Machine learning for weather and climate predictions

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

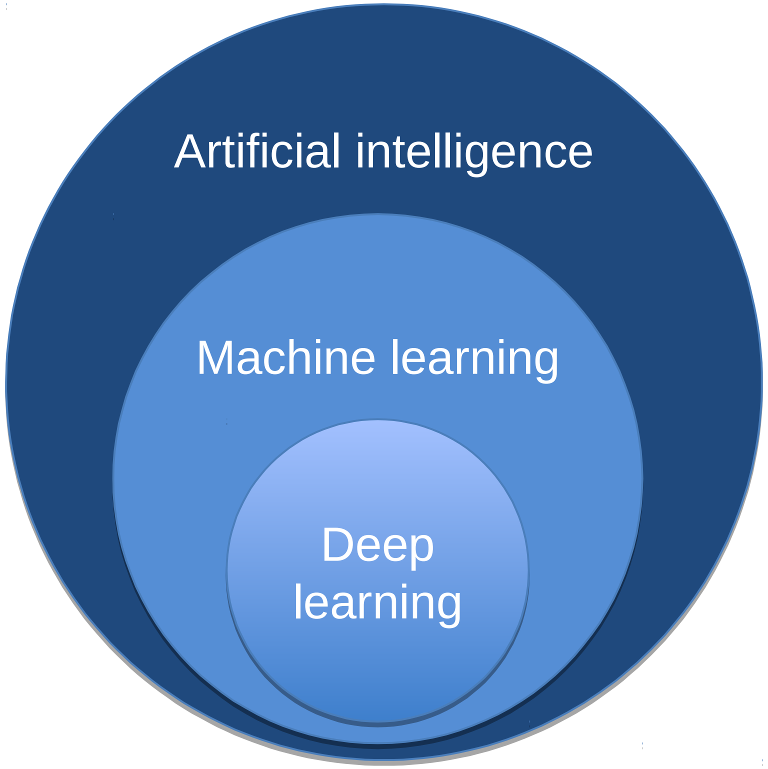
The ESIWACE2 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823988

Outline

- What is machine learning?
- How can we use machine learning to improve weather and climate modelling?
- A couple of examples for the use of machine learning in weather and climate predictions.
- Machine learning and high-performance computing.
- What are challenges for the use of machine learning?

Peter.Dueben@ecmwf.int **@PDueben**

Let’s start with definitions



Artificial intelligence (AI) *is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans (Wikipedia)*
Example: A self-driving car stops as it detects a cyclist crossing

Machine learning (ML) *is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions… (Wikipedia)*
Example: To learn to distinguish between a cyclist and other things from data

Deep learning *is part of a broader family of machine learning methods based on artificial neural networks (Wikipedia)*
Example: The technique that is used to detect a cyclist in a picture

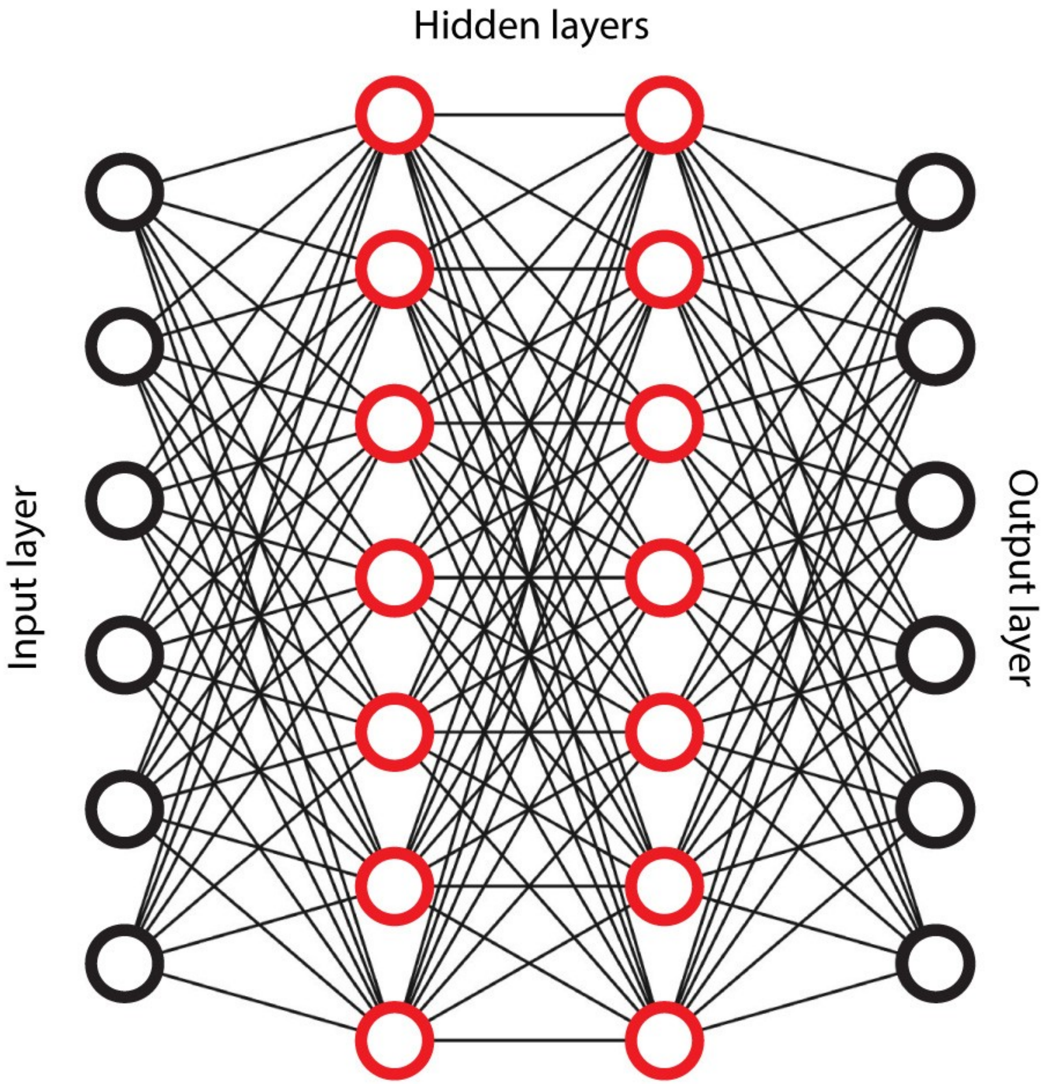
Deep learning and artificial neural networks as one example of machine learning

The concept:
Take input and output samples from a large data set
Learn to predict outputs from inputs
Predict the output for unseen inputs

The key:
Neural networks can learn a complex task as a “black box”
No previous knowledge about the system is required
More data will allow for better networks

The number of applications is increasing by day:
Image recognition
Speech recognition
Healthcare
Gaming
Finance
Music composition and art
…

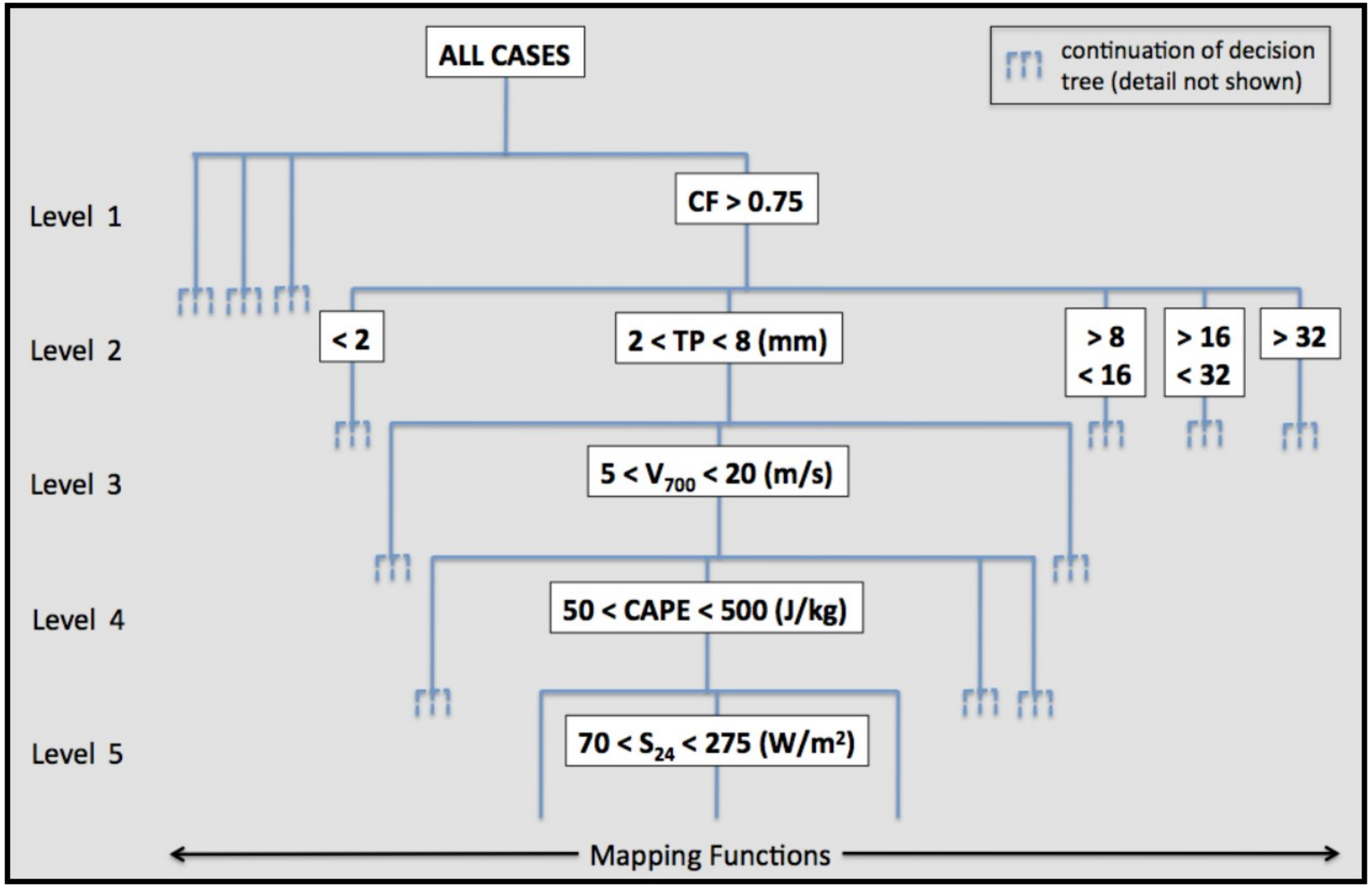
And weather?



Decision trees and random forrests

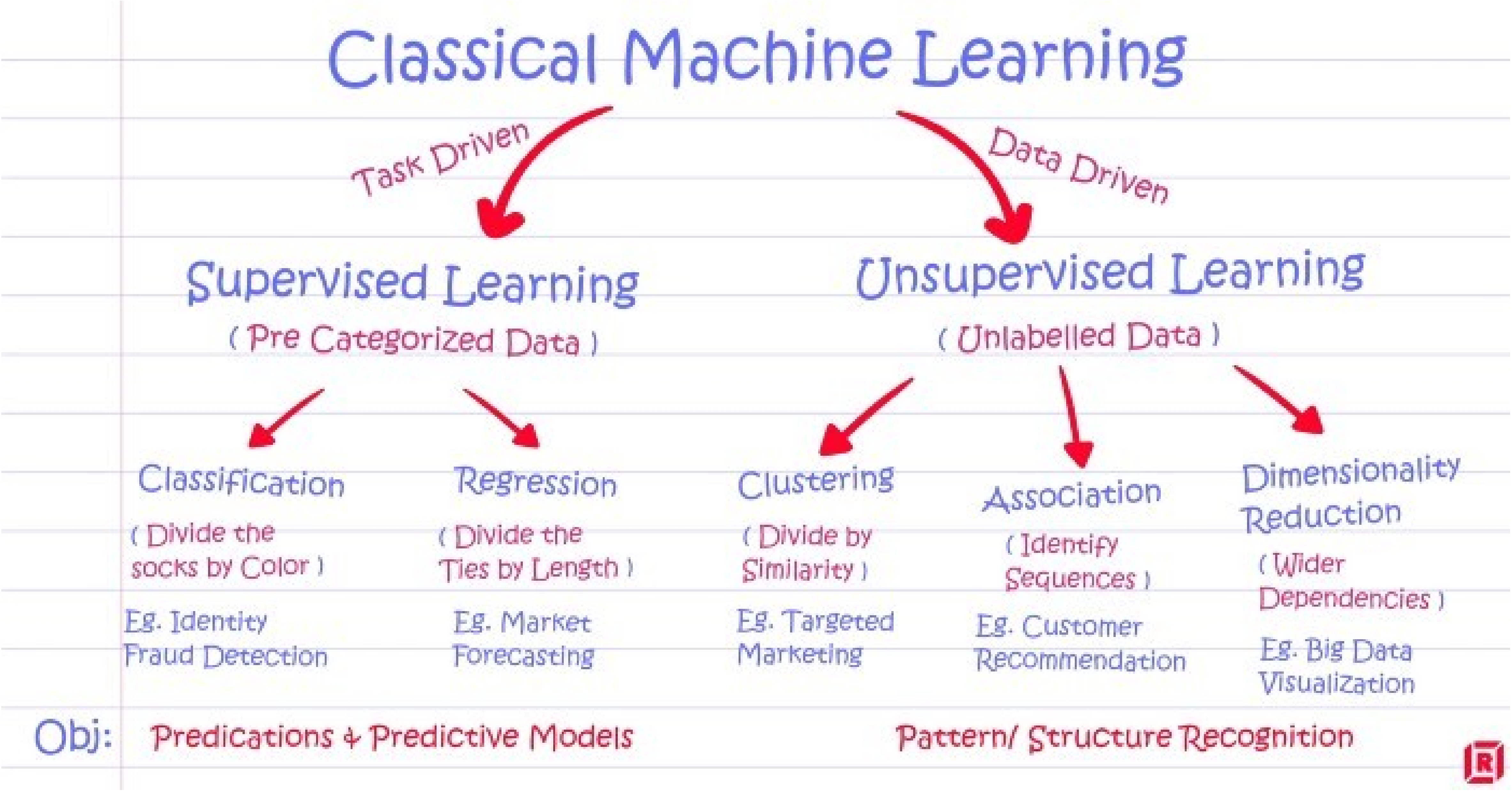
- Decisions fork in tree structures until a prediction is made.
- “Random forest” methods are training a multitude of decision trees using a mean predictions or the value with the most hits as a result.
- Decision trees are often fast and accurate and they are able to conserve some of the properties of the system.
- Decision trees often require a lot of memory (as they serve as an efficient look-up table).

An example for ecPoint:



Hewson and Pillosu 2020

Two families of machine learning



Source: https://medium.com/@recrosoft.io/supervised-vs-unsupervised-learning-key-differences-cdd46206cdcb

Why is there a sudden hype about machine learning?

Deep learning is fairly new but machine learning has been used for weather predictions for many years.

Machine learning is no niche application anymore because…

- …of an explosion of supercomputing performance and data availability
- …of success stories of machine learning tools beating humans
  (counting cars in pictures, translating text, playing Jeopardy, playing chess or Go…)
- …of powerful software tools that make it easy for non-experts
  to develop machine learning applications
- …of breath-taking developments in the design of new machine learning tools
- …artificial intelligence is a trillion $ market

Machine learning can learn the behaviour of extremely complex systems if enough data is available.

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

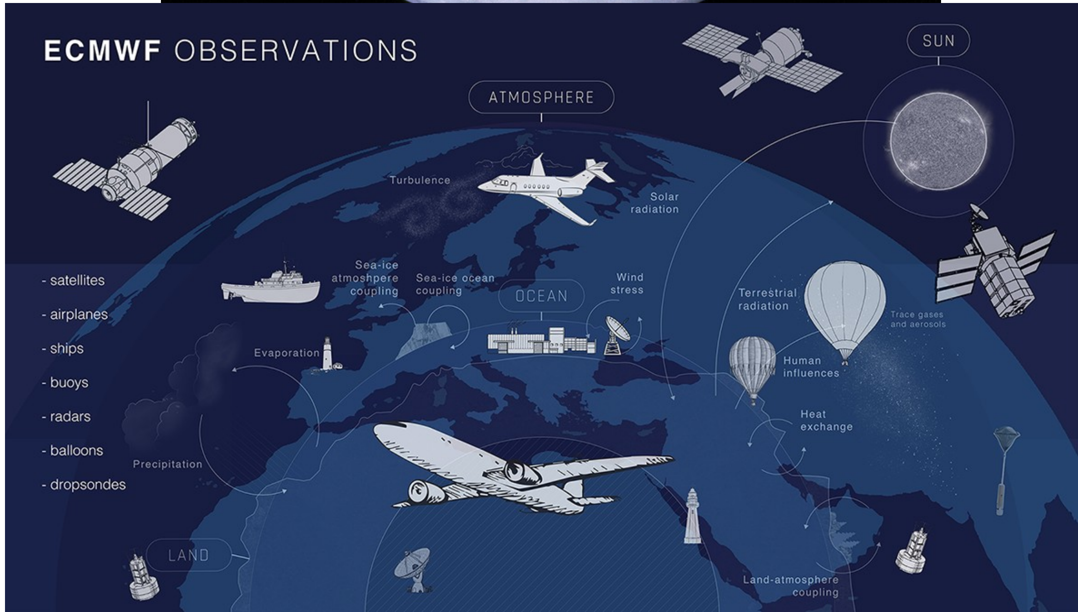
- All Earth System components (atmosphere, ocean, land surface, cloud physics,…) are connected in a non-trivial way

- Some of the processes involved are not well understood

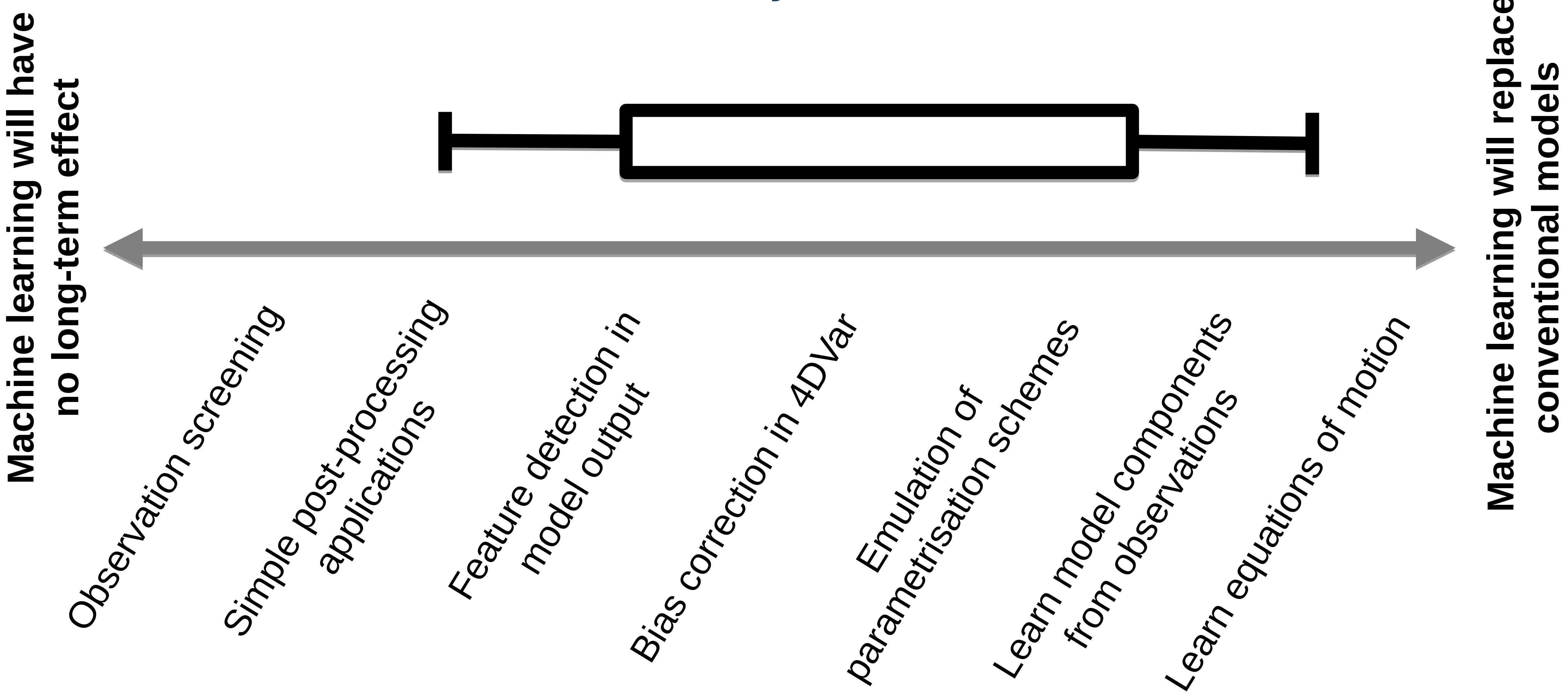
However, we have hundreds of petabytes of Earth System observation and model data available

➢ **There are many application areas for machine learning in numerical weather predictions**

➢ **Machine learning also provides a number of opportunities for high performance computing**



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



The uncertainty range is still very large...

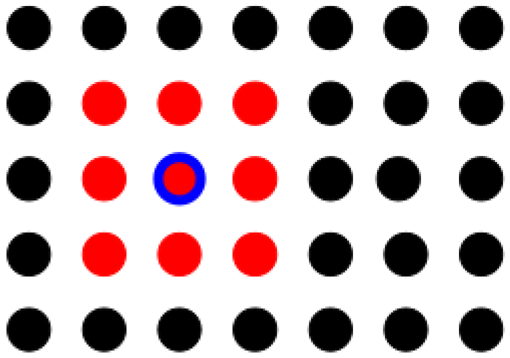
Can we replace conventional weather forecast systems by deep learning?

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 >200 petabyte of data

A simple test configuration:

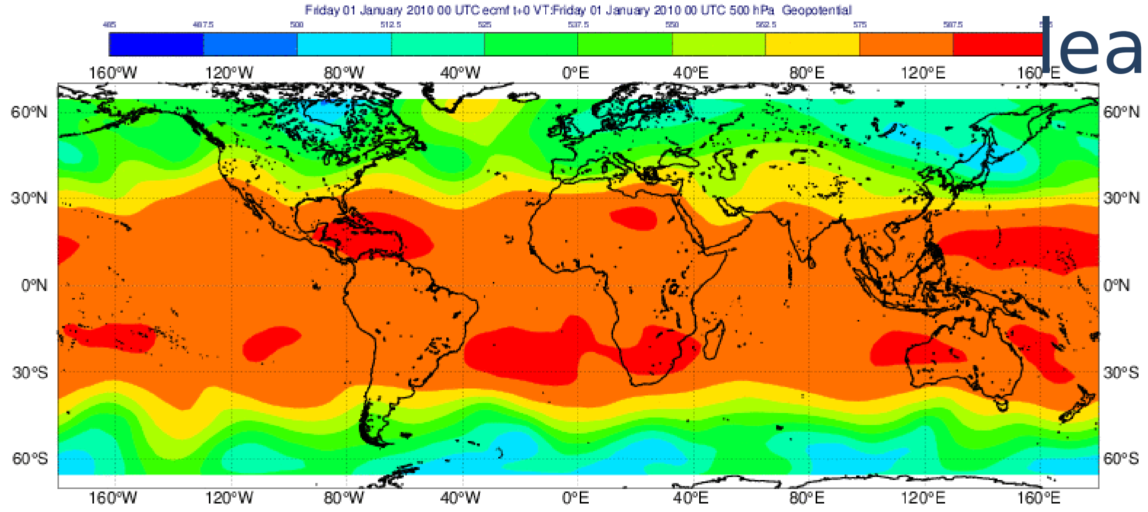
- We retrieve historical data (ERA5) for geopotential at 500 hPa (Z500) for the last decades (>65,000 global data sets)
- We map the global data to a coarse two-dimensional grid (60x31)
- We learn to predict the update of the field from one hour to the next using deep learning
- Once we have learned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

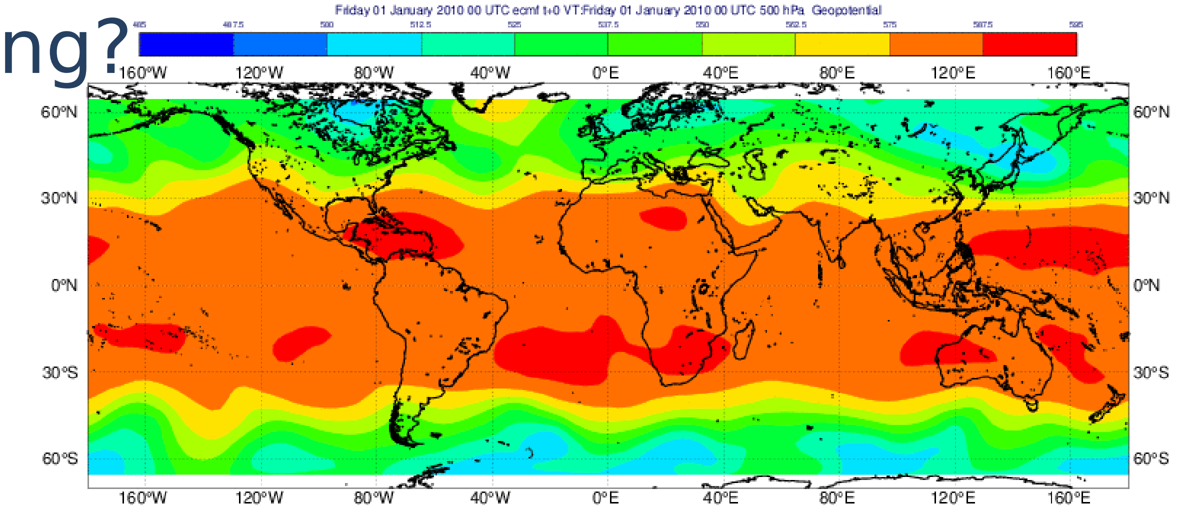
No physical understanding is required!



Dueben and Bauer GMD 2018

Can we replace conventional weather forecast systems by deep learning?





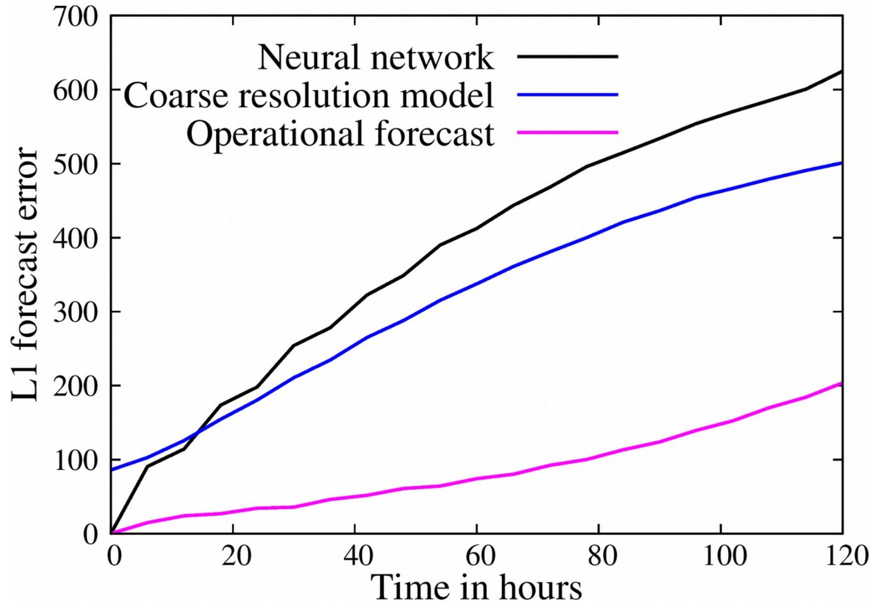
Time evolution of Z500 for historic data and a neural network prediction.
Can you tell which one is the neural network?

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

Unlikely…

- The simulations are unstable 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.

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

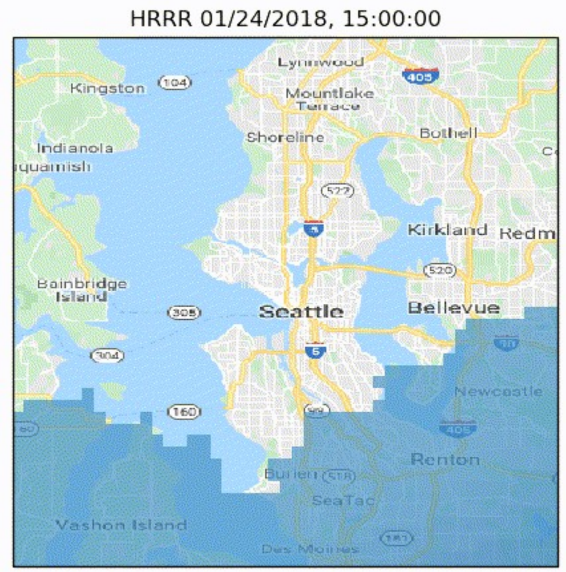
Can we replace conventional weather forecast systems by deep learning?

However, machine learning models are very promising for **now-casting applications** that provide weather predictions for a couple of hours lead time.

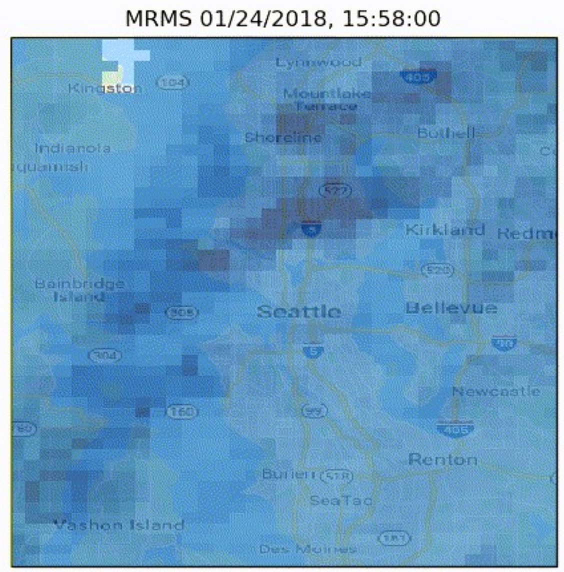
Here…
…conservation is not important as errors have no time to accumulate.
…interactions between weather features are not important (more advection, less physics).
…only local predictions are required → more independent datasets are available for training.

Example: 1-hour predictions of precipitation by Google:

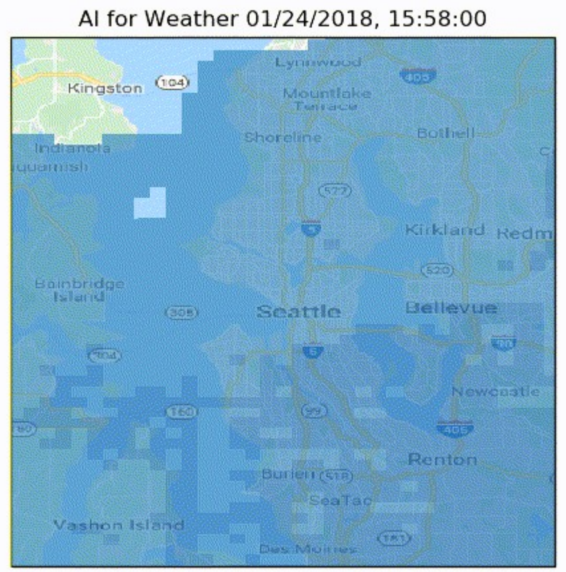
NOAA forecast:



Ground truth:

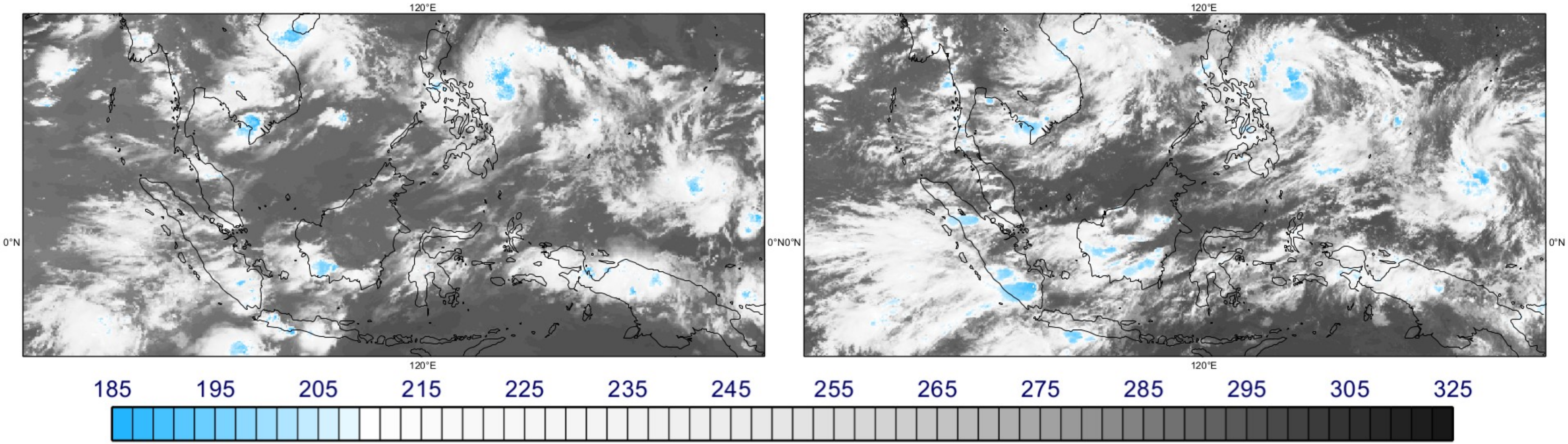


Machine learning solution:



https://ai.googleblog.com/2020/01/using-machine-learning-to-nowcast.html and research of Jason Hickey et al.

Why is it hard to beat conventional weather forecast systems?

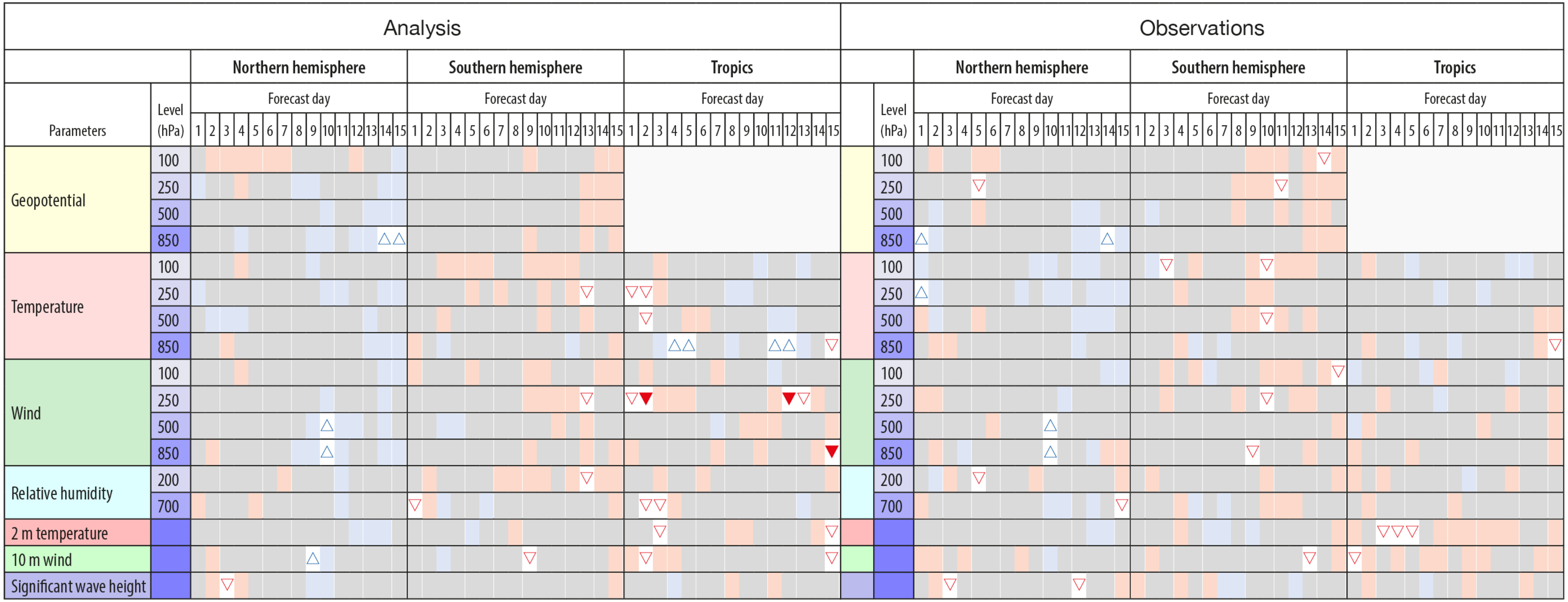


Top-of-the-atmosphere cloud brightness temperature [K] for satellite observations and a simulation of the global atmosphere with 1.45 km resolution on the SUMMIT supercomputer at the Oak Ridge National Laboratory.
Dueben, Wedi, Saarinen and Zeman JSMJ 2020

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.

Earth System models are based on decades of model developments and process understanding.

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

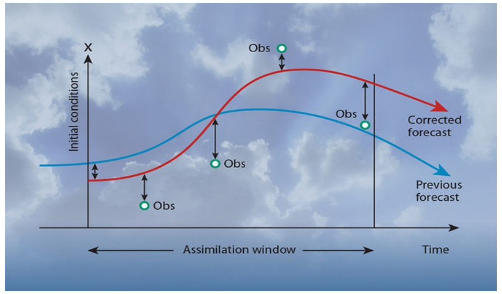
So…using machine learning for weather predictions is useless then?

No!

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

Machine learning applications across the numerical weather prediction workflow

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,

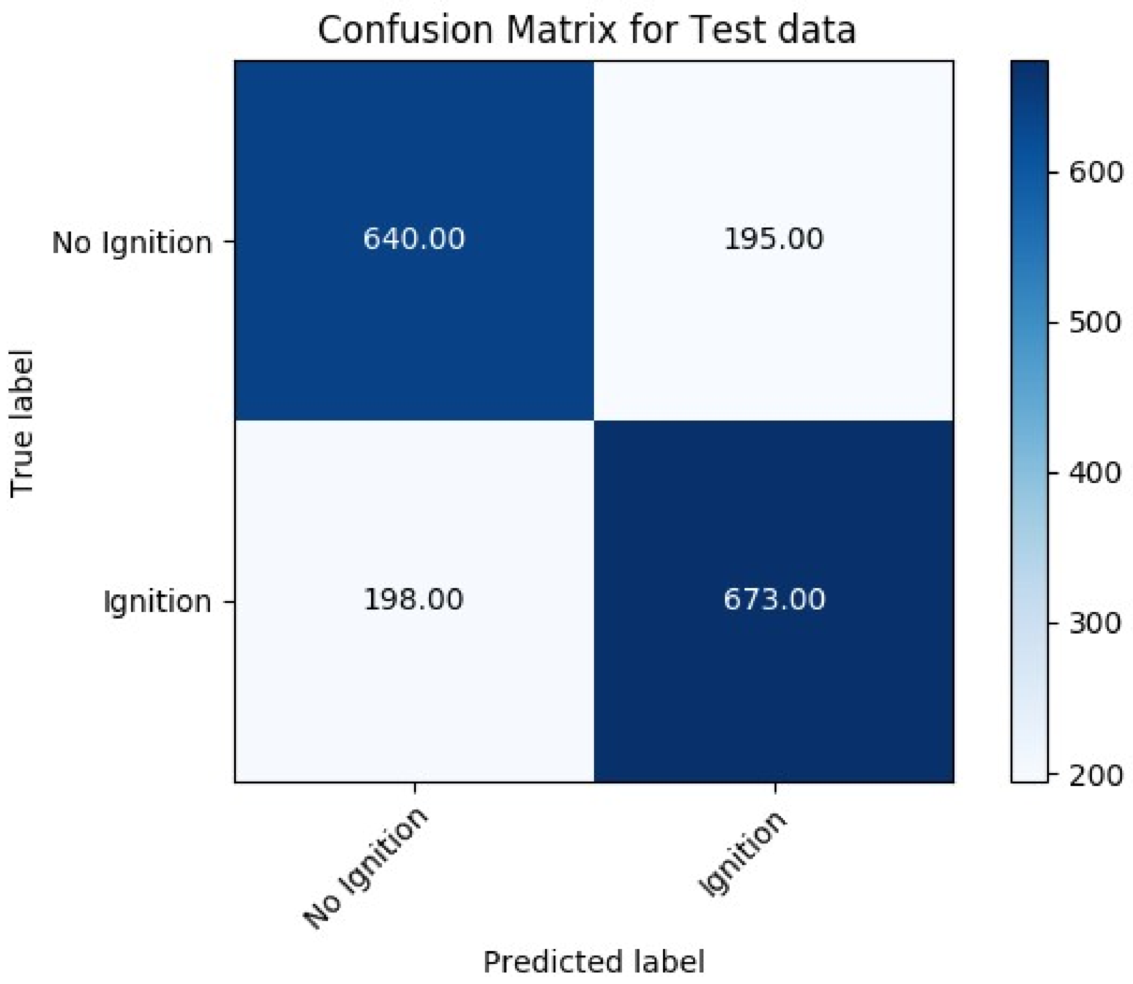
learn governing differential equations, non-linear bias correction, learn operational operators, define optical properties of hydrometeors and aerosols, emulate conventional tools to improve efficiency,

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…

Observations:
Detect the risk for the ignition of wild fires by lightnings

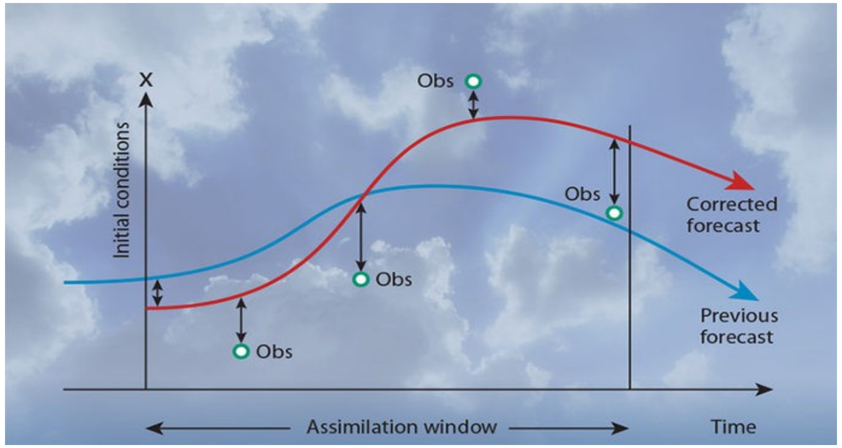


- Observations for 15 variables are used as inputs including soil moisture, 2m temperature, soil type, vegetation cover, relative humidity, and precipitation
- The rate of radiant heat output from the Global Fire Assimilation System (GFAS) of CAMS (monitored by the MODIS satellite) was used to generate a “truth”
- 12,000 data points were used for training
- Different machine learning tools (decision trees, random forest and Ada Boost) are used to classify the cases into “ignition” and “no-ignition”
- The best classifier has an accuracy of 77 %

Ruth Coughlan, Francesca Di Giuseppe, Claudia Vitolo and the SMOS-E project

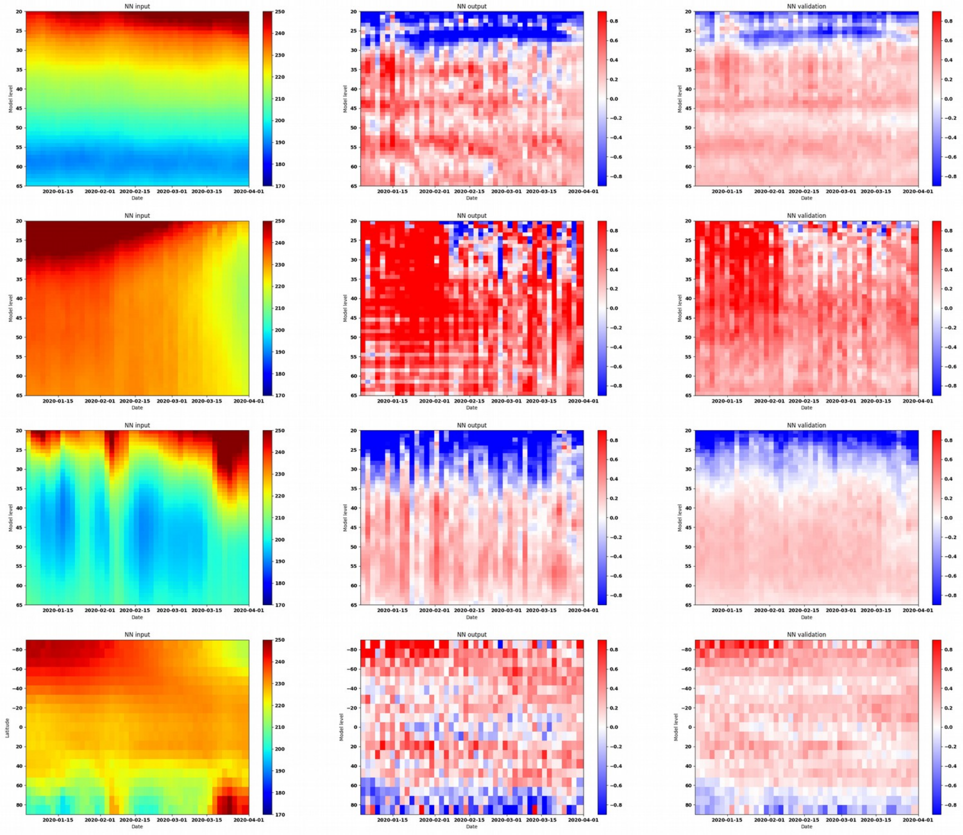
Data assimilation: Bias-correct the forecast model in 4DVar data assimilation

- 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. This can be done with a column-based approach or with three-dimensional machine learning solutions



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

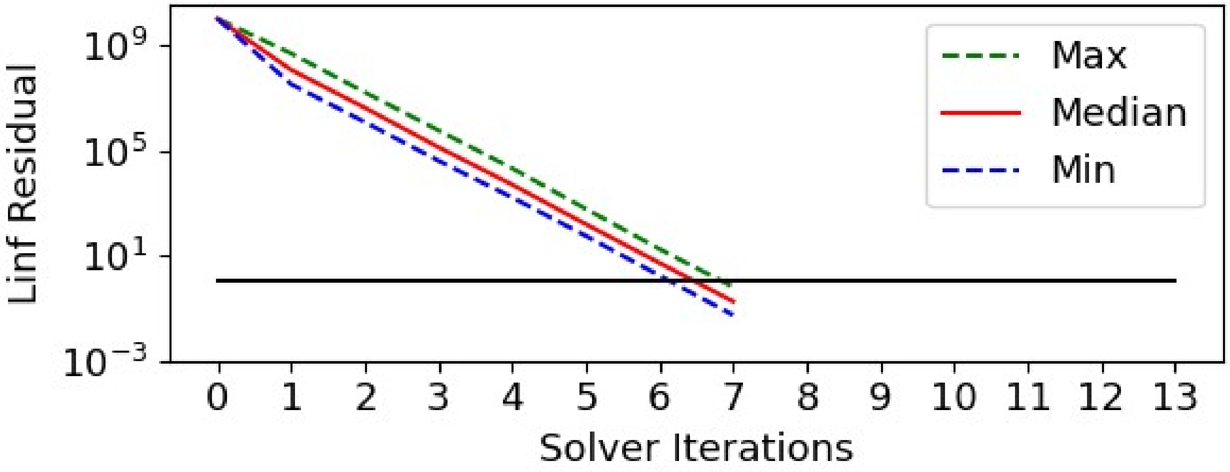
Numerical weather forecasts: To precondition the linear solver

- 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.

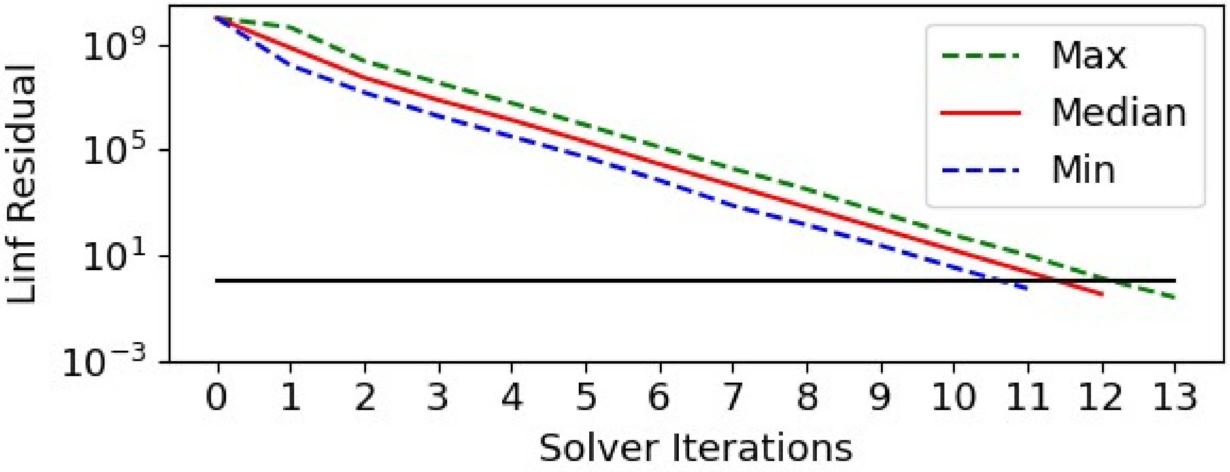
Can we use machine learning for preconditioning, predict the inverse of the matrix and reduce the number of iterations that are required for the solver?

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:



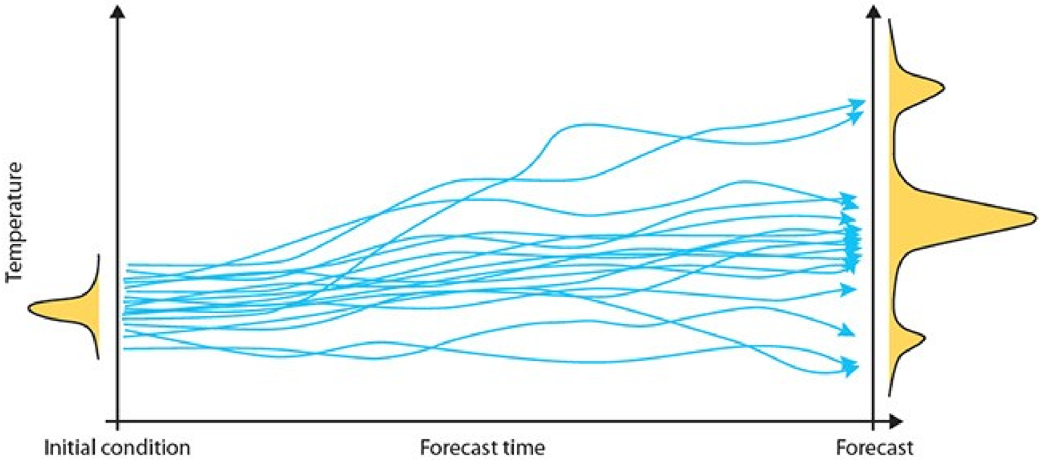
No preconditioner:

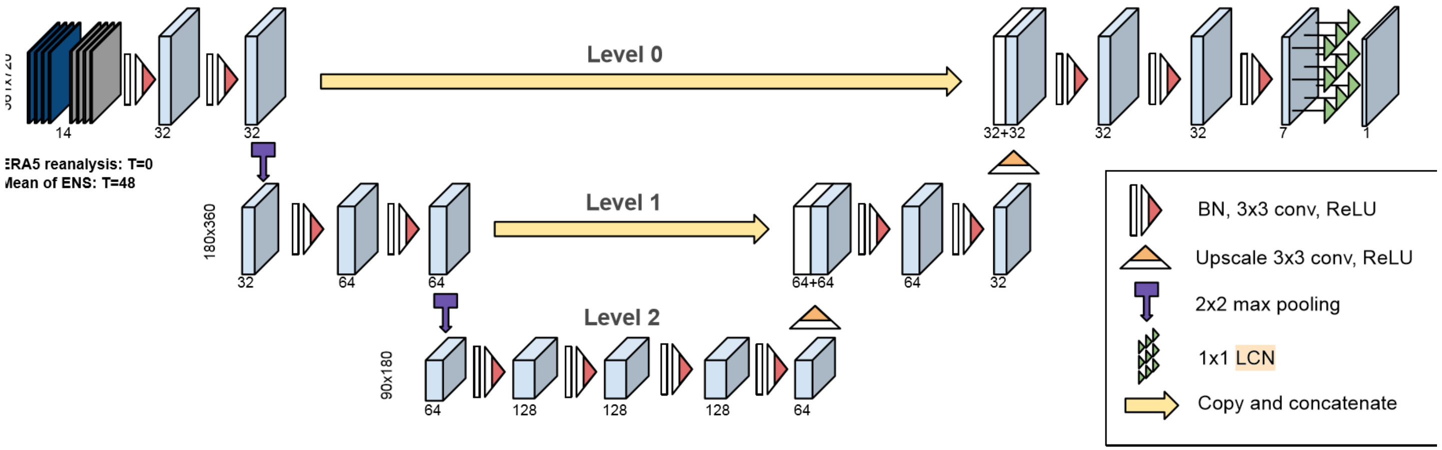


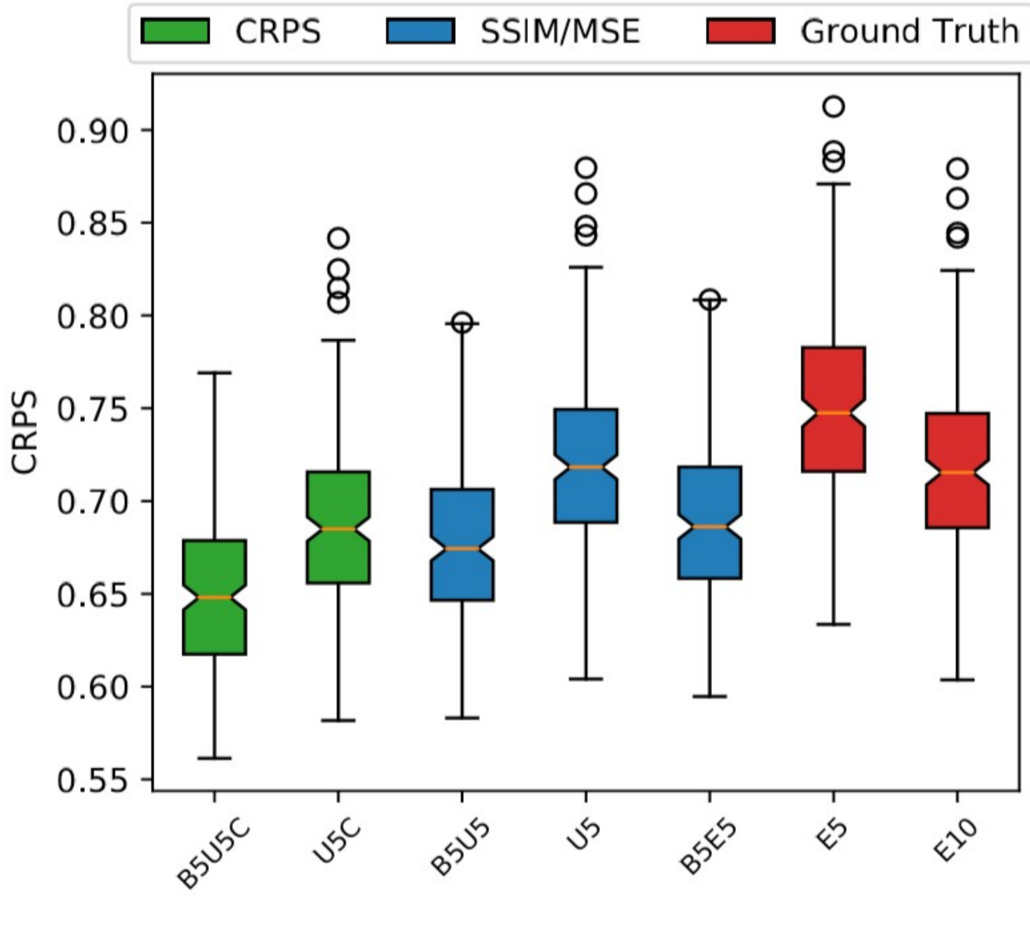
It turns out that the approach (1) is working and cheap, (2) interpretable and (3) easy to implement even if no preconditioner is present.

Ackmann, Dueben, Smolarkieicz and Palmer

-processing and dissemination: Improve ensemble predictions







(a) T850

Ensemble predictions are important but expensive.

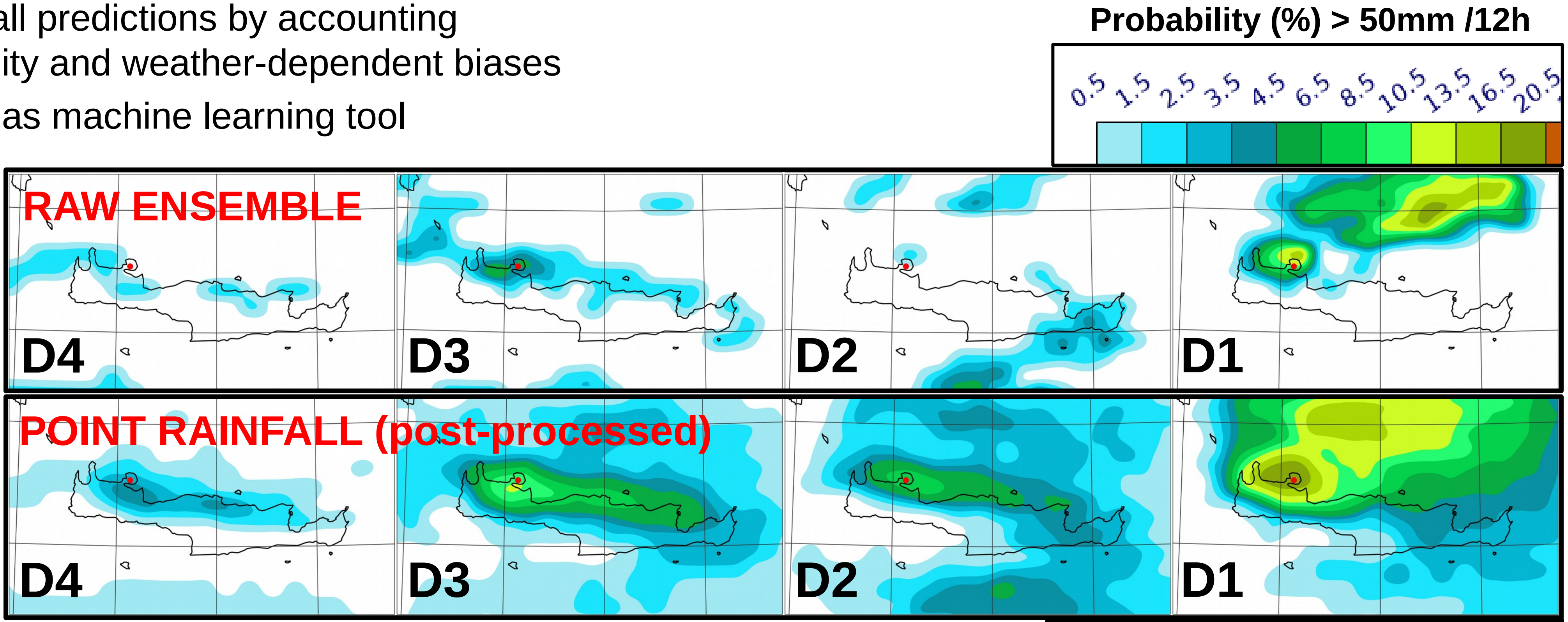
Can we improve ensemble skill scores from a small number of ensemble members via deep learning?

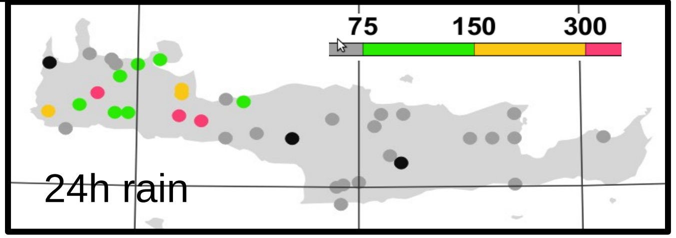
- Use global fields of five ensemble members as inputs.
- Correct the ensemble scores of temperature at 850 hPa and Z500 hPa for a 2-day forecast towards a full 10 member ensemble forecast.

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

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





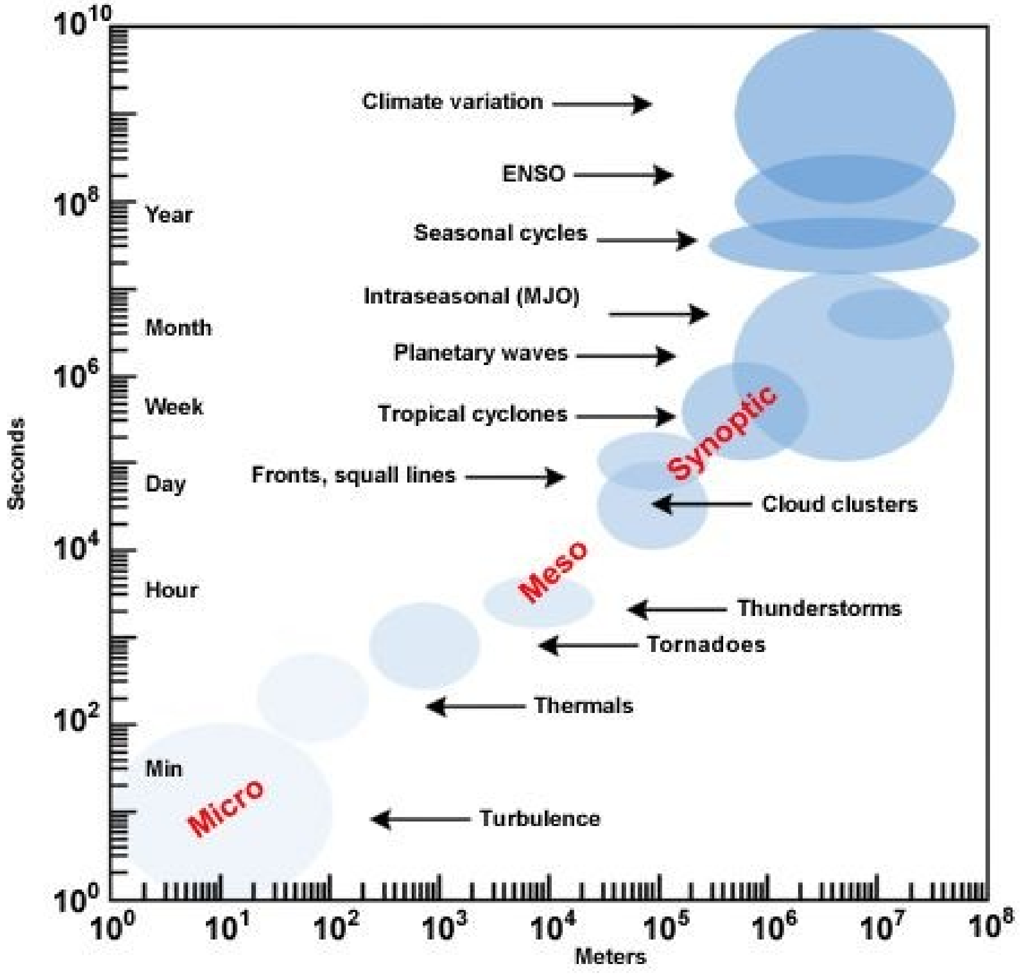
Example: Devastating floods in Crete on 25 February 2019

Benefits: Earlier and more consistent signal with higher probabilities

Timothy Hewson and Fatima Pillosu

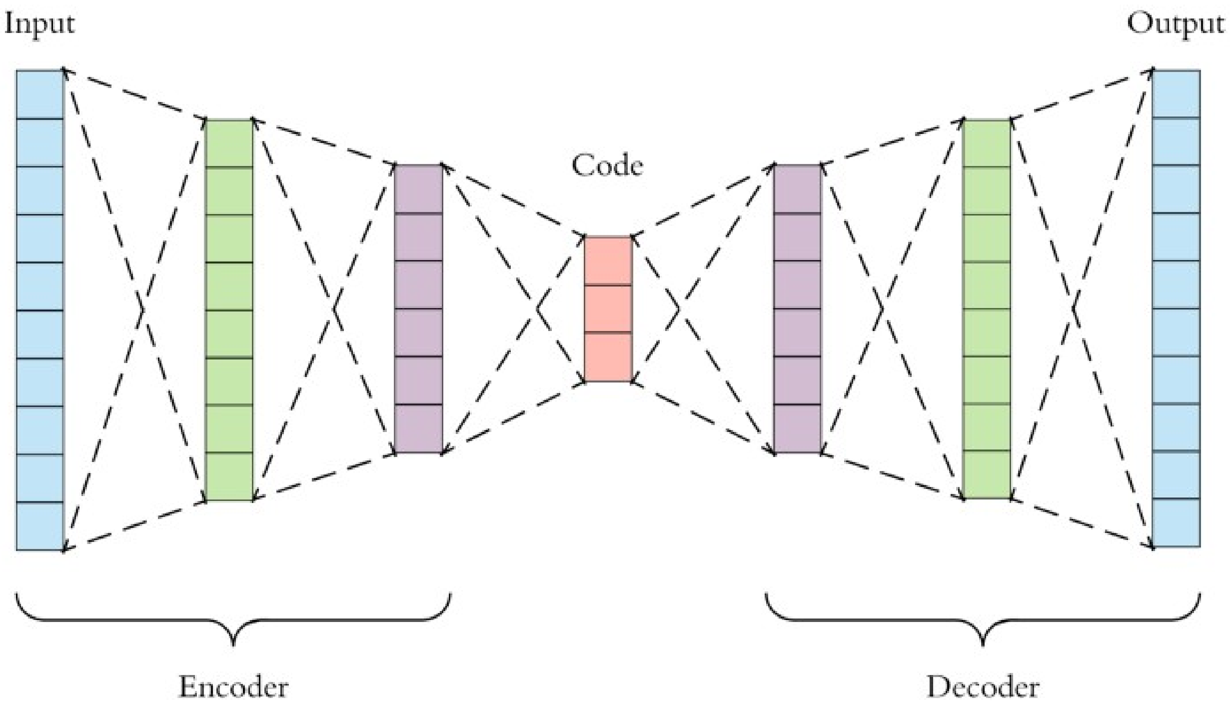
An example of a “co-designed” machine learning solution for weather and climate

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



Source: UCAR

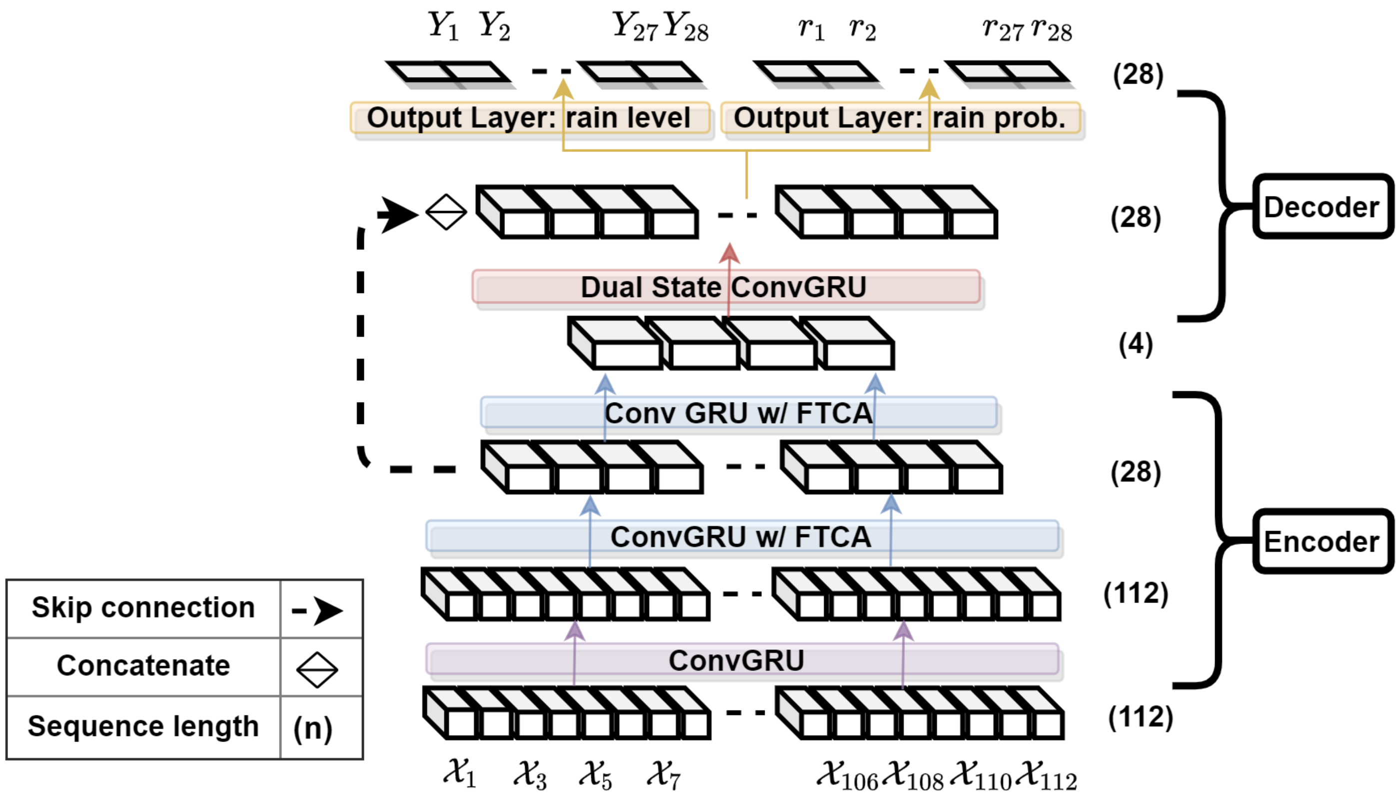
Machine learning:
Neural network tools allow for encoding/decoding structures



Source:
https://towardsdatascience.com

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

-processing and dissemination: Precipitation down-scaling



Model name	RMSE
IFS	3.627
HCGRU	3.268
HCGRU+CC	3.272
T-NET	3.106
T-NET+CC	**3.005**

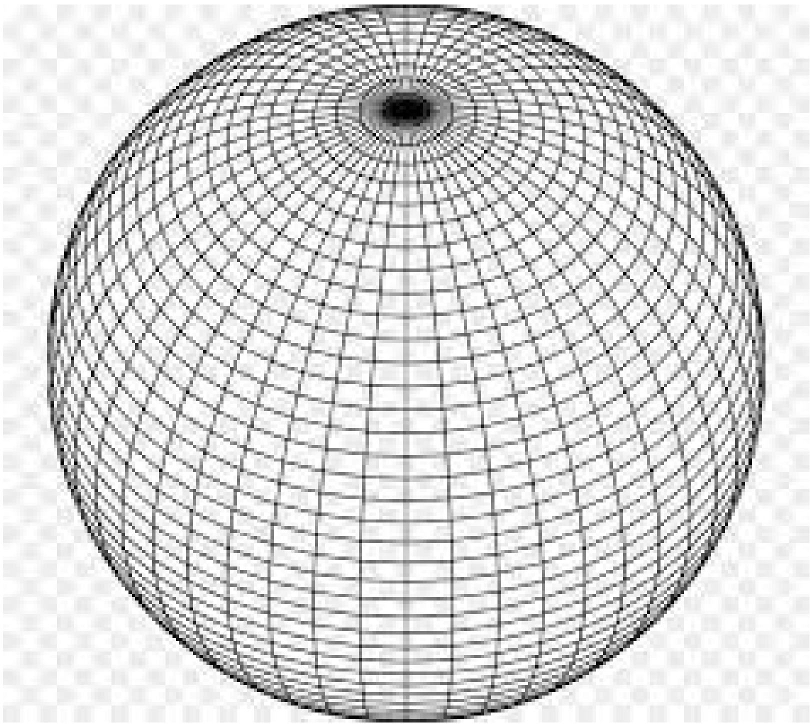
Problem: Learn to map precipitation predictions from ERA5 reanalysis data at ~50 km resolution to E-OBS local precipitation observations at ~10 km resolution over the UK.

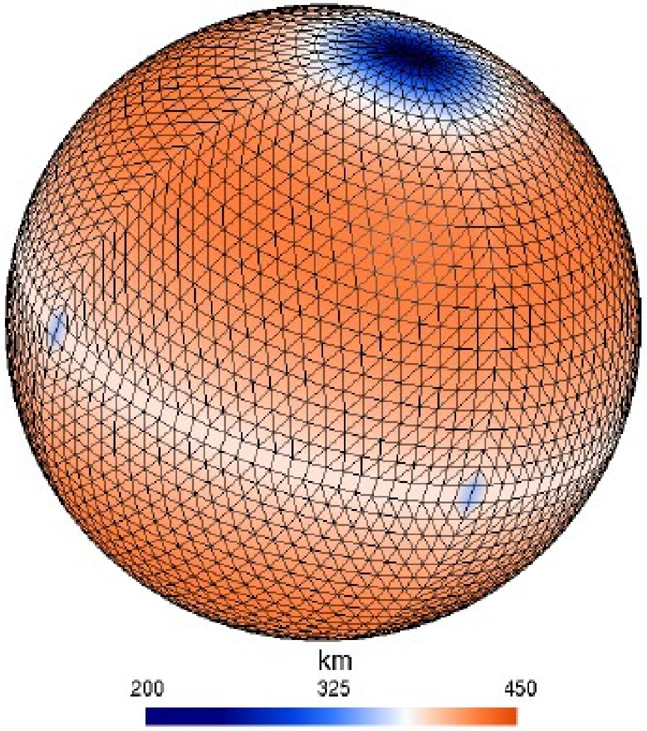
Use case: Eventually, apply the tool to climate predictions to understand changes of local precipitation pattern due to climate change.

Method: Use T-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.

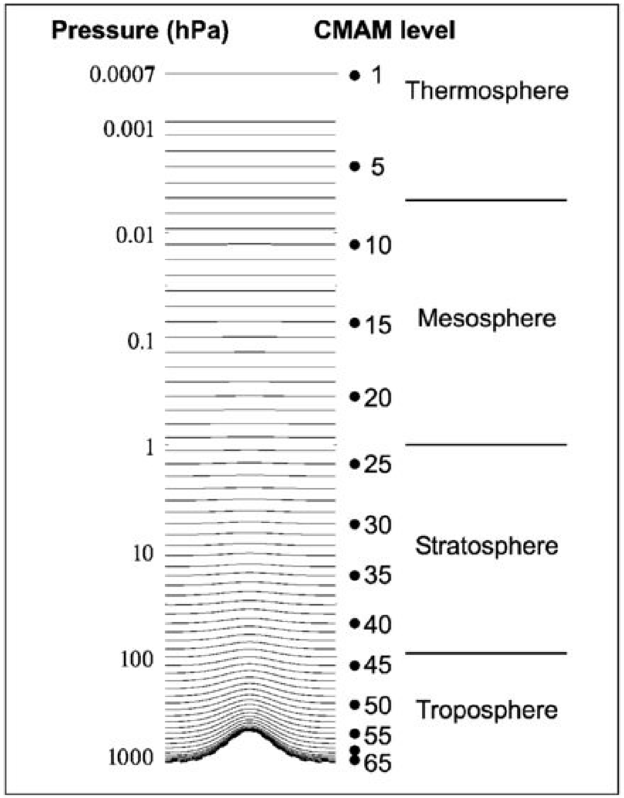
Rilwan A. Adewoyin, Ritabrata Dutta, Peter Watson, Peter Dueben, Yulan He in preparation

How to do multi-scale modelling on unstructured grids?





Source: Willem Deconinck



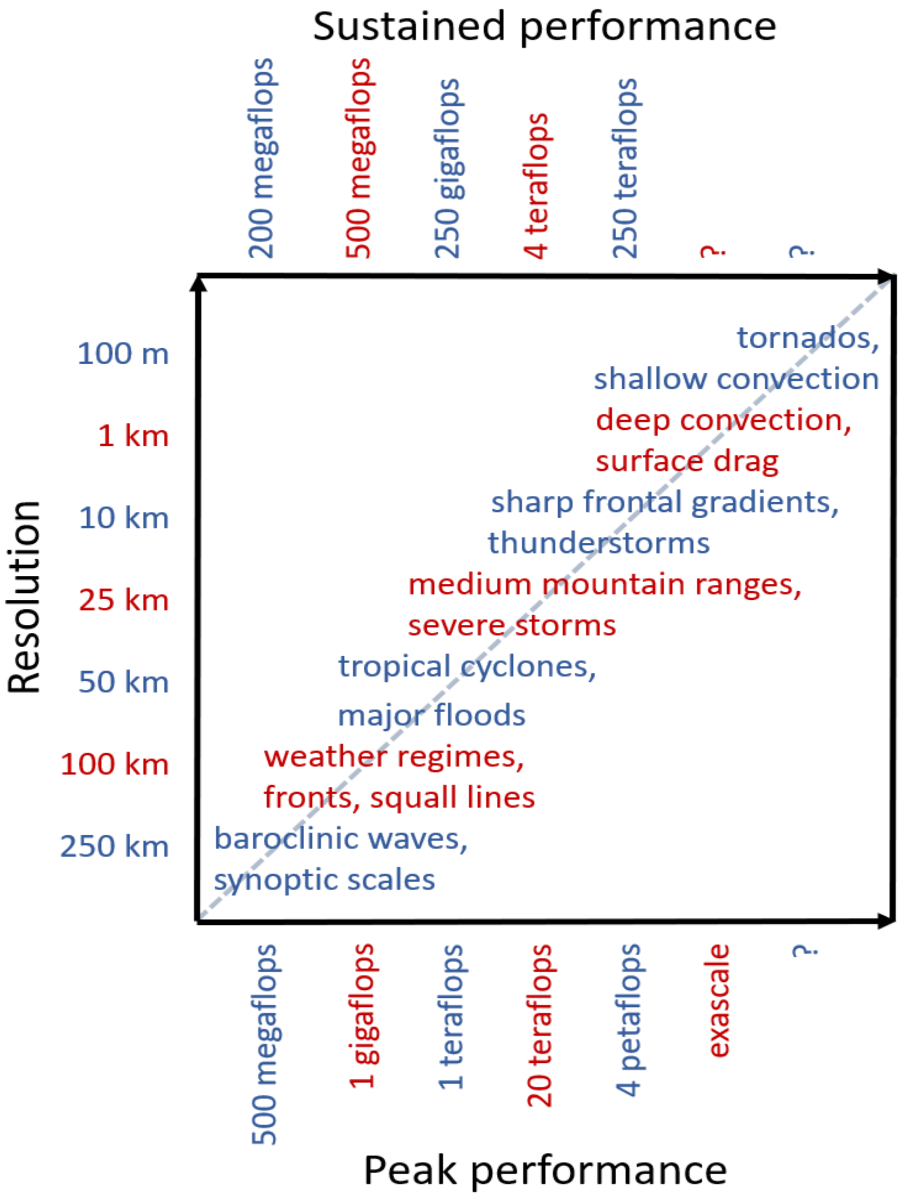
Source: Polavarapu et al. 2005

Longitude/latitude (easy but inefficient and un-isotropic) **vs. reduce 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

What about high performance computing?



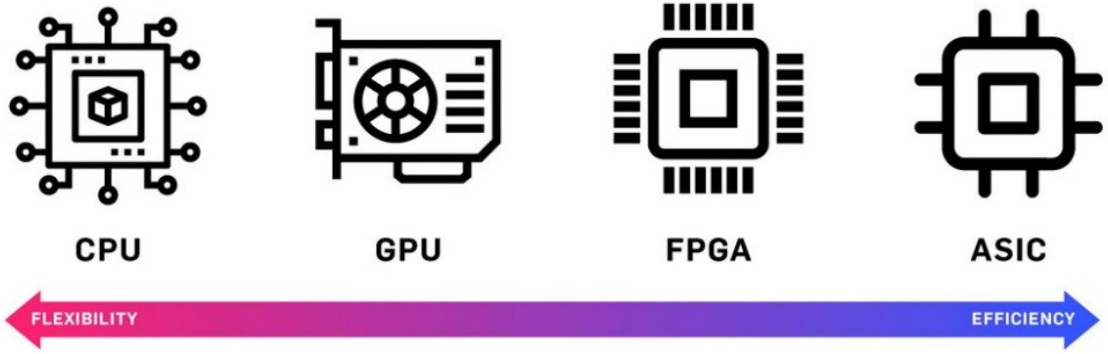
Adapted from Neumann, Dueben et al. Phil Trans A 2018

- Individual processors will not be faster
  → Parallelisation / power consumption / hardware faults

- Hardware will be more heterogeneous
  → CPUs / GPUs / FPGAs / ASICs

- Machine learning has strong impact on hardware development
  → High floprate at low precision

- I/O will become a nightmare



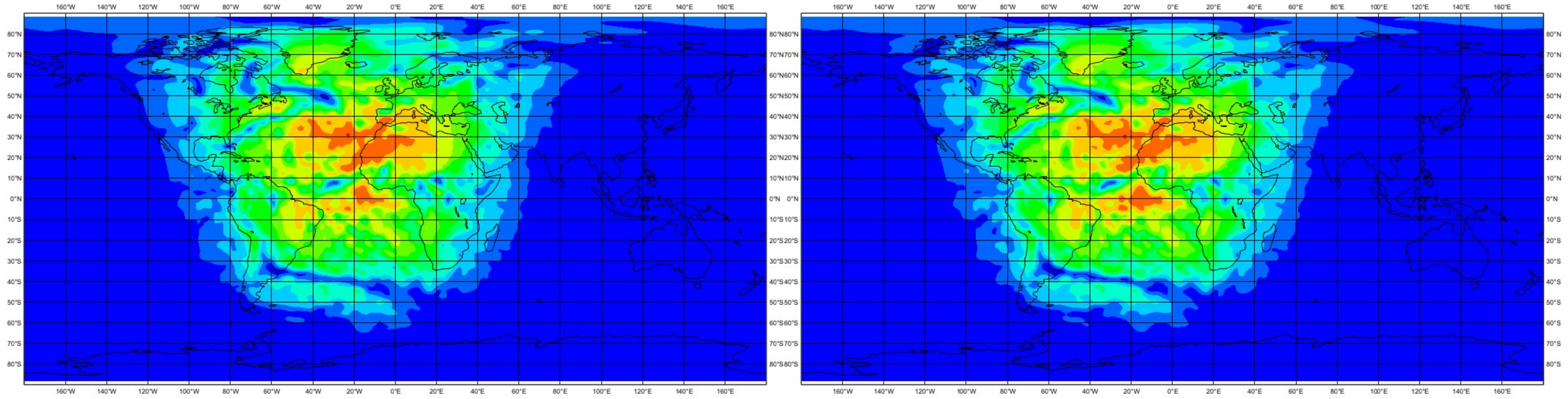
Source: venturebeat.com

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.

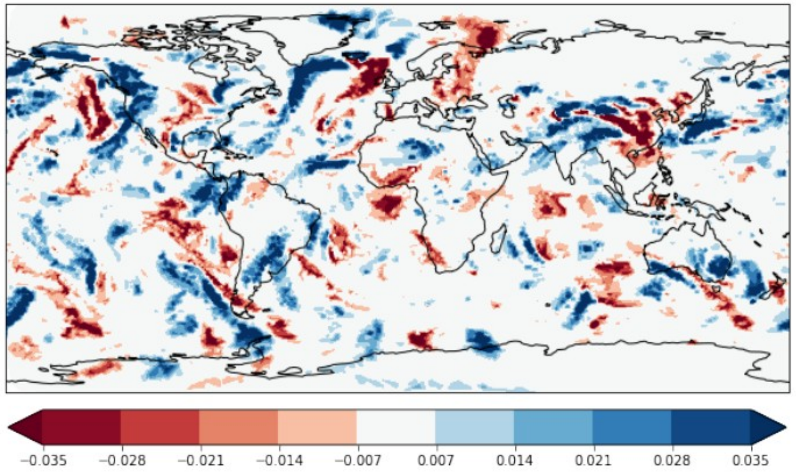
Dueben, Hogan, Bauer @ECMWF and Progsch, Angerer @NVIDIA

Numerical weather forecasts: To emulate gravity wave drag

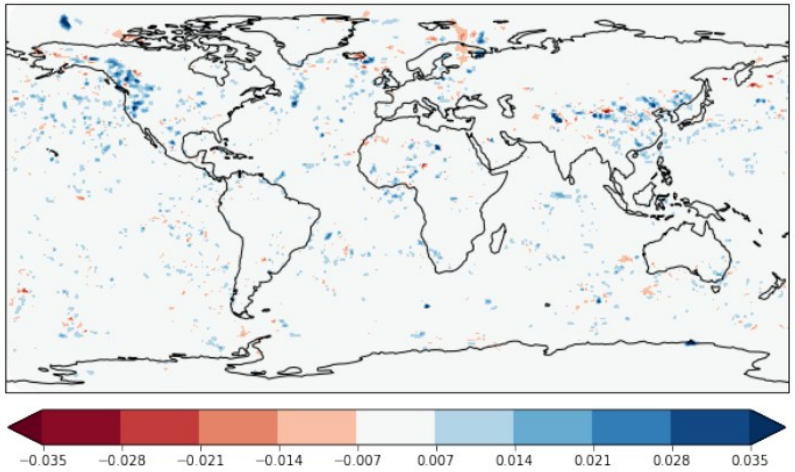
- Repeat the same approach for the gravity wave drag scheme of IFS
- Start with non-orographic and continue with orographic wave drag

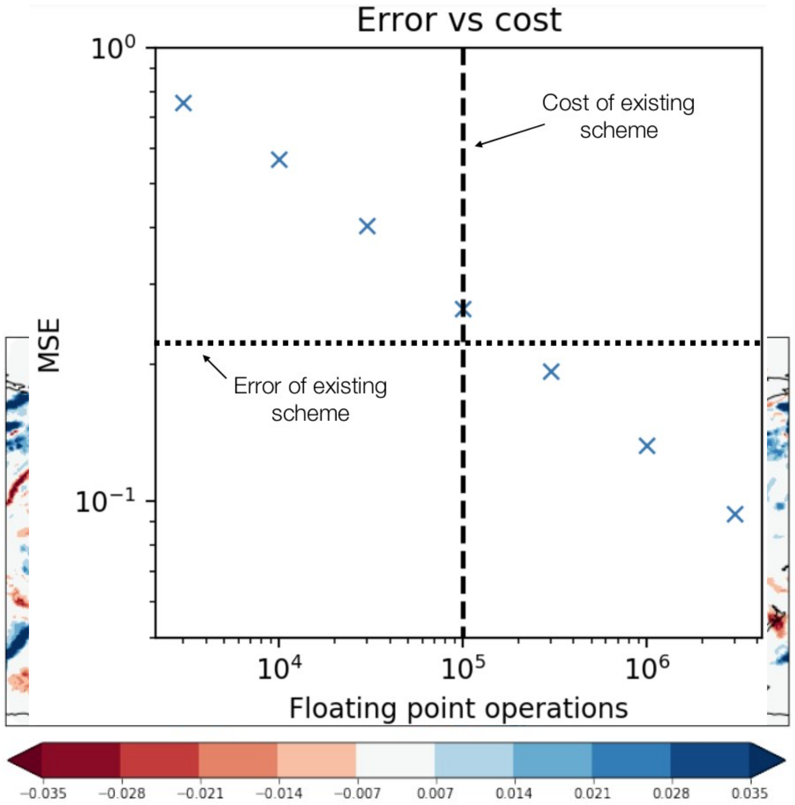
Results for the non-orographic gravity wave drag are promising.

Original scheme



Difference





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

Can we use deep learning hardware for conventional models?

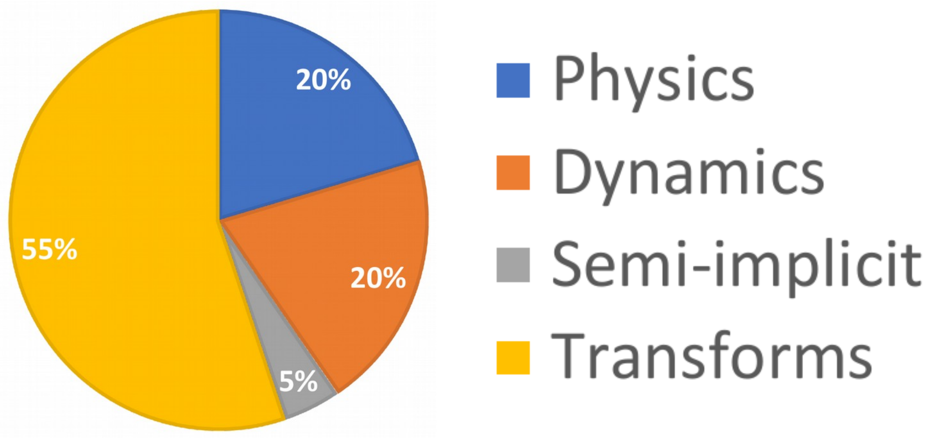
Machine learning accelerators are focussing on low numerical precision and high floprats.

Example: TensorCore accelerators on NVIDIA Volta GPUs are optimised for half-precision matrix-matrix multiplications with single precision output.

→ 7.8 TFlops for double precision vs. 125 TFlops for half precision

Can we use TensorCores within our models?

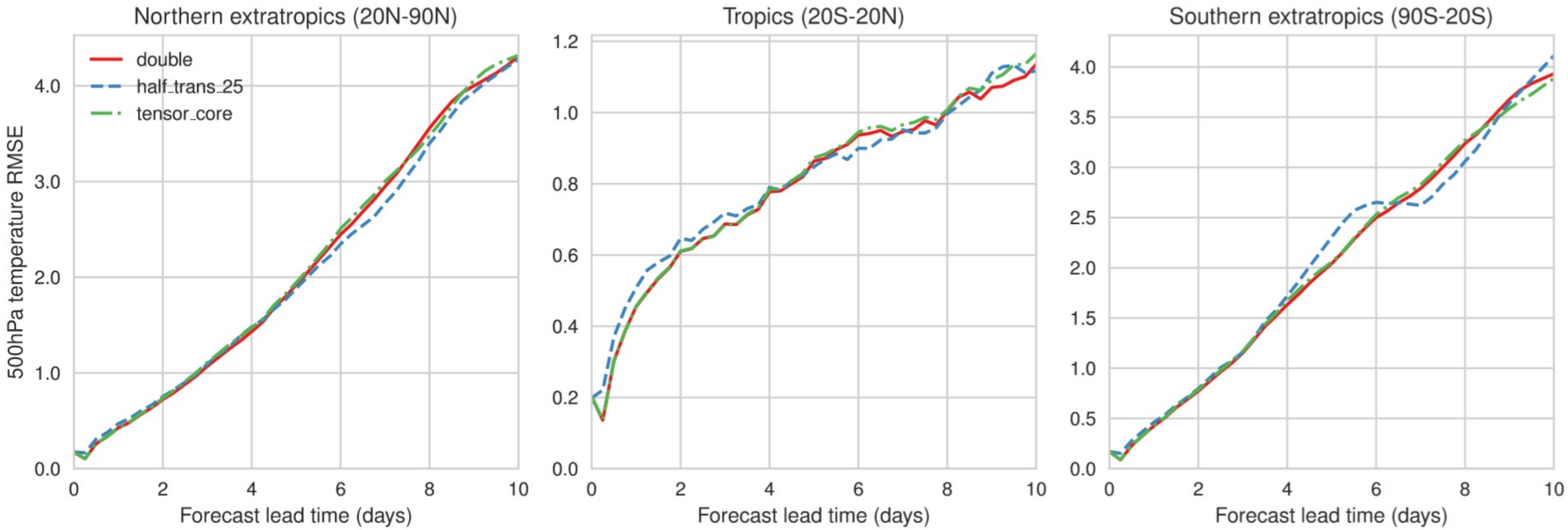
Relative cost for model components for a non-hydrostatic model at 1.45 km resolution:



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.

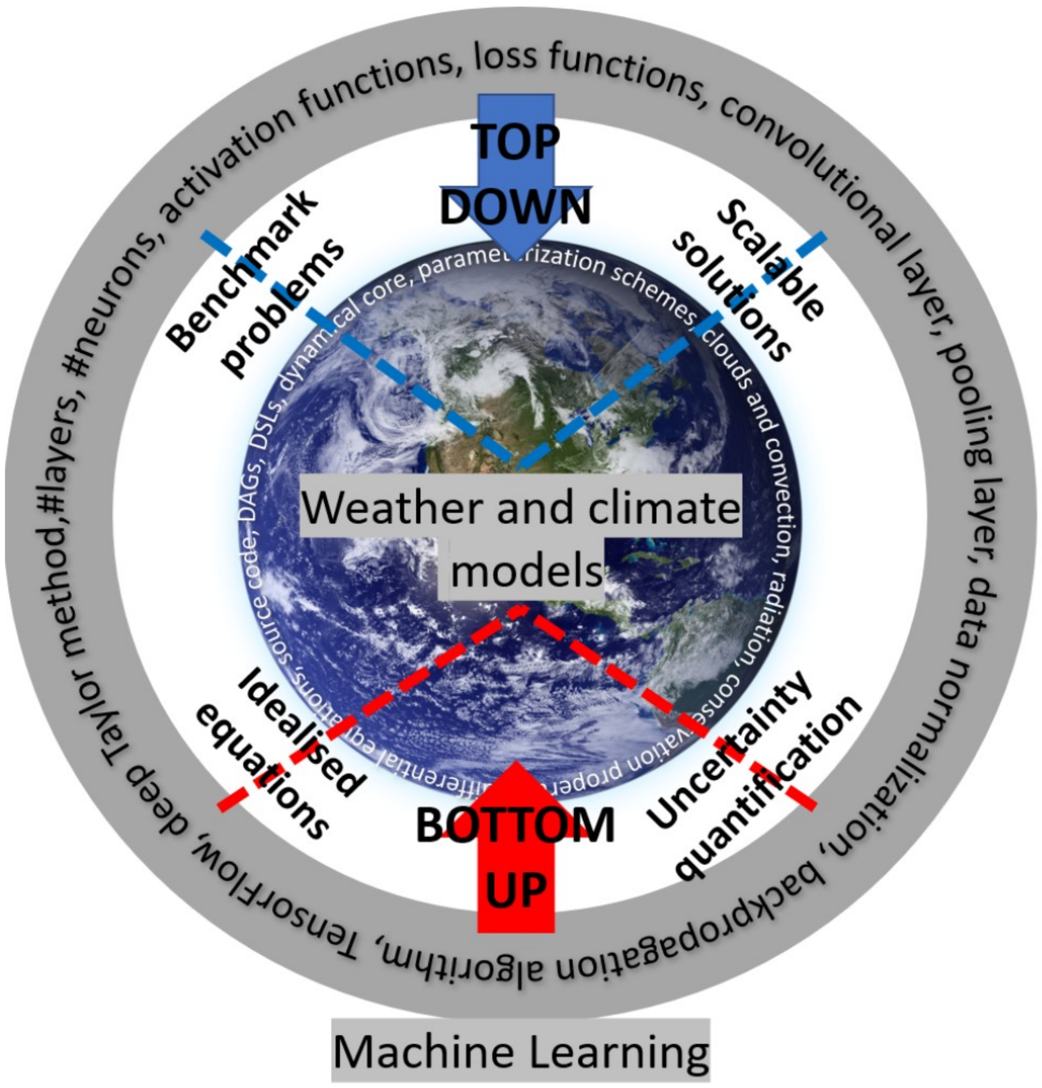
Scientific challenges for machine learning in numerical weather predictions

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.

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.

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*)

We need to focus on useful tools that can serve as beacons.

This will require machine learning solutions that are customised to weather and climate science and a "co-designed" approach between domain and machine learning science.

Some room for interactions with machine learning efforts at ECMWF

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

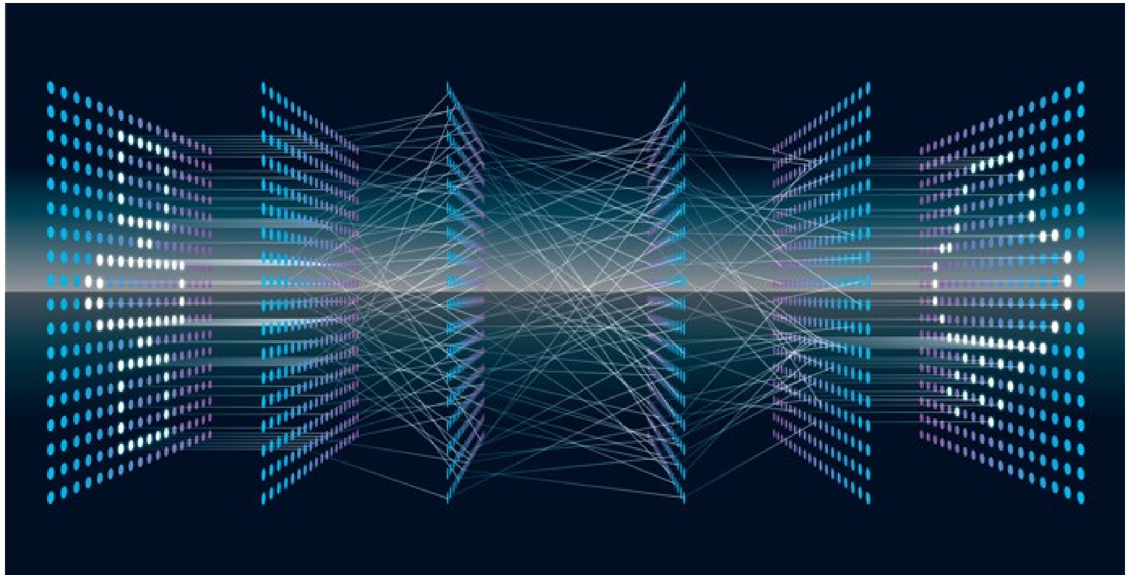
We have also started a special seminar series on Machine Learning that is broadcasted.

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 soon.

MAELSTROM



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 will have a significant impact on future high-performance computing hardware and therefore also on conventional weather and climate models.

- There are challenges for the application of black-box machine learning solutions within weather and climate models that need to be addressed.

- There is still a lot of work to be done regarding the development of customised machine learning solutions for weather and climate predictions which will require “co-designed” solutions.

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

ECMWF | EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

The strength of a common goal