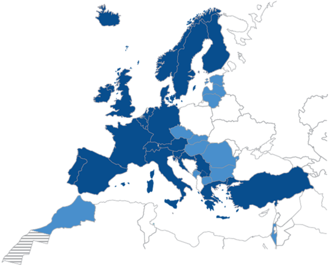
Deep learning for weather and climate

Peter Düben

European Centre for Medium-Range Weather Forecasts (ECMWF)

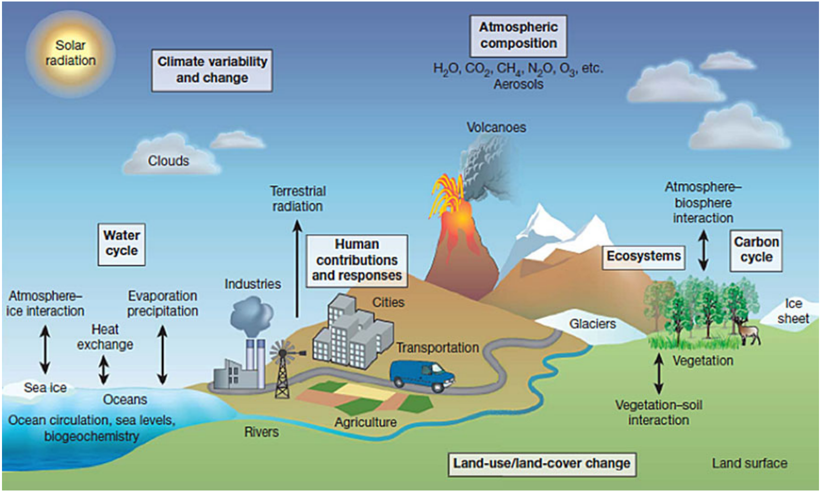
The European Centre for Medium-Range Weather Forecasts (ECMWF)



www.ecmwf.int

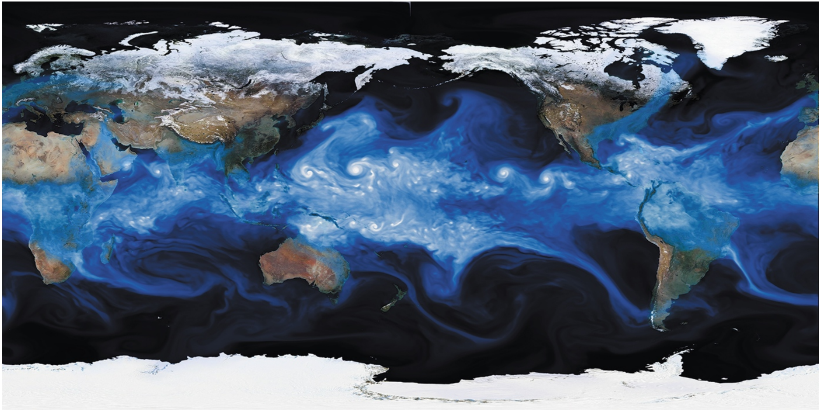
- Research institute and 24/7 operational weather service for medium-range, monthly and seasonal forecasts.
- Independent, intergovernmental organization supported by 34 states.
- Based close to Oxford in the UK; $\approx$ 350 member of staff.
- Home of two supercomputers.
- Home of the Integrated Forecast System (IFS).

Predicting weather and climate: Why is it so hard?



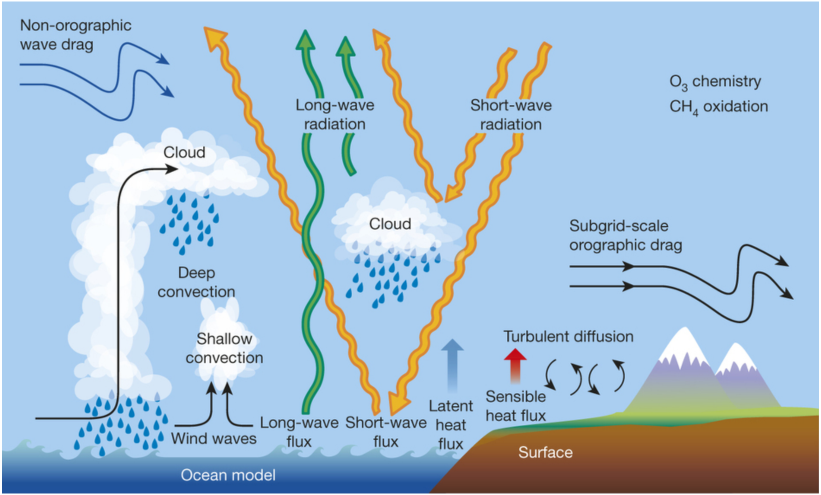
www.gfdl.noaa.gov

Predicting weather and climate: Why is it so hard?



Wehner and Prabhat

Predicting weather and climate: Why is it so hard?



Bauer et al. Nature 2015

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?

Clouds in a global weather simulation at 1 km resolution.

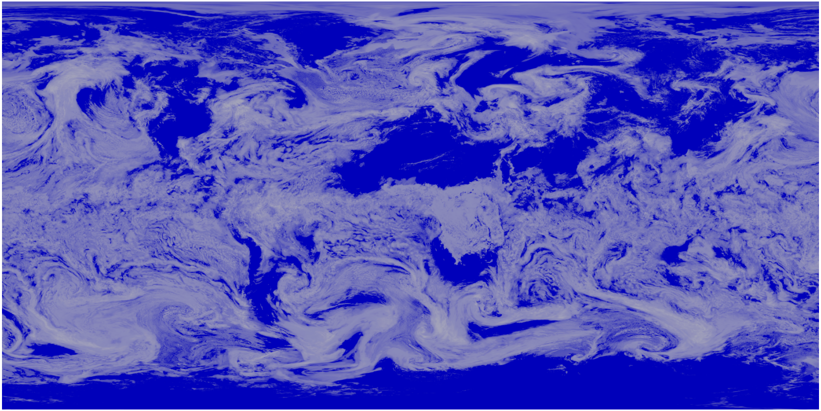


Figure courtesy of Nils Wedi.

Predicting weather and climate: Why is it so hard?

Clouds in a global weather simulation at 1 km resolution.

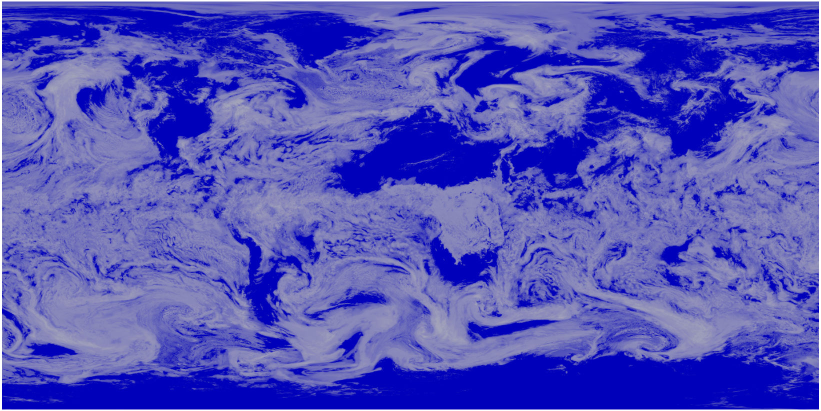


Figure courtesy of Nils Wedi.

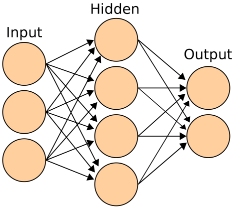
Global simulations show a breath-taking level of complexity and can represent many details of the Earth System.

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.

Deep learning for weather and climate

Outline:

- **Emulate existing model components.**
- Improve existing model components.
- Learn the equations of motion.
- Improve post-processing.
- Use machine learning hardware.
- Challenges for deep learning in weather and climate models.

Emulate existing model components

- 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.

Emulate existing model components

- 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.

Emulate existing model components

- 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.

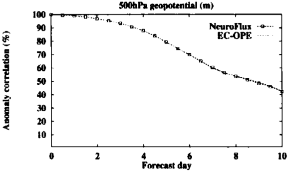
Why would you do this?

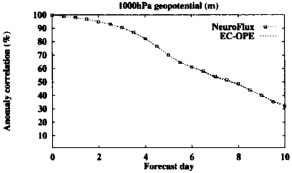
- A large fraction of the computational cost is generated by parametrisation scheme.
- Parametrisation schemes cause $>$ 90% of model code.
- Optimization of this code is very difficult
  ($\rightarrow$ less than 5% peak performance).
- Neural Networks are highly optimized and can even use co-designed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

Neural Networks to replace the radiation scheme at ECMWF in the year 2000

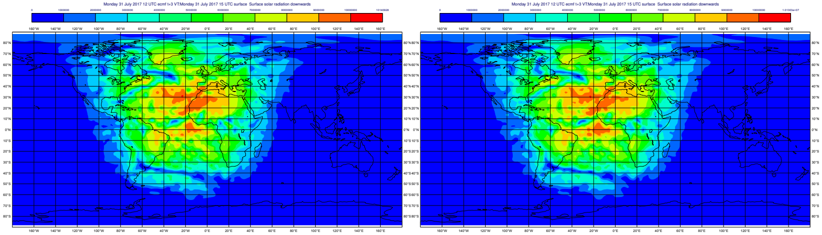
- 20-30 hidden neurons.
- Trained on 80,000 vertical profiles.
- Accuracy of the new scheme was comparable.
- The new scheme was seven times faster.
- The network could be used to generate tangent linear and adjoint code for 4DVar data assimilation.
- However, Neural Networks are currently not used in operational models.





Chevallier et al. QJRMS 2000.

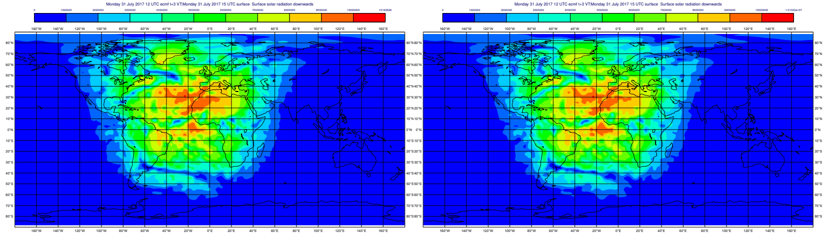
A neural network emulator for the state-of-the-art model configuration with 137 vertical levels



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

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

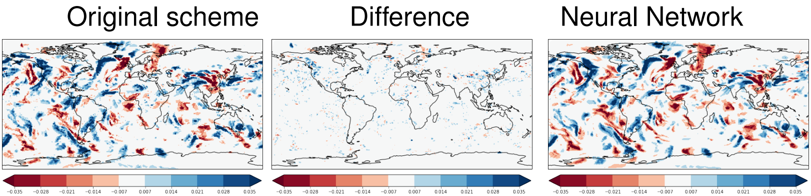


Progsch, Ko, Angerer @NVIDIA and Dueben, Hogan, Bauer @ECMWF

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

However, we still need to stabilize free-running model simulations with the Neural Network and more work is required.

A neural network emulator for gravity wave drag



Chantry, Abdelrahman, Desai, Dueben, Palem, Palmer.

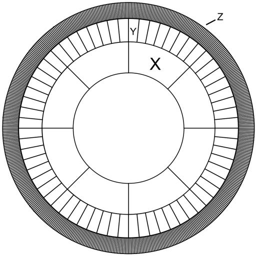
Tendency output for the non-orographic gravity wave drag parametrisation scheme for the standard scheme and a neural network emulator.

Deep learning for weather and climate

Outline:

- Emulate existing model components.
- **Improve existing model components.**
- Learn the equations of motion.
- Improve post-processing.
- Use machine learning hardware.
- Challenges for deep learning in weather and climate models.

Lorenz'95 and parametrisation schemes: A toy model as surrogate for the Earth System



Truth:

$$\frac{dX_k}{dt} = X_{k-1}\left(X_{k+1} - X_{k-2}\right) - X_k + F - \frac{hc}{b}\sum_{j=1}^{J} Y_{j,k}$$

$$\frac{dY_{j,k}}{dt} = -cbY_{j+1,k}\left(Y_{j+2,k} - Y_{j-1,k}\right) - cY_{j,k} + \frac{hc}{b}X_k - \frac{he}{d}\sum_{i=1}^{I} Z_{i,j,k}$$

$$\frac{dZ_{i,j,k}}{dt} = edZ_{i-1,j,k}\left(Z_{i+1,j,k} - Z_{i-2,j,k}\right) - g_Z eZ_{i,j,k} + \frac{he}{d}Y_{j,k}$$

Model:

$$\frac{dX_k}{dt} = X_{k-1}\left(X_{k+1} - X_{k-2}\right) - X_k + F + U(X_k)$$

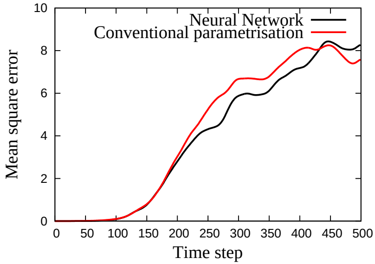
- We use the three-level Lorenz'95 model (Thornes et al. QJRMS 2017) to study scale interactions in a non-linear environment.
- Three levels are the truth, one level is the model.
- To find the right parametrisation scheme is tricky ($U(X_k) \approx -\frac{hc}{b}\sum_{j=1}^{J} Y_{j,k}$).

Learn parametrisation schemes
Results from John Griffith

- We perform a long run with the Lorenz'95 truth and diagnose the parametrisation term $U = -\frac{hc}{b}\sum_{j=1}^{J} Y_{j,k}$.
- We train a Neural Network to learn $U(X_k)$ for the coarse resolution model.
- We run a parametrised model using the trained $U(X_k)$ and compare results against the conventional method to fit a polynomial.

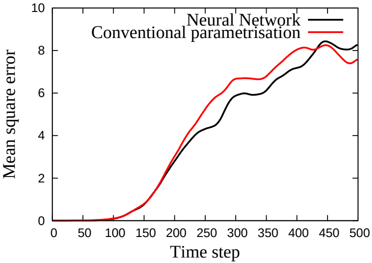
Learn parametrisation schemes
Results from John Griffith

- We perform a long run with the Lorenz'95 truth and diagnose the parametrisation term $U = -\frac{hc}{b}\sum_{j=1}^{J} Y_{j,k}$.
- We train a Neural Network to learn $U(X_k)$ for the coarse resolution model.
- We run a parametrised model using the trained $U(X_k)$ and compare results against the conventional method to fit a polynomial.



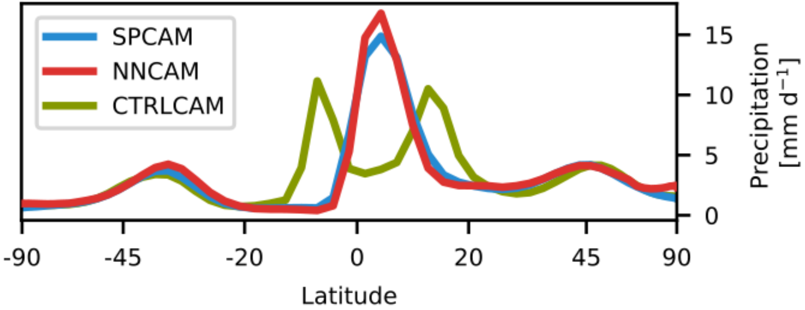
Learn parametrisation schemes
Results from John Griffith

- We perform a long run with the Lorenz'95 truth and diagnose the parametrisation term $U = -\frac{hc}{b}\sum_{j=1}^{J} Y_{j,k}$.
- We train a Neural Network to learn $U(X_k)$ for the coarse resolution model.
- We run a parametrised model using the trained $U(X_k)$ and compare results against the conventional method to fit a polynomial.



Results are promising.

Recent results for a superparametrised model



- *Rasp, Pritchard and Gentine (arXiv 2018)* have trained a Neural Network to emulate the parametrisation schemes from a superparametrised model.
- They could replicate the benefits of superparametrisation in comparison to the standard model using a Neural Network that was 10 times faster.

Deep learning for weather and climate

Outline:

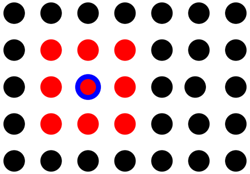
- Emulate existing model components.
- Improve existing model components.
- **Learn the equations of motion.**
- Improve post-processing.
- Use machine learning hardware.
- Challenges for deep learning in weather and climate models.

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

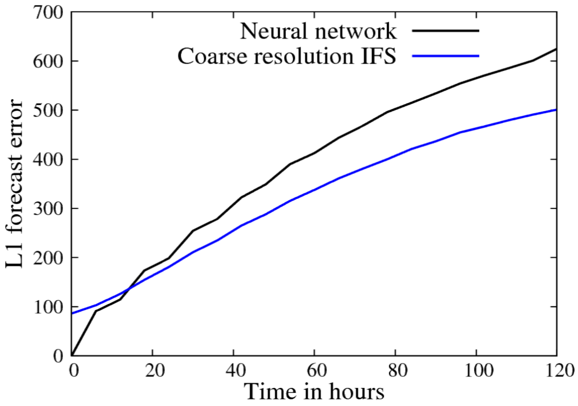
- We know the equations of motion but we cannot solve them.
- Discretisation and sub-grid-scale variability generates significant errors.
- The data handling system of ECMWF provides access to over 210 petabyte of primary data and the data archive of ECMWF grows by about 233 terabyte per day.
- Data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

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

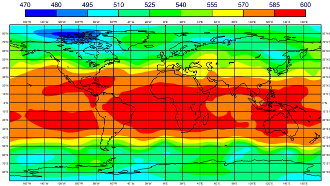


The Neural Network model can compete with a dynamical model of similar complexity.

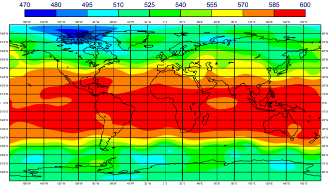
Dueben and Bauer GMD 2018

Global weather forecast based on Neural Networks

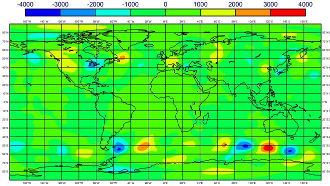
24 hours; Analysis



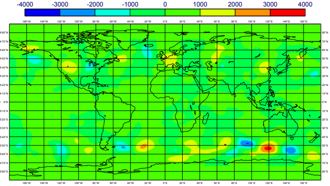
24 hours; Local Neural Network



24 hours; Analysis

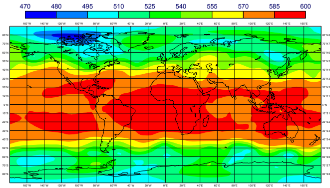


24 hours; Local Neural Network

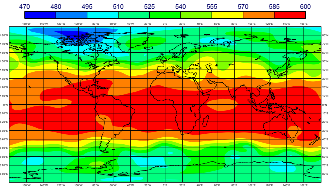


Global weather forecast based on Neural Networks

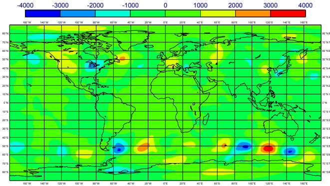
24 hours; Analysis



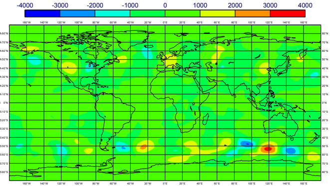
24 hours; Local Neural Network



24 hours; Analysis



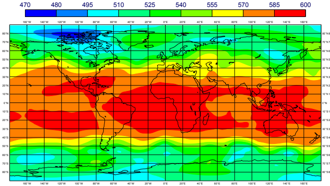
24 hours; Local Neural Network



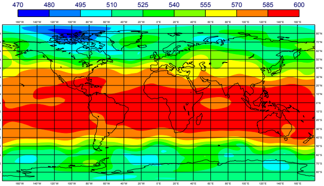
The simulations show reasonable dynamics.

Global weather forecast based on Neural Networks

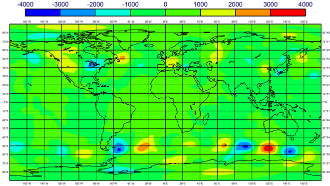
24 hours; Analysis



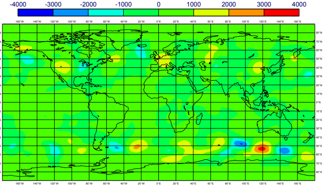
24 hours; Local Neural Network



24 hours; Analysis



24 hours; Local Neural Network

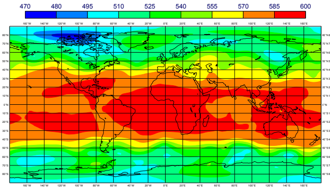


The simulations show reasonable dynamics.

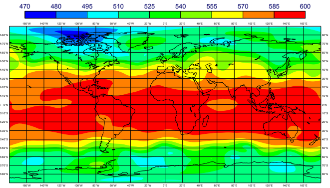
Just adding further inputs does not necessarily help.

Global weather forecast based on Neural Networks

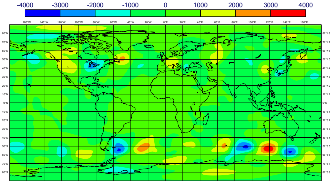
24 hours; Analysis



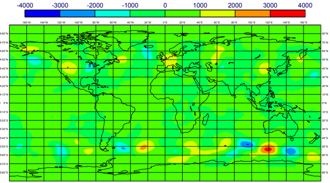
24 hours; Local Neural Network



24 hours; Analysis



24 hours; Local Neural Network



The simulations show reasonable dynamics.

Just adding further inputs does not necessarily help.

Model runs crash after a couple of weeks.

Deep learning for weather and climate

Outline:

- Emulate existing model components.
- Improve existing model components.
- Learn the equations of motion.
- **Improve post-processing.**
- Use machine learning hardware.
- Challenges for deep learning in weather and climate models.

Improve post-processing

Ensemble simulations are important but expensive.

Improve post-processing

Ensemble simulations are important but expensive.

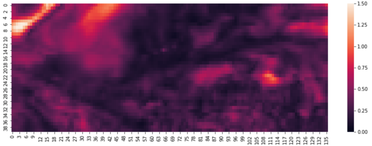
We use one model trajectory, the ensemble mean and the ensemble spread field at initialisation to predict the ensemble spread of a 10 member ensemble six hours into the forecast for an area over Europe (40W-30E and 40N-60N).

Improve post-processing

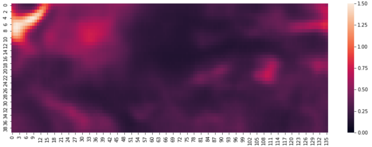
Ensemble simulations are important but expensive.

We use one model trajectory, the ensemble mean and the ensemble spread field at initialisation to predict the ensemble spread of a 10 member ensemble six hours into the forecast for an area over Europe (40W-30E and 40N-60N).

Spread after 6 hours:



Prediction from neural network:



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

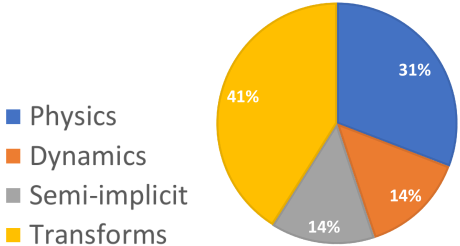
Deep learning for weather and climate

Outline:

- Emulate existing model components.
- Improve existing model components.
- Learn the equations of motion.
- Improve post-processing.
- **Use machine learning hardware.**
- Challenges for deep learning in weather and climate models.

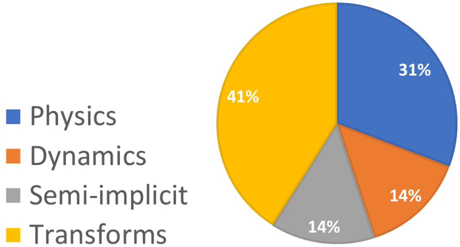
Use machine learning hardware

Relative cost for model components at 1.25 km for a spectral model:



Use machine learning hardware

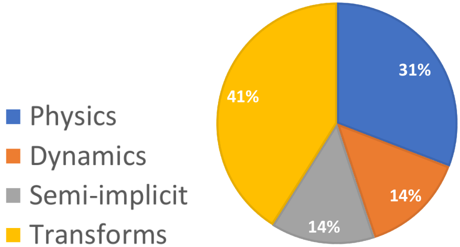
Relative cost for model components at 1.25 km for a spectral model:



The Legendre transforms are the killer (as expected). They are standard matrix-matrix multiplications.

Use machine learning hardware

Relative cost for model components at 1.25 km for a spectral model:

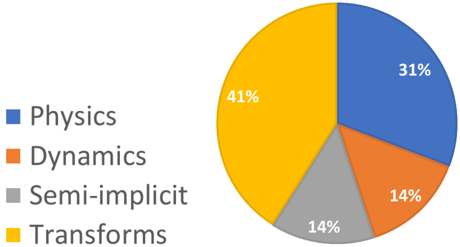


The Legendre transforms are the killer (as expected). They are standard matrix-matrix multiplications.

If we can re-scale the input and output fields, we can use half precision arithmetic (low zonal wave numbers need to be secured).

Use machine learning hardware

Relative cost for model components at 1.25 km for a spectral model:

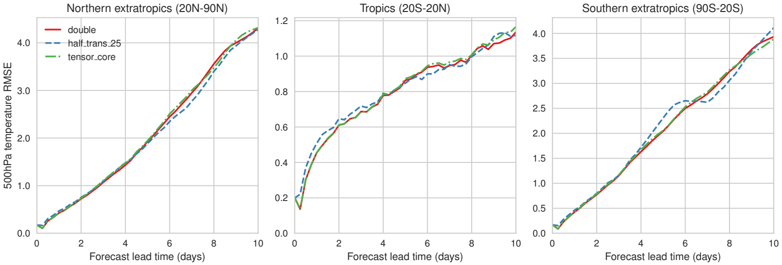


The Legendre transforms are the killer (as expected). They are standard matrix-matrix multiplications.

If we can re-scale the input and output fields, we can use half precision arithmetic (low zonal wave numbers need to be secured).

Tensor Cores on NVIDIA Volta GPUs are optimized 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.

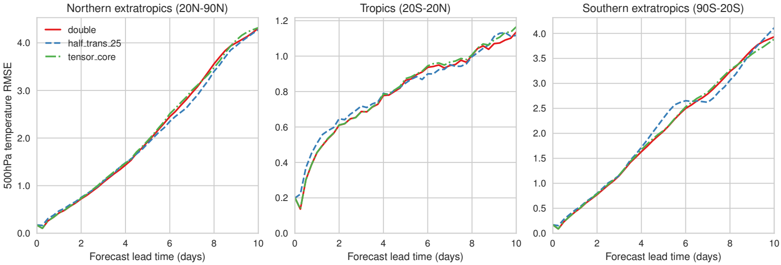
Half precision Legendre Transformations



Root-mean-square error for Z500 at TCo1279 resolution averaged over multiple start dates.

Hatfield, Chantry, Dueben, Palmer, submitted to PASC2019.

Half precision Legendre Transformations



Root-mean-square error for Z500 at TCo1279 resolution averaged over multiple start dates.

Hatfield, Chantry, Dueben, Palmer, submitted to PASC2019.

The simulations are using an emulator to reduce precision.

Dawson and Dueben GMD 2017

Deep learning for weather and climate

Outline:

- Emulate existing model components.
- Improve existing model components.
- Learn the equations of motion.
- Improve post-processing.
- Use machine learning hardware.
- **Challenges for deep learning in weather and climate models.**

To improve a multi-dimensional, non-linear system...

To improve a multi-dimensional, non-linear system...

You may need to run 100 years of a coupled climate model to identify a response to a forcing...

Challenges for deep learning

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

Challenges for deep learning

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?

Challenges for deep learning

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?

Challenges for deep learning

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?

Challenges for deep learning

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?
- How can we secure conservation laws?

Challenges for deep learning

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?
- How can we secure conservation laws?
- How can we hyper-parameters?

Challenges for deep learning

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?
- How can we secure conservation laws?
- How can we hyper-parameters?
- How can we guarantee reproducibility during training?

Challenges for deep learning

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?
- How can we secure conservation laws?
- How can we hyper-parameters?
- How can we guarantee reproducibility during training?
- How can we get beyond “dense” networks but still take local properties into account?

Challenges for deep learning

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?
- How can we secure conservation laws?
- How can we 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?

Challenges for deep learning

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?
- How can we secure conservation laws?
- How can we 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 design good training data (short time steps and high resolution)?

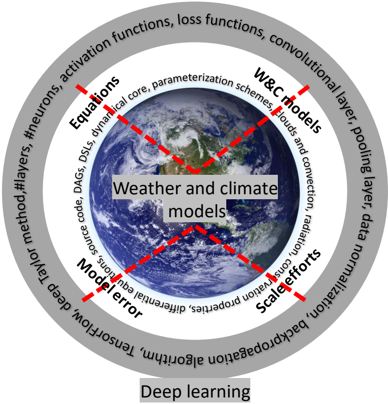
Challenges for deep learning

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?
- How can we secure conservation laws?
- How can we 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 design good training data (short time steps and high resolution)?
- How can we we explore the full phase space (all weather regimes) during training?

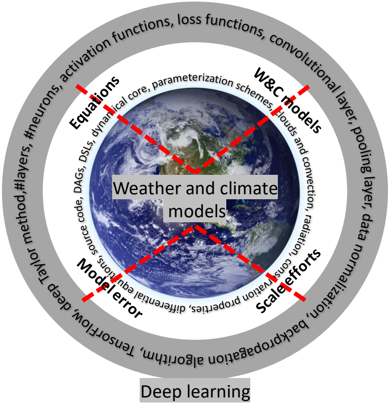
The way forward

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



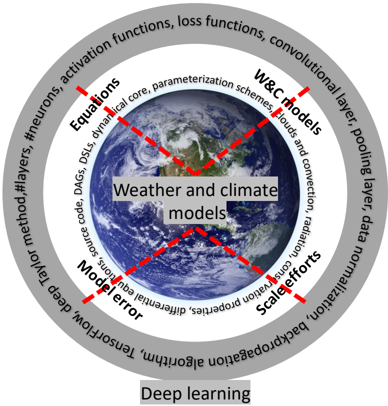
The way forward

- To study known differential equations to learn how to derive blueprints for neural network architectures.
- To study model source code to learn how to derive blueprints for the design of network architectures.



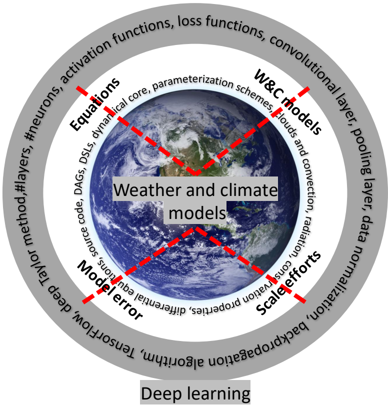
The way forward

- To study known differential equations to learn how to derive blueprints for neural network architectures.
- To study model source code to learn how to derive blueprints for the design of network architectures.
- To study the representation of sub-grid-scale processes and systematic errors for neural networks.



The way forward

- To study known differential equations to learn how to derive blueprints for neural network architectures.
- To study model source code to learn how to derive blueprints for the design of network architectures.
- To study the representation of sub-grid-scale processes and systematic errors for neural networks.
- To scale the application of neural networks in W&C models beyond today's limits.



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

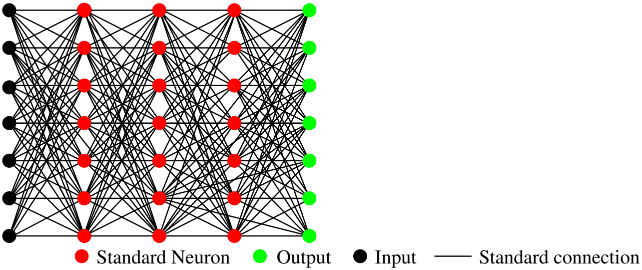
Let's represent a non-linear system that is approximated by the Burgers' equation:

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

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

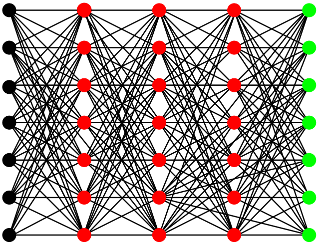
Let's represent a non-linear system that is approximated by the Burgers' equation:

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

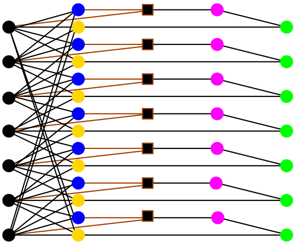
The conventional approach:

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

The data-science approach:



The way forward:



● Standard Neuron ● Output ● Input —— Standard connection

ECMWF's Summer of Weather Codes (ESoWC)

It will be essential to develop **benchmark tests for deep-learning applications in weather and climate science** to address the challenges ahead.

(conservation and bias, hyper-parameter-zoo, physical consistency, black-box-character, dense connectivity...)

→ I have proposed a challenge to ESoWC.

Apply until 17 April and win £5000 to work on this!

https://www.ecmwf.int/en/learning/workshops/ecmwf-summer-weather-code-2019

Conclusions

- It is likely that deep learning will play an important role in future weather and climate models.
- There are many different potential application areas for neural networks.
- This requires a better understanding how knowledge of the physical system can be projected into the network configurations and how to “debug” biases etc..
- Deep learning hardware may also be useful for weather and climate models.
- Challenges are similar when compared to the development of conventional models (Earth System complexity, non-linearity and scale interactions, exponential error growth, numerical instabilities, the sphere, conservation properties, model biases, uncertainty and insufficient coverage of observations,...)

Funding

Funded by the European Union

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