Machine learning for weather and climate predictions

Peter Düben

European Centre for Medium-Range Weather Forecasts (ECMWF)

Royal Society University Research Fellow

Predicting weather and climate: Why is it so hard?

Earth as seen from Apollo 17

Predicting weather and climate: Why is it so hard?

Earth as seen from Apollo 17

The Earth System is complex, huge and chaotic and we do not have sufficient resolution to resolve all important processes.

Predicting weather and climate: Why is it so hard?

Earth as seen from Apollo 17

The Earth System is complex, huge and chaotic and we do not have sufficient resolution to resolve all important processes.

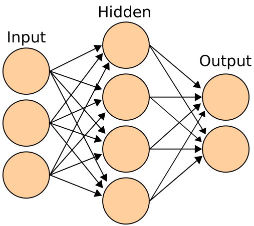
Use cases for machine learning are all over the place...

Deep learning for weather and climate

The Earth System has many components that show non-linear dynamics, we have plenty of observations and often need to apply rough approximations to formulate our models.

Deep learning for weather and climate

The Earth System has many components that show non-linear dynamics, we have plenty of observations and often need to apply rough approximations to formulate our models.



www.wikipedia.org

- Neural Networks learn from input/output pairs.
- Neurons have weighted connections to each other and the weights are trained to produce the optimal results.

Neural networks can emulate non-linear systems.

Future perspective for machine learning in weather forecasts

- ## Use cases in quality control and automated alarm systems

  weather data monitoring, model and data assimilation systems, real-time quality control for observational data, interpret anomalies and guide quality assignment and decision making

- ## Use cases in data assimilation and use of observations

  derive information on the governing differential equations, non-linear bias correction, bias predictors, operational operators, define optical properties of hydrometeors and aerosols

- ## Use cases in numerical modelling

  emulate model components, develop improved parametrisation schemes, provide better error models, learn the underlying equations of motion, generate tangent linear or adjoint code from machine learning emulators)

- ## Use cases in forecast and climate reanalysis outputs post-processing

  real time adjustments of forecast products, feature detection, uncertainty quantification, error corrections for seasonal predictions, development of low-complexity models, plenty of business opportunities

- ## Changes of the infrastructure will be required

  different use of data, data mining and data fusion; more/larger data requests; need for use of deep learning hardware; user products within the forecast model

Future perspective for machine learning in weather forecasts

- **Use cases in quality control and automated alarm systems**
  weather data monitoring, model and data assimilation systems, real-time quality control for observational data, interpret anomalies and guide quality assignment and decision making
- **Use cases in data assimilation and use of observations**
  derive information on the governing differential equations, non-linear bias correction, bias predictors, operational operators, define optical properties of hydrometeors and aerosols
- **Use cases in numerical modelling**
  emulate model components, develop improved parametrisation schemes, provide better error models, learn the underlying equations of motion, generate tangent linear or adjoint code from machine learning emulators)
- **Use cases in forecast and climate reanalysis outputs post-processing**
  real time adjustments of forecast products, feature detection, uncertainty quantification, error corrections for seasonal predictions, development of low-complexity models, plenty of business opportunities
- **Changes of the infrastructure will be required**
  different use of data, data mining and data fusion; more/larger data requests; need for use of deep learning hardware; user products within the forecast model

The next slides: I will show example applications and discuss challenges.

To learn the equations

Neural networks enable us to learn complex non-linear dynamics as a black box.

To learn the equations

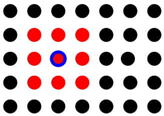
Neural networks enable us to learn complex non-linear dynamics as a black box.

We could base the entire model on Neural Networks.
Who needs Navier Stokes?

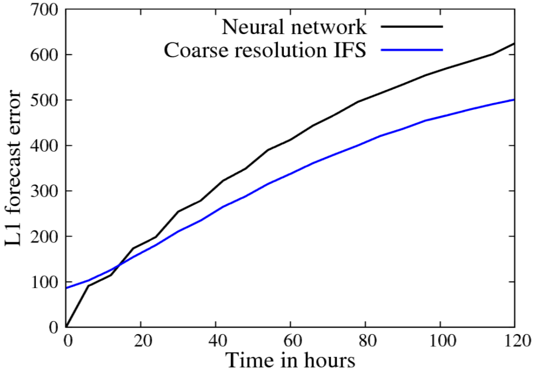
- We know the equations of motion but we cannot solve them.
- Discretisation and sub-grid-scale variability generates significant errors.
- The data handling system of ECMWF provides access to over 210 petabyte of primary data and the data archive of ECMWF grows by about 233 terabyte per day.

Global weather forecast based on Neural Networks

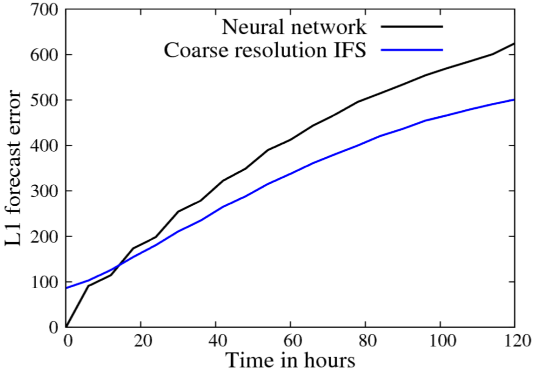
- Retrieve hourly data of geopotential height at 500 hPa from ERA5 re-analysis for training ($>$ 65000 global data sets).
- Map the data to a coarse lon/lat grid (60x31).
- Use the state of the model at timestep *i* as input and the state of the model at timestep $i + 1$ as output.
- Use a 9×9 stencil around the grid point that should be predicted.
- Add time of day and year as well as the coordination of a gridpoint (lon+lat) as input variables to the network.
- The Pole needs special treatment.



Global weather forecast based on Neural Networks



Global weather forecast based on Neural Networks

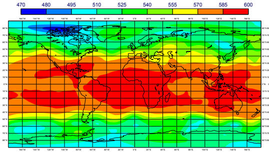


The Neural Network model can compete with a dynamical model of similar complexity.

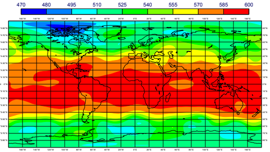
Dueben and Bauer GMD 2018

Global weather forecast based on Neural Networks

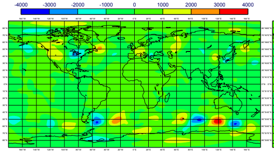
24 hours; Analysis



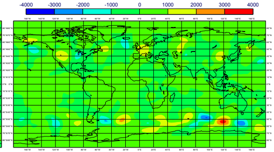
24 hours; Local Neural Network



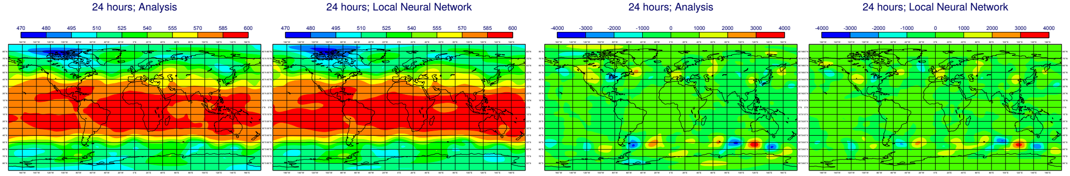
24 hours; Analysis



24 hours; Local Neural Network

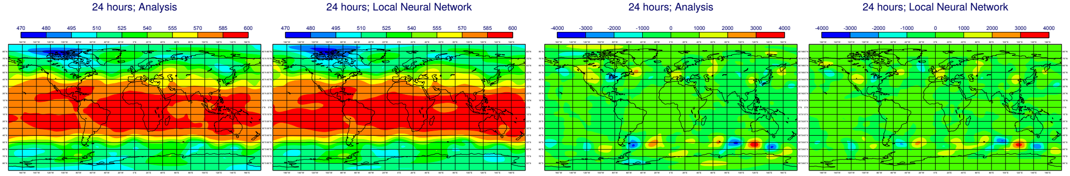


Global weather forecast based on Neural Networks



The simulations show reasonable dynamics.

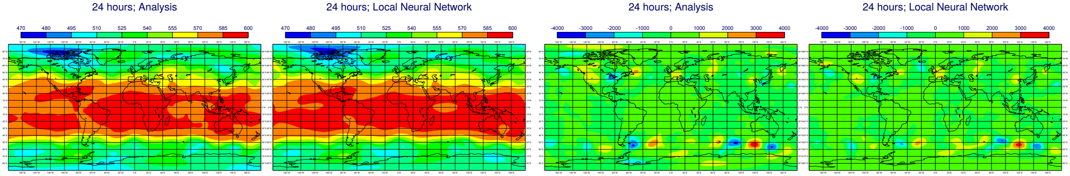
Global weather forecast based on Neural Networks



The simulations show reasonable dynamics.

Just adding further inputs does not necessarily help.

Global weather forecast based on Neural Networks



The simulations show reasonable dynamics.

Just adding further inputs does not necessarily help.

Model runs crash after a couple of weeks.

Dueben and Bauer GMD 2018

Improve post-processing

Ensemble simulations are important but expensive.

Improve post-processing

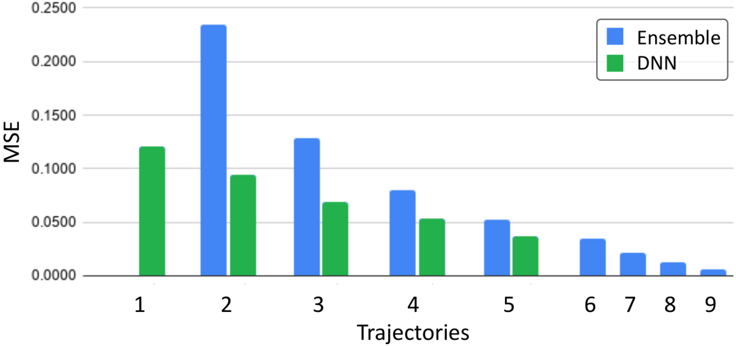
Ensemble simulations are important but expensive.

Use 3D fields of a small number of ensemble members as inputs and try to predict ensemble spread of temperature at 850 hPa for a 24h forecast of a full 10 member ensemble forecast for an area over Europe (40W-30E and 40N-60N).

Improve post-processing

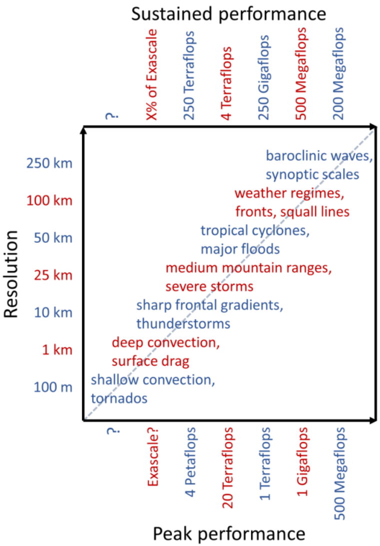
Ensemble simulations are important but expensive.

Use 3D fields of a small number of ensemble members as inputs and try to predict ensemble spread of temperature at 850 hPa for a 24h forecast of a full 10 member ensemble forecast for an area over Europe (40W-30E and 40N-60N).



Grönquist, Ben-Nun, Taranov, Höfler @ ETH and Dueben and Bauer @ ECMWF

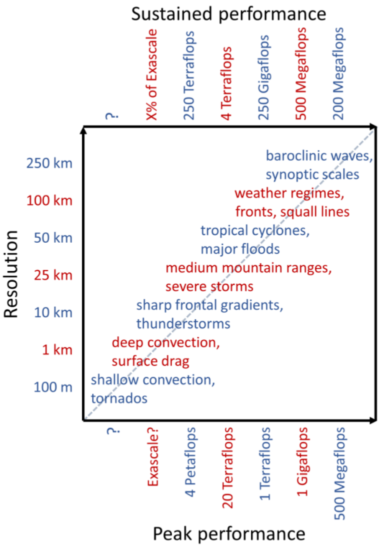
Machine learning, HPC, and weather and climate predictions



Weather and climate models are high performance computing applications:

- More resolution $\rightarrow$ more processes resolved.
- Ratio sustained/peak is going down.
- Machine learning has a very strong impact on hardware developments at the moment.

Machine learning, HPC, and weather and climate predictions



Weather and climate models are high performance computing applications:

- More resolution $\rightarrow$ more processes resolved.
- Ratio sustained/peak is going down.
- Machine learning has a very strong impact on hardware developments at the moment.

Can we emulate model components with neural networks?

Can we use deep learning hardware for conventional models?

Let’s use neural networks to emulate existing parametrisation schemes

- Store input/output pairs of parametrisation schemes.
- Use this data to train a neural network to do the same job.
- Replace the parametrisation scheme by the neural network.

Let's use neural networks to emulate existing parametrisation schemes

- Store input/output pairs of parametrisation schemes.
- Use this data to train a neural network to do the same job.
- Replace the parametrisation scheme by the neural network.

Why would you do this?

- A large fraction of the computational cost is generated by parametrisation scheme.
- Parametrisation schemes cause $>$ 90% of model code.
- Optimization of this code is very difficult
  ($\rightarrow$ less than 5% peak performance).
- Neural Networks are highly optimized and can even use co-designed hardware.
  $\rightarrow$ Portability comes for free.

Let's use neural networks to emulate existing parametrisation schemes

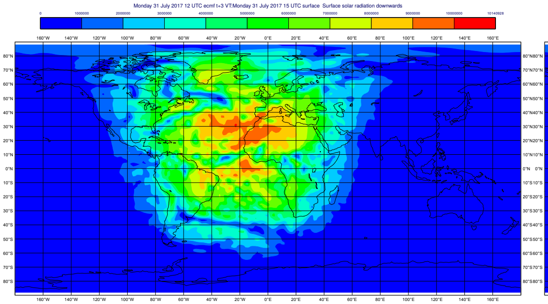
- Store input/output pairs of parametrisation schemes.
- Use this data to train a neural network to do the same job.
- Replace the parametrisation scheme by the neural network.

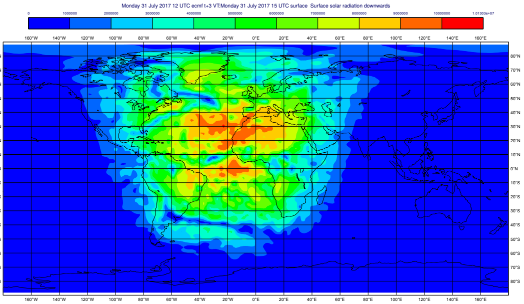
Why would you do this?

- A large fraction of the computational cost is generated by parametrisation scheme.
- Parametrisation schemes cause $> 90\%$ of model code.
- Optimization of this code is very difficult
  ($\rightarrow$ less than 5% peak performance).
- Neural Networks are highly optimized and can even use co-designed hardware.
  $\rightarrow$ Portability comes for free.

We hope that deep Neural Networks will be almost as good as the original parametrisation schemes but much more efficient.

A neural network emulator for the state-of-the-art model configuration with 137 vertical levels

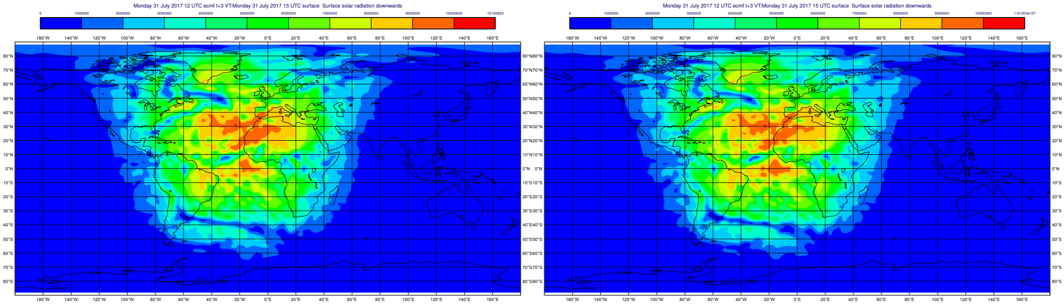




Dueben, Hogan, Bauer @ECMWF and Progsch, Angerer @NVIDIA

Downward solar radiation at the surface for the original radiation scheme and the Neural Network emulator.

A neural network emulator for the state-of-the-art model configuration with 137 vertical levels

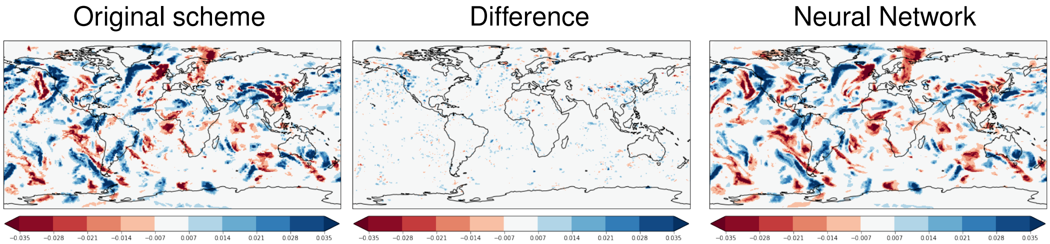


Dueben, Hogan, Bauer @ECMWF and Progsch, Angerer @NVIDIA

Downward solar radiation at the surface for the original radiation scheme and the Neural Network emulator.

Free-running model simulations are stable and we get a factor of O(10) speed-up for radiation.

A neural network emulator for gravity wave drag

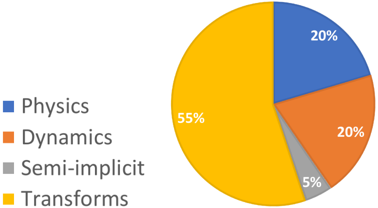


Chantry, Dueben, Palmer.

Tendency output for the non-orographic gravity wave drag parametrisation scheme for the standard scheme and a neural network emulator.

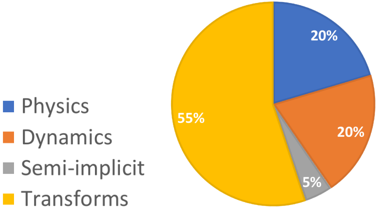
Use machine learning hardware for the Legendre transform

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



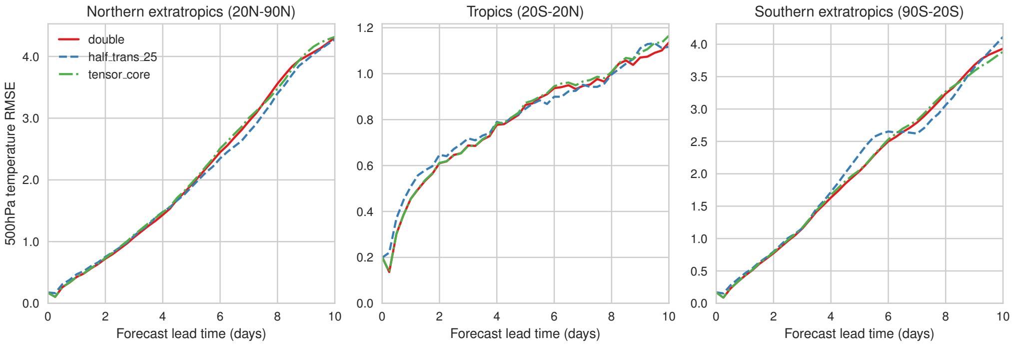
Use machine learning hardware for the Legendre transform

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.

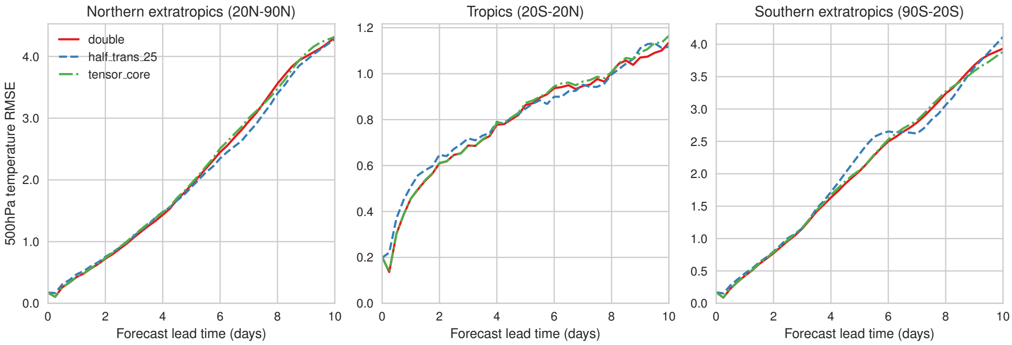
Half precision Legendre Transformations



Root-mean-square error for Z500 at 9 km resolution averaged over multiple start dates.

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

Half precision Legendre Transformations



Root-mean-square error for Z500 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

The challenge for the machine-learning community

The challenge for the machine-learning community

The time of nice pictures and videos is over...

The challenge for the machine-learning community

The time of nice pictures and videos is over...

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

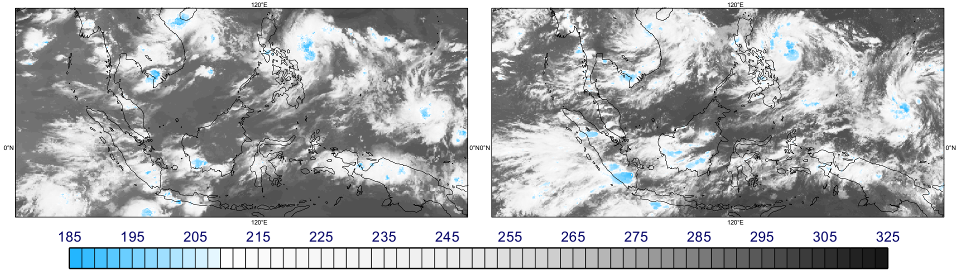
The challenge for the machine-learning community

The time of nice pictures and videos is over...

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.*

Why is this challenge so hard?

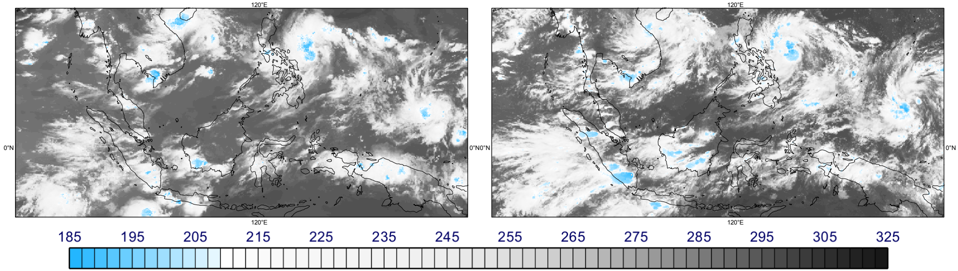


Dueben, Wedi, Saarinen submitted

Top-of-the-atmosphere cloud brightness temperature [K] for satellite observations and a simulation of the atmosphere with 1.45 km resolution.

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

Why is this challenge so hard?



Dueben, Wedi, Saarinen submitted

Top-of-the-atmosphere cloud brightness temperature [K] for satellite observations and a simulation of the atmosphere with 1.45 km resolution.

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

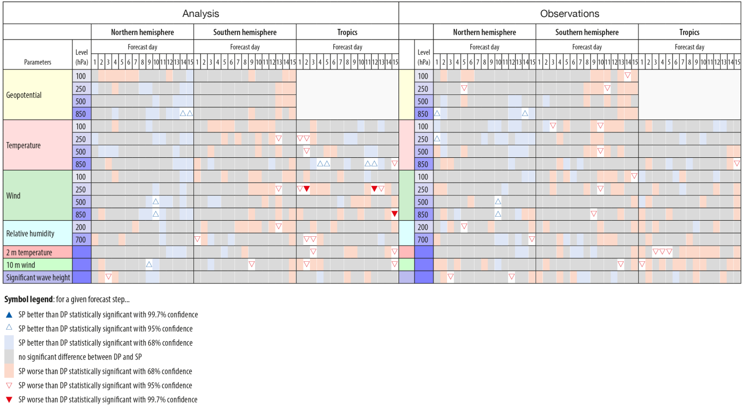
Global simulations show a breath-taking level of complexity and can represent many details of the Earth System.

Why is this challenge so hard?

Dueben and Palmer MWR 2015:
Single precision runs in IFS are possible and time-to-solution is reduced by 40%.

Why is this challenge so hard?

Dueben and Palmer MWR 2015:
Single precision runs in IFS are possible and time-to-solution is reduced by 40%.



Why is this challenge so hard?

Real-world models have a myriad of compensating errors.

Why is this challenge so hard?

Real-world models have a myriad of compensating errors.

You may need to run 20x10x15 days of weather forecasts or 10 years of a coupled climate model to identify a response to a forcing...

Why is this challenge so hard?

Real-world models have a myriad of compensating errors.

You may need to run 20x10x15 days of weather forecasts or 10 years of a coupled climate model to identify a response to a forcing...

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

Questions to be answered...

There is no fundamental reasons not to use a black box within weather and climate models. However,...

Questions to be answered...

There is no fundamental reasons not to use a black box within weather and climate models. However,...

- How can we use our knowledge about the Earth System?

Questions to be answered...

There is no fundamental reasons not to use a black box within weather and climate models. However,...

- How can we use our knowledge about the Earth System?
- How can we diagnose physical knowledge from the network?

Questions to be answered...

There is no fundamental reasons not to use a black box within weather and climate models. However,...

- How can we use our knowledge about the Earth System?
- How can we diagnose physical knowledge from the network?
- How can we remove a bias from neural networks and secure conservation laws?

Questions to be answered...

There is no fundamental reasons not to use a black box within weather and climate models. However,...

- How can we use our knowledge about the Earth System?
- How can we diagnose physical knowledge from the network?
- How can we remove a bias from neural networks and secure conservation laws?
- How can we pick hyper-parameters?

Questions to be answered...

There is no fundamental reasons not to use a black box within weather and climate models. However,...

- How can we use our knowledge about the Earth System?
- How can we diagnose physical knowledge from the network?
- How can we remove a bias from neural networks and secure conservation laws?
- How can we pick hyper-parameters?
- How can we guarantee reproducibility during training?

Questions to be answered...

There is no fundamental reasons not to use a black box within weather and climate models. However,...

- How can we use our knowledge about the Earth System?
- How can we diagnose physical knowledge from the network?
- How can we remove a bias from neural networks and secure conservation laws?
- How can we pick hyper-parameters?
- How can we guarantee reproducibility during training?
- How can we get beyond “dense” networks but still take local properties into account?

Questions to be answered...

There is no fundamental reasons not to use a black box within weather and climate models. However,...

- How can we use our knowledge about the Earth System?
- How can we diagnose physical knowledge from the network?
- How can we remove a bias from neural networks and secure conservation laws?
- How can we pick hyper-parameters?
- How can we guarantee reproducibility during training?
- How can we get beyond "dense" networks but still take local properties into account?
- How can we fix interactions between model components?

Questions to be answered...

There is no fundamental reasons not to use a black box within weather and climate models. However,...

- How can we use our knowledge about the Earth System?
- How can we diagnose physical knowledge from the network?
- How can we remove a bias from neural networks and secure conservation laws?
- How can we pick hyper-parameters?
- How can we guarantee reproducibility during training?
- How can we get beyond “dense” networks but still take local properties into account?
- How can we fix interactions between model components?
- How can we interface neural networks with legacy codes and how can we use AI and conventional methods efficiently on the same hardware?

Questions to be answered...

There is no fundamental reasons not to use a black box within weather and climate models. However,...

- How can we use our knowledge about the Earth System?
- How can we diagnose physical knowledge from the network?
- How can we remove a bias from neural networks and secure conservation laws?
- How can we pick hyper-parameters?
- How can we guarantee reproducibility during training?
- How can we get beyond “dense” networks but still take local properties into account?
- How can we fix interactions between model components?
- How can we interface neural networks with legacy codes and how can we use AI and conventional methods efficiently on the same hardware?
- How can we design good training data (short time steps and high resolution)?

Questions to be answered...

There is no fundamental reasons not to use a black box within weather and climate models. However,...

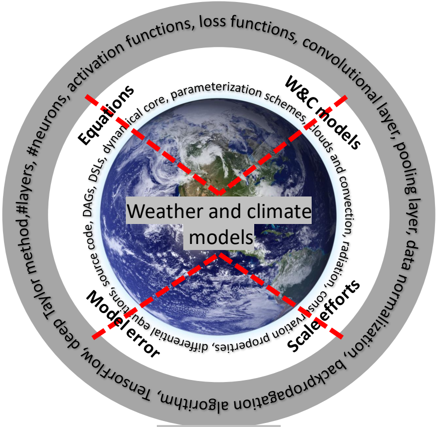
- How can we use our knowledge about the Earth System?
- How can we diagnose physical knowledge from the network?
- How can we remove a bias from neural networks and secure conservation laws?
- How can we pick hyper-parameters?
- How can we guarantee reproducibility during training?
- How can we get beyond “dense” networks but still take local properties into account?
- How can we fix interactions between model components?
- How can we interface neural networks with legacy codes and how can we use AI and conventional methods efficiently on the same hardware?
- How can we design good training data (short time steps and high resolution)?
- How can we explore the full phase space (all weather regimes) during training?

Questions to be answered...

There is no fundamental reasons not to use a black box within weather and climate models. However,...

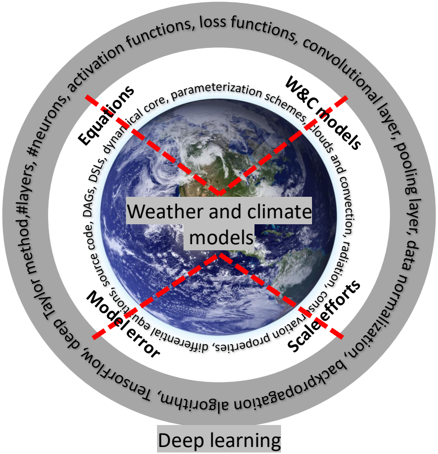
- How can we use our knowledge about the Earth System?
- How can we diagnose physical knowledge from the network?
- How can we remove a bias from neural networks and secure conservation laws?
- How can we pick hyper-parameters?
- How can we guarantee reproducibility during training?
- How can we get beyond “dense” networks but still take local properties into account?
- How can we fix interactions between model components?
- How can we interface neural networks with legacy codes and how can we use AI and conventional methods efficiently on the same hardware?
- How can we design good training data (short time steps and high resolution)?
- How can we explore the full phase space (all weather regimes) during training?
- How can we generate large labelled datasets?

My personal vision of the way forward...



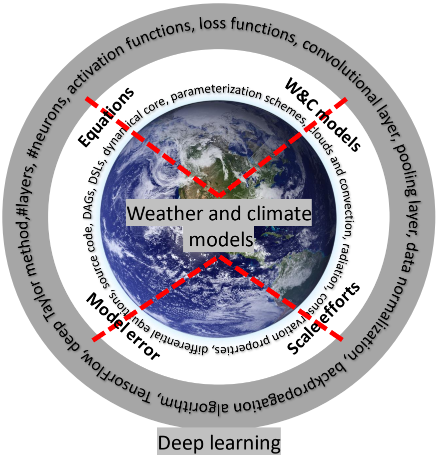
Deep learning

My personal vision of the way forward...



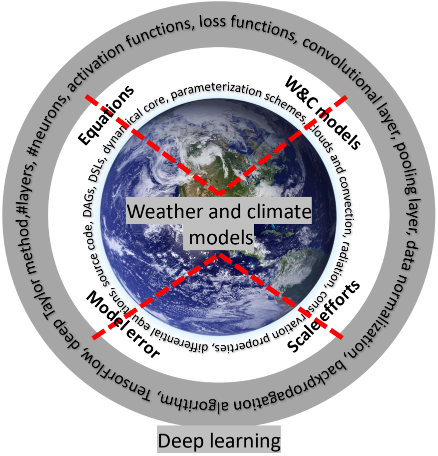
- To study known differential equations to learn how to derive blueprints for neural network architectures.

My personal vision of the way forward...



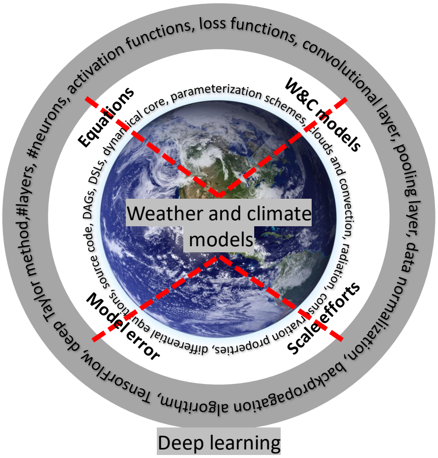
- To study known differential equations to learn how to derive blueprints for neural network architectures.
- To learn how to scale neural networks to O(1,000,000) inputs for 3D fields on the sphere.

My personal vision of the way forward...



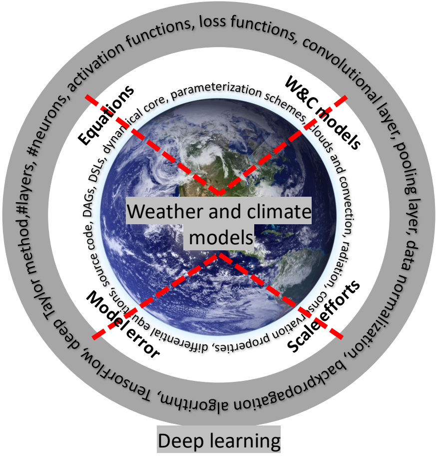
- To study known differential equations to learn how to derive blueprints for neural network architectures.
- To learn how to scale neural networks to O(1,000,000) inputs for 3D fields on the sphere.
- To study the representation of variability and the correction of systematic errors for neural networks.

My personal vision of the way forward...



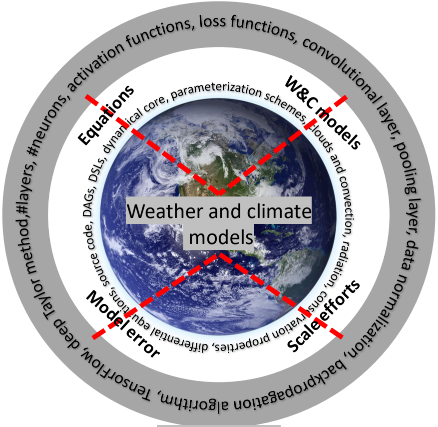
- To study known differential equations to learn how to derive blueprints for neural network architectures.
- To learn how to scale neural networks to O(1,000,000) inputs for 3D fields on the sphere.
- To study the representation of variability and the correction of systematic errors for neural networks.
- To build community tools in the form of benchmark problems.

My personal vision of the way forward...



- To study known differential equations to learn how to derive blueprints for neural network architectures.
- To learn how to scale neural networks to O(1,000,000) inputs for 3D fields on the sphere.
- To study the representation of variability and the correction of systematic errors for neural networks.
- To build community tools in the form of benchmark problems.
- To focus on *useful tools* that can serve as beacon.

My personal vision of the way forward...

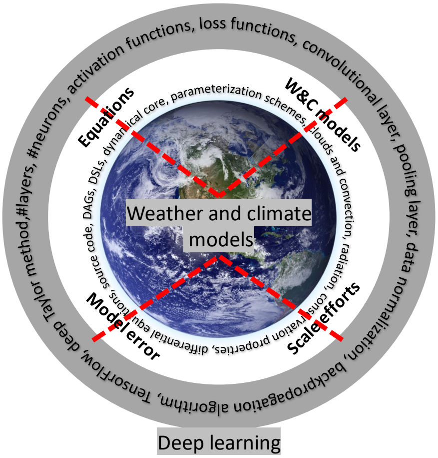


Deep learning

- To study known differential equations to learn how to derive blueprints for neural network architectures.
- To learn how to scale neural networks to O(1,000,000) inputs for 3D fields on the sphere.
- To study the representation of variability and the correction of systematic errors for neural networks.
- To build community tools in the form of benchmark problems.
- To focus on *useful tools* that can serve as beacon.

This will be hard work.

My personal vision of the way forward...



- To study known differential equations to learn how to derive blueprints for neural network architectures.
- To learn how to scale neural networks to O(1,000,000) inputs for 3D fields on the sphere.
- To study the representation of variability and the correction of systematic errors for neural networks.
- To build community tools in the form of benchmark problems.
- To focus on *useful tools* that can serve as beacon.

This will be hard work.

We will require machine learning solutions that are customized to weather and climate models.

An example: The Burgers equation

Let's represent a non-linear system that is approximated by the Burgers' equation:

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial x^2} - u \frac{\partial u}{\partial x} + p.$$

An example: The Burgers equation

Let's represent a non-linear system that is approximated by the Burgers' equation:

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial x^2} - u \frac{\partial u}{\partial x} + p.$$

The conventional approach:

$$\frac{\partial u_i}{\partial t} = \nu \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} - u_i \frac{u_{i+1} - u_{i-1}}{2\Delta x} + c_0 + c_1 \cdot u_i + c_2 \cdot u_i + c_3 \cdot u_i \cdot \zeta.$$

An example: The Burgers equation

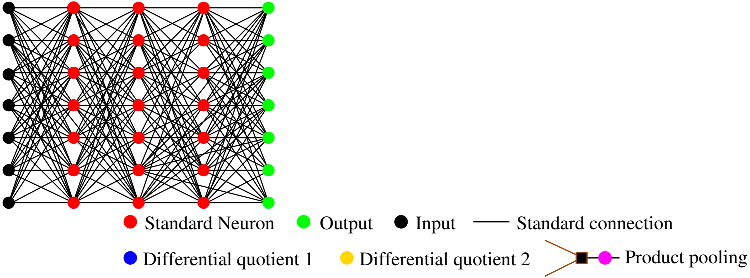
Let's represent a non-linear system that is approximated by the Burgers' equation:

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial x^2} - u \frac{\partial u}{\partial x} + p.$$

The conventional approach:

$$\frac{\partial u_i}{\partial t} = \nu \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} - u_i \frac{u_{i+1} - u_{i-1}}{2\Delta x} + c_0 + c_1 \cdot u_i + c_2 \cdot u_i + c_3 \cdot u_i \cdot \zeta.$$

The data-science approach:



An example: The Burgers equation

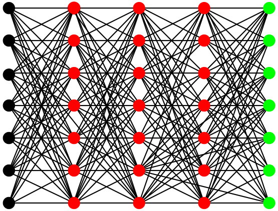
Let's represent a non-linear system that is approximated by the Burgers' equation:

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial x^2} - u \frac{\partial u}{\partial x} + p.$$

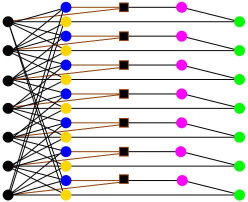
The conventional approach:

$$\frac{\partial u_i}{\partial t} = \nu \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} - u_i \frac{u_{i+1} - u_{i-1}}{2\Delta x} + c_0 + c_1 \cdot u_i + c_2 \cdot u_i + c_3 \cdot u_i \cdot \zeta.$$

The data-science approach:



The way forward:



Standard Neuron   Output   Input   Standard connection

Differential quotient 1   Differential quotient 2   Product pooling

Conclusions

- Weather and climate models offer a large number of promising applications for machine learning that span the entire workflow of numerical weather predictions.
- We can learn the equations of motion using deep learning and build medium-complexity models.
- We need to develop useful tools based on machine learning within the next two years.
- We need to learn how to emulate known differential equations, represent sub-grid-scale variability and systematic errors, and learn how to scale machine learning solutions to O(1,000,000) input fields.
- We need *benchmark datasets and beacons*.
- Machine learning hardware can be used to speed-up conventional model simulations.

Conclusions

- Weather and climate models offer a large number of promising applications for machine learning that span the entire workflow of numerical weather predictions.
- We can learn the equations of motion using deep learning and build medium-complexity models.
- We need to develop useful tools based on machine learning within the next two years.
- We need to learn how to emulate known differential equations, represent sub-grid-scale variability and systematic errors, and learn how to scale machine learning solutions to O(1,000,000) input fields.
- We need *benchmark datasets and beacons*.
- Machine learning hardware can be used to speed-up conventional model simulations.

Many thanks for your attention.

Funding

Funded by the
European Union

The presenter gratefully acknowledges the funding a Royal Society University Research Fellowship and from the ESIWACE and ESIWACE2 project. The ESIWACE2 projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823988.