Machine Learning at the ECMWF

Peter Dueben

Royal Society University Research Fellow & ECMWF's Coordinator for Machine Learning and AI Activities

This research used resources of the Oak Ridge Leadership Computing Facility (OLCF), which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725.

The strength of a common goal

The ESIWACE2 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823988.

Some views on Machine Learning for numerical weather predictions

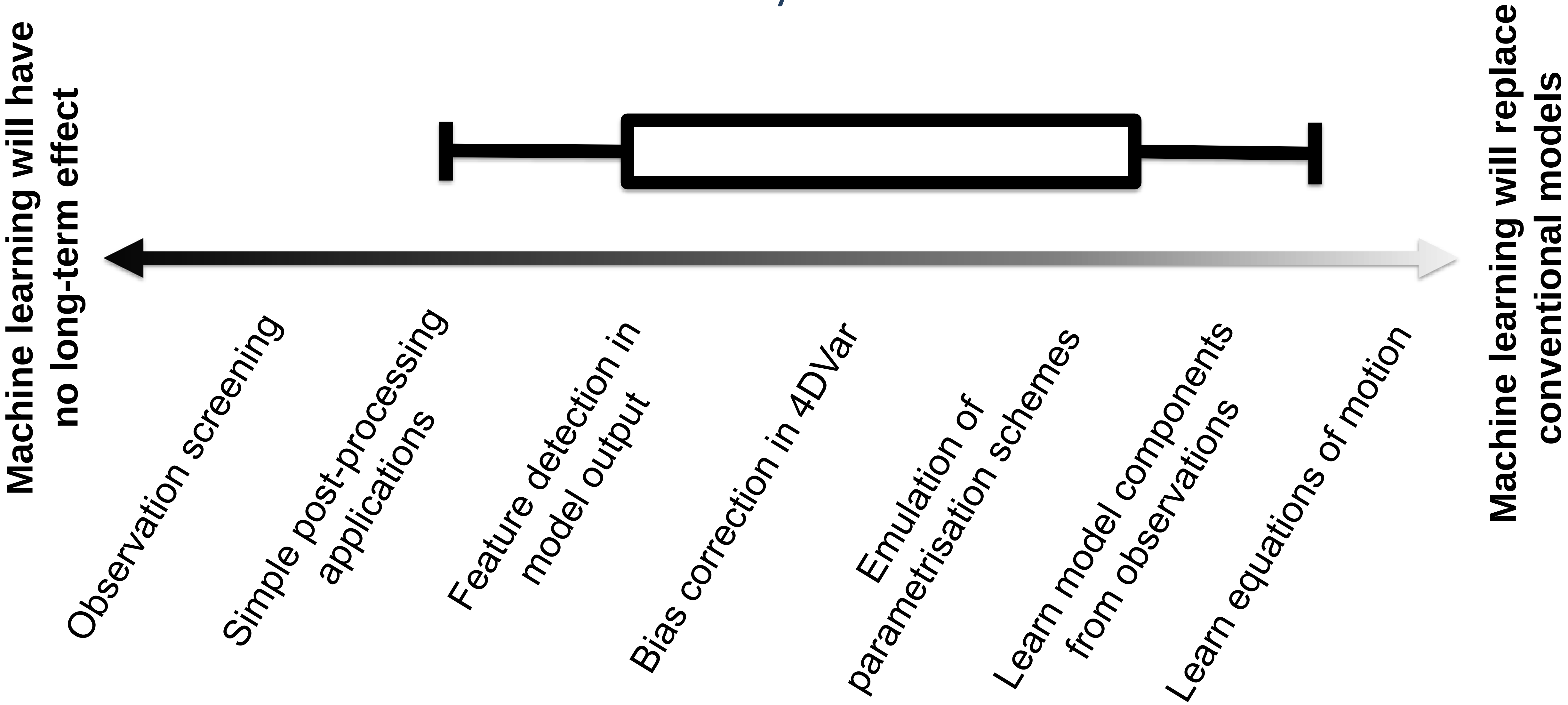
Two years into the “hype” of machine learning…

Conservative domain scientist:
Machine learning is a wave passing by and many of the “new” methods are not really new!
Please consider machine learning as an amazing expansion to your toolbox.

Proactive domain scientist:
Machine learning can only be useful if it’s embracing process understanding and if it turns into a grey box.
Please have a look at developments in other domains.

High Performance Computing scientist:
Think in petabytes and exaflop. Sooner or later machine learning will replace all conventional tools.
Please consider decades of process understanding and model development, and the complexity of the problem.

What will machine learning for numerical weather predictions look like in 10 years from now?



The uncertainty range is still very large...

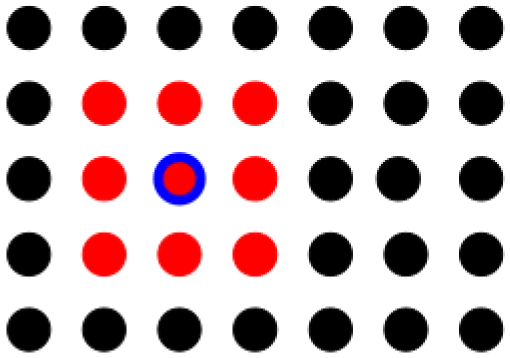
Can we replace conventional weather forecast systems by deep learning?

We could base the entire model on neural networks and trash the conventional models.?
There are limitations for existing models and ECMWF provides access to hundreds of petabytes of data

A simple test configuration:

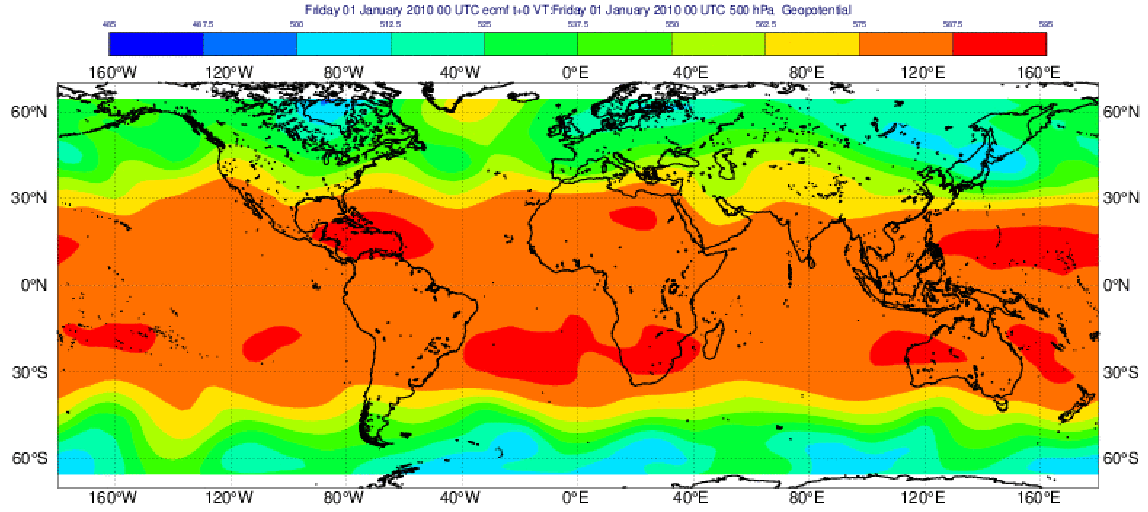
- We retrieve historical data (ERA5) for geopotential at 500 hPa (Z500) for the last decades (>65,000 global data sets)
- We map the global data to a coarse two-dimensional grid (60x31)
- We learn to predict the update of the field from one hour to the next using deep learning
- Once we have learned the update, we can perform predictions into the future

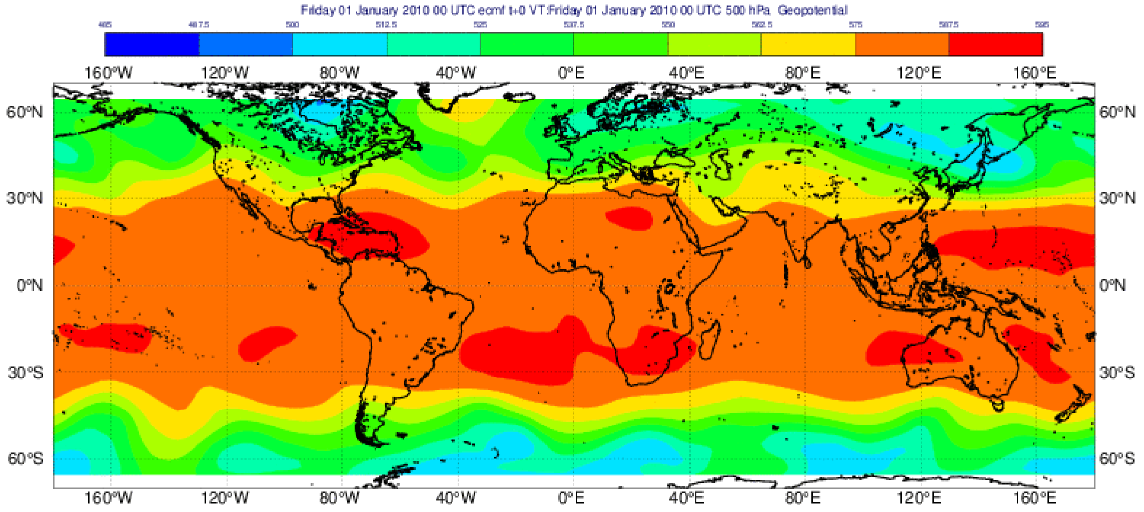
No physical understanding is required!



Dueben and Bauer GMD 2018

Can we replace conventional weather forecast systems by deep learning?

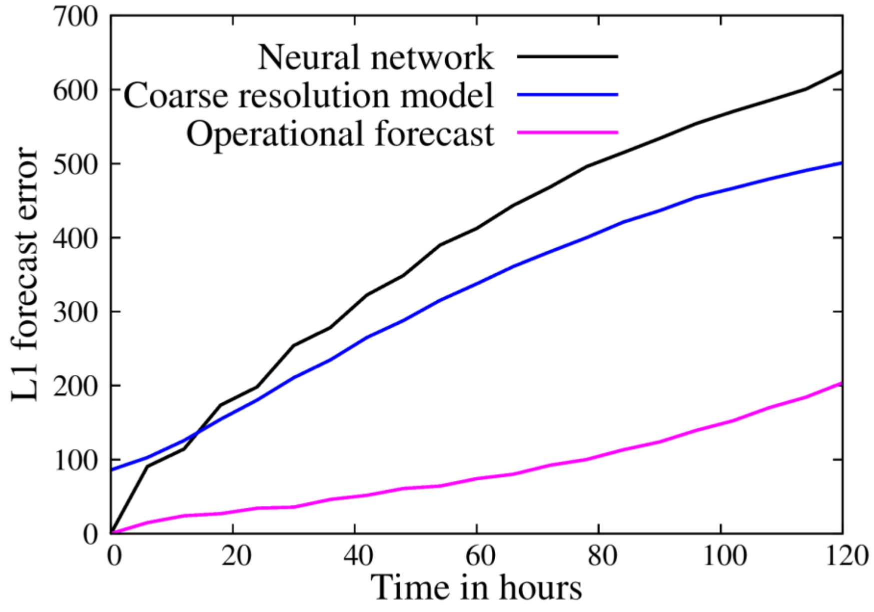




Time evolution of Z500 for historic data and a neural network prediction.
Can you tell which one is the neural network?

- The neural network is picking up the dynamics nicely.
- Forecast errors are comparable if we compare like with like.
- There is a lot of progress at the moment.
  Scher and Messori GMD 2019; Weyn, Durran, and Caruana JAMES 2019; Rasp and Thuerey 2020…
- Is this the future for medium-range weather predictions?

Unlikely…
- The simulations change dynamics in long integrations and it is unclear how to fix conservation properties.
- It is unknown how to increase complexity and how to fix feature interactions.
- There are only ~40 years of data available.

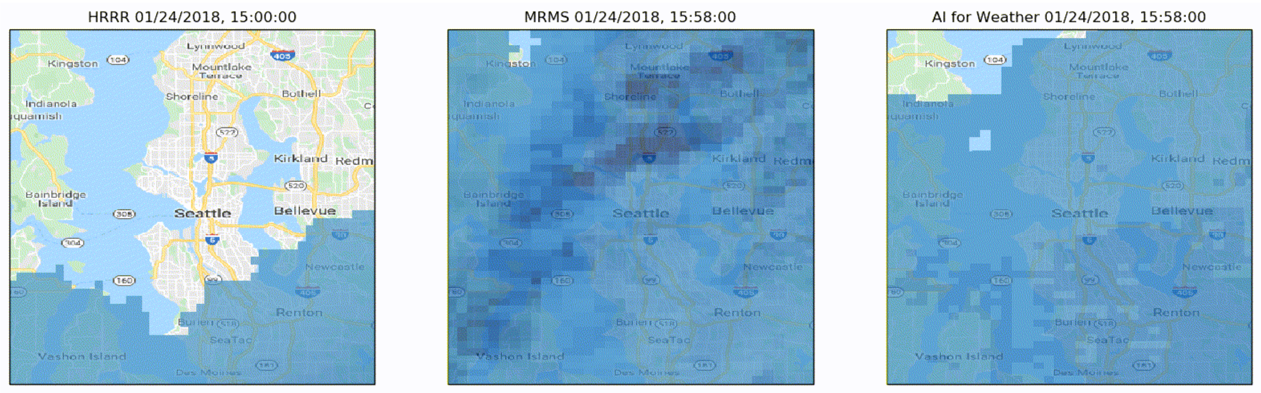


Dueben and Bauer GMD 2018

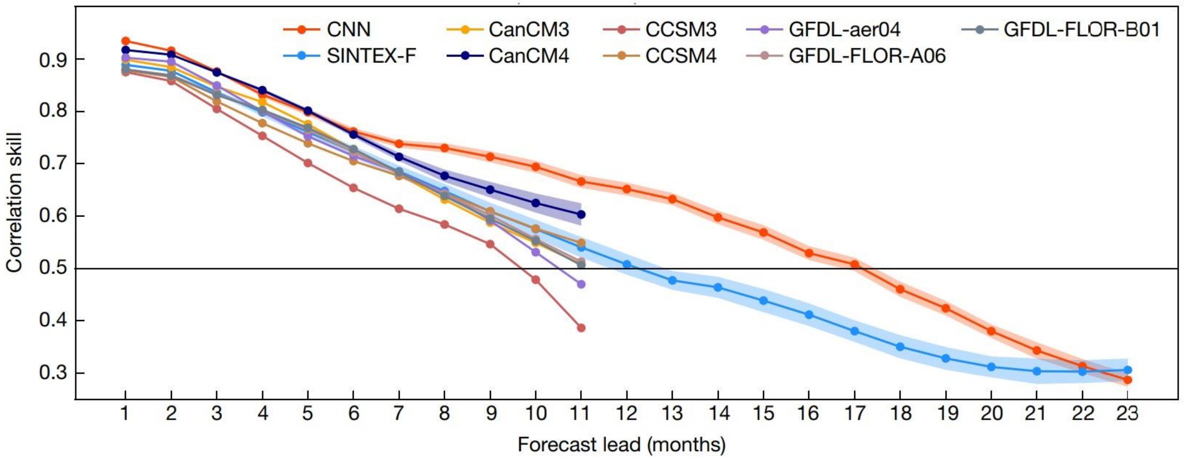
Can we replace conventional Earth System models by deep learning?

1-hour MetNET precipitation predictions by from Google: Agrawal, Barrington, Bromberg, Burge, Gazen, Hickey arXiv:1912.12132

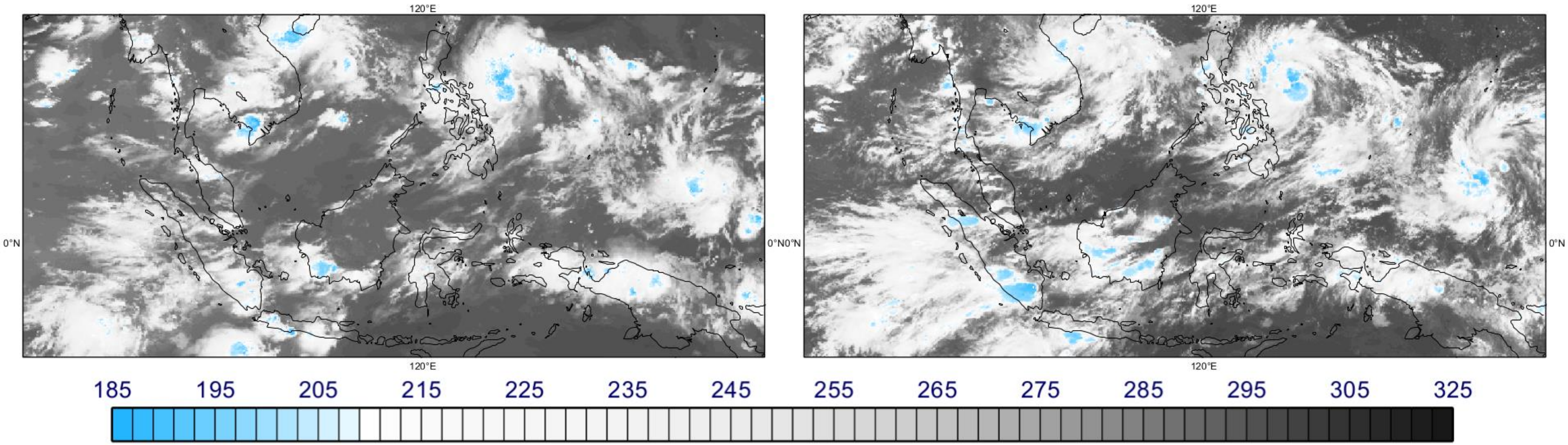
NOAA forecast Ground truth Machine learning:



Deep learning for multi-year ENSO forecasts: Ham, Kim, Luo *Nature* 2019



Why is it hard to beat conventional forecast systems in the medium range?



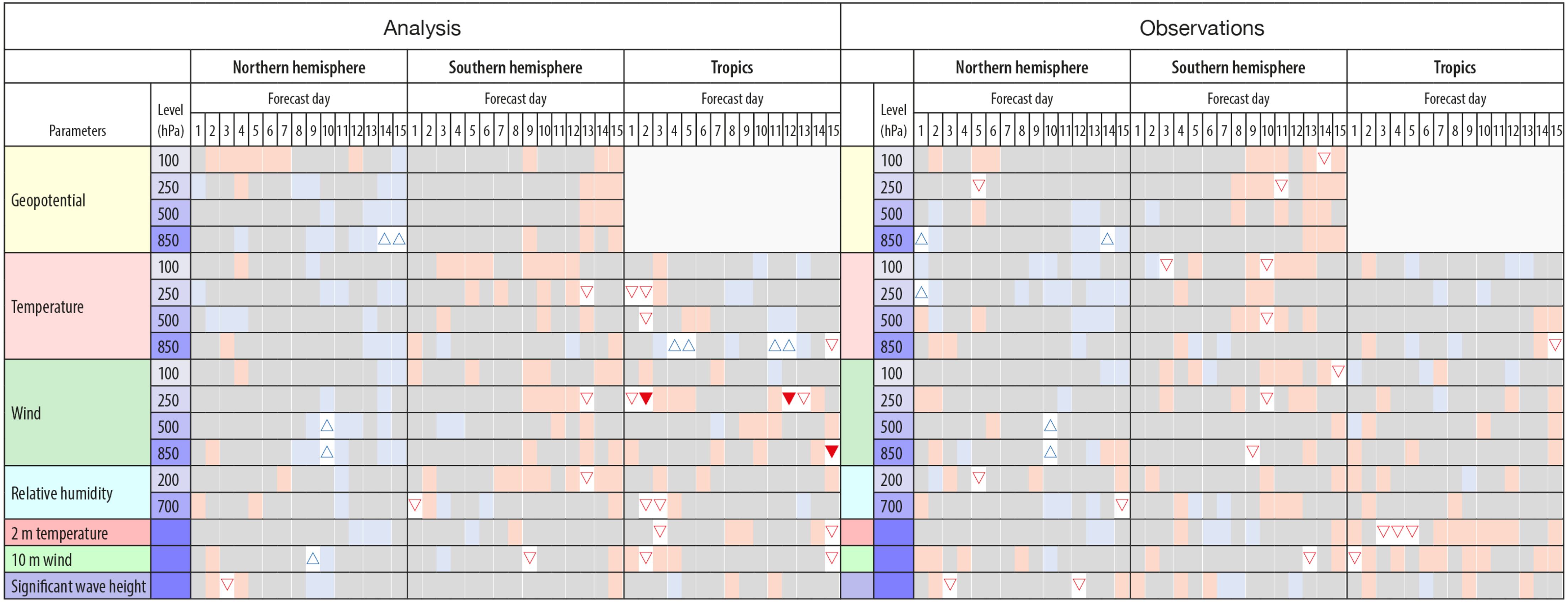
Top-of-the-atmosphere cloud brightness temperature [K] for satellite observations and a simulation of the atmosphere with 1.45 km resolution.
Dueben, Wedi, Saarinen and Zeman JSMJ 2020

Today, global weather forecast simulations have O(1,000,000,000) degrees-of-freedom, can represent many details of the Earth System, and show a breath-taking level of complexity.

They are based on decades of model developments and process understanding.

summit

Why is it hard to beat conventional weather forecast systems?



Symbol legend: for a given forecast step...

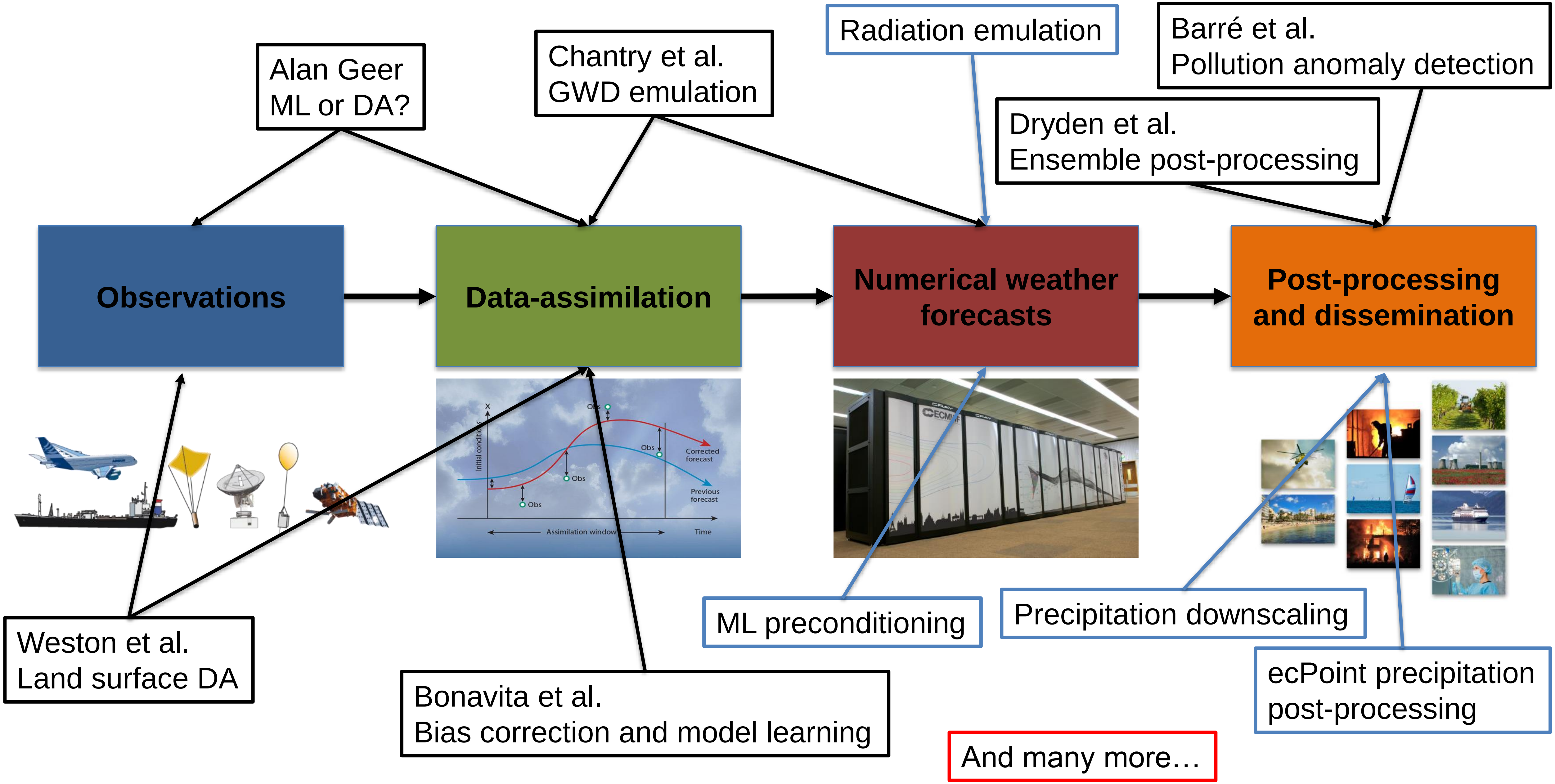
- ▲ SP better than DP statistically significant with 99.7% confidence
- △ SP better than DP statistically significant with 95% confidence
- SP better than DP statistically significant with 68% confidence
- no significant difference between DP and SP
- SP worse than DP statistically significant with 68% confidence
- ▽ SP worse than DP statistically significant with 95% confidence
- ▼ SP worse than DP statistically significant with 99.7% confidence

So…using machine learning for weather predictions is useless then?

No!

Dueben, Diamantakis, Lang, Saarinen, Sandu, Wedi, Wilhelmsson ECMWF Newsletter 2018

Machine learning applications across the numerical weather prediction workflow

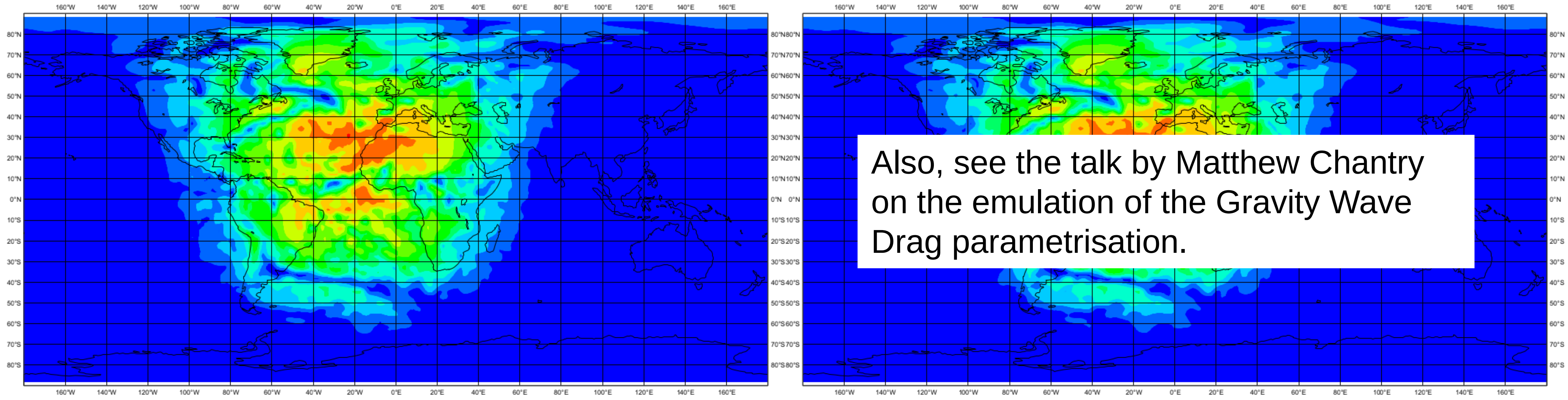


Numerical weather forecasts: To emulate the radiation scheme

- Store input/output data pairs of the radiation schemes
- Use this data to train a neural network
- Replace the radiation scheme by the neural network within the model

This is a very active area of research:
Poster by Belochitski and Krasnopolsky
Poster by Ukkonen et al.
Rasp, Pritchard, Gentine PNAS 2018
Brenowitz and Bretherton GRL 2018
…

Why would you do this?
Neural networks are likely to be much more efficient and portable to heterogenous hardware



Also, see the talk by Matthew Chantry on the emulation of the Gravity Wave Drag parametrisation.

Surface downward solar radiation for the original scheme and the neural network emulator (based on a ResNet).

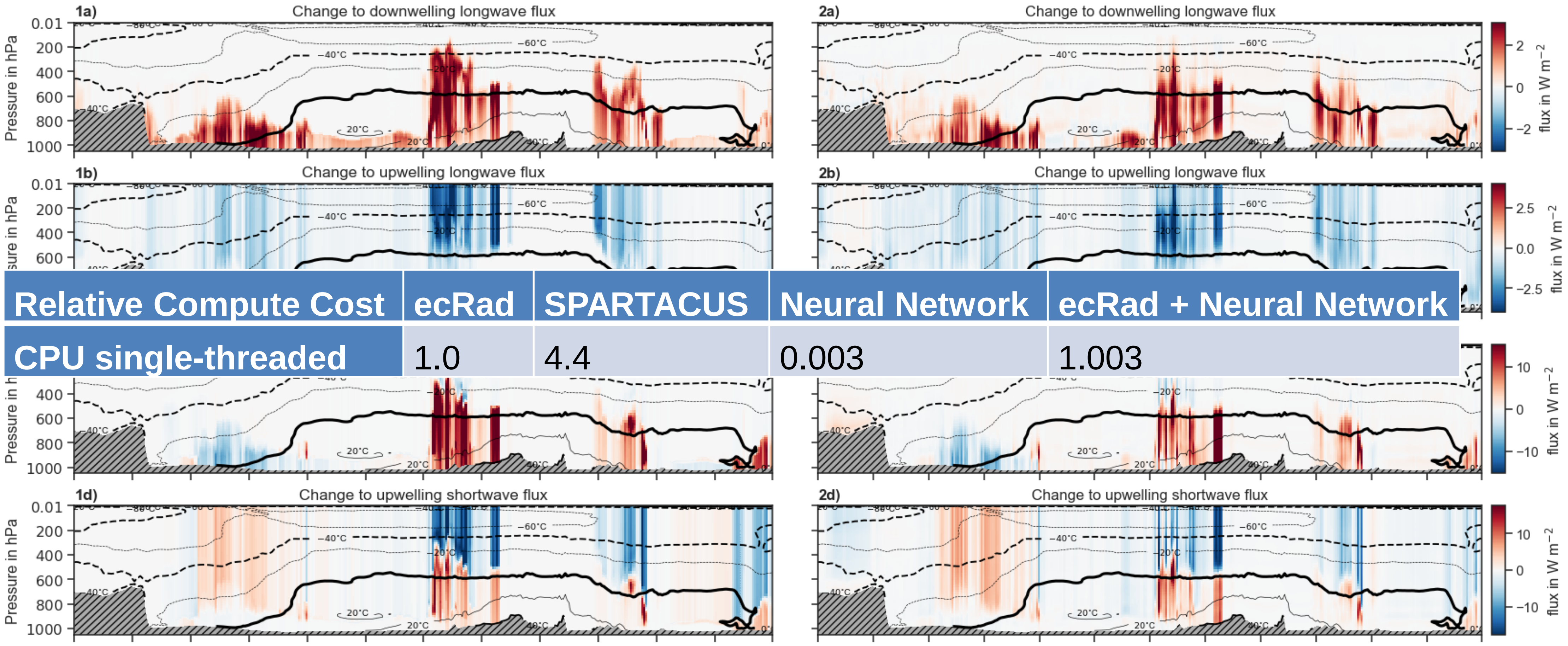
The approach is working and the neural network is ~10 times faster than the original scheme. However, model results are still degraded.

Dueben, Hogan, Bauer @ECMWF and Progsch, Angerer @NVIDIA

Numerical weather forecasts: To emulate the 3D cloud effects in radiation

To represent 3D cloud effects for radiation (SPARTACUS) within simulations of the Integrated Forecast Model is four time slower than the standard radiation scheme (ecRad)

Can we emulate the difference between ecRad and SPARTACUS using neural networks?



Relative Compute Cost	ecRad	SPARTACUS	Neural Network	ecRad + Neural Network
CPU single-threaded	1.0	4.4	0.003	1.003

Preliminary results look good.

Meyer, Hogan, Dueben

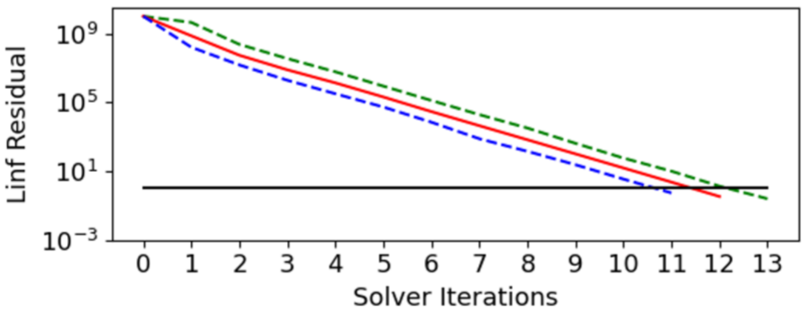
Numerical weather forecasts: To precondition the linear solver

- Linear solvers are important to build efficient semi-implicit time-stepping schemes for atmosphere and ocean models.
- However, the solvers are expensive.
- The solver efficiency depends critically on the preconditioner that is approximating the inverse of a large matrix.

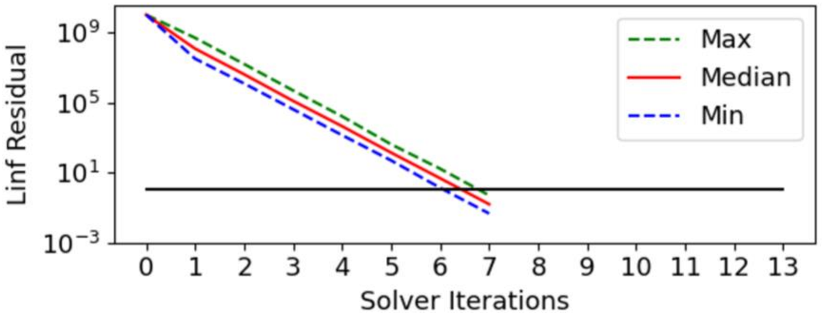
Can we use machine learning for preconditioning, predict the inverse of the matrix and reduce the number of iterations that are required for the solver?

Testbed: A global shallow water model at 5 degree resolution but with real-world topography.
Method: Neural networks that are trained from the model state and the tendencies of full timesteps.

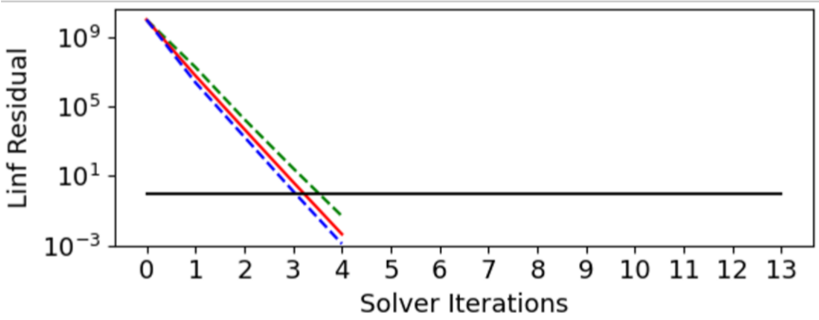
No preconditioner:



Machine learning preconditioner:



Implicit Richardson preconditioner:



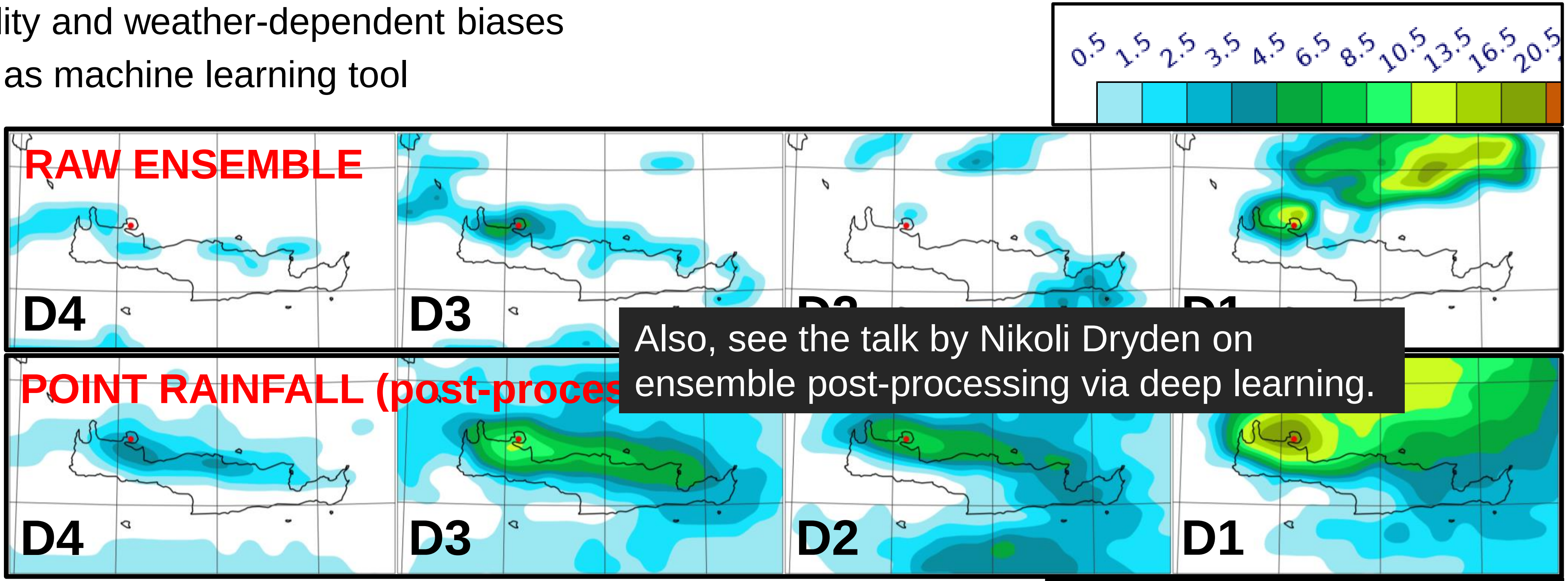
It turns out that the approach (1) is working and cheap, (2) interpretable and (3) easy to implement even if no preconditioner is present.

Ackmann, Dueben, Smolarkieicz and Palmer

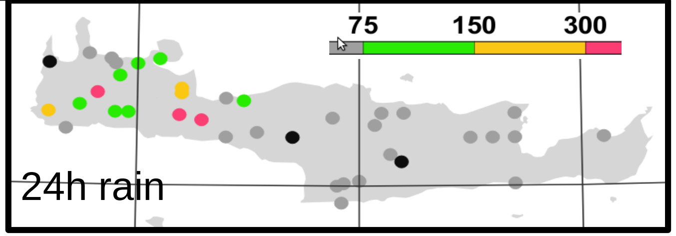
Post-processing and dissemination: ***ecPoint*** to post-process rainfall predictions

- Use forecast data as inputs
- Train against worldwide rainfall observations
- Improve local rainfall predictions by accounting for sub-grid variability and weather-dependent biases
- Use decision trees as machine learning tool

Probability (%) > 50mm /12h



Also, see the talk by Nikoli Dryden on ensemble post-processing via deep learning.



Example: Devastating floods in Crete on 25 February 2019

Benefits: Earlier and more consistent signal with higher probabilities

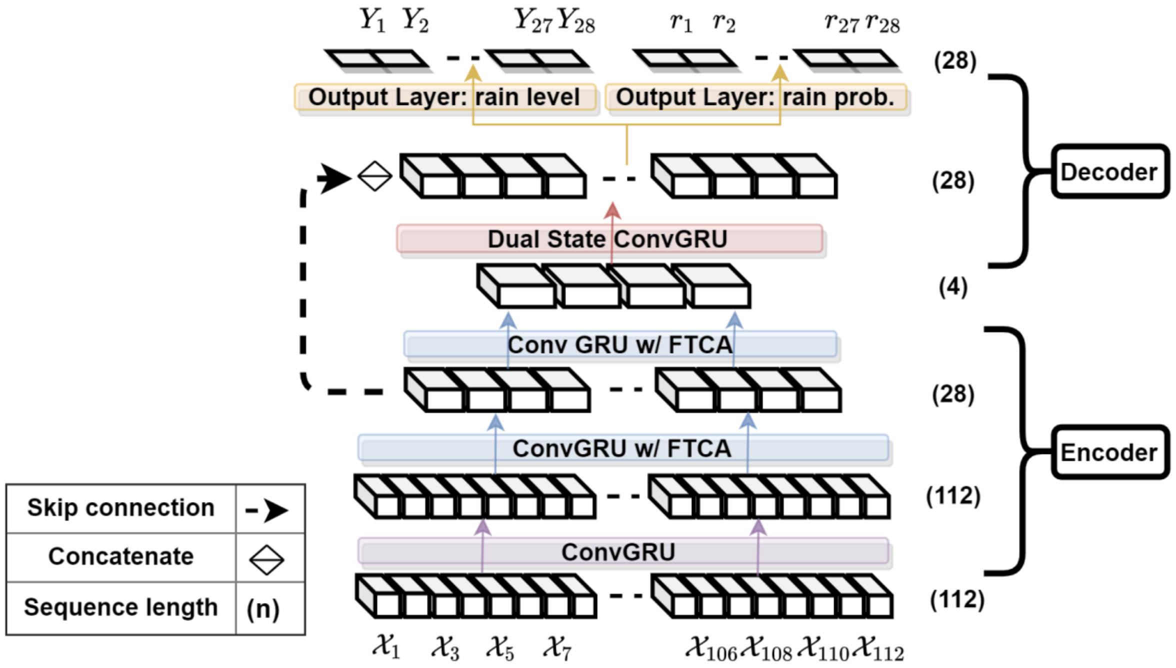
Timothy Hewson and Fatima Pillosu

Post-processing and dissemination: Precipitation down-scaling

Problem: Learn to map precipitation predictions from ERA5 reanalysis data at ~50 km resolution to E-OBS local precipitation observations at ~10 km resolution over the UK.

Use case: Eventually, apply the tool to climate predictions to understand changes of local precipitation pattern due to climate change.

Method: Use Tru-NET with a mixture of ConvGru layers to represent spatial-temporal scale interactions and a novel Fused Temporal Cross Attention mechanism to improve time dependencies.



Model	RMSE
IFS within ERA5	3.627
Hierarchical Convolutional GRU	3.266
Tru-Net	3.081

Adewoyin, Dueben, Watson, He, Dutta http://arxiv.org/abs/2008.09090

How to progress?

Follow a **bottom-up approach** and work on physics informed machine learning, trustworthy AI, hybrid models and uncertainty quantification.

Follow a **top-down approach** and build machine learning solutions that scale neural networks to millions of inputs for 3D fields on the sphere.

Focus on **useful tools** that can serve as beacons.

Build **customised machine learning solutions** for weather and climate applications.

Build **benchmark problems** similar to ImageNet.
(*see WeatherBench in Rasp, Dueben, Scher, Weyn, Mouatadid and Thureey JAMES 2020*)

Non of the above will be possible without a close **collaboration** between domain scientists and machine learning experts!

Future of Machine Learning at ECMWF

Foster collaborations with experts in machine learning.

Build infrastructure to allow domain scientists to easily explore machine learning applications.

Train staff and talk about opportunities and limits of machine learning.

Support weather and climate modelling community.

1. New Center of Excellence in Weather & Climate Modelling between ATOS and ECMWF
+ AMD, Mellanox, Nvidia and DDN

Machine learning project to…

… develop vanilla solutions for the emulation of physical parametrisation schemes.
… develop vanilla solutions for machine learning applications that take the three-dimensional state of the atmosphere on unstructured grids as well as scale interactions in both space and time into account.
… develop vanilla solutions for feature detection in three-dimensional IFS model output.
… develop infrastructure to enable the use of machine learning libraries that are called within IFS on the new HPC.
… enhance data workflow to facilitate machine learning applications that use IFS model output and re-analysis products.

Future of Machine Learning at ECMWF

Foster collaborations with experts in machine learning.

Build infrastructure to allow domain scientists to easily explore machine learning app

Train staff and talk about opportunities and limits of machine learning.

Support weather and climate modelling community.

MAELSTROM

2. MAELSTROM 4m€ EuroHPC_JU project coordinated by ECMWF

Compute system designs

Benchmark datasets and ML solutions

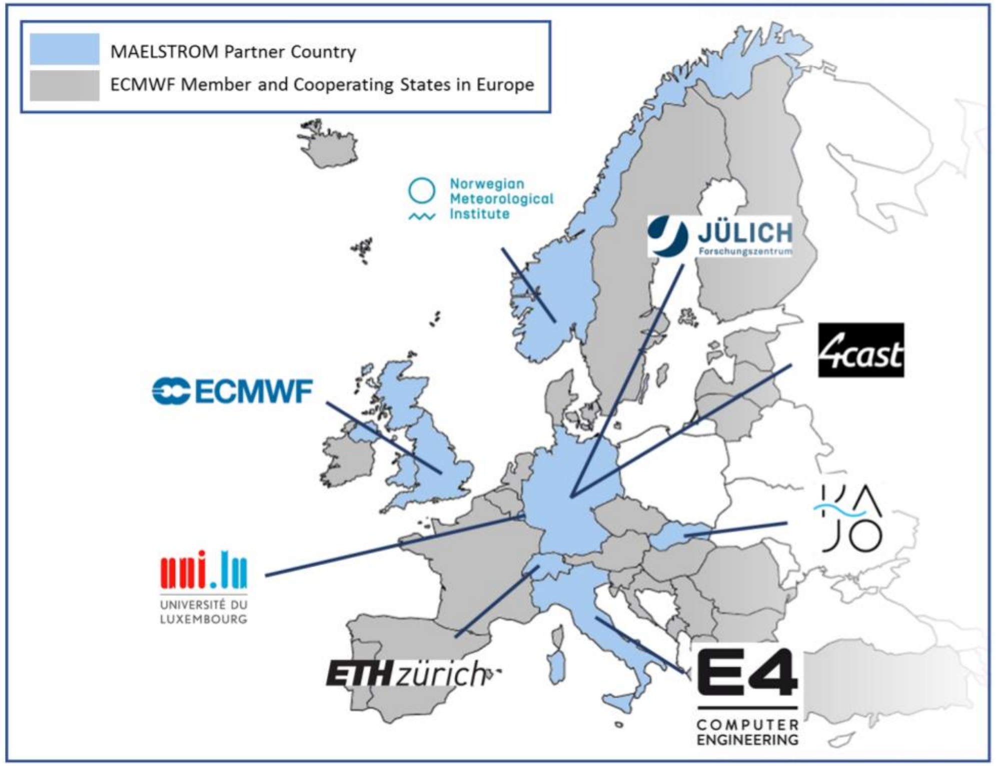
ML workflow tools

W&C applications and ML solution development

Hardware benchmarking and bespoke system design

Co-design cycle

ML workflow and software benchmarking



We are hiring!

Future of Machine Learning at ECMWF

Foster collaborations with experts in machine learning.

Build infrastructure to allow domain scientists to easily explore machine learning applications.

Train staff and talk about opportunities and limits of machine learning.

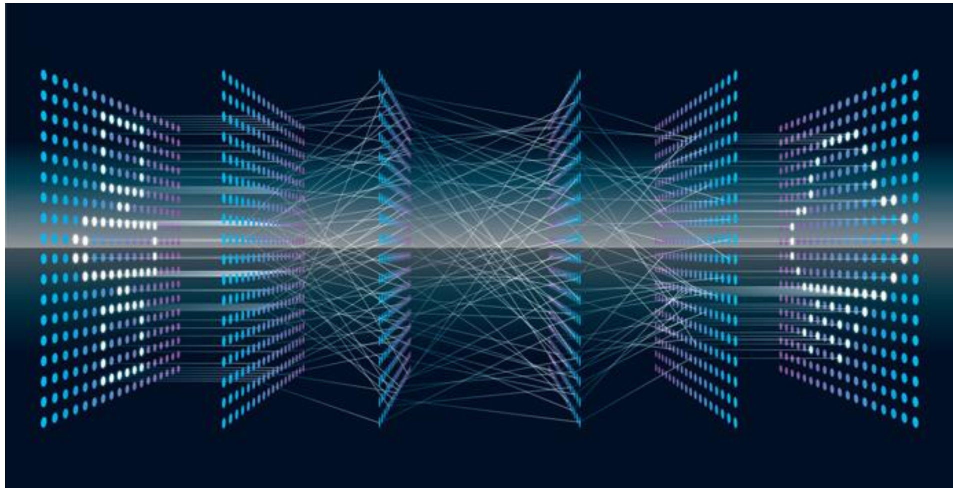
Support weather and climate modelling community.

3. ECMWF is developing the European Weather Cloud in collaboration with EUMETSAT.

4. Significant GPU power on the new ECMWF high performance computer via A100 GPUs.

5. Opening of the data archive at ECMWF.

6. Seminar series on Machine Learning.



Future of Machine Learning at ECMWF

Foster collaborations with experts in machine learni

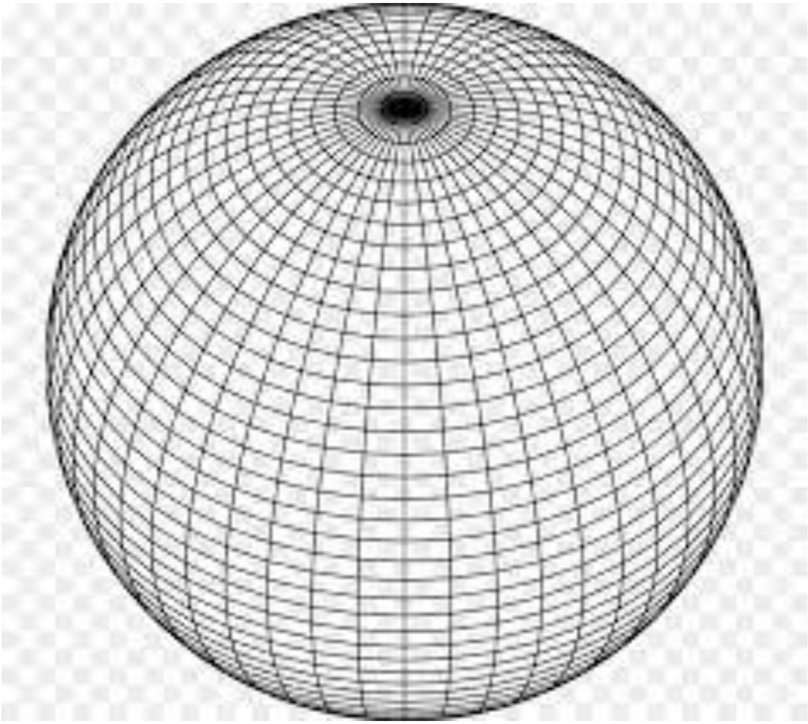
Build infrastructure to allow domain scientists to eas

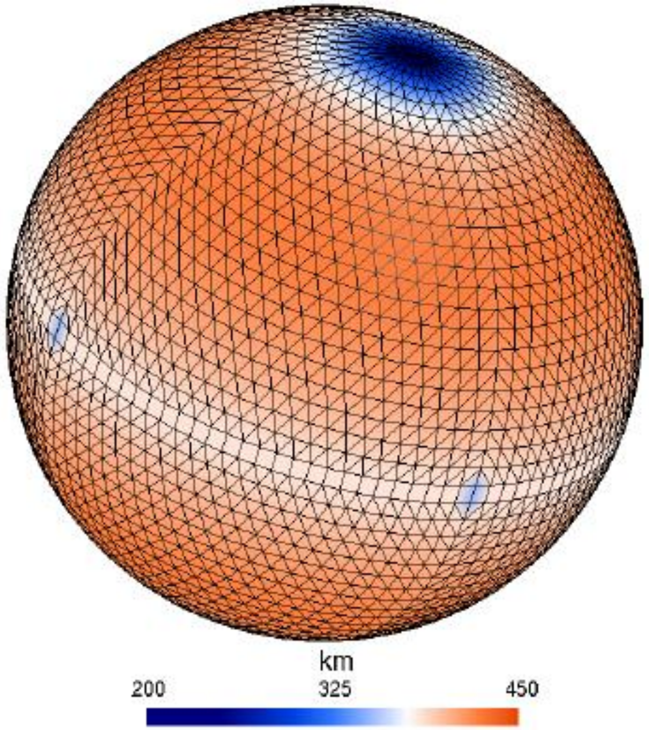
Train staff and talk about opportunities and limits of

Support weather and climate modelling community.

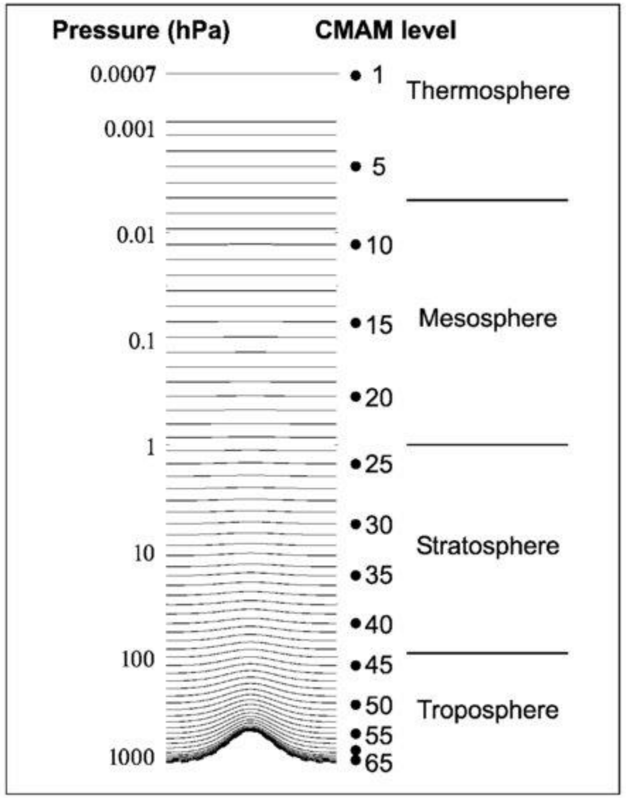
7-10: Destination Earth
and the [Digital Twin of the Earth]

Progress needed! How to do multi-scale modelling on unstructured grids?





Source: Willem Deconinck



Source: Polavarapu et al. 2005

Longitude/latitude (easy but inefficient) **vs. reduced Gaussian cubic octahedral** (unstructured) grid

Problem: Find a three-dimensional machine learning solution to that can work on unstructured grids.

Solution: ???
Maybe Geometric deep learning and Graph Neural Networks,
see Master Thesis of Icíar Lloréns Jover @ EPFL (https://infoscience.epfl.ch/record/278138)

Conclusions

- There are a large number of application areas throughout the prediction workflow in weather and climate modelling for which machine learning could really make a difference.

- The weather and climate community is still only at the beginning to explore the potential of machine learning (and in particular deep learning).

- Machine learning could not only be used to improve models, it could also be used to make them more efficient on future supercomputers.

Many thanks! **Peter.Dueben@ecmwf.int** **@PDueben**

ECMWF | EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

The strength of a common goal

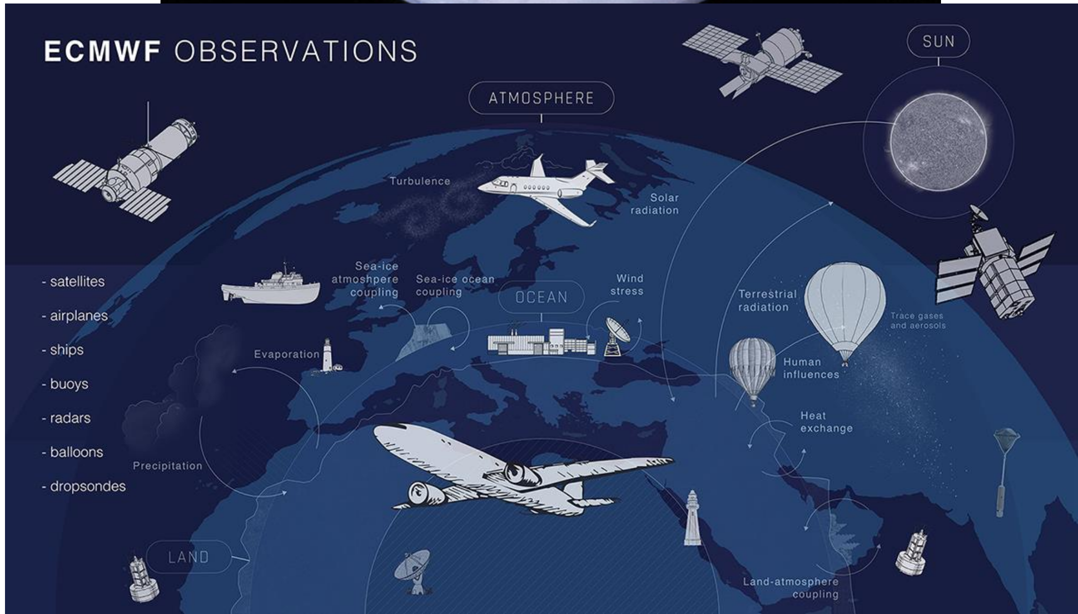
Why would machine learning help in weather predictions?

Predictions of weather and climate are difficult:

- The Earth is huge, resolution is limited and we cannot represent all important processes within model simulations
- The Earth System shows “chaotic” dynamics which makes it difficult to predict the future based on equations
- Some of the processes involved are not well understood
- All Earth System components (atmosphere, ocean, land surface, cloud physics,…) are connected in a non-trivial way

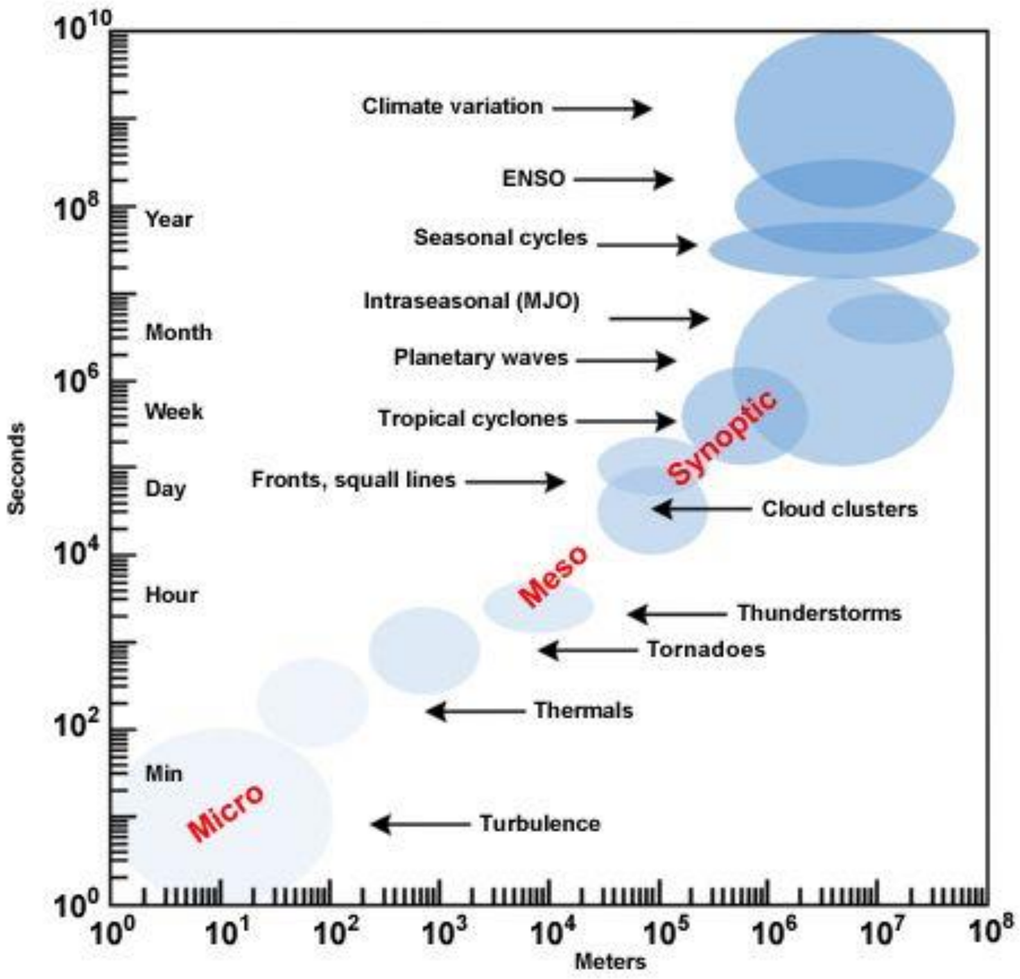
However, we have a huge number of observations and Earth System data

- ➢ **There are many application areas for machine learning in numerical weather predictions**
- ➢ **Machine learning also provides a number of opportunities for high performance computing**



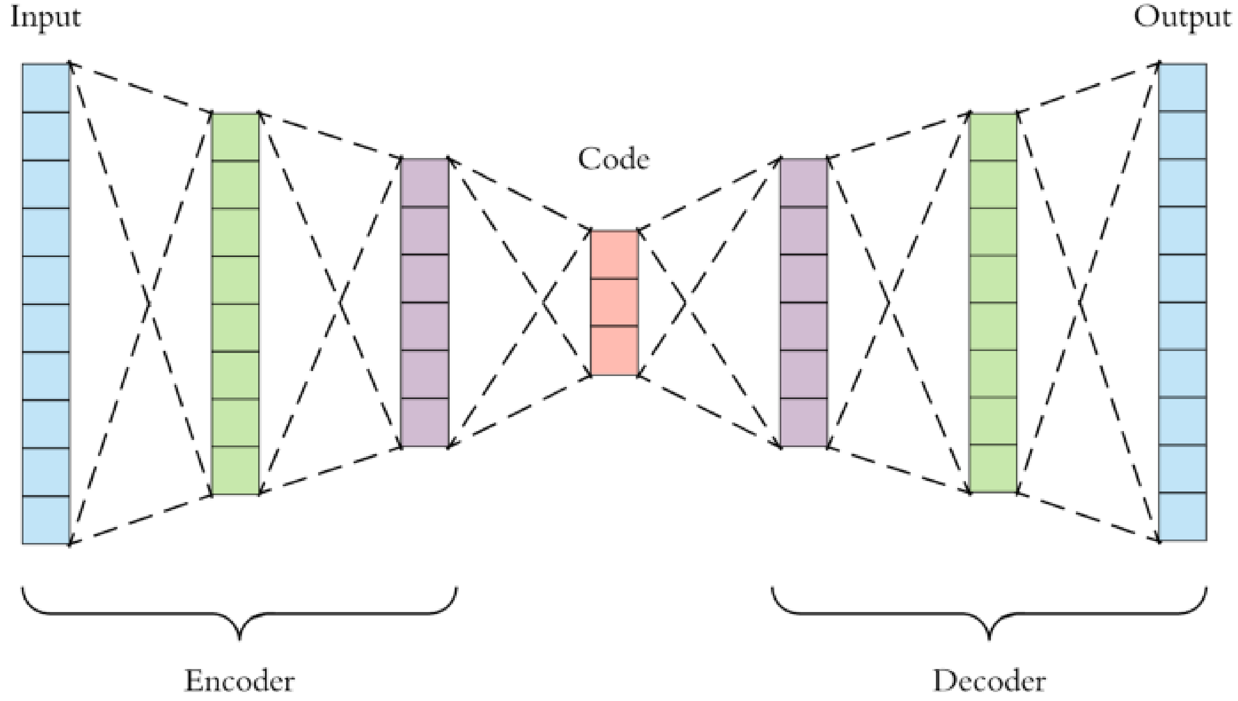
Scale interactions of machine learning solution for weather and climate

Weather and climate modelling:
Tools need to allow for scale interactions



Source: UCAR

Machine learning:
Neural network tools allow for encoding/decoding structures



Source: https://towardsdatascience.com

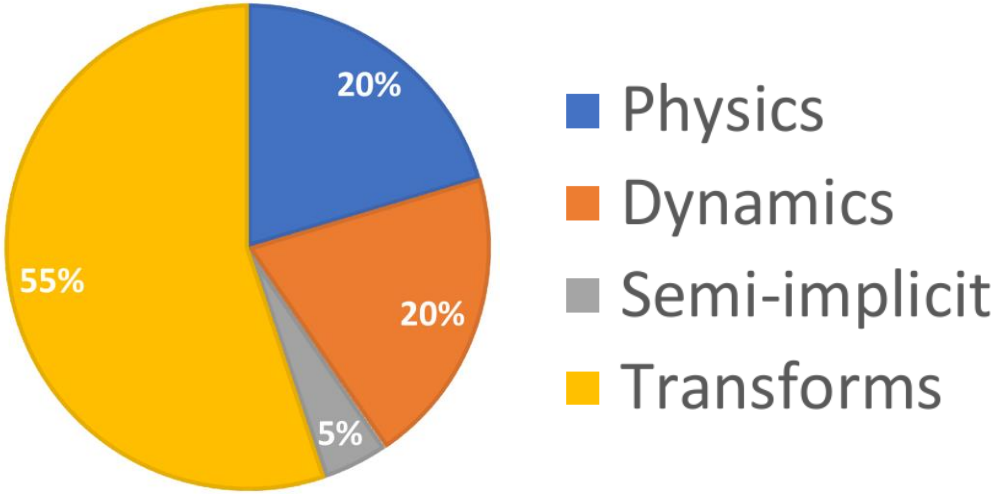
Can we use encoder/decoder networks to represent scale interactions?

Can we use deep learning hardware for conventional models?

- Machine learning accelerators are focussing on low numerical precision and high floprats.
- Example: TensorCores on NVIDIA Volta GPUs are optimised for half-precision matrix-matrix calculations with single precision output.
  → 7.8 TFlops for double precision vs. 125 TFlops for half precision

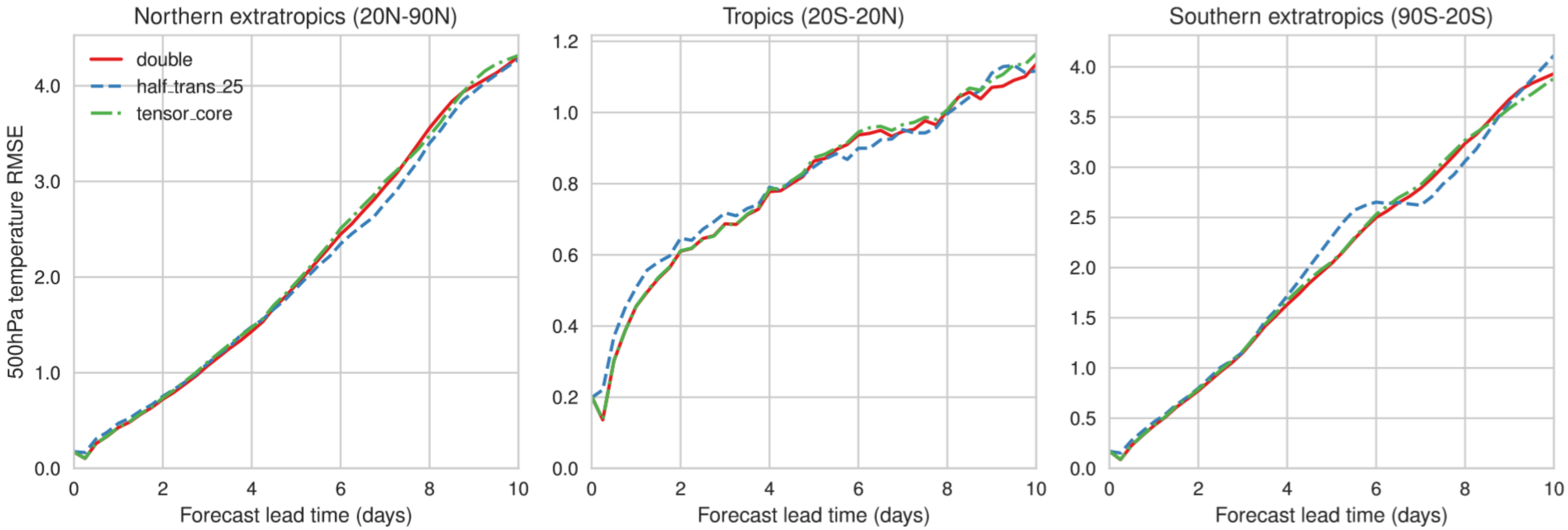
Can we use TensorCores within our models?

Relative cost for model components for a non-hydrostatic model at 1.45 km resolution:



- The Legendre transform is the most expensive kernel. It consists of a large number of standard matrix-matrix multiplications.
- If we can re-scale the input and output fields, we can use half precision arithmetic.

Half precision Legendre Transformations



Root-mean-square error for geopotential height at 500 hPa at 9 km resolution averaged over multiple start dates. *Hatfield, Chantry, Dueben, Palmer Best Paper Award PASC2019*

The simulations are using an emulator to reduce precision (*Dawson and Dueben GMD 2017)* and more thorough diagnostics are needed.

Some room for interactions with machine learning efforts at ECMWF

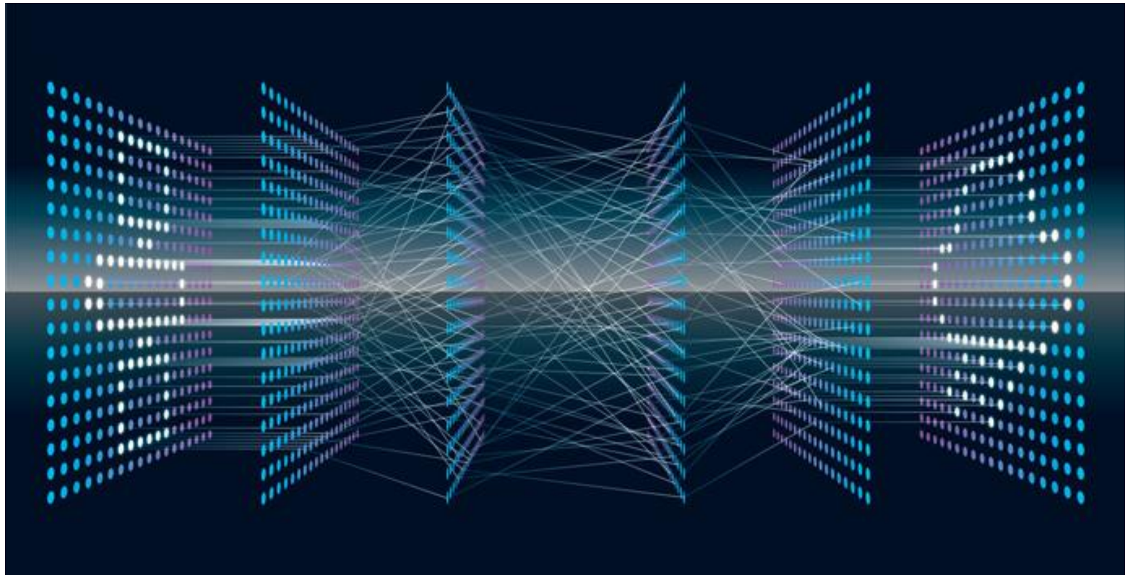
We have also started a special seminar series on Machine Learning that is broadcasted.

ECMWF is developing the European Weather Cloud in collaboration with EUMETSAT.

Our MAELSTROM EuroHPC proposal was successful which will allow us to develop customised machine learning solutions for weather and climate models.

We are hiring!

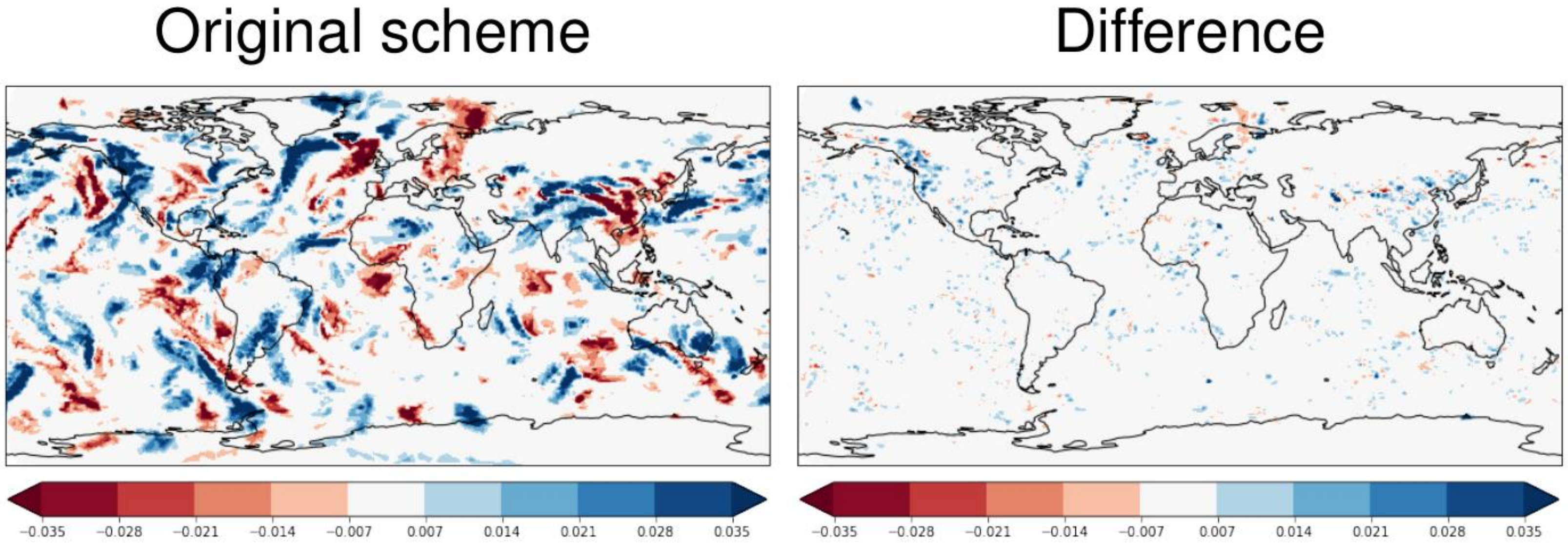
MAELSTROM

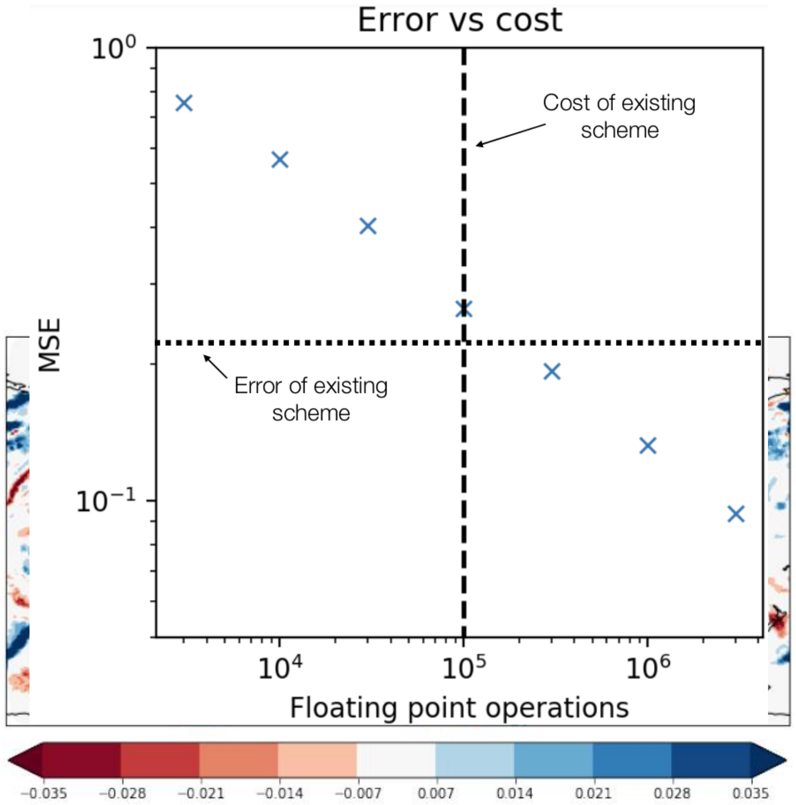


Numerical weather forecasts: To emulate gravity wave drag

- Repeat the same approach for the gravity wave drag scheme of IFS
- Start with non-orographic and continue with orographic wave drag

Results for the non-orographic gravity wave drag are promising.





There is also a nice relation between network size and accuracy.

However, it is still questionable whether computational performance of the Neural Nets is better when compared to the conventional scheme.

Results are not as good for the orographic gravity wave drag scheme.

Chantry, Dueben, Palmer

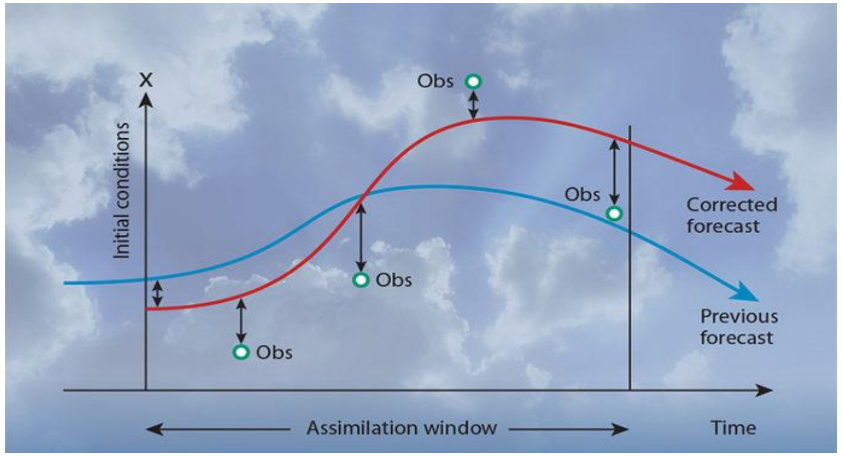
Scientific challenges for machine learning in numerical weather predictions

There is no fundamental reason not to use a black box within weather and climate models but there are unanswered questions.

- Can we use our knowledge about the Earth System to improve machine learning tools?
- Can we diagnose physical knowledge from the machine learning tools?
- Can we remove errors from neural networks and secure conservation laws?
- Can we guarantee reproducibility?
- Can we find the optimal hyper-parameters?
- Can we efficiently scale machine learning tools to high performance computing applications?
- Can we interface machine learning tools with conventional models?
- Can we design good training data (short time steps and high resolution, labelled datasets)?
- Can we explore the full phase space (all weather regimes) during training?

Many scientists are working on these challenges as we speak.

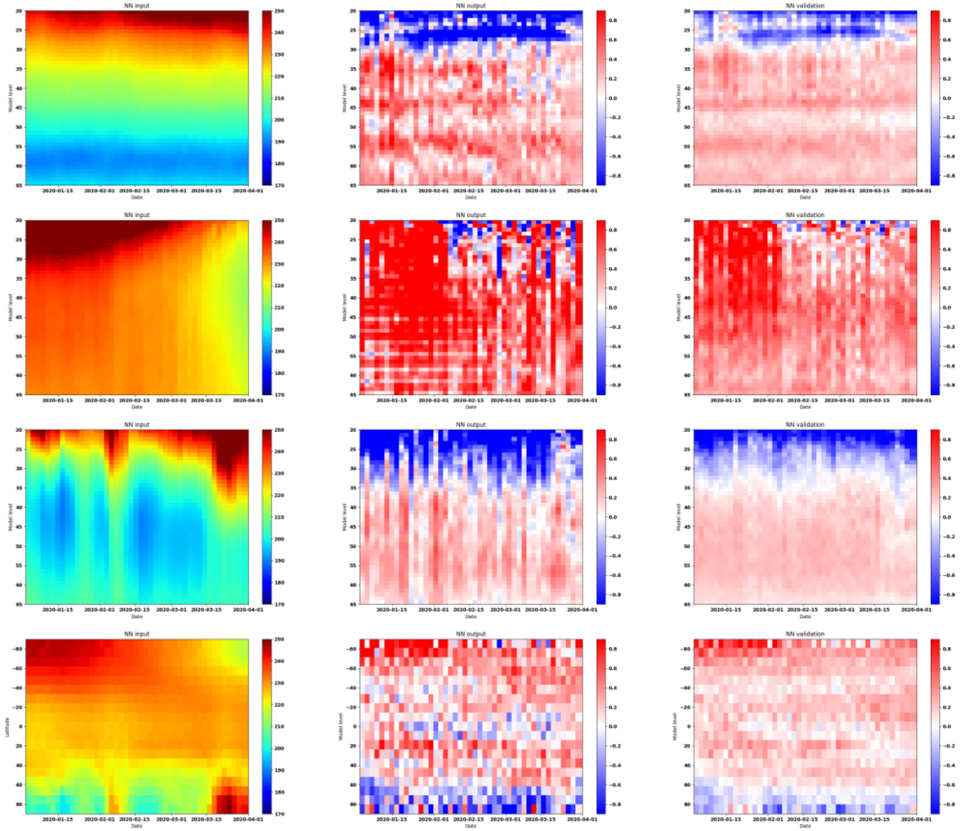
Data assimilation: Bias-correct the forecast model in 4DVar data assimilation



- Data-assimilation blends observations and the forecast model to generate initial conditions for weather predictions
- During data-assimilation the model trajectory is "synchronised" with observations for the same weather regimes
- It is possible to learn model error when comparing the model with (trustworthy) observations

Two approaches:

- Learn model error within the 4DVar data-assimilation framework for so-called "weak-constraint 4D-Var"
- Learn model error from a direct comparison of the model trajectory to observations or analysis increments using deep learning (column-based or with three-dimensional)



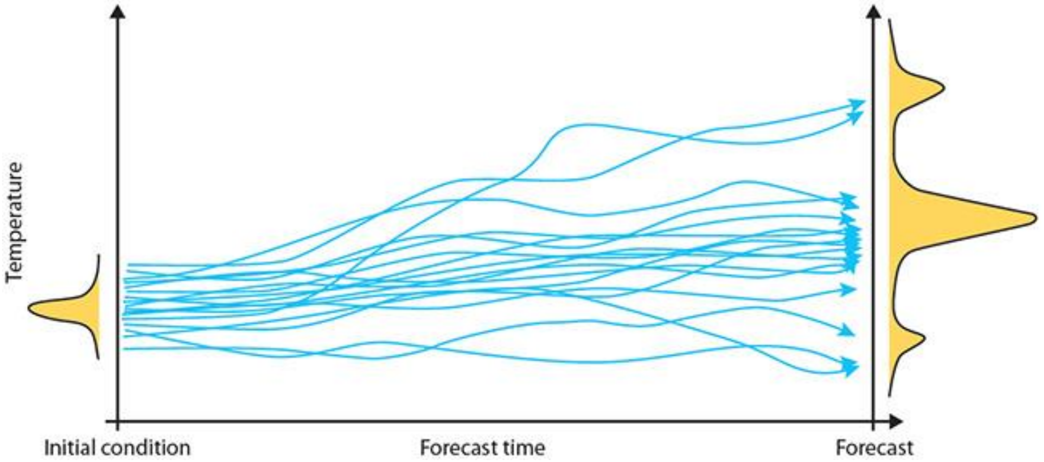
Benefit:

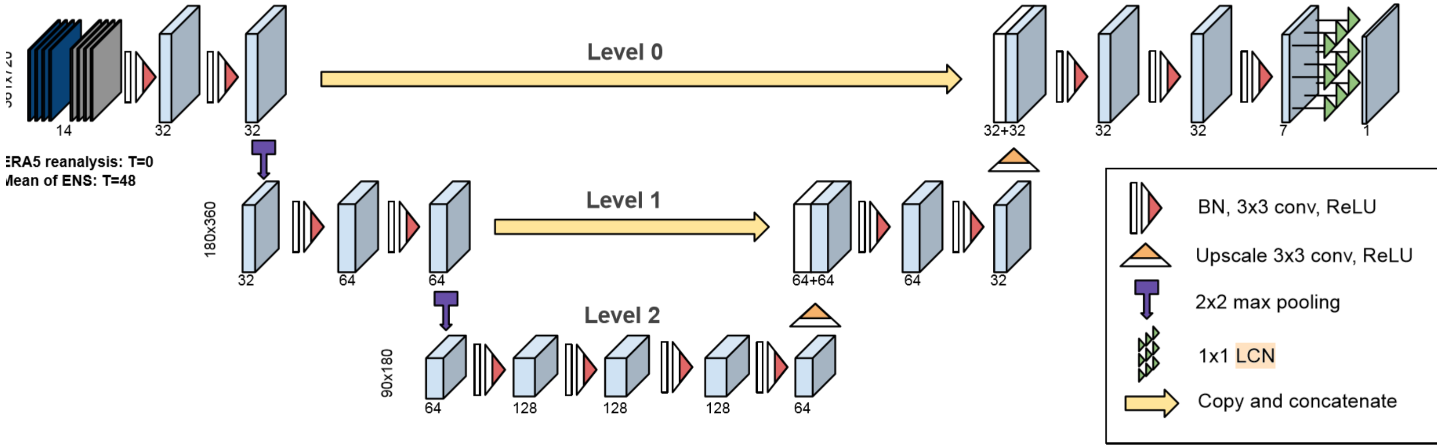
When the bias is learned, it can be used to:

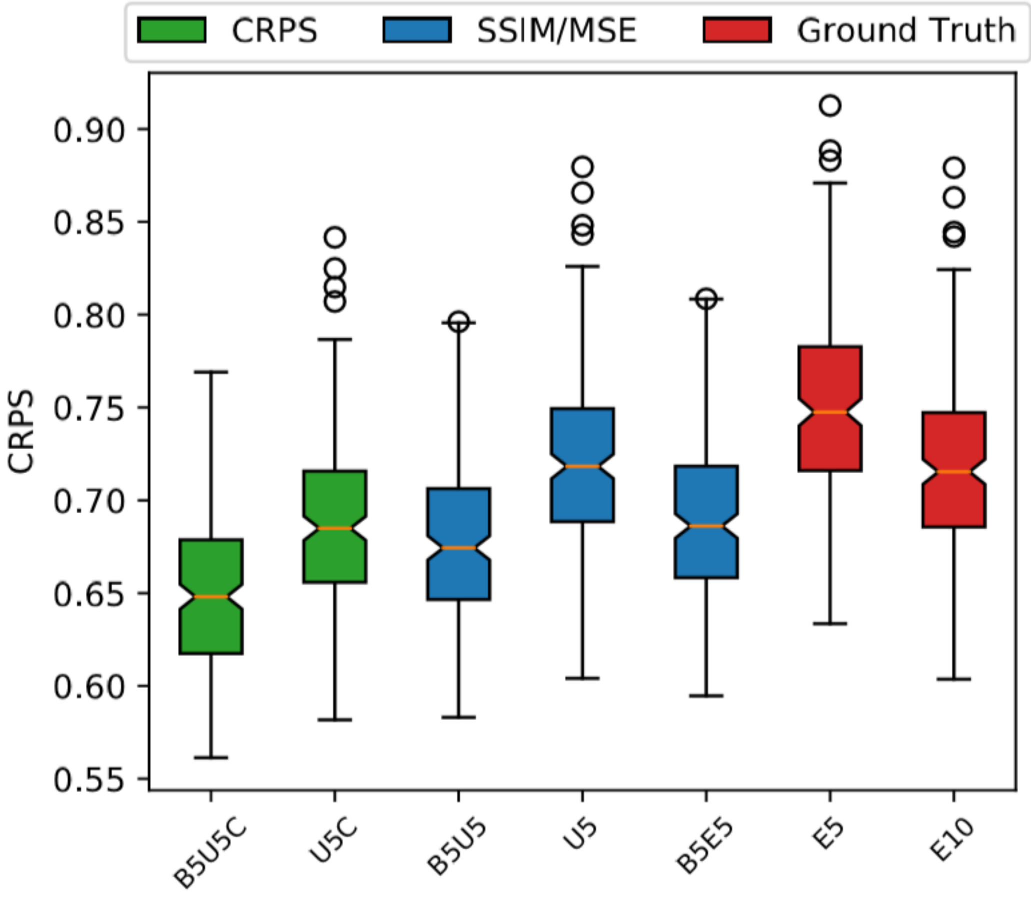
- Correct for the bias during data-assimilation to improve initial conditions
- Correct for the bias in forecast simulations to improve predictions (discussed controversially)
- Understand model deficiencies

Patrick Laloyaux, Massimo Bonavita and Peter Dueben @ ECMWF + Thorsten Kurth and David Matthew Hall @ NVIDIA

Post-processing and dissemination: Improve ensemble predictions







(a) T850

Ensemble predictions are important but expensive.

Can we improve ensemble skill scores from a small number of ensemble members via deep learning?

- Use global fields of five ensemble members as inputs.
- Correct the ensemble scores of temperature at 850 hPa and Z500 hPa for a 2-day forecast towards a full 10 member ensemble forecast.

Grönquist, Yao, Ben-Nun, Dryden, Dueben, Lavarini, Li, Hoefler https://arxiv.org/abs/2005.08748