Machine Learning for weather predictions at ECMWF

Peter Dueben

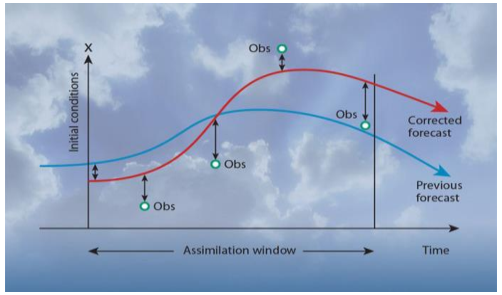
Royal Society University Research Fellow & ECMWF's Coordinator for Machine Learning and AI Activities

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.

Machine learning applications across the numerical weather prediction workflow

Observations → **Data-assimilation** → **Numerical weather forecasts** → **Post-processing and dissemination**

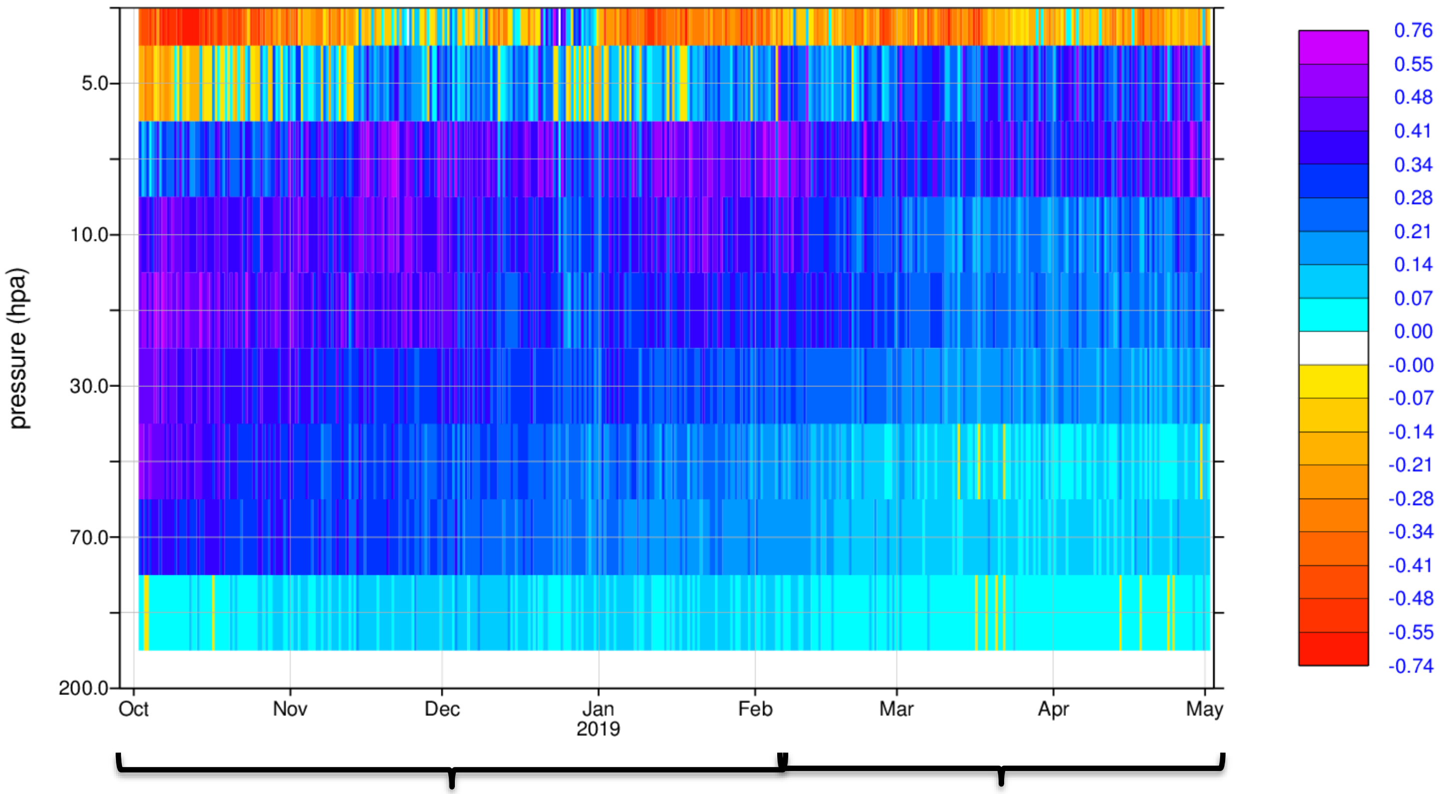


Application areas for machine learning are spread over the entire workflow:
weather data monitoring, real-time quality control for observational data, anomaly interpretation, guided quality assignment and decision making, data fusion from different sources, correction of observation error,
learn governing differential equations, non-linear bias correction, bias predictors, learn operational operators, define optical properties of hydrometeors and aerosols, emulate conventional tools improve efficiency,
emulate model components, develop improved parametrisation schemes, build better error models, learn the underlying equations of motion, generate tangent linear or adjoint code from machine learning emulators,
real-time adjustments of forecast products, feature detection, uncertainty quantification, error corrections for seasonal predictions, development of low-complexity models, bespoke products for business opportunities,
and many more…

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
- This requires estimates of errors of observations and the forecast model
- The new weak-constraint 4D-Var algorithm learns that the model consistently underestimates temperature between 100hPa and 10hPa
- We learn a forcing to correct for the systematic model error
- We still use fairly simple machine learning techniques but we have started to investigate deep learning approaches together with NVIDIA

Mean first-guess departure with respect to GPS-RO temperature retrievals



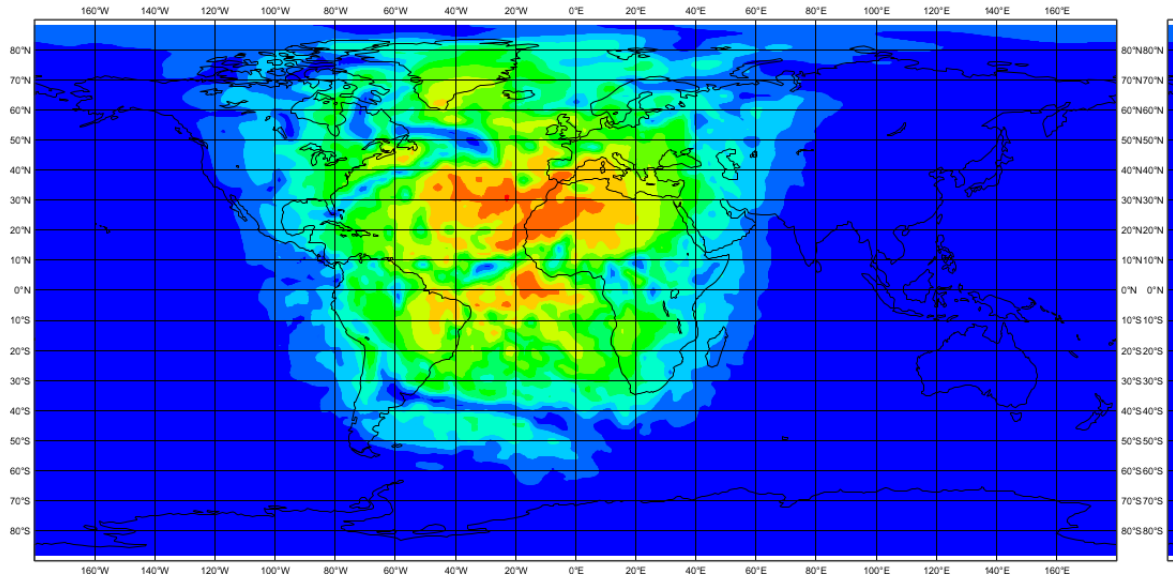
Patrick Laloyaux and Massimo Bonavita

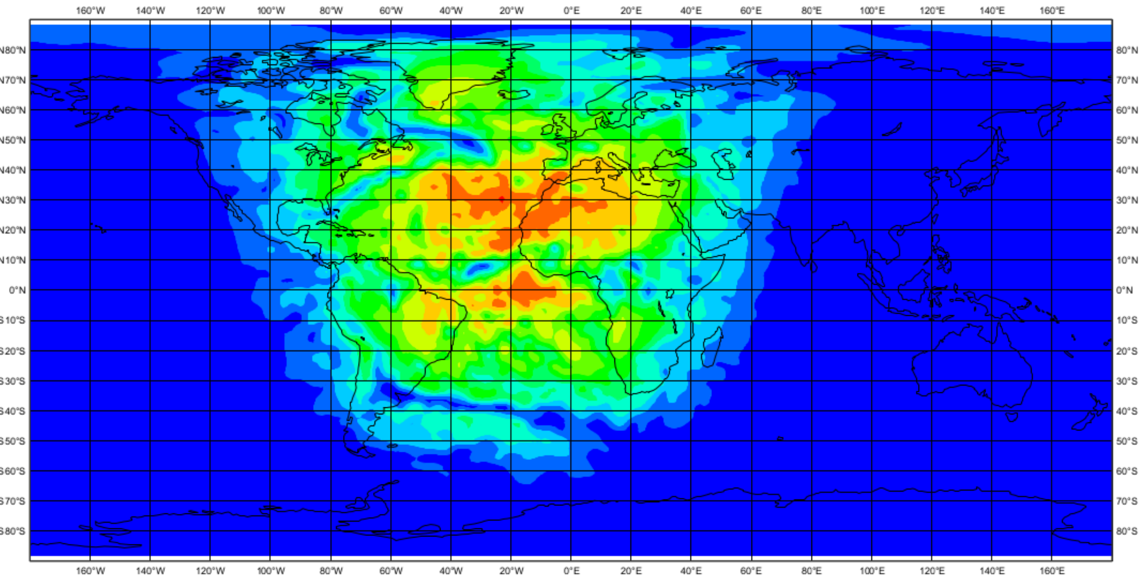
Numerical weather forecasts: To emulate the radiation scheme

- 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:
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





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

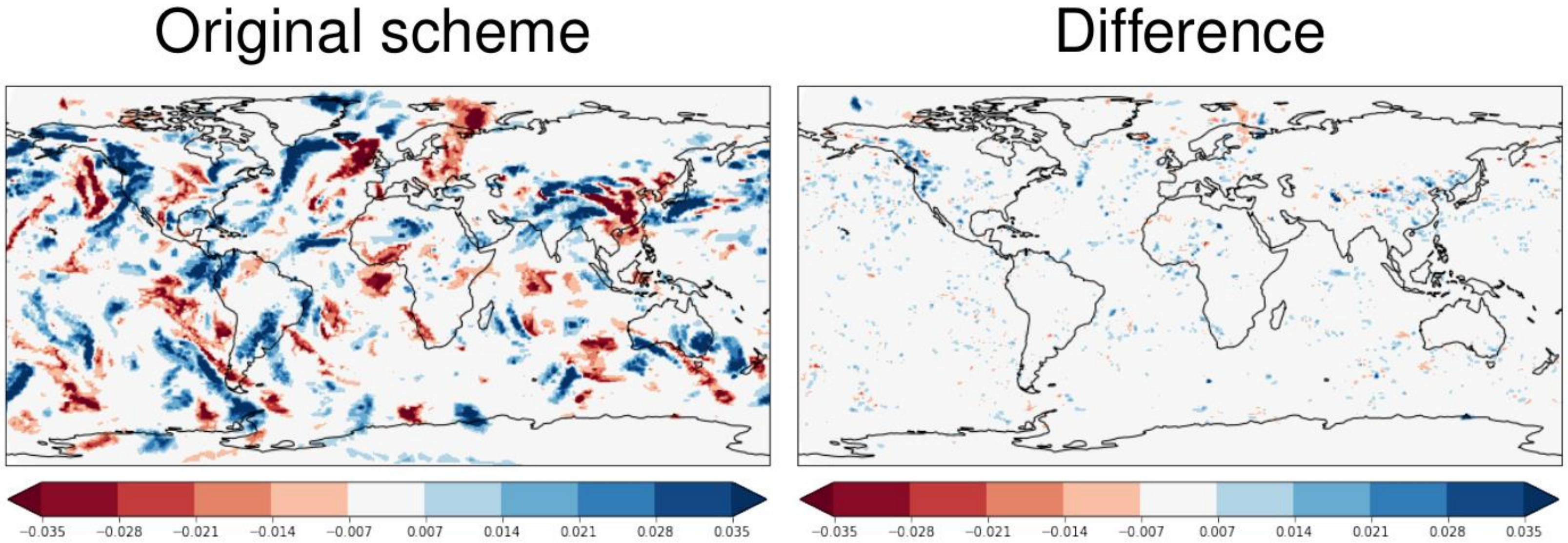
The approach is working and the neural network is ~10 times faster than the original scheme.
However, model results are still degraded.

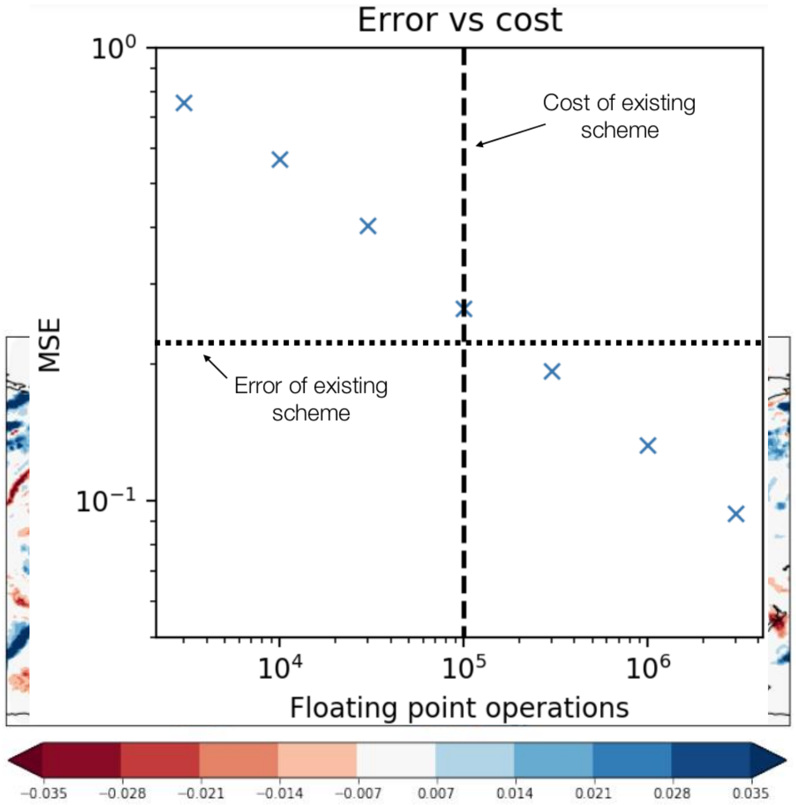
Dueben, Hogan, Bauer @ECMWF and Progsch, Angerer @NVIDIA

Numerical weather forecasts: To emulate 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

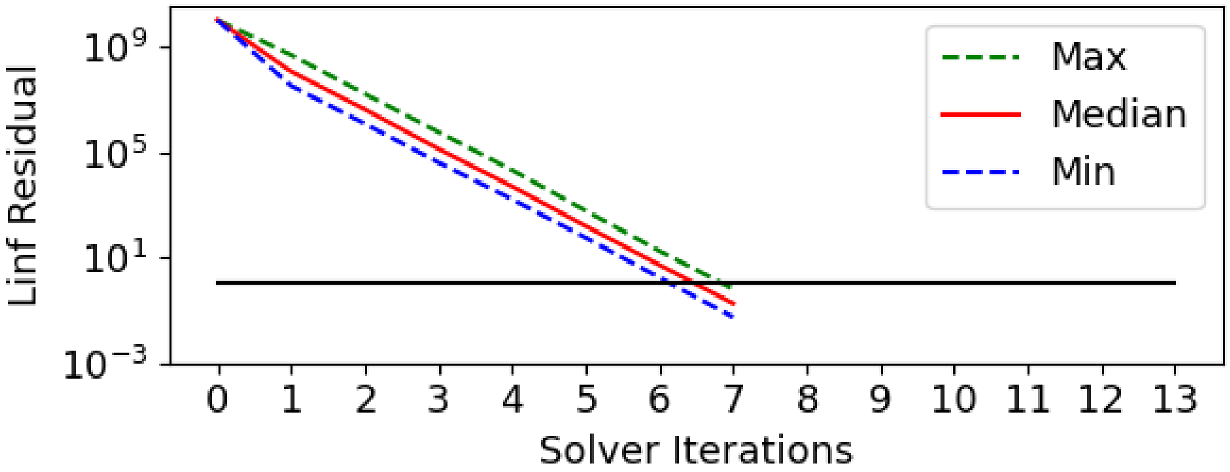
Numerical weather forecasts: To precondition the linear solver

- 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.

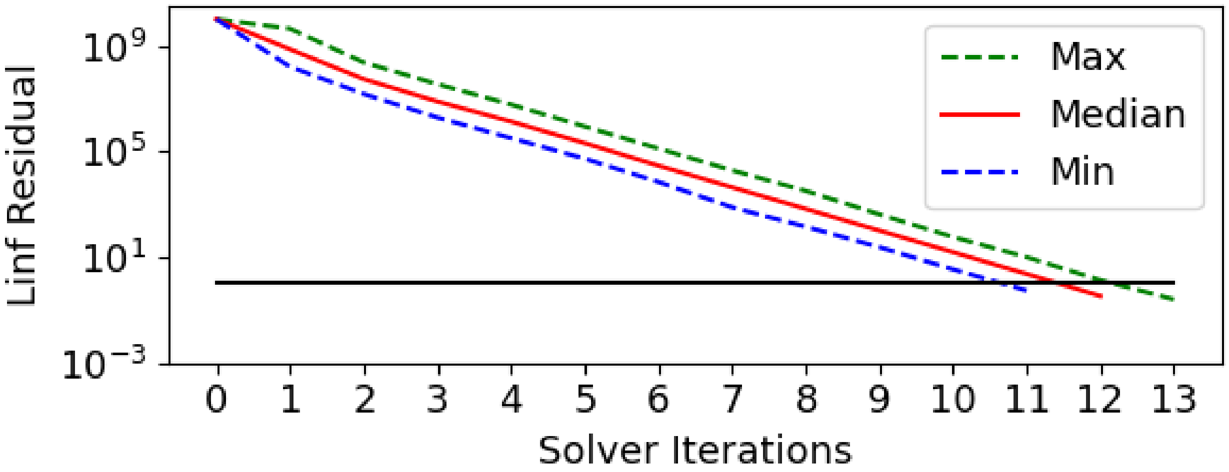
Can we use machine learning for preconditioning, predict the inverse of the matrix and reduce the number of iterations that are required for the solver?

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.

Machine learning preconditioner:



No 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

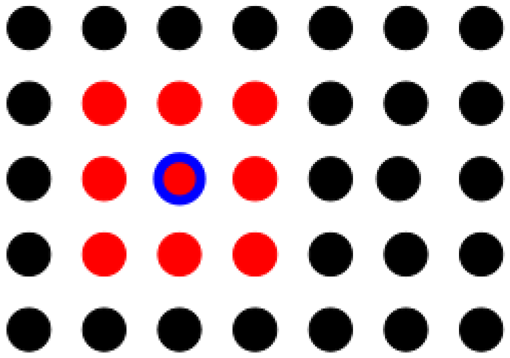
What is the limit?
Can we replace the entire forecast system?

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 210 petabyte of data

A simple test configuration:

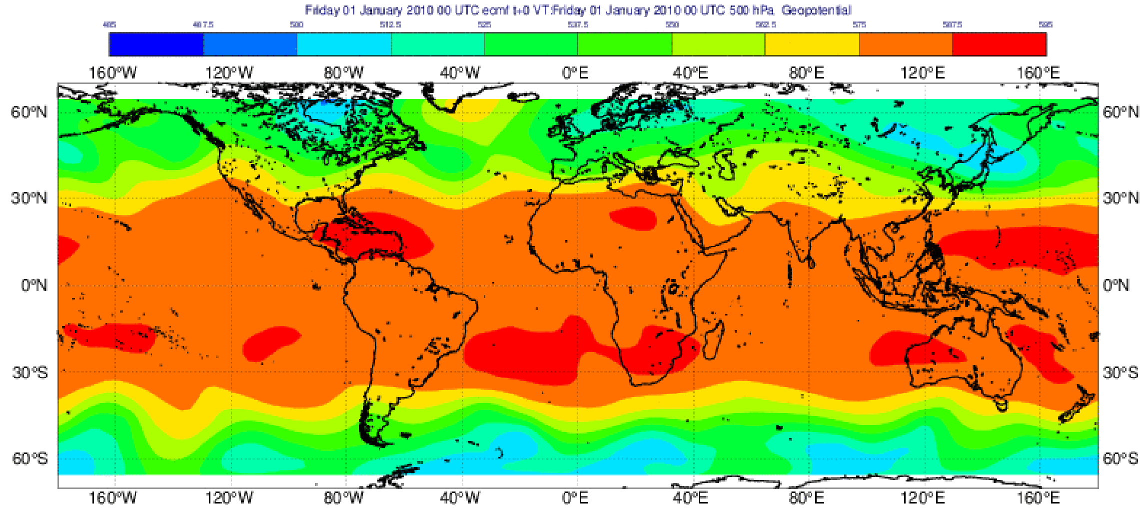
- We retrieve historical data (ERA5) for geopotential at 500 hPa (Z500) for the last decades (>65,000 global data sets)
- We map the global data to a coarse two-dimensional grid (60x31)
- We learn to predict the update of the field from one hour to the next using deep learning
- Once we have learned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

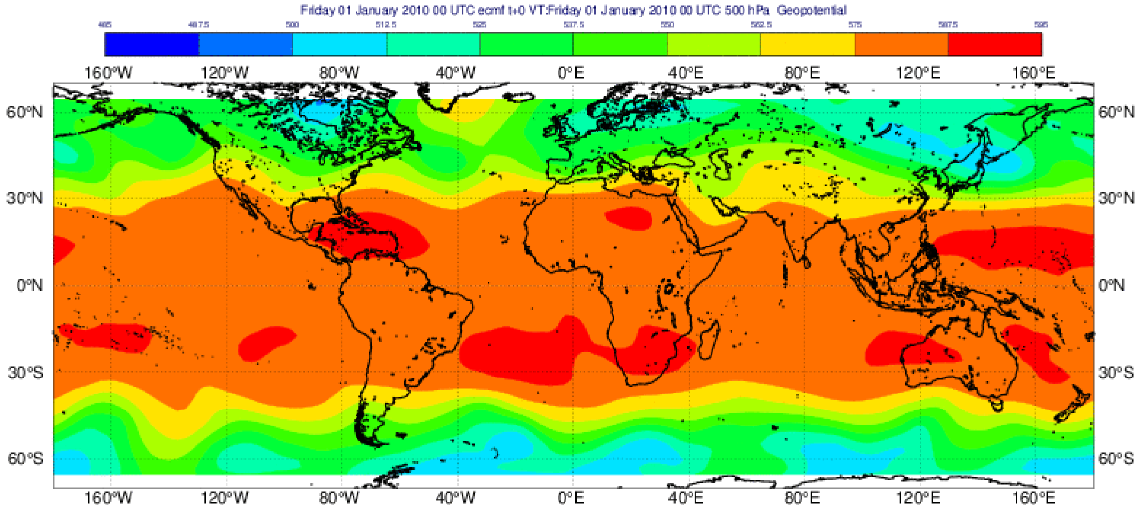
No physical understanding is required!



Dueben and Bauer GMD 2018

What is the limit? Can we replace the entire forecast system?



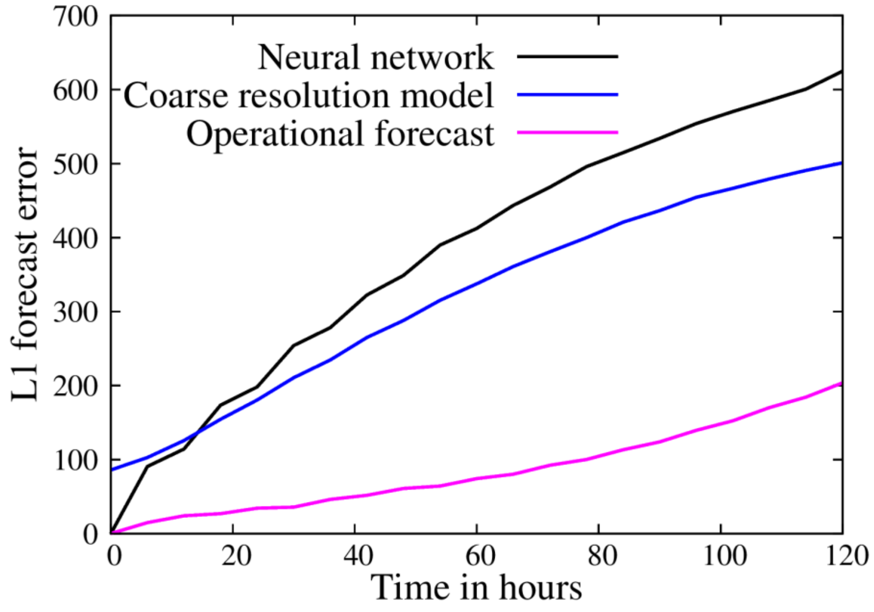


Time evolution of Z500 for historic data and a neural network prediction.
Can you tell which one is the neural network?

- The neural network is picking up the dynamics nicely.
- Forecast errors are comparable if we compare like with like.
- Is this the future?

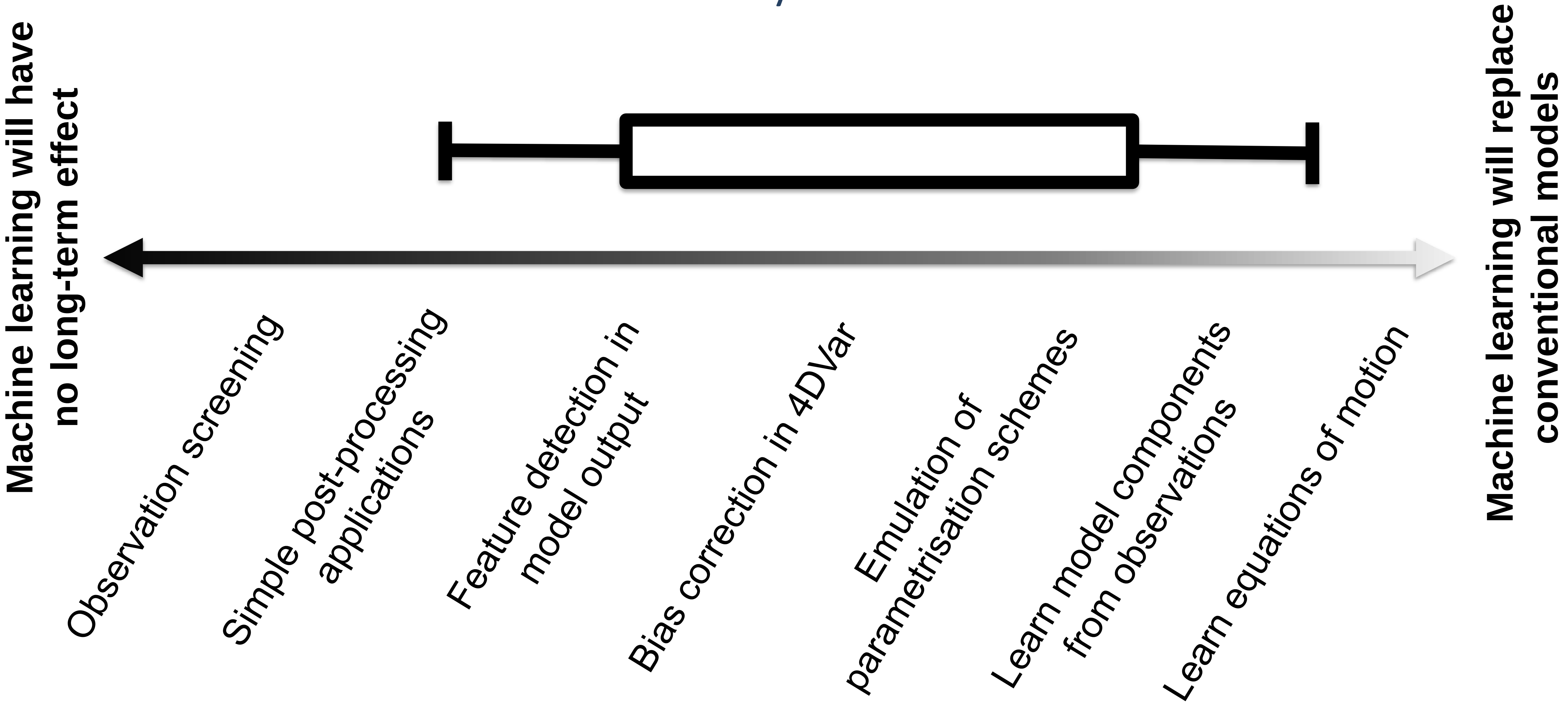
Unlikely…
The simulations are unstable.
It is unknown how to increase complexity.
There are only ~40 years of data available.

However, there is a lot of progress at the moment:
Scher and Messori GMD 2019; Weyn, Durran, and Caruana JAMES 2019; **…**



Dueben and Bauer GMD 2018

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



The uncertainty range is still very large...

Numerical weather forecasts: Low dimensional ocean models

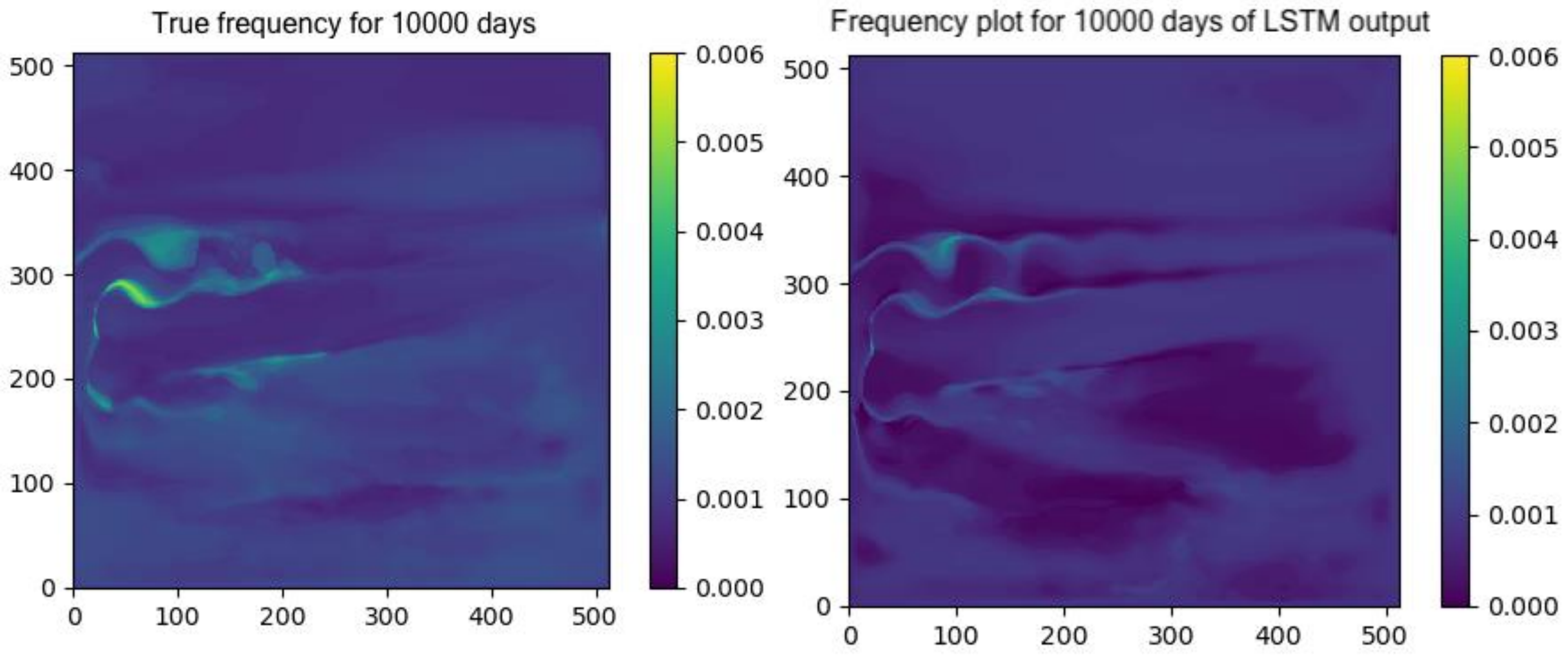
Motivation: We would like to build a low-dimensional ocean model for medium-range weather forecasts.

Testbed: 3-layer double gyre quasi-geostrophic model in rectangular domain.

Data: 40 year time series for the coefficients of the first eight Principle Components of the surface layer (daily data).

Approach: Use neural networks to learn the equations of motion.

Neural Network setup: LSTM with 2 hidden layers, 50 neurons/layer, Sigmoid activation and Adam optimizer.

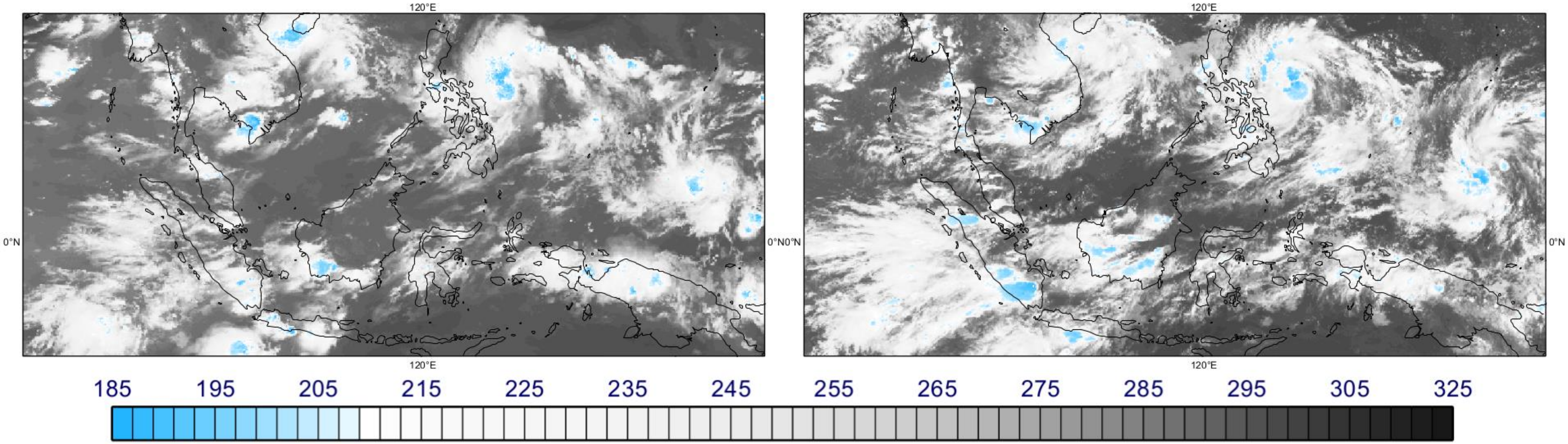


Agarwal, Dueben, Berloff, Ryzhov, Kondrashov

Please note that "conventional machine learning" with linear regression + red-noise performs even better.

Why is hard for machine learning tools to compete?

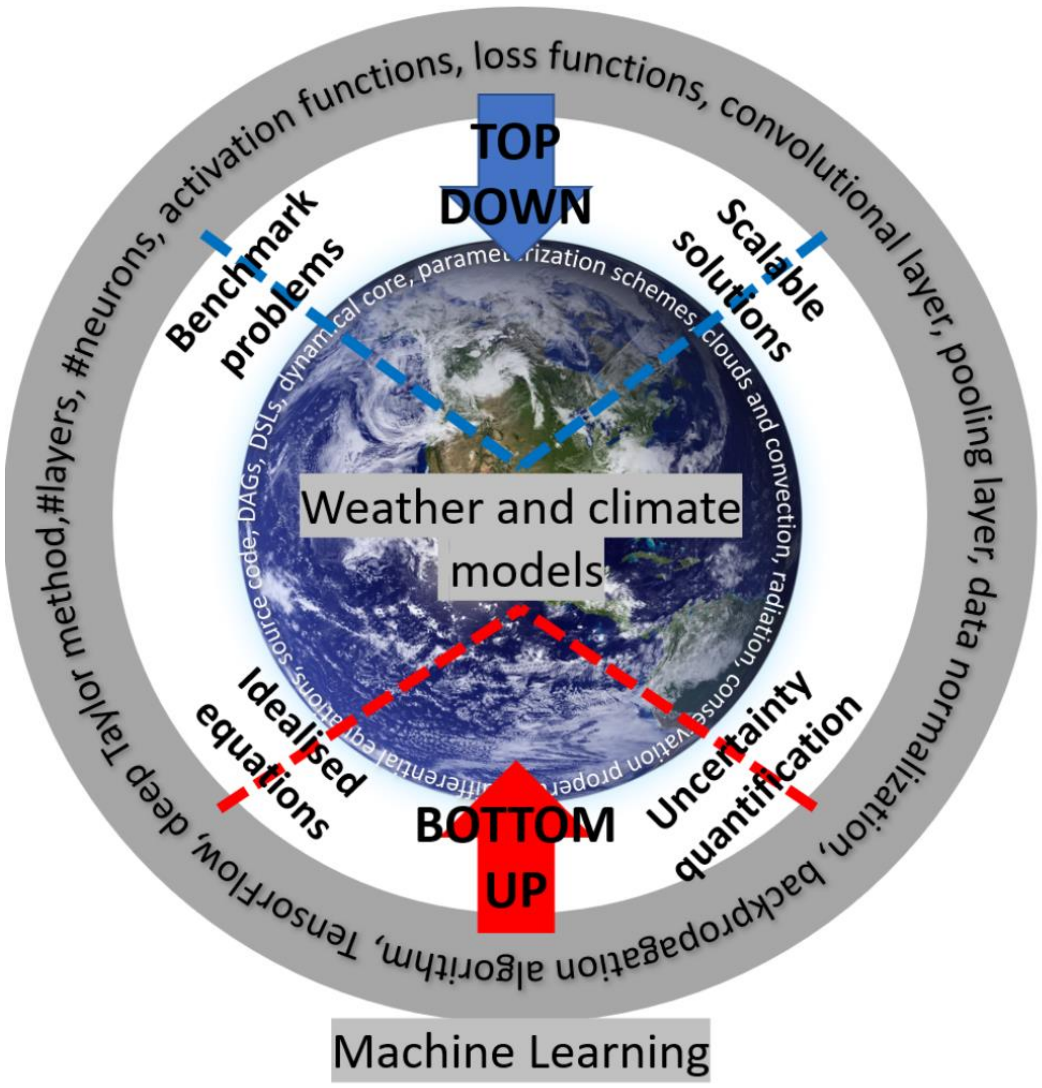
Because our models are astonishing!



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

A weather forecast simulation has O(1,000,000,000) degrees-of-freedom.

My personal vision of the way forward...



Idealised equations: To study known differential equations to learn how to derive blueprints for neural network architectures.

Uncertainty quantification: To study the representation of variability and the correction of systematic errors for neural networks.

Scalable solutions: To learn how to scale neural networks to millions of inputs for 3D fields on the sphere.

Benchmark problems: To build benchmark problems similar to ImageNet (*see WeatherBench in Rasp, Dueben, Scher, Weyn, Mouatadid and Thureey 2020*)

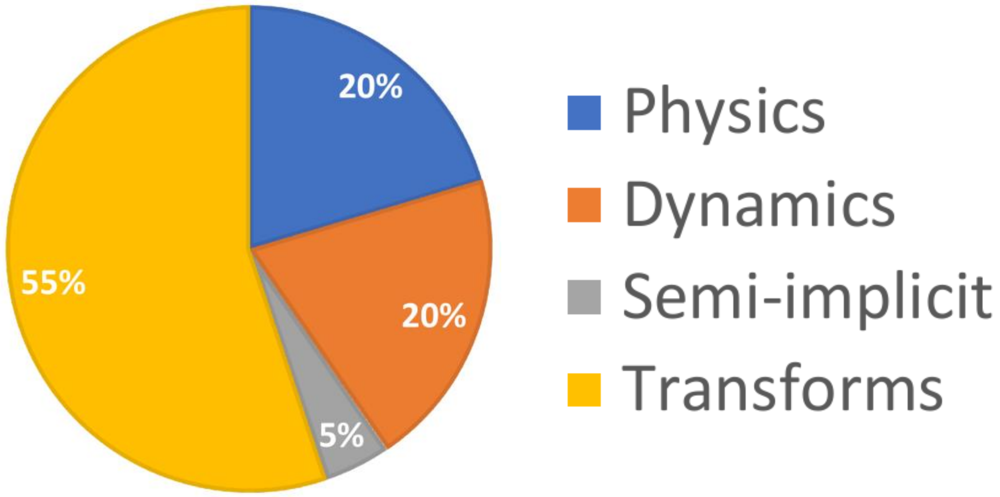
This will require machine learning solutions that are customised to weather and climate models.

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

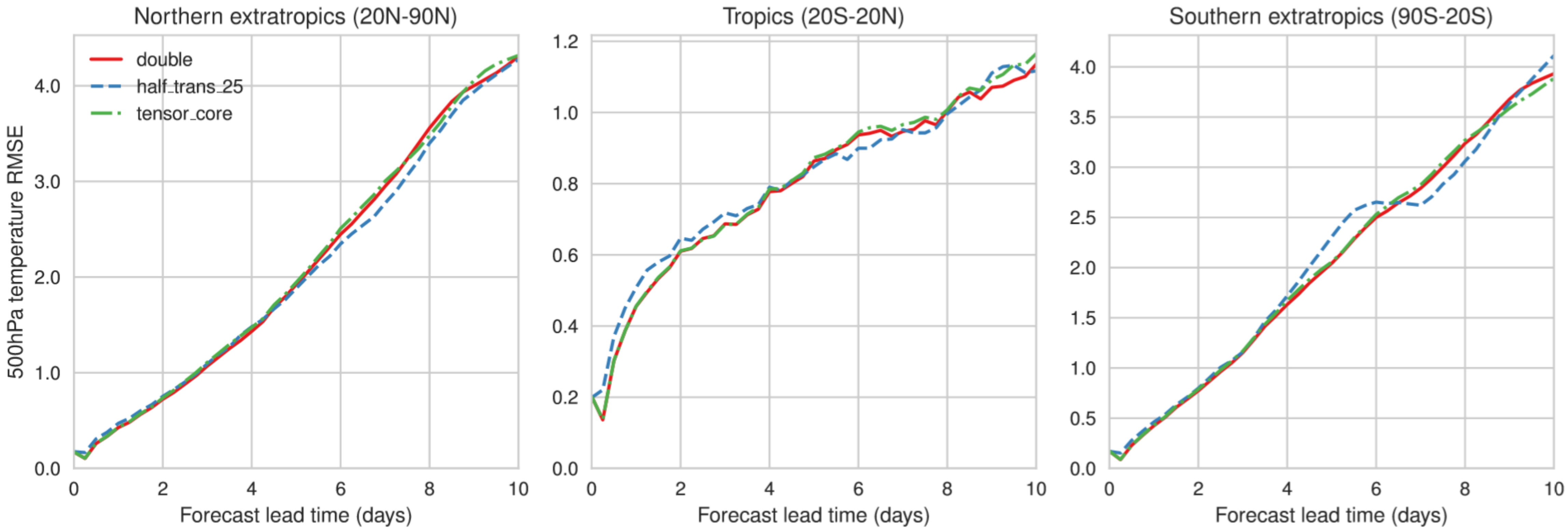
Can we use TensorCores within our models?

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.

Conclusions

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

- The weather and climate community is still 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.

- Machine learning accelerators could be useful to speed-up components of weather and climate models.

- However, there are limitations for the application of black-box solutions within weather and climate models and challenges that need to be addressed.

ECMWF-ESA Workshop on Machine Learning for Earth System Observation and Prediction at ECMWF 5-8 October 2020. More information is here.

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

Many thanks.

ECMWF | EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

The strength of a common goal