Machine Learning for Weather Forecasts

Peter Dueben

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

The strength of a common goal

The ESIWACE project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 823988.
European Centre for Medium-Range Weather Forecasts (ECMWF)

- Research institute and 24/7 operational weather service for medium-range, monthly and seasonal forecasts.
- Independent, intergovernmental organisation supported by 34 states.
- Based close to London in the UK; ≈350 member of staff.
- Home of two supercomputers.

www.ecmwf.int
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 throughout the workflow of numerical weather predictions
- Machine learning also provides a number of opportunities for high performance computing
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…
State-of-play and outline of the talk

There are many interesting application areas for machine learning to improve weather and climate predictions.

We are only at the very beginning to explore the potential of machine learning in the different areas.

I will present a couple of example applications of machine learning at ECMWF in the following.

I will name the main challenges that we are facing when using machine learning today.
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) 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.
- 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 deep learning will be investigated soon.

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

**Why would you do this?**

Neural networks are likely to be much more efficient and portable to heterogenous hardware

This is a very active area of research:
Rasp, Pritchard, Gentine PNAS 2018
Brenowitz and Bretherton GRL 2018

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:
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
Ensemble predictions are important but expensive.
Can we correct ensemble spread calculated from a small number of ensemble members via deep learning?
• Use 3D fields of a small number of ensemble members as inputs.
• Predict ensemble spread of temperature at 850 hPa for a 1-day forecast of a full 10 member ensemble forecast.
• Focus on an area over Europe (40W-30E and 40N-60N)
• A global training is a high-performance computing application.

➢ First results are promising.
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
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!
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; …
The main challenge for the machine-learning community

We need to prove that machine learning tools are better than state-of-the-art within the next two years.

We need to build *useful tools* that improve weather and climate predictions and/or help to improve our understanding of the Earth System and are able to convince the conservative weather and climate scientist.

Sanity check for Dueben and Bauer GMD 2018:
Is it new? - Yes;  Has is impact? – Yes;  Does it produce nice pictures and videos? – Yes
Does it produce a *useful tool*? – No;  Will it convince the weather and climate community? – No

Why?
*We do not know how to efficiently scale the approach to $O(1,000,000)$ degrees of freedom and the black-box approach will not convince conservative domain scientists.*
Develop useful tools: Why is this challenge so hard?

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.

Global simulations show a breath-taking level of complexity and can represent many details of the Earth System.
Develop useful tools: Why is this challenge so hard?

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Northern hemisphere</th>
<th>Southern hemisphere</th>
<th>Tropics</th>
<th>Northern hemisphere</th>
<th>Southern hemisphere</th>
<th>Tropics</th>
<th>Northern hemisphere</th>
<th>Southern hemisphere</th>
<th>Tropics</th>
<th>Northern hemisphere</th>
<th>Southern hemisphere</th>
<th>Tropics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecast day</td>
<td>Forecast day</td>
<td>Forecast day</td>
<td>Forecast day</td>
<td>Forecast day</td>
<td>Forecast day</td>
<td>Forecast day</td>
<td>Forecast day</td>
<td>Forecast day</td>
<td>Forecast day</td>
<td>Forecast day</td>
<td>Forecast day</td>
</tr>
<tr>
<td></td>
<td>Level (hPa)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Geopotential</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>▲▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>250</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>500</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>850</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>Temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>▲▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>250</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>500</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>850</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>Wind</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>▲▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>250</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>500</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>850</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>Relative humidity</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>200</td>
<td>▲▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>700</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>2 m temperature</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>10 m wind</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>Significant wave height</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
</tbody>
</table>

**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
Develop useful tools: Why is this challenge so hard?

Real-world models show complex feedbacks on a number of timescales.

Changes to the models require thorough testing and significant HPC resources.

It is really difficult to compare computational speed for machine learning and conventional methods in a fair way.
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.
Challenges for ECMWF

• ECMWF will need to prepare for different pattern of user requests (in particular for data).

• ECMWF will need to keep up with developments of supercomputing hardware that are currently driven by machine learning and artificial intelligence (e.g. low numerical precision).

• ECMWF will need to support the community infrastructure for machine learning.

The European Weather Cloud will be a great tool to address some of the challenges above.

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)

To focus on useful tools that can serve as beacons.

This will be hard work and it will require machine learning solutions that are customised to weather and climate models.
What will machine learning for numerical weather predictions look like in 10 years from now?

Machine learning will have no long-term effect

Observation screening
Simple post-processing applications
Feature detection in model output
Bias correction in 4DVar

Emulation of parametrisation schemes
Learn model components from observations
Learn equations of motion

Machine learning will replace conventional models

The uncertainty range is still very large...
Can we use deep learning hardware for conventional 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.
- Tensor Cores 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 on the Tensor Core.
Half precision Legendre Transformations

Root-mean-square error for geopotential height at 500 hPa at 9 km resolution averaged over multiple start dates.

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

Hatfield, Chantry, Dueben, Palmer Best Paper Award PASC2019
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.

• Machine learning accelerators may 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.

Many thanks.
The strength of a common goal