Machine Learning at the Met Office

Niall Robinson

Deputy Head, Met Office Informatics Lab
Global Systems Institute, Exeter University,
Innovation Manager, ERDF Environmental Futures and Big
Data Impact Lab

Thanks to:
Rachel Prudden, Alberto Arribas, Sam Adams,
Simon Jackson, Tom Dunstan et al from the Met Office
Dmitri Kangin from Exeter University
Shakir Mohamed, Suman Ravuri from Google DeepMind
Talk menu:

- Some things we've done at the Met Office
- Some things we're doing at the Met Office
- Some things we could do as a community
- Some things we could do to make ML easier to do in our community...

...but first
Quick! We need to do ML! Everyone else is doing it! Hire some data scientists!

2016 – establishment of the ML Forum. Informal community of interest, but spawned some interesting prototype projects

2019 – establishment of the ML Strategy Working Group develop a ML Strategy for inclusion in the next Met Office Science Strategy
Quick! We need to do ML! Everyone else is doing it!

Hire some data scientists!

Everyone be calm

We've got a good track record, it's just that historically we often talk about:

machine learning = statistics
data scientists = physicists

however - there are lots more opportunities.
Quick! We need to do ML! Everyone else is doing it!
Hire some data scientists!

"An algorithm that gets better at doing something as it experiences more data"...me
Quick! We need to do ML! Everyone else is doing it! Hire some data scientists!

"An algorithm that gets better at doing something as it experiences more data"...me

Modern ML can be highly non-linear now
"How can we trust a black box?"

- Validation validation validation...obviously.
"How can we trust a black box?"

- Validation validation validation...obviously...but

Need to be confident we right for the right reasons.
"How can we trust a black box?"

- Validation validation validation.
- Traditional models are black boxes - how do we validate them?
"How can we trust a black box?"

- Validation validation validation.
- Traditional models are black boxes - how do we validate them?
- Drive the stats with physics i.e. embedded Physics
"How can we trust a black box?"

- Validation validation validation.
- Traditional models are black boxes - how do we validate them?
- Drive the stats with physics i.e. embedded Physics
- What about embedding physics in stats?
"How can we trust a black box?"

• Validation validation validation.
• Traditional models are black boxes - how do we validate them?
• Drive the stats with physics i.e. embedded Physics
• What about embedding physics in stats?

Clever loss functions
e.g. insisting on conservation

Curating input information

Neural ODEs

---

Computer Science > Machine Learning

**Neural Ordinary Differential Equations**

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, David Duvenaud
(Submitted on 19 Jun 2018 (v1), last revised 15 Jan 2019 (this version, v4))
Broad areas of application

• Emulating: "what could we do it we had unlimited compute"
  interesting discussion this week about "beating Moore's law"
Broad areas of application

• Emulating: "what could we do if we had unlimited compute"
• Mapping: what we couldn't do if we had unlimited compute
Broad areas of application

• Emulating: "what could we do if we had unlimited compute"
• Mapping: what we couldn't do if we had unlimited compute
• Downscaling: from deterministic to non-deterministic regimes
Broad areas of application

- Emulating: "what could we do if we had unlimited compute"
- Mapping: what we couldn't do if we had unlimited compute
- Downscaling: from deterministic to non-deterministic regimes
- Finding: interesting stuff in the (big) data we can't find
Broad areas of application

- Emulating: "what could we do if we had unlimited compute"
- Mapping: what we couldn't do if we had unlimited compute
- Downscaling: from deterministic to non-deterministic regimes
- Finding: interesting stuff in the (big) data we can't find
- "Hand of god": black-magic, Move 37 type stuff
Broad areas of application

• Emulating: "what could we do if we had unlimited compute"
• Mapping: what we couldn't do if we had unlimited compute
• Downscaling: from deterministic to non-deterministic regimes
• Finding: interesting stuff in the (big) data we can't find
• "Hand of god": black-magic, Move 37 type stuff

Machine Learning for the Geosciences: Challenges and Opportunities

Anuj Karpate, Imme Ebert-Uphoff, Sai Ravala, Hassan Ali Babaie, and Yipin Kumar

Abstract—Geosciences is a field of great societal relevance that requires solutions to several urgent problems facing our humanity and the planet. As geosciences enters the era of big data, machine learning (ML)—that has been widely successful in commercial domains—offers immense potential to contribute to problems in geosciences. However, problems in geosciences have several unique
Neural Networks 101 - "it's not magic"  
David Hall, yesterday

Numbers in (3 of them)

Input

Hidden

Output

Numbers out (2 of them)
Neural Networks 101 - "it's not magic"  
David Hall, yesterday
Neural Networks 101

Numbers are a weighted sum of numbers from the preceding layer...

\[ x_i' = \sum_i m x + c \]
Neural Networks 101

...but the values go through an activation fn. This creates non-linear magic.
Neural Networks 101

```
<table>
<thead>
<tr>
<th></th>
<th>NN guess</th>
<th>Real answer</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>12</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>42</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
Neural Networks 101

"Deep" in deep learning refers to the bits in the middle.

It means we can learn hierarchies of information.
Neural Networks 102 - ConvNets

ConvNets find "things" in fields

where is in =
Neural Networks 102 - ConvNets

ConvNets find "things" in fields
They find hierarchies of things - i.e. thing in the heat maps of things
Neural Networks 102 - ConvNets

ConvNets find "things" in fields
They find hierarchies of things - i.e. thing in the heat maps of things

in $\Rightarrow$ "ear-ness"

in $\Rightarrow$ "eye-ness"

in $\Rightarrow$ then "nose-ness"

then $\Rightarrow$ "cat-ness"
Neural Networks 102 - ConvNets

ConvNets find "things" in fields
They find hierarchies of things - i.e. thing in the heat maps of things
Back prop means they can learn the things they are looking for!
Neural Networks

Training in progress...
Can we extrapolate?
Can we extrapolate?
Now - a smorgasbord of Met Office ML stuff
Some things we've done

1. Better post-processing

Simon Jackson et al
Better post processing

temp (min/max/instant)
wind speed
(min/max/instant)
relative humidity
weather symbol (!)
forecast lead time
distance to coast
altitude

% within 2°C improvement

Day1: 3.0%
Day3: 2.3%
Day5: 1.5%
Day7: 0.8%
Some things we've done

2. Emulating Physics schemes

Tom Dunstan et al.
Emulating Physics Schemes

Profiling the Met Office Unified Model

60% dynamics : 35% physics

Physics schemes in order of cost:
  1) Convection
     =) Boundary layer inc. surface
  2) Microphysics
  3) Orographic GW drag
  4) Radiation
Emulating Physics Schemes

Radiation scheme is a good candidate for emulation:

• Expensive: only called hourly
• Training data: SOCRATES can be run offline
• Variable complexity: clear-sky, no aerosols / all-sky
• Well understood physics: easier to interpret results

RT emulation has been done before
Chevallier et. al. (2000) Q.J.R. Met. Soc., 126: 761-776
Emulating Physics Schemes
More success with Deep ConvNets

- Included clouds and aerosols
- Increased train/test data size (~1x10^6 samples)
- Moved to NN (Keras/Tensorflow)
- Minimal no. of targets:
  - LW/SW net flux profiles
  - LW/SW net flux divergence profiles
  - LW/SW net surface fluxes
- Training:
  - mini-batch gradient descent, batch size = 32, epochs = 50, loss function = MAE
  - time ≈ 24 hrs on 48 CPUs
Emulating Physics Schemes

NN emulator is more accurate and less noisy than standard radiation scheme
Emulating Physics Schemes

Is the NN emulator fast enough?

<table>
<thead>
<tr>
<th></th>
<th>NB</th>
<th>GA7</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spice (1 x CPU, Xeon E5-2690 v3 @ 2.60GHz)</td>
<td>33.75</td>
<td>0.75</td>
<td>4.16</td>
</tr>
</tbody>
</table>
Emulating Physics Schemes
Is the NN emulator fast enough?

<table>
<thead>
<tr>
<th>Scheme Description</th>
<th>NB</th>
<th>GA7</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spice (1 x CPU, Xeon E5-2690 v3 @ 2.60GHz)</td>
<td>33.75</td>
<td>0.75</td>
<td>4.16</td>
</tr>
<tr>
<td>GW4-Isambard Power 9 (1 x CPU @ 3.8GHz)</td>
<td>-</td>
<td>-</td>
<td>2.86</td>
</tr>
<tr>
<td>GW4-Isambard Power 9 (1 x GPU NVIDIA V100)</td>
<td>-</td>
<td>-</td>
<td>0.12</td>
</tr>
</tbody>
</table>
Emulating Physics Schemes

Can we get in into the UM?...the hardest part?

• Calling Python from Fortran is a terrible idea!
• Why not just re-code it? Time consuming, inflexible
• CFFI (creates a C library which initialises and runs Python code? Maybe OK for tests, probably not very fast
• Use Tensorflow C API directly
Some things we're doing

1. Nowcasting

Dmitri Kangin, Rachel Prudden, Shakir Mohamed, Suman Ravuri
We need to assimilate observations

- A forecast uses about $1 \times 10^{12}$ data points, and assimilates $1 \times 10^7$ observation points.
- 4DVar approach: use error covar matrix to mix last forecast and fresh obs
- Hybrid 4DVar used for deterministic
- Developing En-4DEnVar (?) for ensemble forecast
We need to assimilate observations...

But that presents a couple of problems:
1. It takes an hour or so (and half the operational HPC load!), so the obs are old by the time the forecast goes out
We need to assimilate observations...

But that presents a couple of problems:

1. It takes an hour or so (and half the operational HPC load!), so the obs are old by the time the forecast goes out

2. It's imperfect, so the model gets "shock" which takes model world time to dissipate.
Hence "nowcasting"

Take high res obs e.g. radar

Advect with the wind
My latent model:
Rain is created by moving air up so it cools by:
1. Pushing it up and over a mountain
2. Pushing it over some other air (i.e. a "front")
3. Convection
   ...plus moving this stuff sideways
Future work, we want a neural network based nowcasting system that
• is skilful because it takes advantage of prior physical knowledge
• lets us make use of the latest obs
• lets us predict for relatively large areas
• lets us make probabilistic predictions
Some things we're doing

2. Downscaling at the convective scale

Rachel Prudden et al.
Downscaling at the convective scale

- Use Gaussian Random Fields
- Conditioning on spatial averages, find covariance of high-res field given low-res field
- Generate new distribution for hi-res
Downscaling at the convective scale

It works! Next we need to:

• Need measure of similarity of generated hi-res distribution to truth (surprisingly difficult)
• Extend to whole UK
• Deal with time-evolution (non-stationarity)
Some things we're doing

3. Self organising maps for resolving cloud feedbacks

Sam Adams and Mark Webb
Self organising maps

• They're a type of neural network inspired by cortical maps in brains
Self organising maps

• They're a type of neural network inspired by cortical maps in brains
Self organising maps to understand cloud-circulation couplings

• They're a type of neural network inspired by cortical maps in brains

• An **unsupervised** competitive learning strategy (as opposed to most NN which use error-correcting supervised learning – i.e. back prop with gradient descent)
Self organising maps to understand cloud-circulation couplings

• They're a type of neural network inspired by cortical maps in brains
• An unsupervised competitive learning strategy (as opposed to most NN which use error-correcting supervised learning – i.e. back prop with gradient descent)
• Essentially a dimension reduction technique
Self organising maps to understand cloud-circulation couplings

• They're a type of neural network inspired by cortical maps in brains
• An unsupervised competitive learning strategy (as opposed to most NN which use error-correcting supervised learning – i.e. back prop with gradient descent)
• Essentially a dimension reduction technique
• SOMs retain the interrelations and structure of the input dataset
Self organising maps to understand cloud-circulation couplings

• They're a type of neural network inspired by cortical maps in brains
• An unsupervised competitive learning strategy (as opposed to most NN which use error-correcting supervised learning – i.e. back prop with gradient descent)
• Essentially a dimension reduction technique
• SOMs retain the interrelations and structure of the input dataset
• SOMs uncover correlations that wouldn't be otherwise easily identifiable.
Self-organising maps to understand cloud-circulation couplings

The basic SOM has three components:

- An **input** layer with an input for each variable
- An **output** layer which forms the actual map
- **Weights** which connect them

Training process:
1. Present an input vector to the input layer.
2. Select the output node with weights best matching the input vector.
3. Identify a spatial neighbourhood around the winning output node.
4. Adjust weights for local nodes toward the input using a learning rate.
5. Decrease the neighbourhood size and learning rate.
6. Repeat 1…5 until convergence.
Self organising maps to understand cloud-circulation couplings

- ISCCP FD and ERA Interim
- 27 input variables
- SOM weights indicate relationship between variables
- Qualitative analysis - plot the weights for each variable (‘component planes’)
Self organising maps to understand cloud-circulation couplings
Self-organising maps to understand cloud-circulation couplings

Substantial relationships that are not just plain Pearson's r
Self-organising maps to understand cloud-circulation couplings

SOM relationships that are not in the model

- **Low cloud vs LTS/EIS**
  - LTS - Lower Tropospheric Stability
  - EIS – Estimated Inversion Strength

- **ω700 vs LTS/EIS**
  - W700 – 700 hPa vertical pressure velocity (subsidence)

- **LH vs SH flux**
  - LH Flux – Surface Latent Heat Flux
  - SH Flux – Surface Sensible Heat Flux

- **10m wind vs air-sea ΔT & SH flux**

- **q1000 vs rh1000**
  - q1000 – 1000 hPa specific humidity
  - rh1000 – 1000 hPa relative humidity
Concluding remarks
Concluding remarks - what's holding us back?

"An algorithm that gets better at doing something as it experiences more data"...me
Concluding remarks - what's holding us back?

• We need way better access to data
Concluding remarks - what's holding us back?

• We need way better access to data
• We need better ways to write and read analyses so we can make hierarchies of information
Concluding remarks - what's holding us back?

• We need way better access to data
• We need better ways to write and read analyses so we can make hierarchies of information
• A lot of others have had success with supervised learning. START COLLECTING DATA NOW! INSTRUMENT EVERYTHING!
Concluding remarks - what's holding us back?

• We need way better access to data
• We need better ways to write and read analyses so we can make hierarchies of information
• A lot of others have had success with **supervised** learning. **START COLLECTING DATA NOW! INSTRUMENT EVERYTHING!**
• The (first order) problem IS NOT access to expertise - it's just posh stats
If I asked people what they wanted they would have said faster horses.

NOT Henry Ford!...apparently