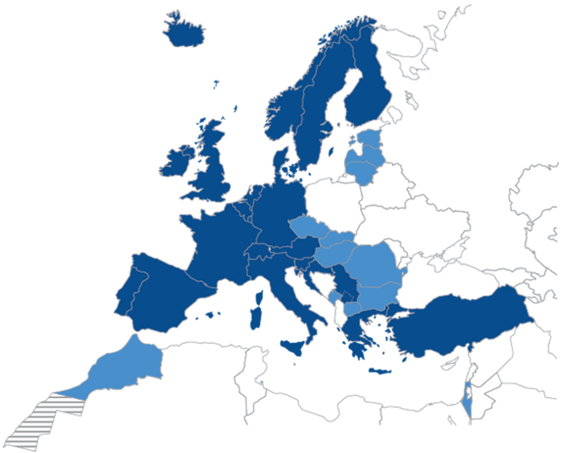
How to Prepare Weather and Climate Models for Future HPC Hardware

Peter Düben

European Weather Centre (ECMWF)

The European Weather Centre (ECMWF)



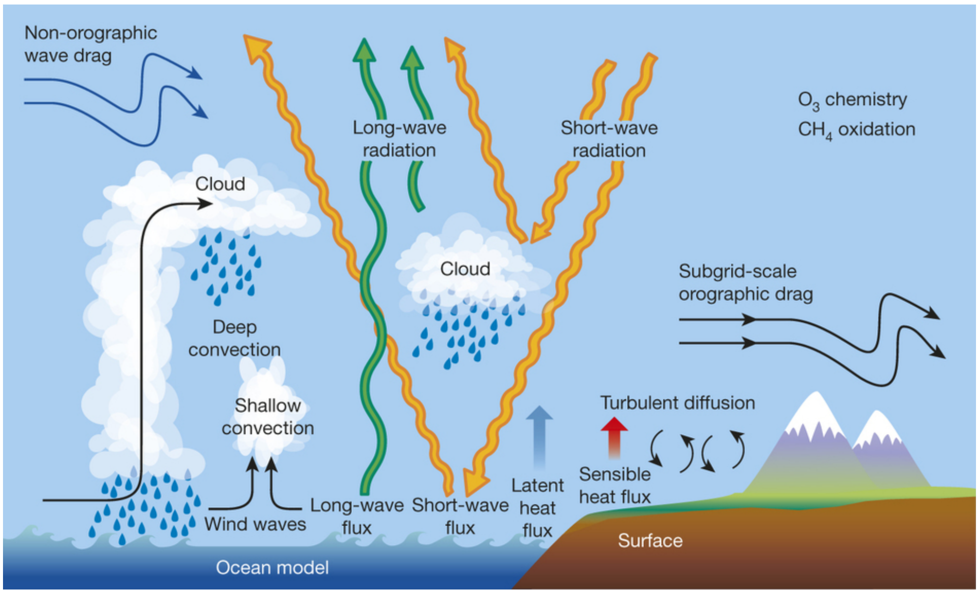
www.ecmwf.int

- Independent, intergovernmental organisation supported by 34 states.
- Research institute and 24/7 operational weather service.
- Weather forecasts cover time frames from medium-range, to monthly and seasonal.
- Based in the UK, $\approx$ 350 member of staff from 30 different countries.

Predicting weather and climate: Why is it so hard?

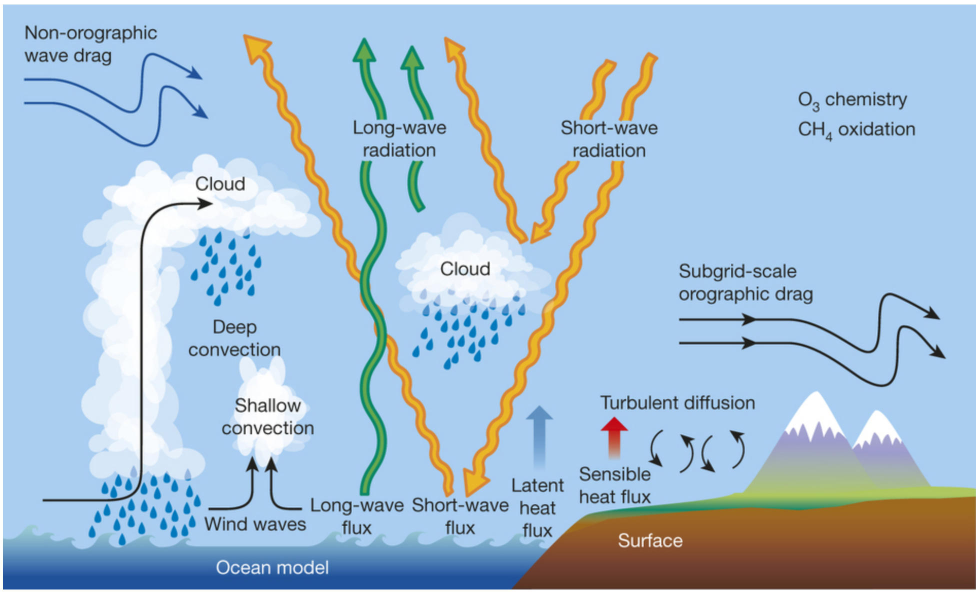
Earth seen from Apollo 17 (NASA 1972)

Predicting weather and climate: Why is it so hard?



Bauer et al. Nature 2015

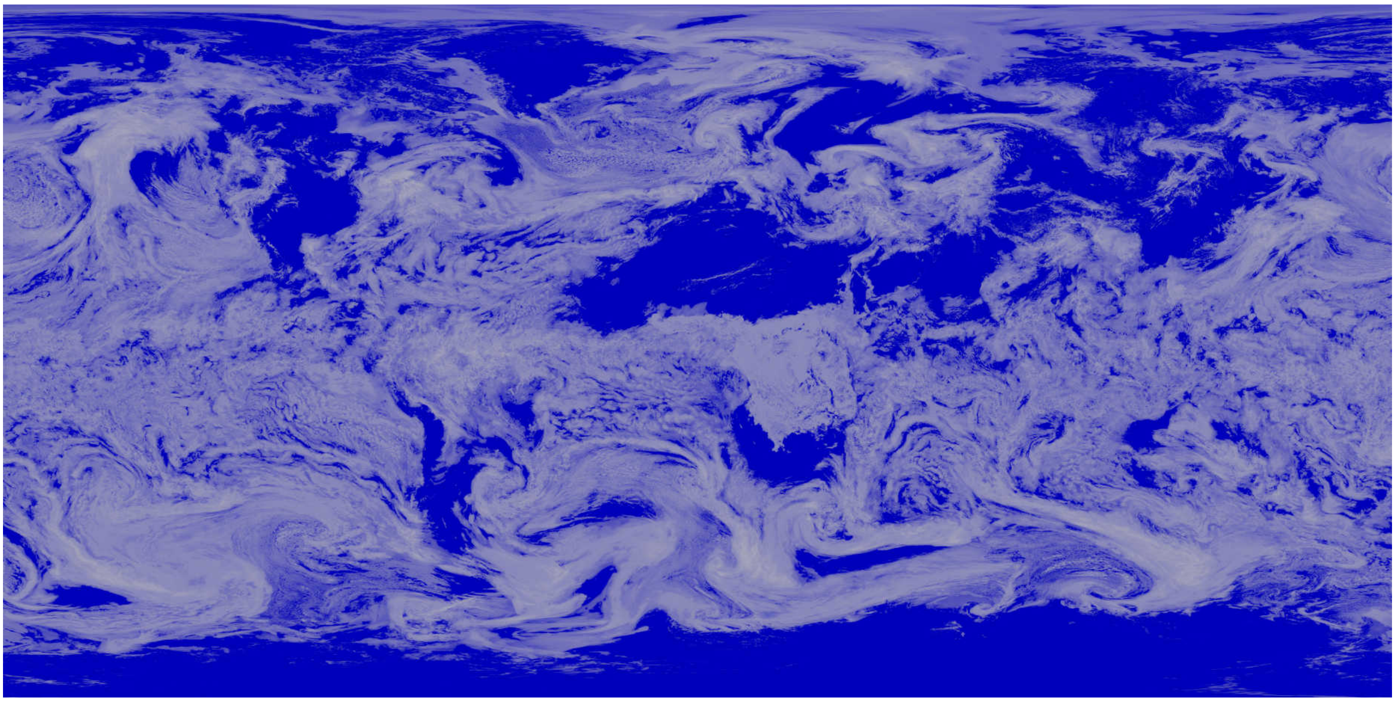
Predicting weather and climate: Why is it so hard?



Bauer et al. Nature 2015

The Earth System is complex, chaotic and huge, 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)

High Performance Computing in Earth System Modelling

Weather and climate models are high performance computing applications.

High Performance Computing in Earth System Modelling

Weather and climate models are high performance computing applications.

Forecast quality depends on resolution and model complexity.

High Performance Computing in Earth System Modelling

Weather and climate models are high performance computing applications.

Forecast quality depends on resolution and model complexity.

Resolution depends on the performance of state-of-the-art supercomputers.

High Performance Computing in Earth System Modelling

Weather and climate models are high performance computing applications.

Forecast quality depends on resolution and model complexity.

Resolution depends on the performance of state-of-the-art supercomputers.

- Individual processors will not be faster.
  $\rightarrow$ Parallelisation ($> 10^6$ parallel processing units).
- Parallelisation and performance will be essential for future model development.
- We fail to operate close to peak performance.
- Power consumption will be a big problem.

High Performance Computing in Earth System Modelling

Weather and climate models are high performance computing applications.

Forecast quality depends on resolution and model complexity.

Resolution depends on the performance of state-of-the-art supercomputers.

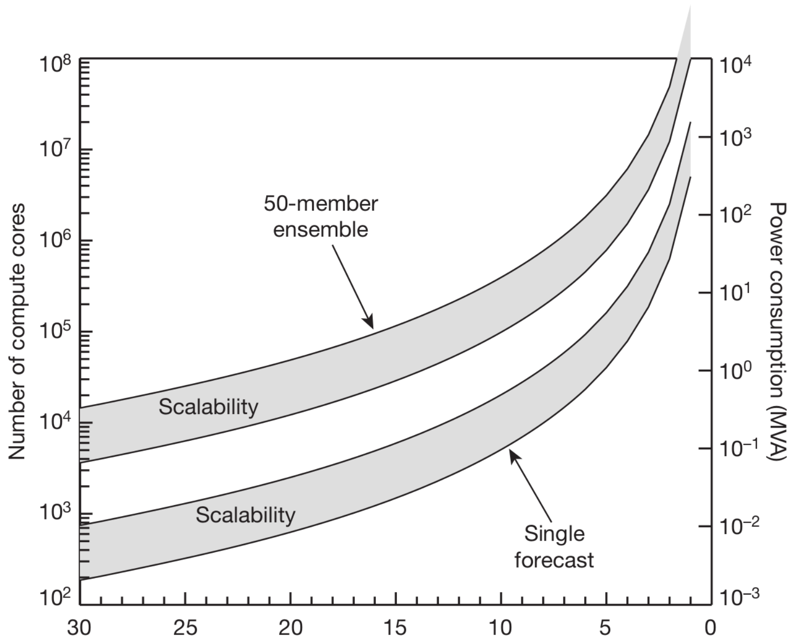
- Individual processors will not be faster.
  $\rightarrow$ Parallelisation ($> 10^6$ parallel processing units).
- Parallelisation and performance will be essential for future model development.
- We fail to operate close to peak performance.
- Power consumption will be a big problem.

The free lunch is over.

ECMWF's scalability project towards exascale supercomputing

Challenges for HPC in Earth System modelling:

- Huge code with $O(10^7)$ lines of code.
  $\rightarrow$ Difficult to port.
- Data intensive.
  $\rightarrow$ Difficult to reach peak performance.
- Global scale interactions and fast waves.
  $\rightarrow$ Difficult to parallelise.
- Operational deadlines.
  $\rightarrow$ Difficult to reduce power.



Bauer et al. Nature 2015

ECMWF’s scalability project towards exascale supercomputing

A community effort to takle the challenges:

- Define and encapsulate the fundamental algorithmic building blocks – ’Weather & Climate Dwarfs’ – to port to accelerators and to allow co-design.
- Introduce domain specific languages.
- Develop new algorithms for use in extreme scale (elliptic solver, spatial discretisation, time stepping methods,...).

ESCAPE

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

The ESCAPE project to test GPUs and other accelerators

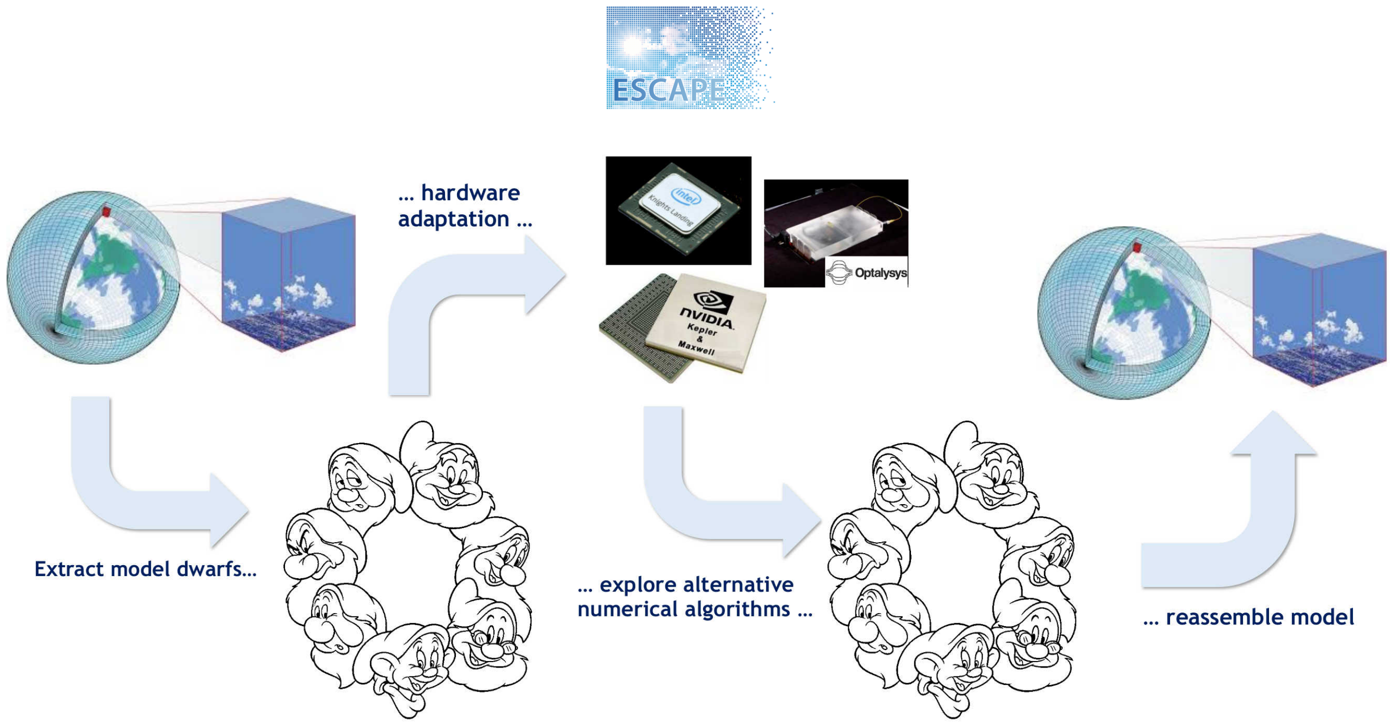


Figure courtesy Peter Bauer

The transform dwarf on GPUs

- At ECMWF we work with a spectral model that describes model fields via global basis functions.
- We need to transform fields between spectral and gridpoint space during every timestep.
- The transformations represent a significant fraction of the computing cost and the relativ cost is increasing with resolution.

The transform dwarf on GPUs

- At ECMWF we work with a spectral model that describes model fields via global basis functions.
- We need to transform fields between spectral and gridpoint space during every timestep.
- The transformations represent a significant fraction of the computing cost and the relativ cost is increasing with resolution.

Can we use GPUs to speed up the transform dwarf?

The transform dwarf on GPUs

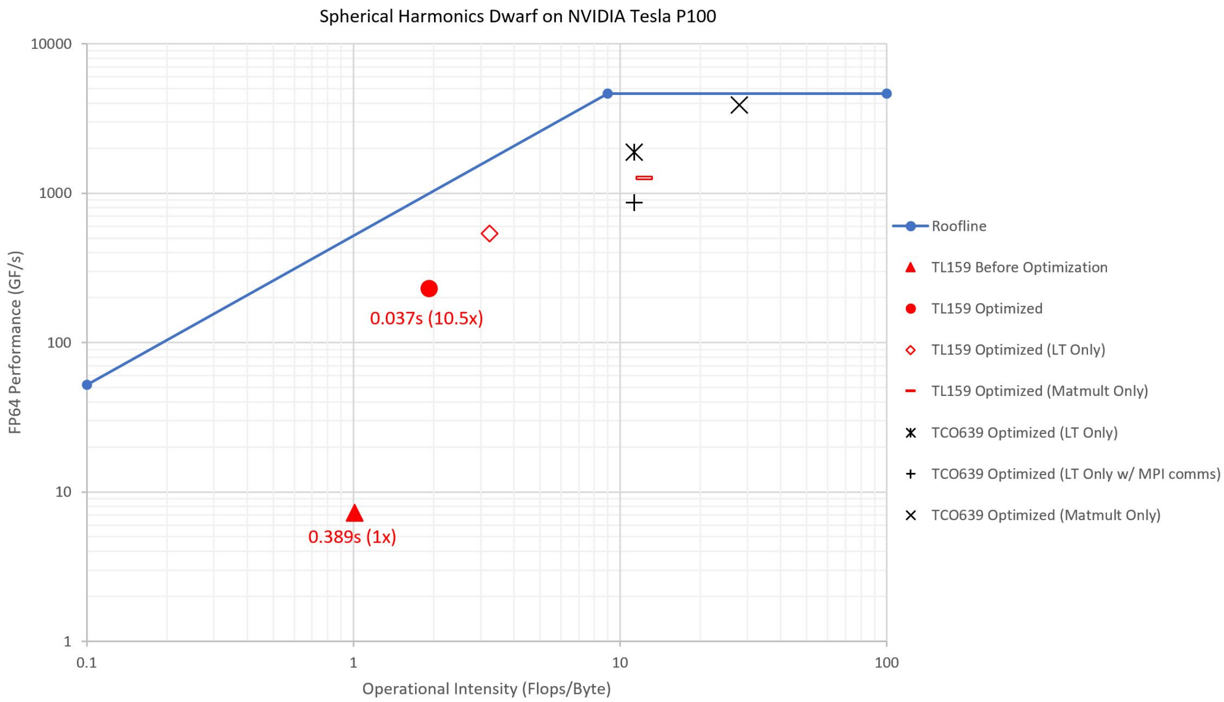


Figure courtesy Alan Gray and Peter Messmer

The transform dwarf on GPUs

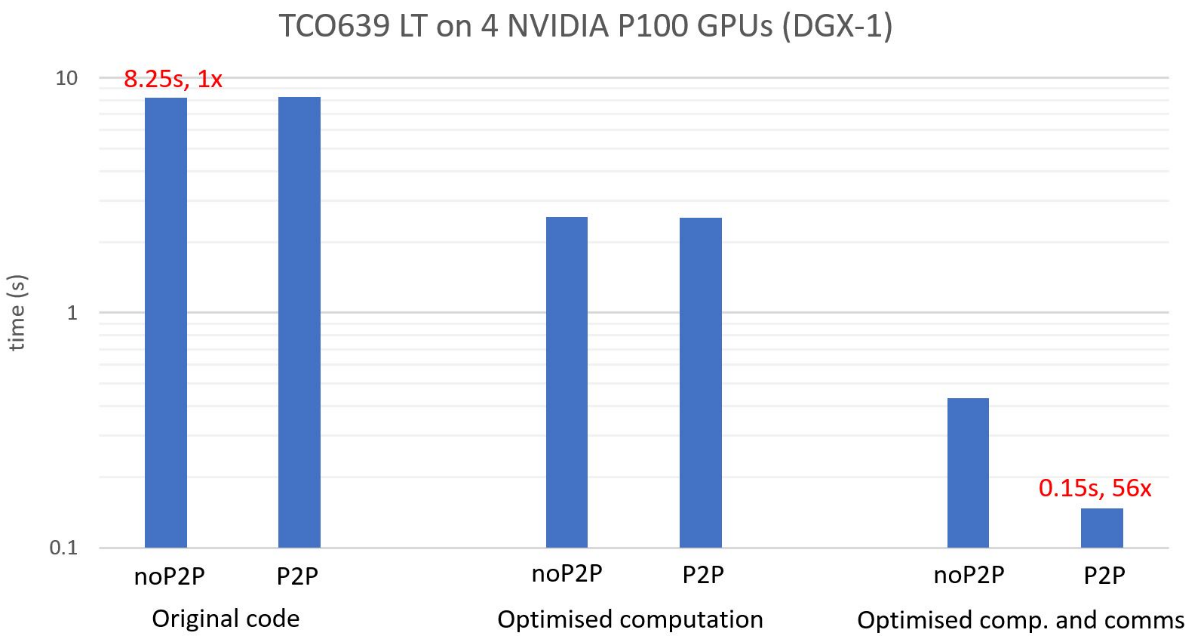


Figure courtesy Alan Gray and Peter Messmer

To speed-up weather forecasts using low numerical precision

The weather and climate community is using double precision as default since decades.

To speed-up weather forecasts using low numerical precision

The weather and climate community is using double precision as default since decades.

Reduce numerical precision

$\rightarrow$ lower power, higher performance.

$\rightarrow$ higher resolution or increased complexity.

$\rightarrow$ more accurate predictions of future weather and climate.

To speed-up weather forecasts using low numerical precision

The weather and climate community is using double precision as default since decades.

Reduce numerical precision

$\rightarrow$ lower power, higher performance.

$\rightarrow$ higher resolution or increased complexity.

$\rightarrow$ more accurate predictions of future weather and climate.

Temperature in Munich:
double precision (64 bits): 14.561192512512207°C
single precision (32 bits): 14.5611925°C
half precision (16 bits): 14.5625°C

How can we trade precision against computing cost?

- double → single → half.

How can we trade precision against computing cost?

- double $\rightarrow$ single $\rightarrow$ half.
- Reduction of precision in data storage.

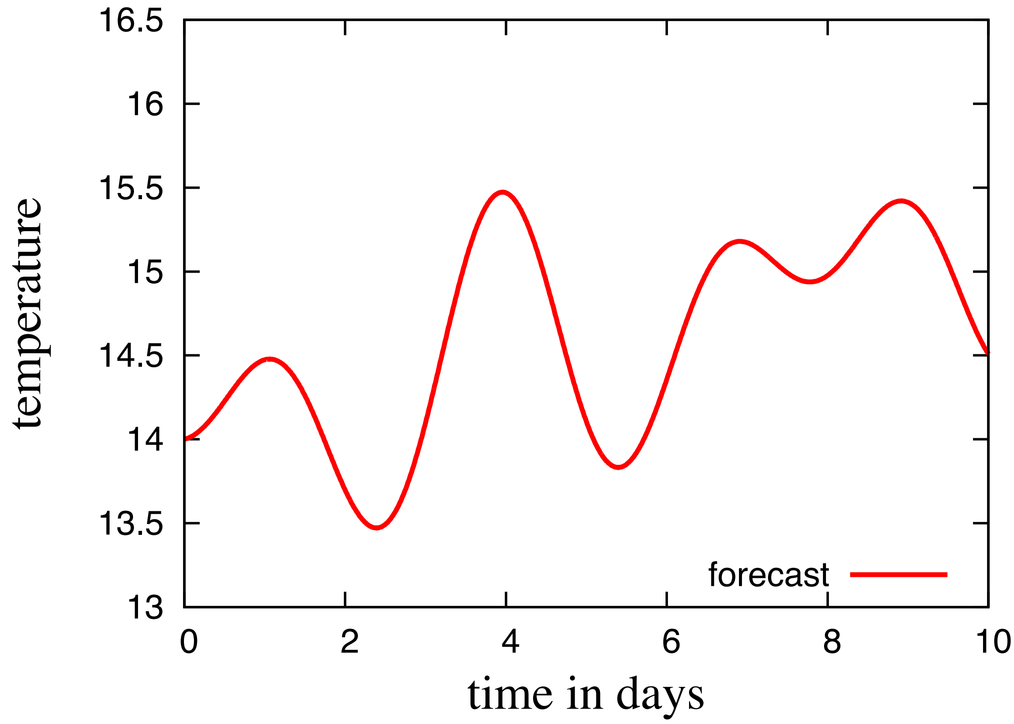
How can we trade precision against computing cost?

- double $\rightarrow$ single $\rightarrow$ half.
- Reduction of precision in data storage.
- Field Programmable Gate Arrays (FPGAs).

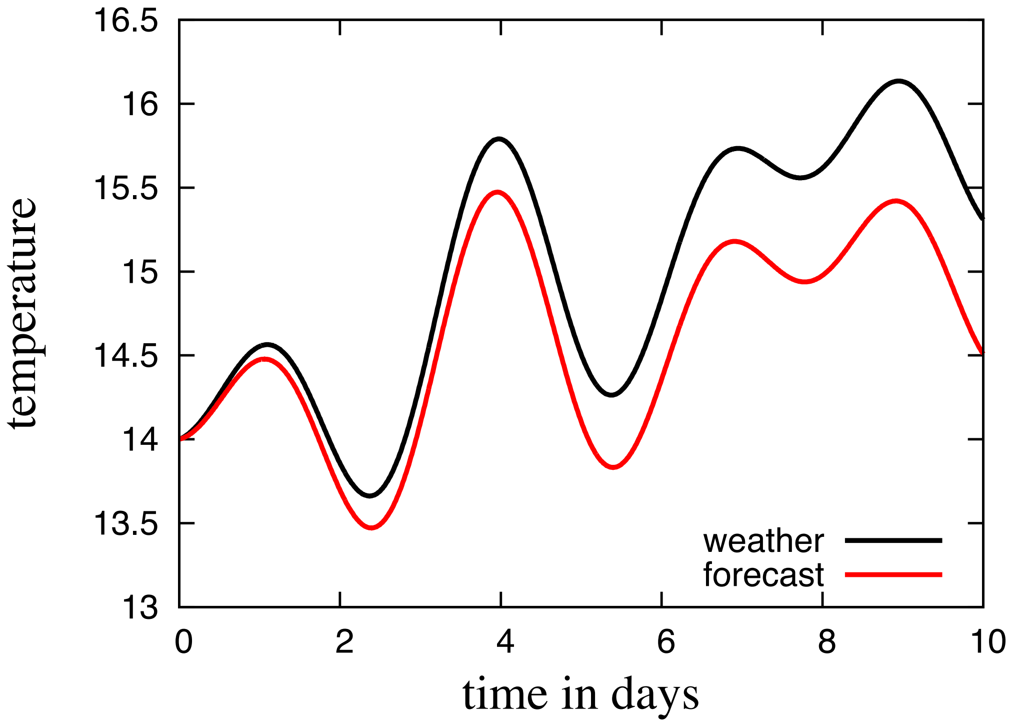
How can we trade precision against computing cost?

- double $\rightarrow$ single $\rightarrow$ half.
- Reduction of precision in data storage.
- Field Programmable Gate Arrays (FPGAs).
- Future perspective: Flexible precision hardware, probabilistic CMOS, pruned hardware, hardware with frequent hardware faults,...

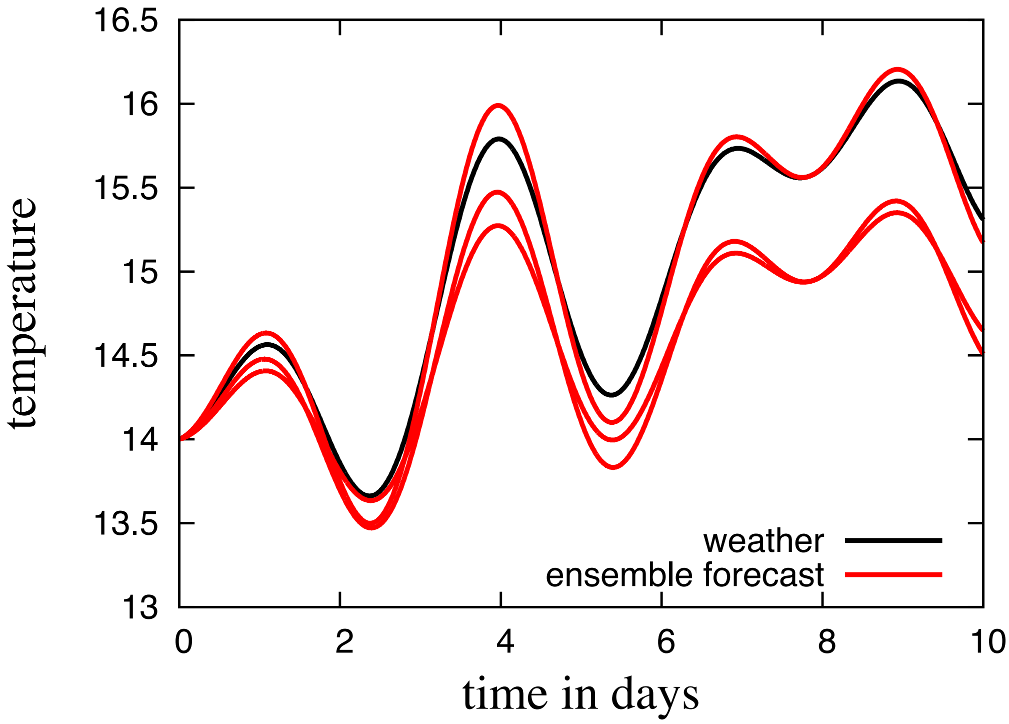
How do we treat uncertainties in weather forecasts?



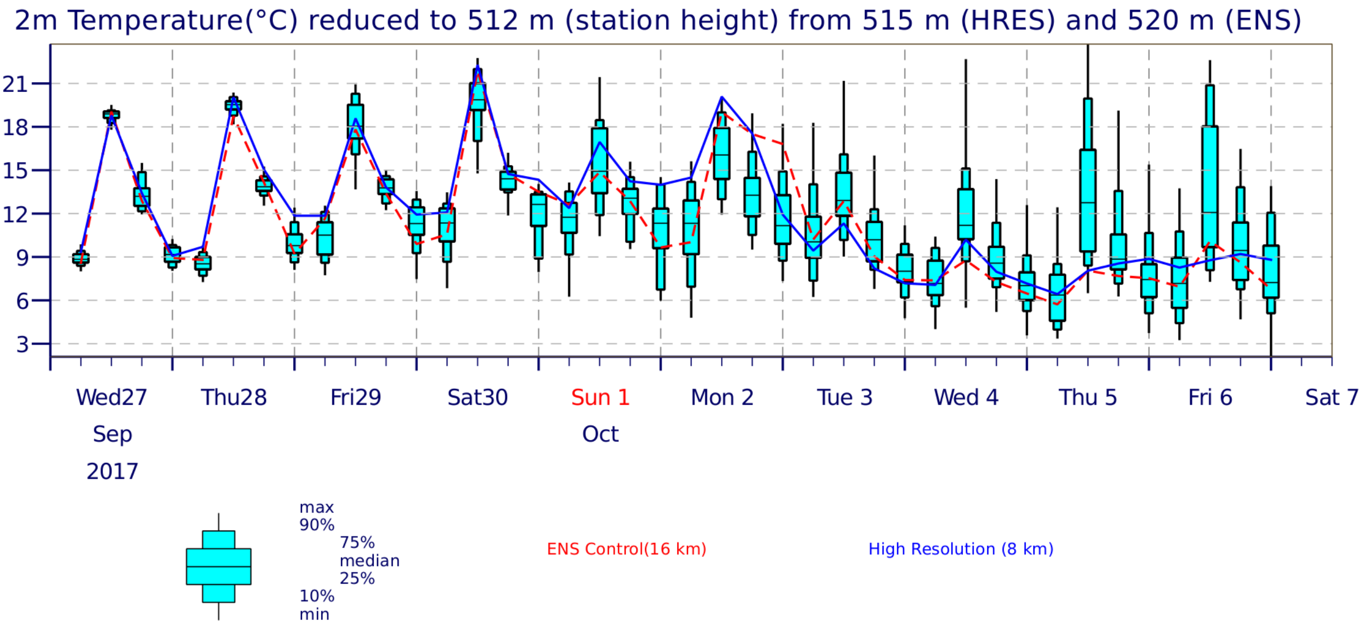
How do we treat uncertainties in weather forecasts?



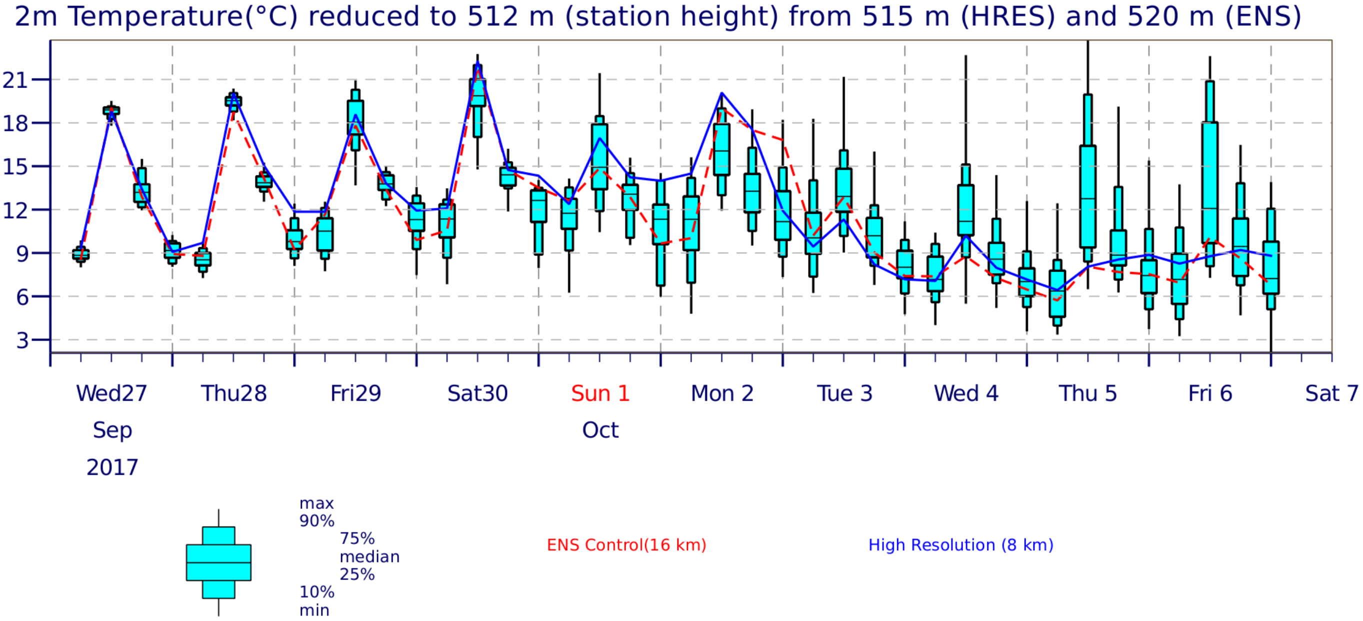
How do we treat uncertainties in weather forecasts?



How do we treat uncertainties in weather forecasts?



How do we treat uncertainties in weather forecasts?



Will a simulation with reduced precision change the ensemble spread?

Reduced precision in an atmosphere model

- We calculate weather forecasts with a spectral dynamical core (full 3D dynamics on the globe but no physics).
- Floating point precision is reduced to 8 bits in the significand using an emulator in almost the entire model.
- We estimate energy savings in cooperation with computer scientists (the groups of Krishna Palem - Rice University, Christian Enz - EPFL and John Augustine - IITM).

Resolution	Number of bits in significand	Normalised Energy Demand	Forecast error Z500 at day 2
235 km	52	1.0	2.3
315 km	52	0.47	4.5
235 km	**8**	0.29	2.5

Reduced precision in an atmosphere model

- We calculate weather forecasts with a spectral dynamical core (full 3D dynamics on the globe but no physics).
- Floating point precision is reduced to 8 bits in the significand using an emulator in almost the entire model.
- We estimate energy savings in cooperation with computer scientists (the groups of Krishna Palem - Rice University, Christian Enz - EPFL and John Augustine - IITM).

Resolution	Number of bits in significand	Normalised Energy Demand	Forecast error Z500 at day 2
235 km	52	1.0	2.3
315 km	52	0.47	4.5
235 km	**8**	0.29	2.5

We should reduce precision to allow simulations at higher resolution.
The IEEE floating point standard is not ideal.

Reduced precision in an atmosphere model

- We calculate weather forecasts with a spectral dynamical core (full 3D dynamics on the globe but no physics).
- Floating point precision is reduced to 8 bits in the significand using an emulator in almost the entire model.
- We estimate energy savings in cooperation with computer scientists (the groups of Krishna Palem - Rice University, Christian Enz - EPFL and John Augustine - IITM).

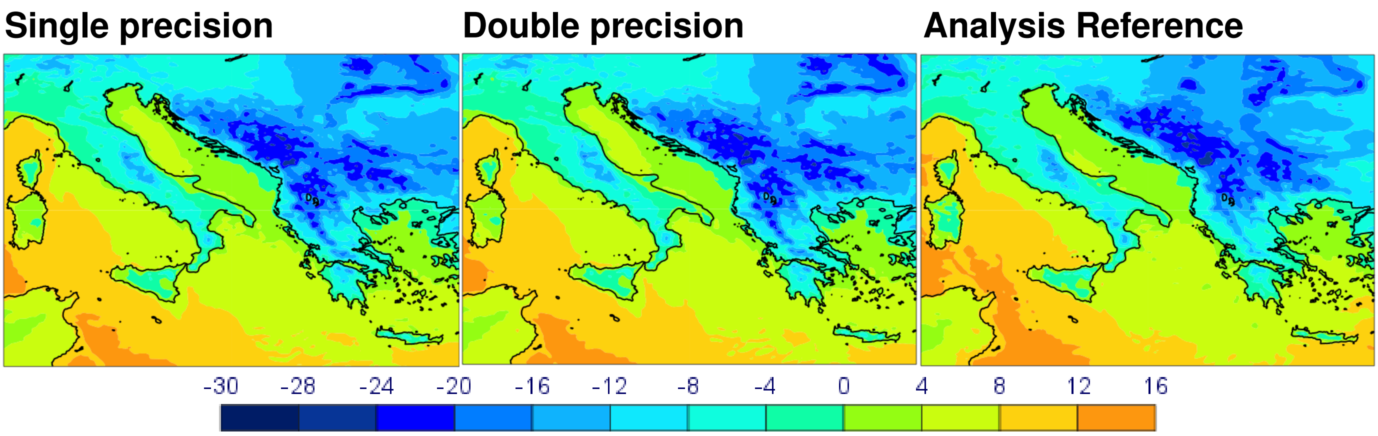
Resolution	Number of bits in significand	Normalised Energy Demand	Forecast error Z500 at day 2
235 km	52	1.0	2.3
315 km	52	0.47	4.5
235 km	**8**	0.29	2.5

We should reduce precision to allow simulations at higher resolution.
The IEEE floating point standard is not ideal.

Studies with real hardware (FPGAs) confirm this result.

Düben et al. MWR 2015; Düben et al. DATE 2015; Düben et al. JAMES 2015; Russel, Düben et al. FCCM 2015.

ECMWF’s weather forecast model in single precision



Surface temperature in °C

- Forecast quality in double and single precision is almost identical.
- 40% speed-up.
- Benefit for global simulations at 1.0 km resolution.

Düben and Palmer MWR 2014; Váňa, Düben et al. MWR 2017

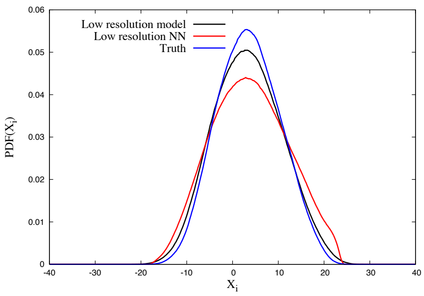
Deep learning for weather forecasts?

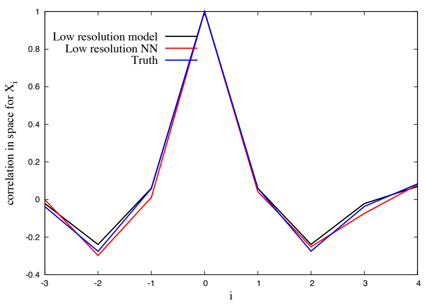
Can Neural Networks (NNs) be used for global weather predictions?

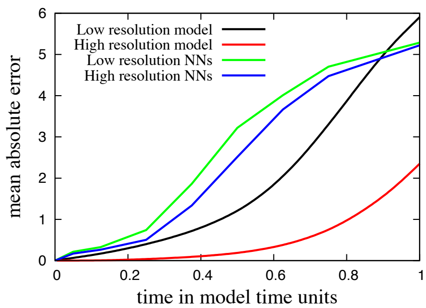
Deep learning for weather forecasts?

Can Neural Networks (NNs) be used for global weather predictions?

We perform tests with a toy model for atmospheric dynamics called Lorenz'95.



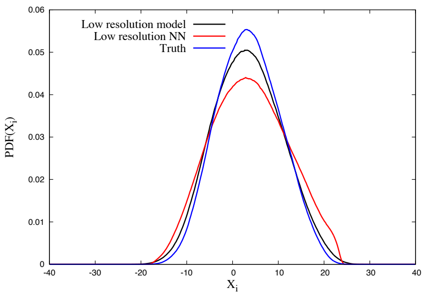


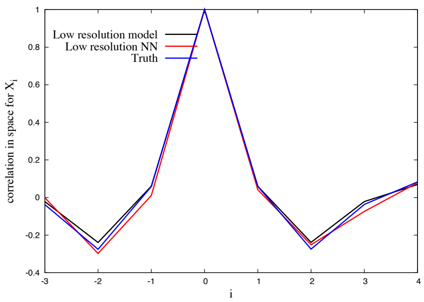


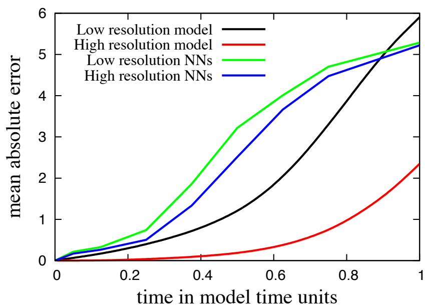
Deep learning for weather forecasts?

Can Neural Networks (NNs) be used for global weather predictions?

We perform tests with a toy model for atmospheric dynamics called Lorenz'95.





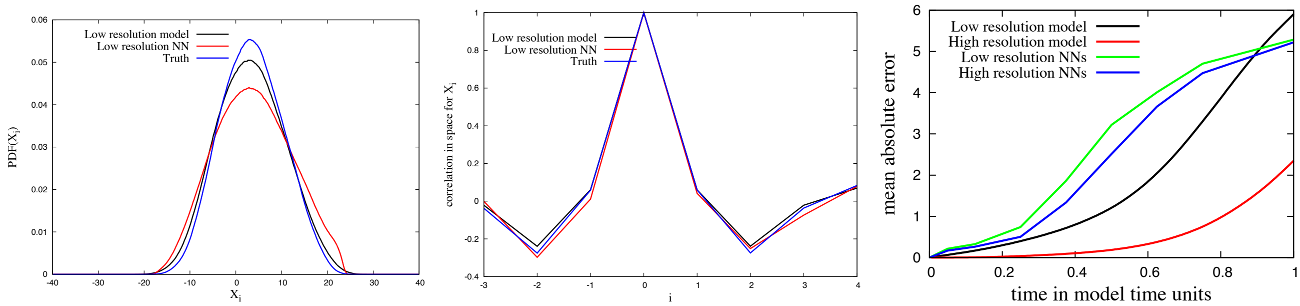


Model dynamics are reasonable but forecast error with NNs is higher compared to dynamical models.

Deep learning for weather forecasts?

Can Neural Networks (NNs) be used for global weather predictions?

We perform tests with a toy model for atmospheric dynamics called Lorenz’95.



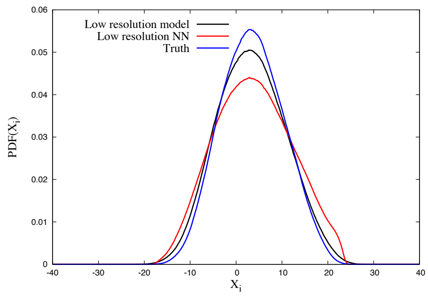
Model dynamics are reasonable but forecast error with NNs is higher compared to dynamical models.

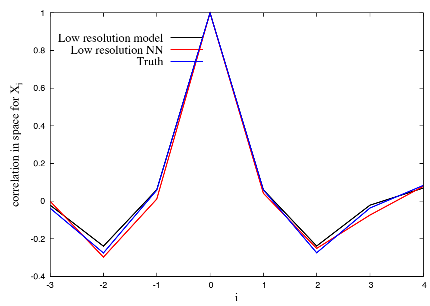
There are also only ≈ 40 years of satellite observations.

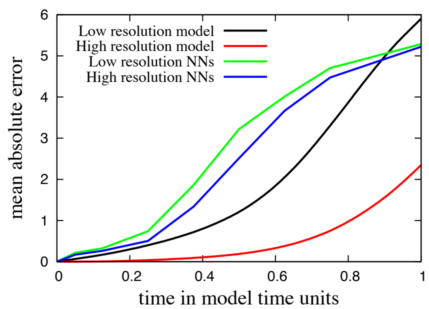
Deep learning for weather forecasts?

Can Neural Networks (NNs) be used for global weather predictions?

We perform tests with a toy model for atmospheric dynamics called Lorenz'95.







Model dynamics are reasonable but forecast error with NNs is higher compared to dynamical models.

There are also only ≈ 40 years of satellite observations.

Google will not make us unemployed but NNs may work well for local predictions.

Deep learning for weather forecasts?

NNs can still be useful for global weather forecasting.

Deep learning for weather forecasts?

NNs can still be useful for global weather forecasting.

NNs can replace existing model components to speed-up simulations.

Deep learning for weather forecasts?

NNs can still be useful for global weather forecasting.

NNs can replace existing model components to speed-up simulations.

NNs were used to replace the radiation scheme in ECMWF weather forecasts in the past. (Chevallier et al. 2000)

Deep learning for weather forecasts?

NNs can still be useful for global weather forecasting.

NNs can replace existing model components to speed-up simulations.

NNs were used to replace the radiation scheme in ECMWF weather forecasts in the past. (Chevallier et al. 2000)

Resources that are saved can be re-invested to improve forecasts.

Deep learning for weather forecasts?

NNs can still be useful for global weather forecasting.

NNs can replace existing model components to speed-up simulations.

NNs were used to replace the radiation scheme in ECMWF weather forecasts in the past. (Chevallier et al. 2000)

Resources that are saved can be re-invested to improve forecasts.

We will now repeat this exercise.

Conclusions

- The Earth System is complex, chaotic and huge, and we do not have sufficient resolution to resolve all important processes. Therefore, weather and climate predictions are difficult.
- Earth System modelling is an HPC application.
- We make a lot of efforts to make the most of state-of-the-art and future supercomputing hardware (dwarfs, domain specific languages, scalable algorithms,...).
- We achieve promising results with the new generation of GPUs.
- A reduction in precision can improve efficiency within our models.
- Neural Networks may help to improve efficiency for existing model components in the future.