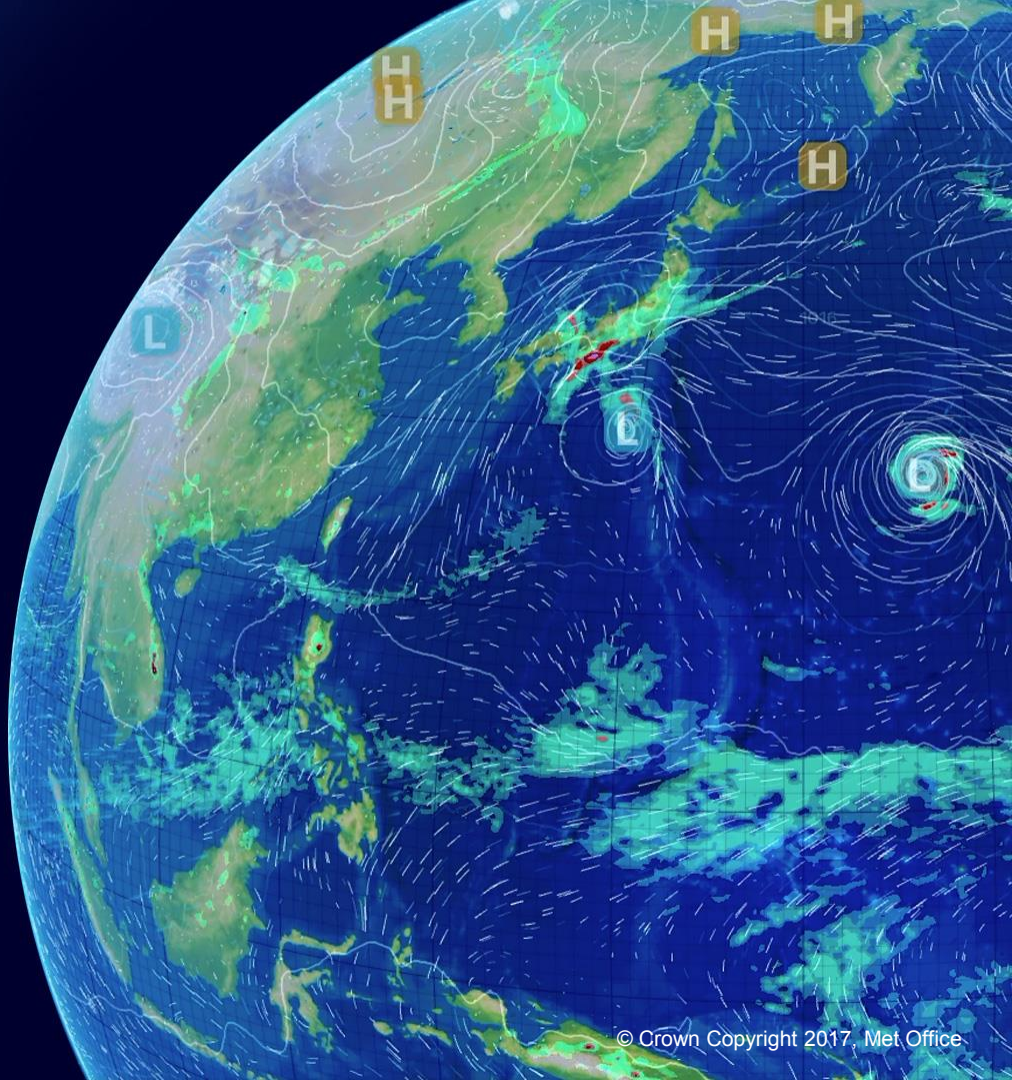
Met Office

Review of machine learning activities in WGNE centres

Keith Williams

WGNE34, 25/09/19



Responses

Responses received from 7 centres of which 5 have active work using machine learning for model development. Only one gave a response related to model evaluation.

In terms of model development, can be mostly split into:

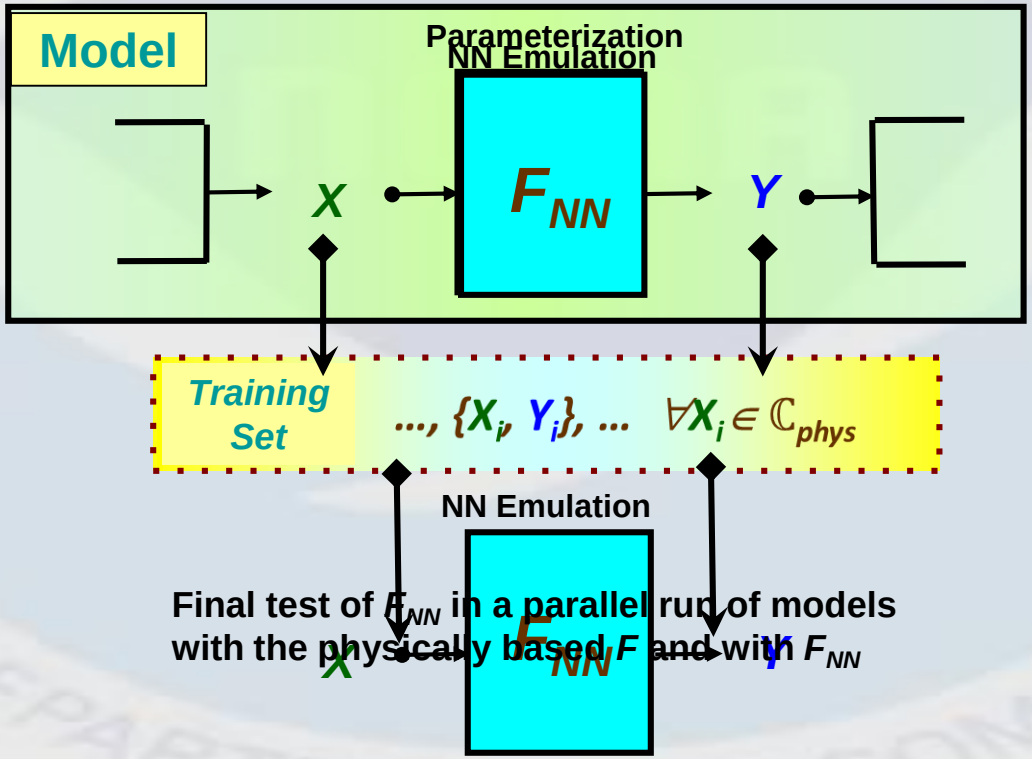
1) Individual parametrization emulation (although some efforts to emulate the full model are in progress).
2) Optimization of model parameters.

Thank you to all who responded.

Parametrization emulation

Fast ML Emulations of Model Physics Parameterizations

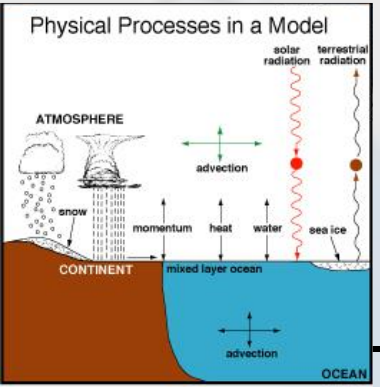
Learning from Data



Final test of F_{NN} in a parallel run of models with the physically based F and with F_{NN}

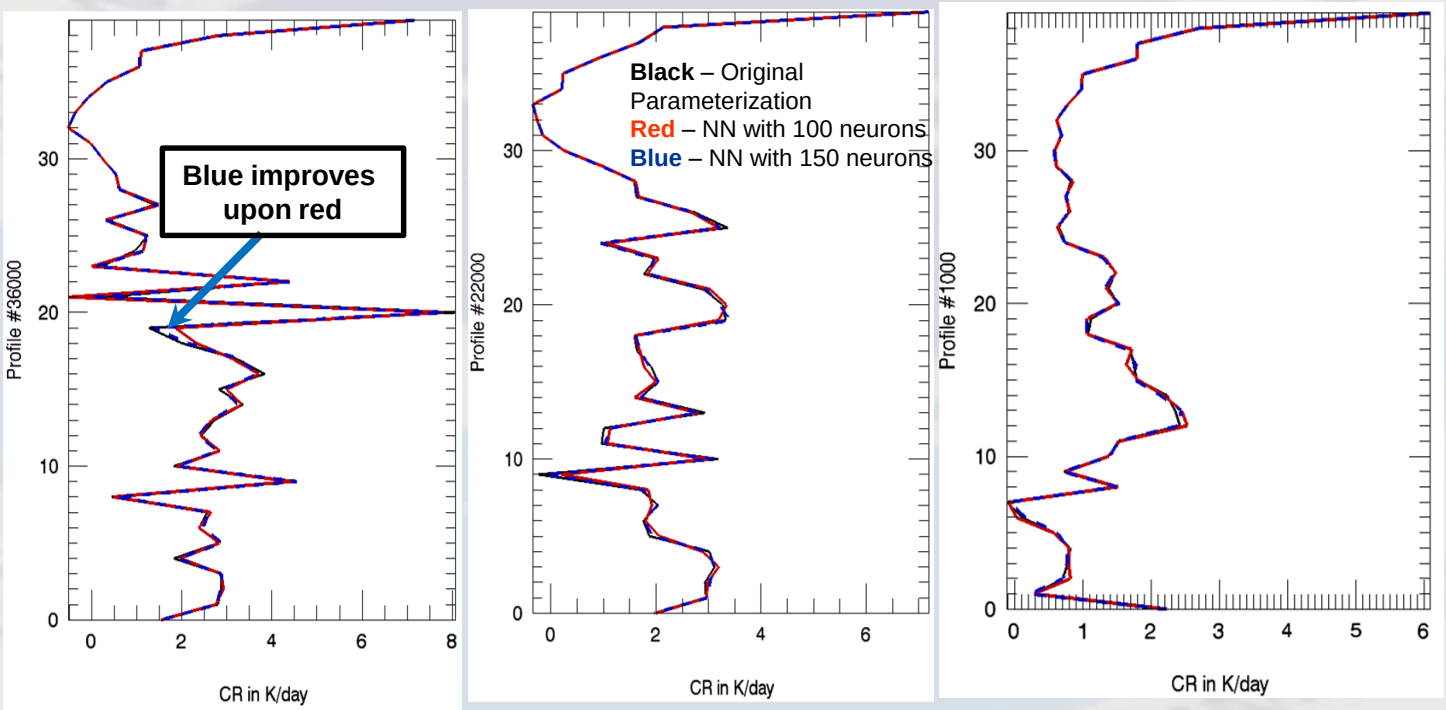
ML for Numerical Model

- **ML Applications developed & under development (red)**
  - *Fast and accurate ML emulations of model physics*
    - Fast NN nonlinear wave-wave interaction for WaveWatch model
      - Tolman, et al.(2005). Neural network approximations for nonlinear interactions in wind wave spectra: direct mapping for wind seas in deep water. *Ocean Modelling*, 8, 253-278
    - Fast NN long and short wave radiation for NCEP CFS, and GFS models and for FV3GFS
      - V. M. Krasnopolsky, M. S. Fox-Rabinovitz, Y. T. Hou, S. J. Lord, and A. A. Belochitski, 2010: "Accurate and Fast Neural Network Emulations of Model Radiation for the NCEP Coupled Climate Forecast System: Climate Simulations and Seasonal Predictions", *Monthly Weather Review*, 138, 1822-1842, doi: 10.1175/2009MWR3149.1
    - Fast NN emulation of super-parameterization (CRM in MMF)
      - Rasp, S., M. S. Pritchard, and P. Gentine, 2018: Deep learning to represent subgrid processes in climate models. *Proceed. National Academy Sci*., 115 (39), 9684–9689, doi:10.1073/pnas.1810286115
    - Fast NN microphysics for FV3GFS and WRF
  - *New ML parameterization*
    - NN convection parameterization for GCM learned by NN from CRM simulated data
      - Brenowitz, N. D., and C. S. Bretherton, 2018: Prognostic validation of a neural network unified physics parameterization. *Geophys. Res. Lett.*, 35 (12), 6289–6298, doi:10.1029/2018GL078510.
  - *ML emulation of simplified GCM*
    - Scher, S., 2018: Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. *Geophys. Res. Lett*., 45 (22), 12,616–12,622, doi:10.1029/2018GL080704.



Individual LWR Heating Rates Profiles

Profile complexity



PRMSE = 0.18 & 0.10 K/day **PRMSE = 0.11 & 0.06 K/day** **PRMSE = 0.05 & 0.04 K/day**

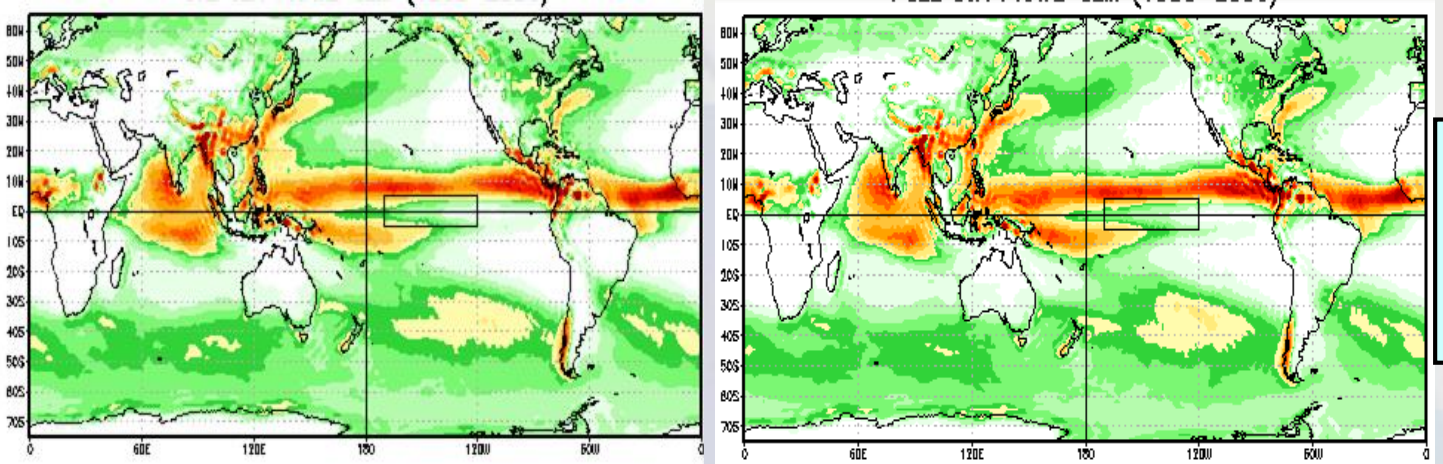
Approximation Statistics and Speedup

Note: **Work in progress to extend radiation emulation for FV3GFS**		**NCEP CFS/GFS (L = 64)**	
		RRTMG LWR	**RRTMG SWR**
Statistics for Differences in Kelvin/day	**Bias**	$2.\cdot 10^{-3}$	$5.\cdot 10^{-3}$
	RMSD	0.49	0.2
Speedup factor, η	***Times***	**Averaged speedup factor: 16** **Speedup factor in cloudy conditions: 20**	**Averaged speedup factor: 60** **Speedup factor in cloudy conditions: 88**

V. M. Krasnopolsky, M. S. Fox-Rabinovitz, Y. T. Hou, S. J. Lord, and A. A. Belochitski, 2010: "Accurate and Fast Neural Network Emulations of Model Radiation for the NCEP Coupled Climate Forecast System: Climate Simulations and Seasonal Predictions", Monthly Weather Review, 138, 1822-1842, doi: 10.1175/2009MWR3149.1

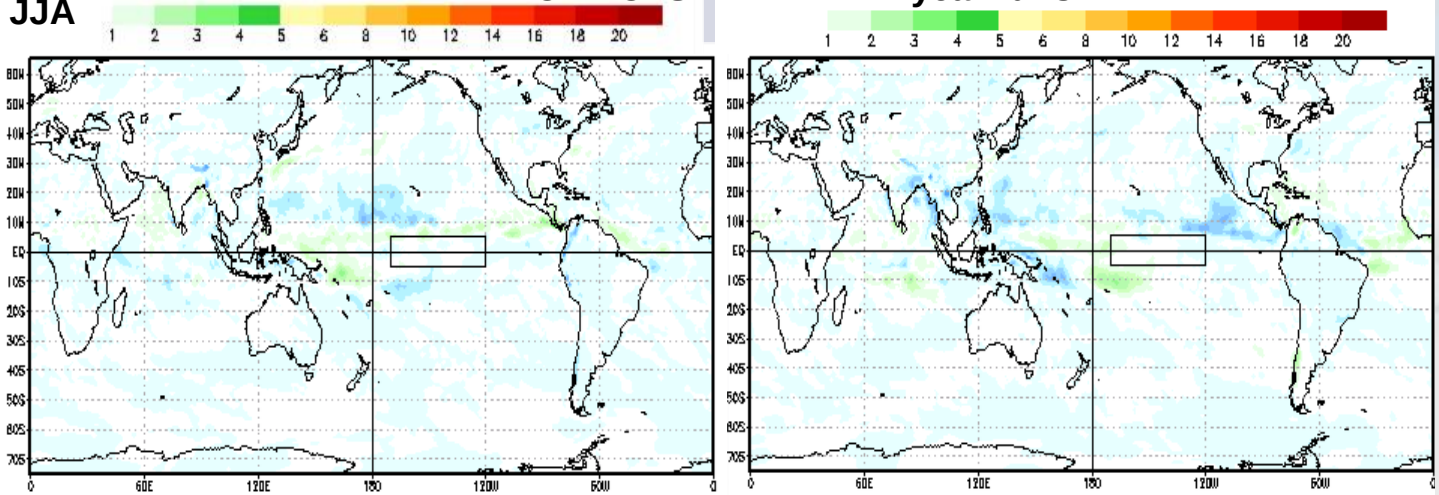
CTL run with RRTMG LW and SW radiations

NN run with NN LW and SW radiations



NCEP CFS PRATE – 17 year runs

JJA



NN – CTL run differences

Differences between two control runs with different versions of FORTRAN compiler

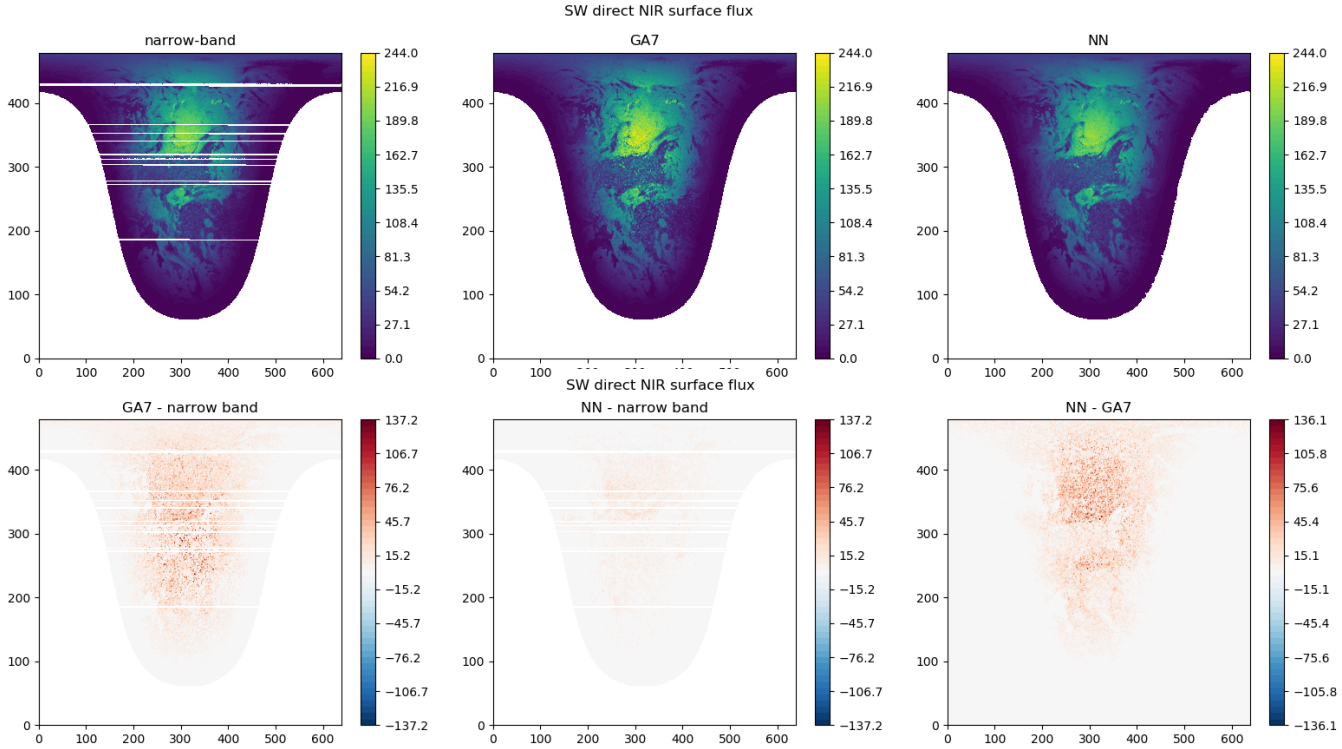
June 13, 2019

SW direct surface fluxes

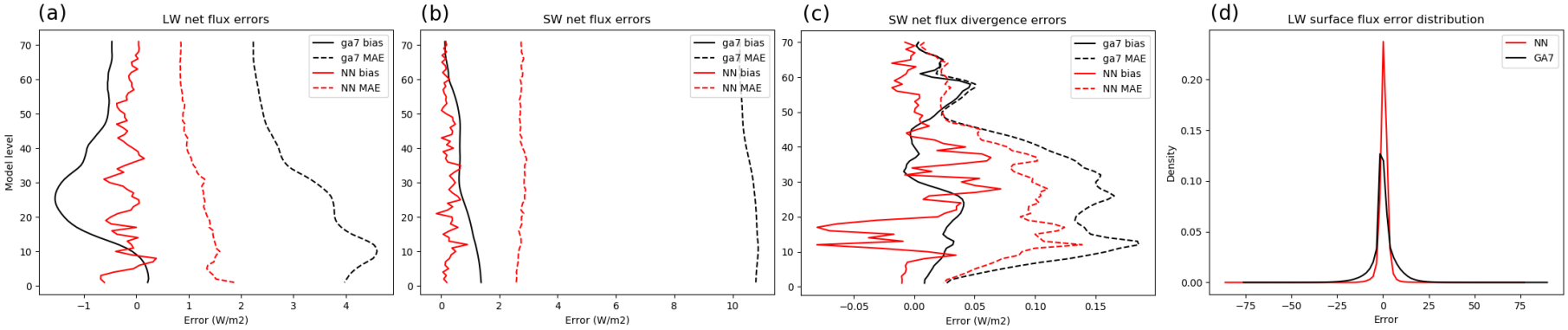
Example of SW direct NIR (near infra-red) surface fluxes

Missing data is due to difficulties running the NB configuration of SOCRATES

NN output is more accurate and less noisy than GA7 configuration compared to NB output



Offline verification



Errors w.r.t narrow band SOCTARES output

a-c: profiles of mean error (bias) and mean absolute error (MAE) for net flux and net flux divergence

d: error distribution (PDF) for LW net surface fluxes

Met Office

The hard part: integrating with the UM

Speed:

Is it fast (enough)?

	NB	GA7	NN
Spice (1 x CPU, Xeon E5-2690 v3 @ 2.60GHz)	33.75	0.75	4.16
GW4-Isambard Power 9 (1 x CPU @ 3.8GHz)	-	-	2.86
GW4-Isambard Power 9 (1 x GPU NVIDIA V100)	-	-	0.12

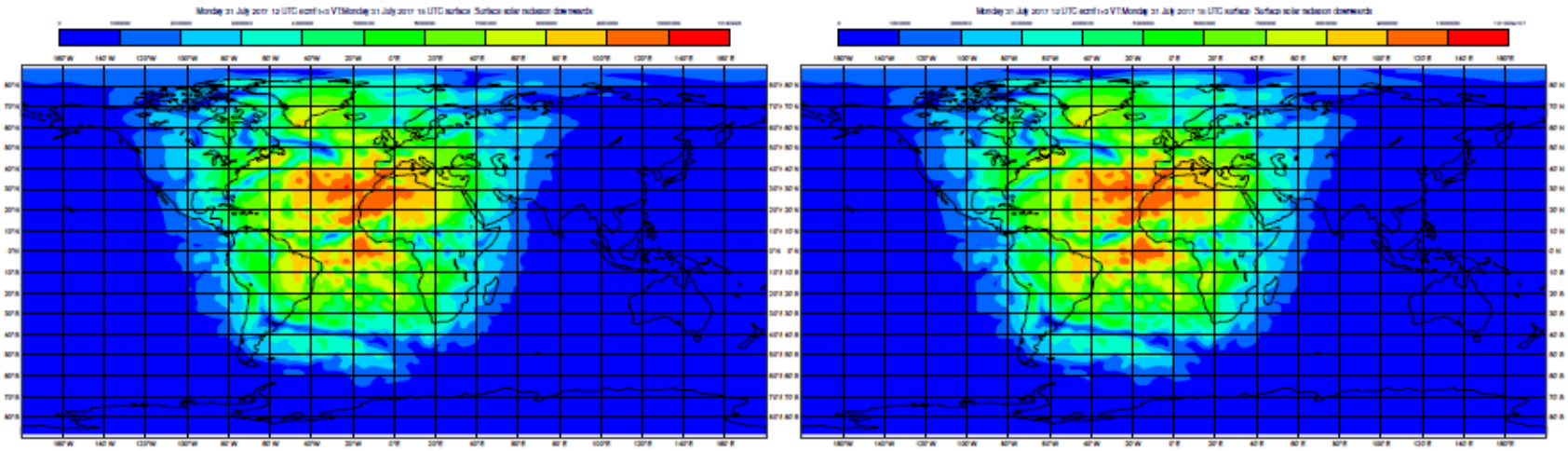
Approximate per-column execution times (ms) on different hardware

Embedding:

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

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

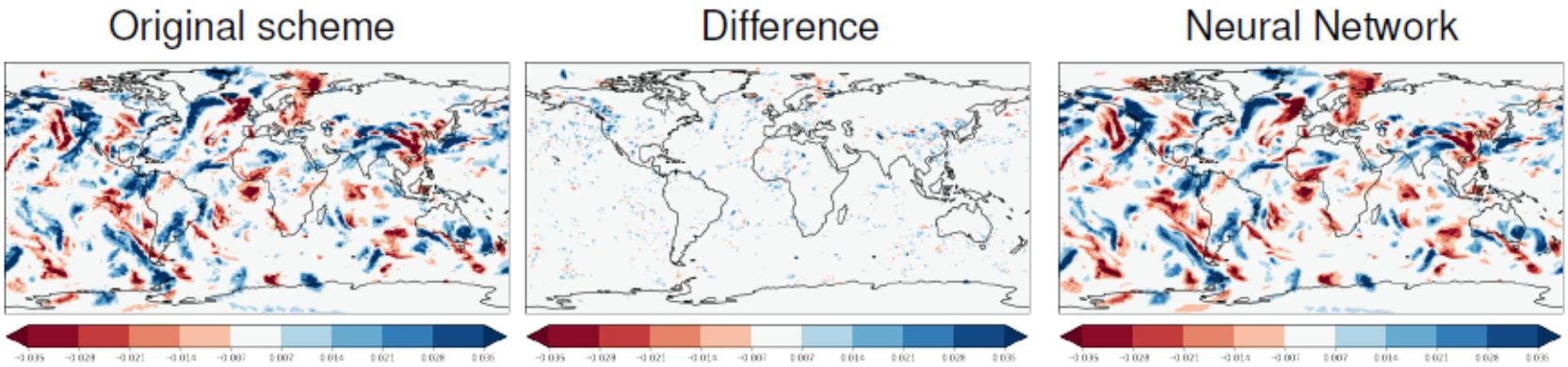


Dueben, Hogan, Bauer @ECMWF and Progsch, Angerer @NVIDIA

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 10 speed-up for radiation.

A neural network emulator for gravity wave drag

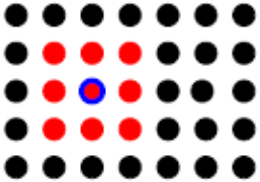


Chantry, Dueben, Palmer.

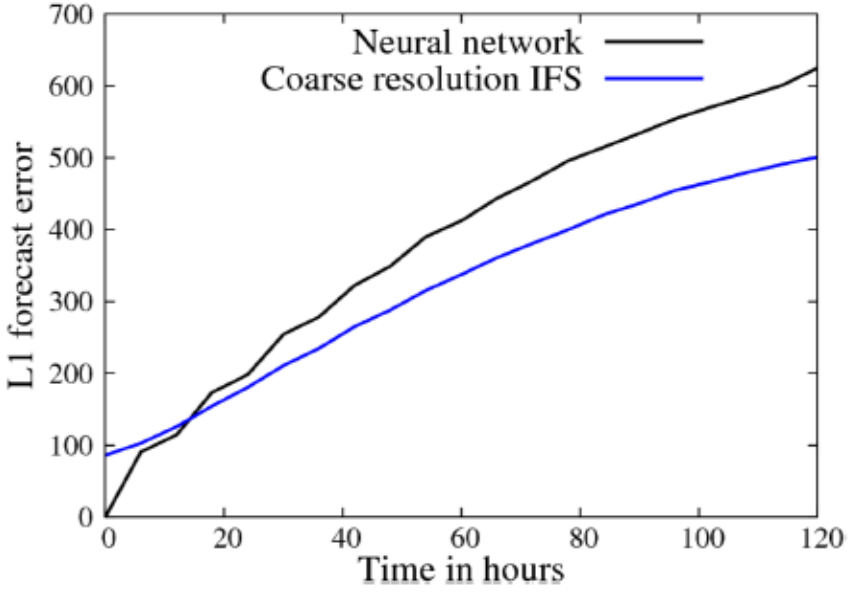
Tendency output for the non-orographic gravity wave drag parametrisation scheme for the standard scheme and a neural network emulator.

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.



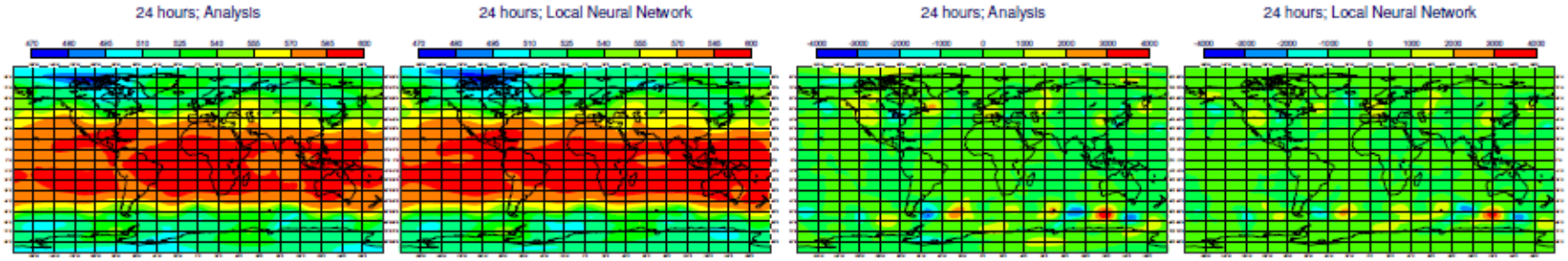
Global weather forecast based on Neural Networks



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



The simulations show reasonable dynamics.

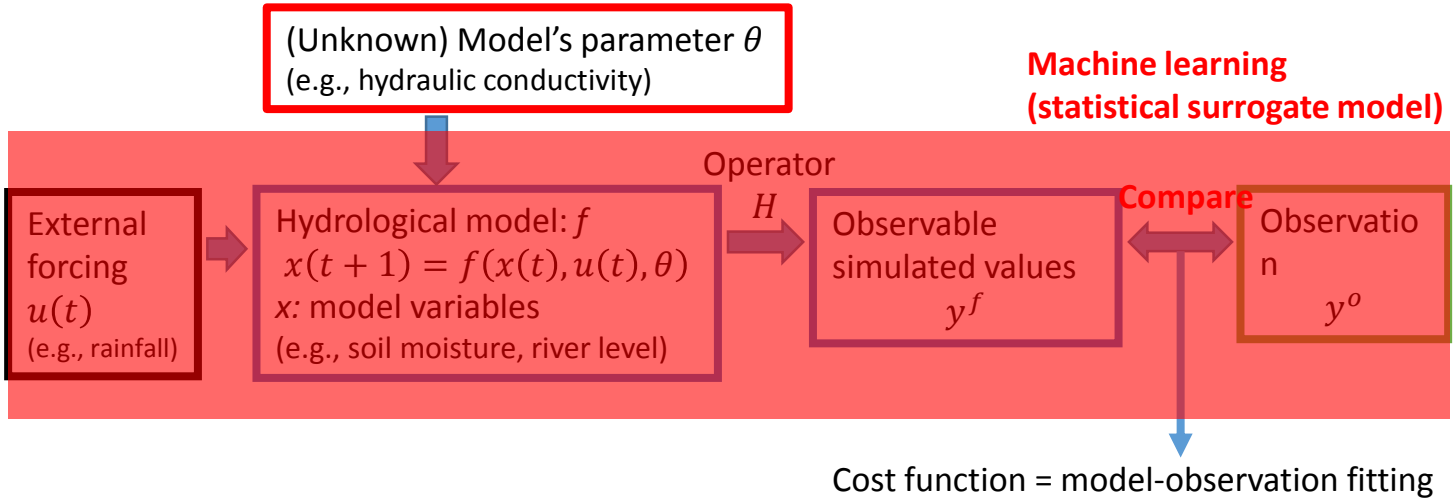
Just adding further inputs does not necessarily help.

Model runs crash after a couple of weeks.

Dueben and Bauer GMD 2018

OPTIMIZATION OF MODEL PARAMETERS

Data-driven approach to speed up Markov Chain Monte Carlo (MCMC)



Algorithm:
MCMC based on a surrogate model

1. Run a hydrological model with 400 random ensemble of parameters in parallel to generate $\theta - cost$ dataset.
2. Learning $\theta - cost$ dataset by Gaussian Process and construct a statistical surrogate model $g(\theta) = cost$
   Note: $g(\theta)$ is much faster than a full hydrological model.
3. Perform MCMC on $g(\theta)$ (10^6 times iteration)

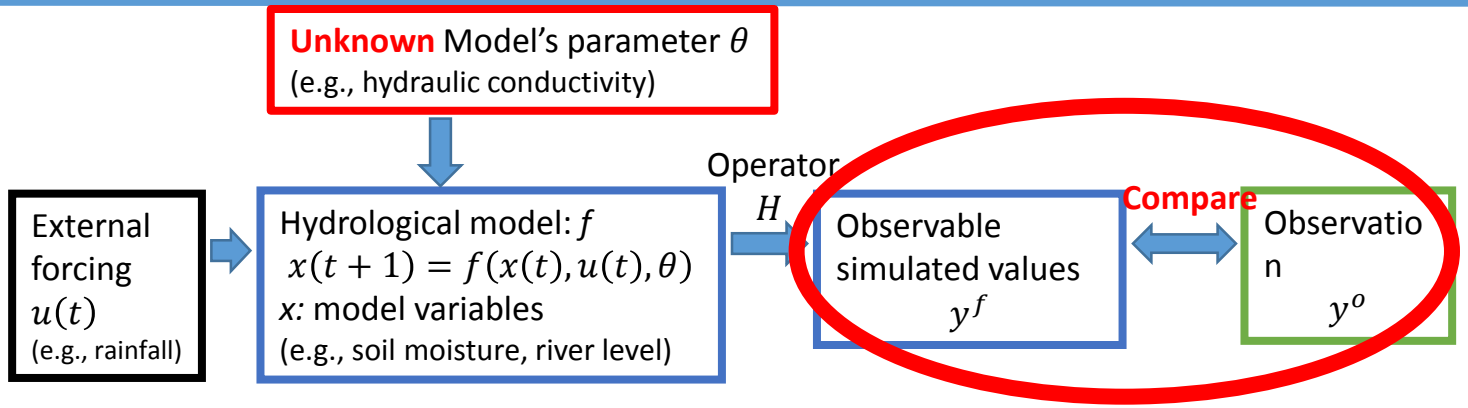
Computational time ~ **10 minutes**

× 400000 faster!!

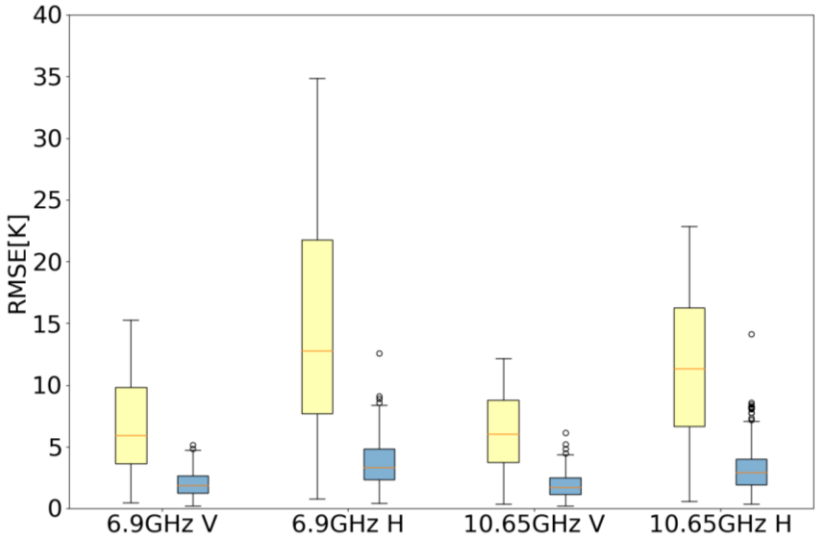
If we run full hydrological model for MCMC, 4 minutes for a single integration
→ $\mathbf{4 \times 10^6}$ **minutes**

Courtesy of Y. Sawada (Univ. Tokyo/MRI JMA)

Skill to simulate observed brightness temperature



Random 400 draws VS. MCMC 400 draws



→ MCMC improves model-observation fitting

Courtesy of Y. Sawada (Univ. Tokyo/MRI JMA)

Ensemble Parameter Estimation

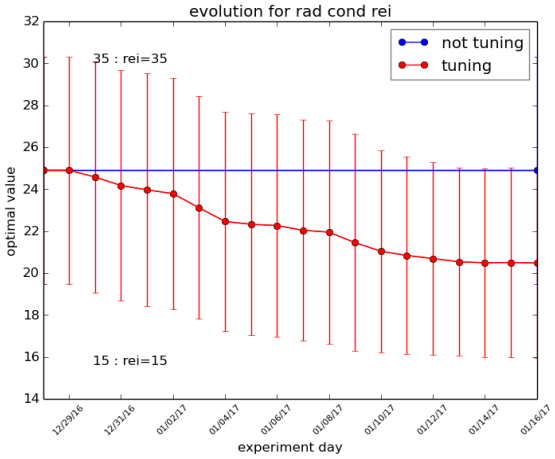
Lead: Pieter Houtekamer (ECCC)

- The goal of this project is to identify optimal parameter values and physics suite configurations in the EnKF assimilation system
- The Ensemble Prediction and Parameter Estimation system (Laine et al. 2012; Jarvinen et al 2012) is adopted:
  1) Begin with a 256 different configurations (each member)
  2) Compute CRPS (upper air) and Brier Scores (precipitation) of background forecasts within the DA system
  3) Successful member remain and poor performers are replaced

Genetic Algorithm

- An evolutionary model derived from natural sciences (Goldberg and Holland 1988).
- Rejected members are replaced with re-sampling of successful members with perturbations to maintain diversity

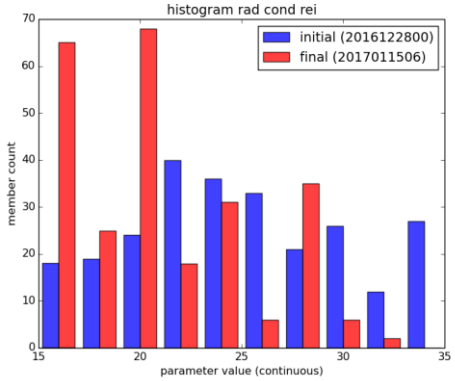
Ensemble Parameter Estimation



The effective radius of ice particles (r_{ei}) provides a good example of the power of the genetic algorithm.

Different constants (15-35 μm range) are used across members with an initial mean of 25 μm, which evolves towards the lower end of the range.

This is consistent with the 15 μm value in the determinstic model.



The right-skewed final distribution suggests that unphysically small <15 μm values might be benificial, perhaps an indication of r_{ei} acting as bias compensation in the model.

Evolution of the r_{ei} distribution (red with +/- 1 standard deviation) over 20 days of integration of the EnKF system (top). The initial mean value is shown with a blue line. The initial (blue) and final (red) distributions of the r_{ei} parameter is shown at the bottom.

Composite Profile Analysis

Lead: Ron McTaggart-Cowan (ECCC)

- The goal of this project is to reduce random model error (error standard deviation) in the boundary layer
- Much of random error arises from systematic model biases that occur under specific conditions:

1) Identify dominant regional boundary layer archetypes in each season
2) Classify each forecast to one of these archetypes
3) Assess forecast error by archetype

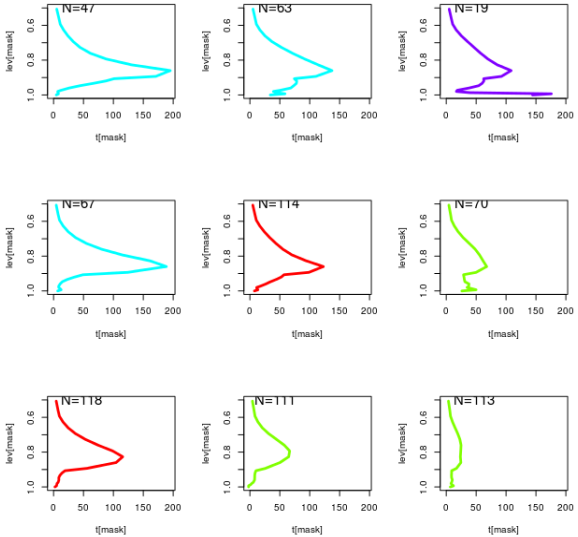
Self Organizing Maps

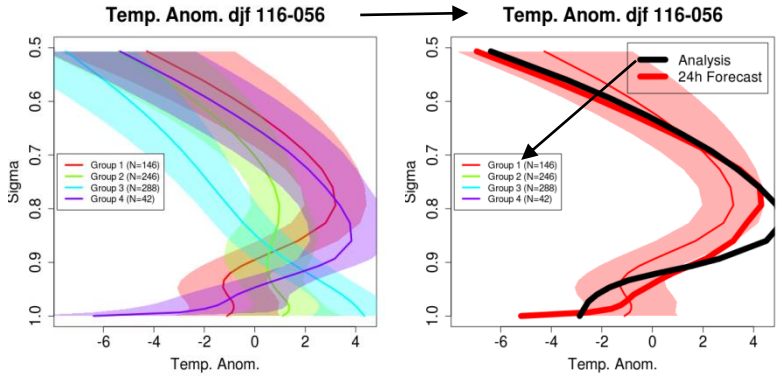
- Unsupervised learning from reanalyses
- Forecast classification

Hierarchical Cluster Analysis

- Synthesis of SOM codes (Herrero and Dopazo 2002)

Composite Profile Analysis

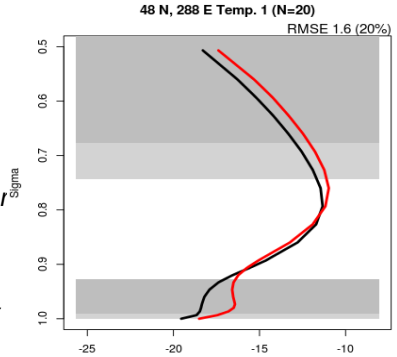




Archetype temperature anomaly profiles for Montreal winter (left) as determined by SOM/HCA (top left). A single 24 h forecast (heavy red) and corresponding analysis (heavy black) are shown against the analysis Group 1 classification (light red with shading) in the top-right panel.

Above: Sample 3x3 map of SOM nodes with corresponding codes plotted as condition profiles based on temperature anomaly profiles in the lower atmosphere over a 2.5º square region for Montreal winter (the full map uses 7x7 nodes).

Right: The mean Group-1 temperature profile over Montreal (black), mean 24-h and 48-h forecast profiles for Group-1 analyses (red), and contribution to total winter RMSE (20% of total). Light and dark grey-shaded regions are significant at 95% and 99%, respectively.



Systematic mixing-related errors that have a large contribution to winter RMSE appear when the profile is conditioned on the ML-derived archetype: **We can fix this!**

Other comments

DWD - Machine learning activities are currently focused on postprocessing, and consideration for data assimilation (for data quality control and related aspects).

NOAA & ECMWF – Also using ML for initialisation and post processing.

Japan - Ito (Univ. of Tokyo) and Mouri (MRI/JMA) has developed by deep learning a surface layer model for wind speeds with short time scales (seconds, when Monin-Obukhov similarity not valid).

Conclusions

Lots of activity going on in some centres.

In terms of model development, can be mostly split into:

1) Parametrization emulation (although some efforts to emulate the full model are in progress)
   - Radiation has been a prime candidate but should it be (we know the right answer and our approximation is already well optimised for cost vs complexity. NN are easy to optimise for GPU, but radiation could also be re-coded for GPU)
   - Putting the trained NN back into the full model seems to be something that many find challenging.
2) Optimization of model parameters.