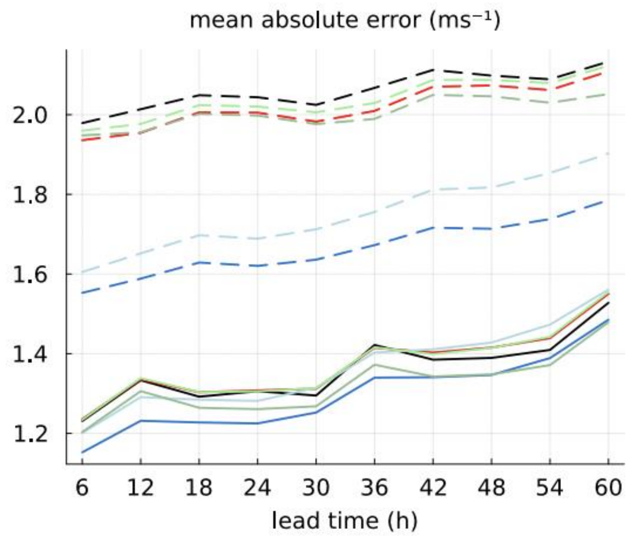
2m temperature



Comparing statistical and ML post-processing of Pangu

Work done at ECMWF

10m wind speed



Post-processed (solid) and raw (dashed) forecasts

Bremnes, J. B., Nipen, T. N., & Seierstad, I. A. (2023). Evaluation of forecasts by a global data-driven weather model with and without probabilistic post-processing at Norwegian stations. arXiv preprint arXiv:2309.01247.

Helpful (hopefully) tips for choosing your post-processing method

How to choose the correct post-processing method for your NWP or data-driven forecast

How much data do I have?

What do I want my output to be? Probabilistic distribution versus scenarios? Who is my end-user?

Do I want to post-process an ensemble forecast or a deterministic forecast?

Do I want to post-process against observations or gridded data?

Do I want to add uncertainty information to a deterministic forecast?

Do I want to increase the temporal/spatial resolution of the forecast?

Machine Learning vs Classical Methods: Pros and Cons?

	PROS	CONS
MACHINE LEARNING	• Often more accurate • Increase resolution • Can consider multiple input features and non-linear correlations	• Data hungry • Can be difficult to understand skill
CLASSICAL METHODS	• Generally more interpretable due to their simplicity	• Tend to be simpler models which cannot capture all of the errors

References

- *Bouallegue Z. B., Weyn J., Clare M., Dramsch J., Düben P. & Chantry M (2023). Improving medium-range ensemble weather forecasts with hierarchical ensemble transformers. arXiv preprint arXiv:2303.17195.*
- *Bouallègue, Z. B., Cooper, F., Chantry, M., Düben, P., Bechtold, P., & Sandu, I. (2023). Statistical Modeling of 2-m Temperature and 10-m Wind Speed Forecast Errors. Monthly Weather Review, 151(4), 897-911.*
- *Bremnes, J. B., Nipen, T. N., & Seierstad, I. A. (2023). Evaluation of forecasts by a global data-driven weather model with and without probabilistic post-processing at Norwegian stations. arXiv preprint arXiv:2309.01247*
- *Hagedorn, R. (2008). Using the ECMWF reforecast dataset to calibrate EPS forecasts. ECMWF Newsletter, 117, 8-13.*
- *Harris, L., McRae, A. T., Chantry, M., Dueben, P. D., & Palmer, T. N. (2022). A generative deep learning approach to stochastic downscaling of precipitation forecasts. Journal of Advances in Modeling Earth Systems, 14(10), e2022MS003120.*
- *Hewson, T. D., & Pillosu, F. M. (2021). A low-cost post-processing technique improves weather forecasts around the world. Communications Earth & Environment, 2(1), 132.*