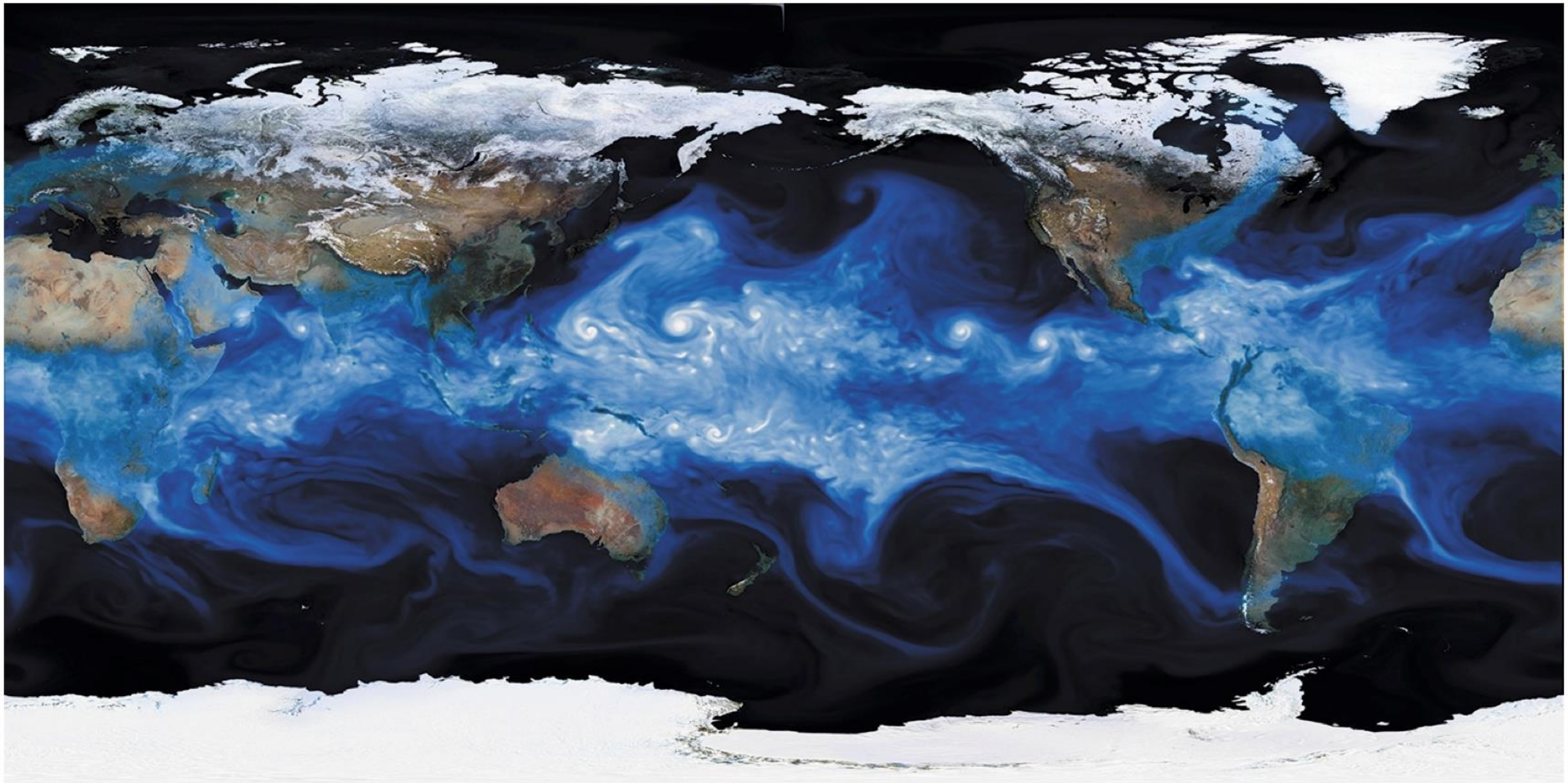
Reduced numerical precision guided by physics-dynamics coupling

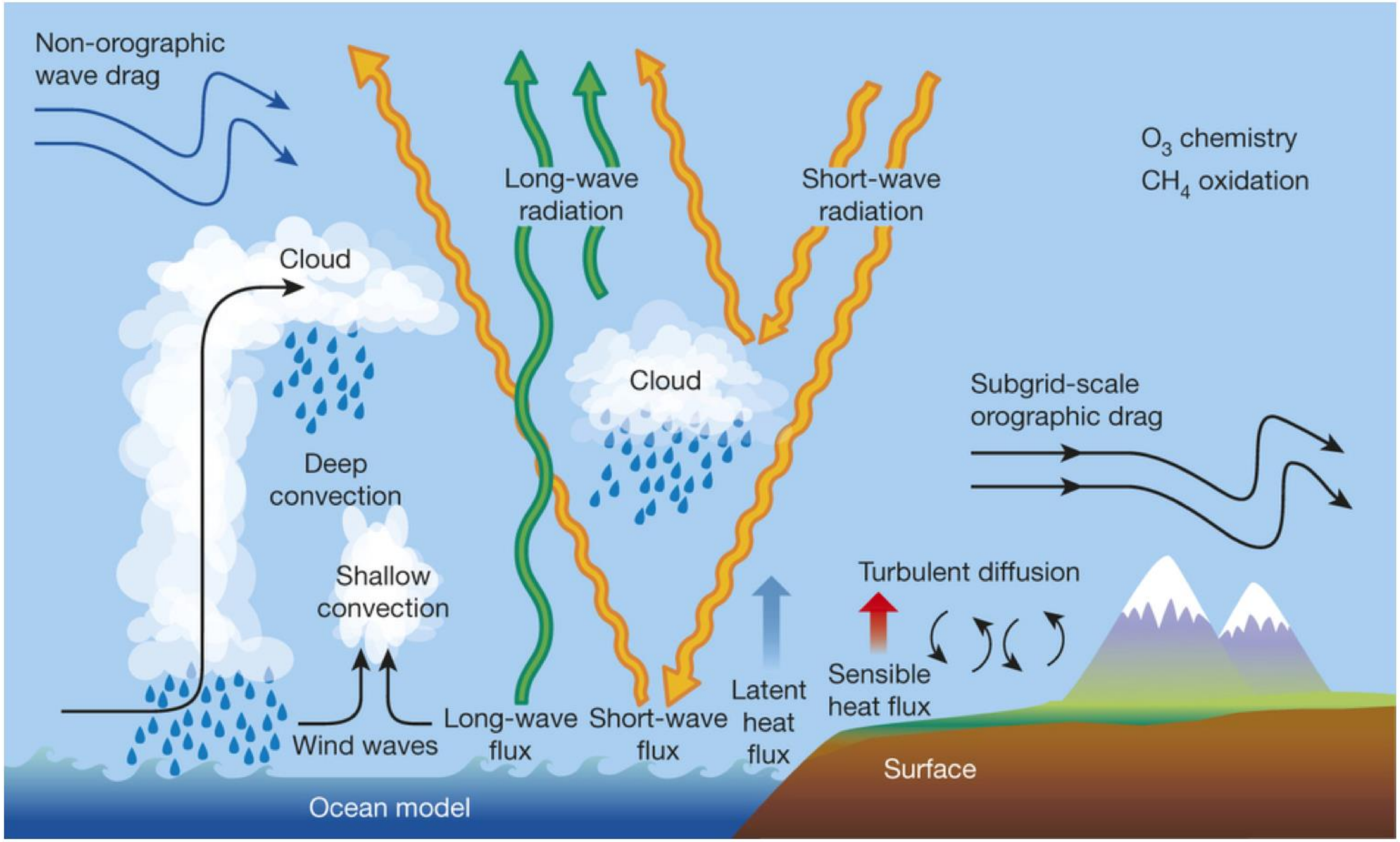
Matthew Chantry
Tim Palmer, Peter Duben
Jan Ackmann, Sam Hatfield, Milan Kloewer,
Andrew McRae, Leo Saffin, Tobias Thornes

Weather prediction: Unresolved scales



Lawrence Berkeley Natl Lab./Data: Michael Wehner (LBNL)/Visualization: Prabhat (LBNL)

Weather prediction: Imperfect parameterisations



Bauer et al. Nature 2015

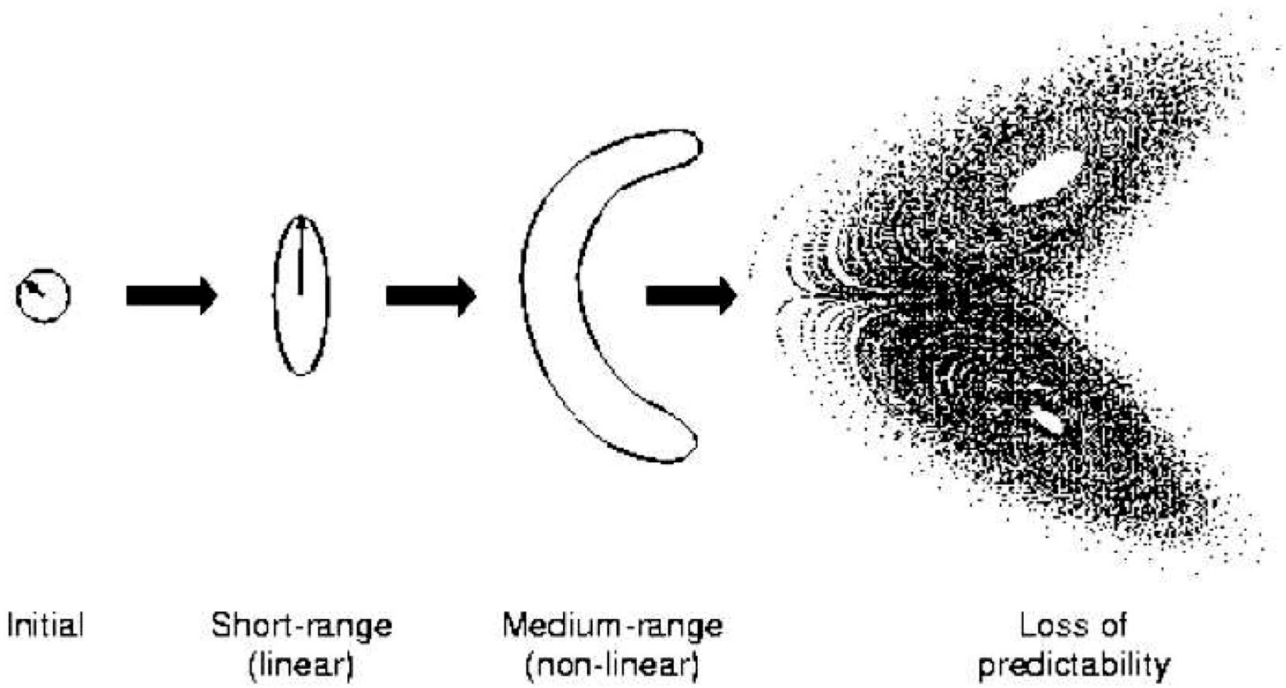
Ensembles and stochasticity

Problem

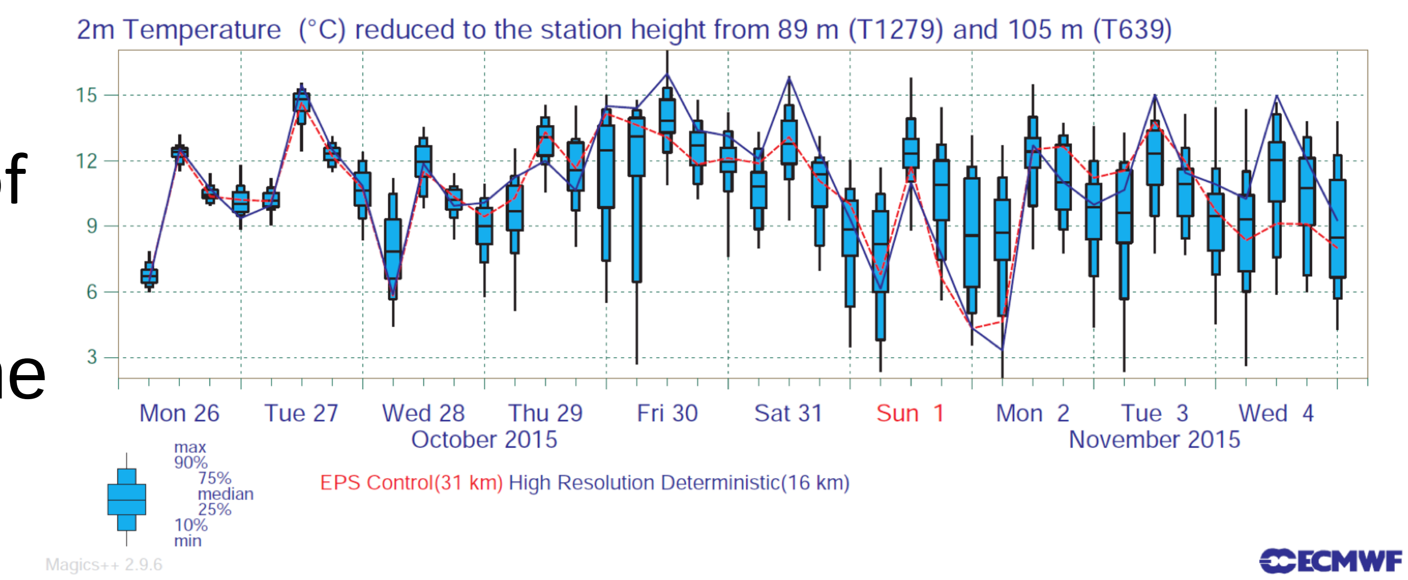
- Weather forecasting attempts to predict a highly chaotic dynamical system.
- Initial condition errors will grow exponentially.

Solution

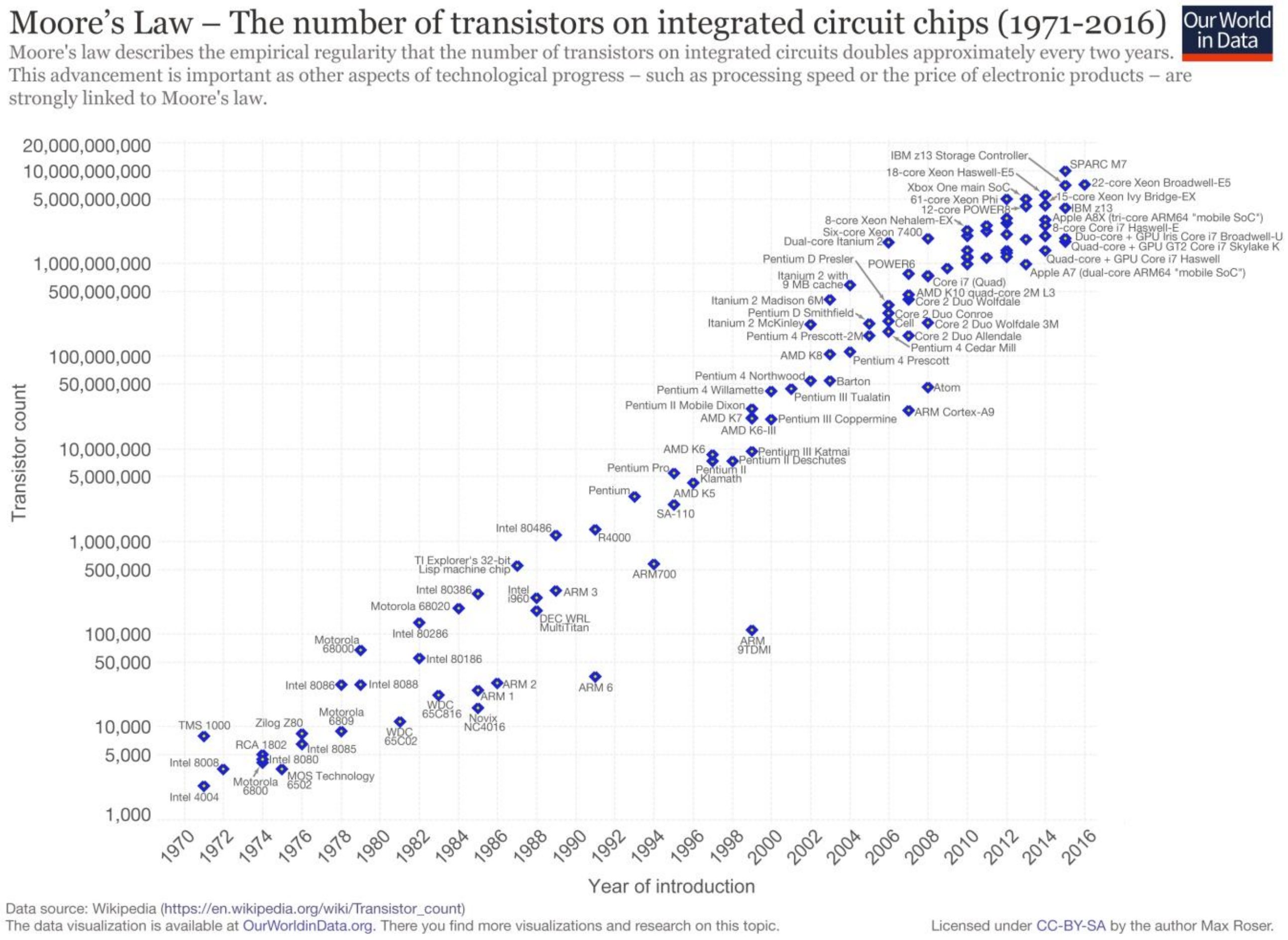
- Propagate an ensemble of initial conditions forward to (hopefully) include the truth in the distribution of possible answers.
- Random elements are added to the model to increase spread. e.g. SPPT, SKEBS, SPT.



Source: ECMWF IFS documentation



Why care about precision?



Moore's "law":
twice as many transistors per chip every 2 years

New computers are bigger but not faster.

- Reaching physical limits of transistor size.
- Parallel computing is the main route to higher grid resolution.

Energy consumption

- MetOffice supercomputer: 2.7 MW of electricity.

Looking for any possible paths to faster/more efficient code.

Floating point numbers

Method to encode numbers in binary

Significand Exponent

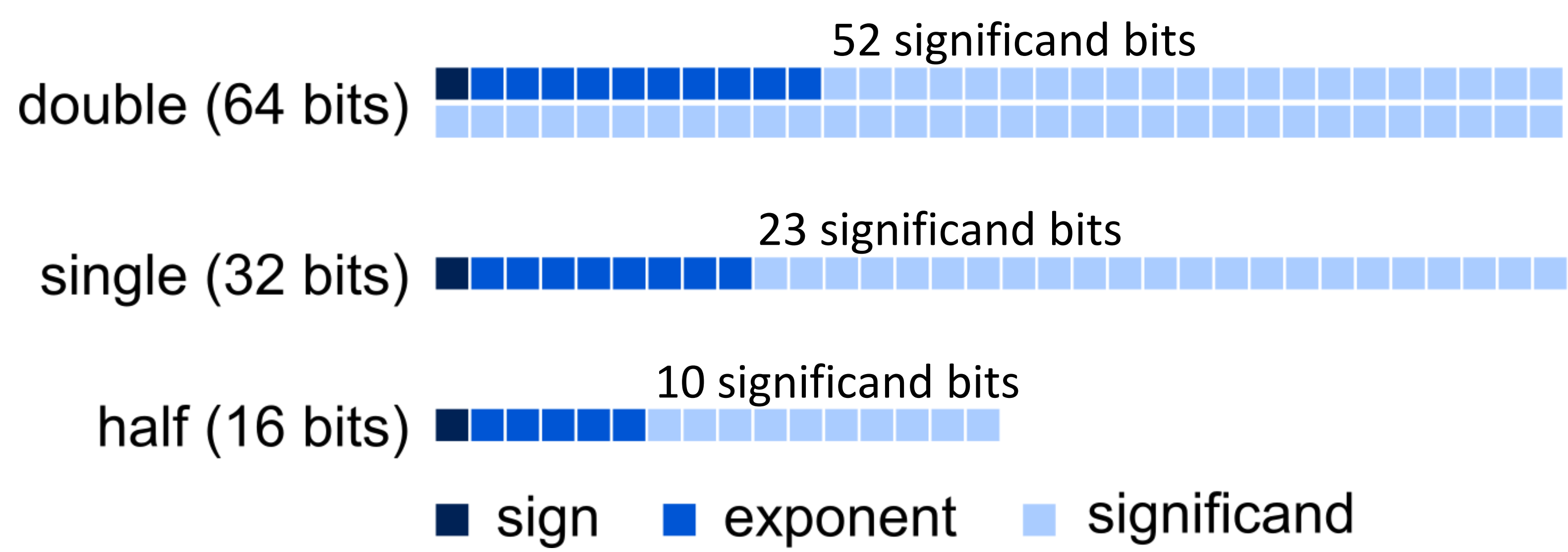
$$x = (-1)^{\text{sign}} \times \left(1 + \sum_{n=1}^{N} s_n 2^{-n}\right) \times 2^{e}$$

Precision Magnitude

Think of

$$65504 = 6.5504 \times 10^{4}$$

Computers have standards layouts for these numbers



This talk: focus on the significand (precision).

“New” types of computers

Lower precision, parallel computations

GPU - Graphs processing unit

- Massively parallel.
- Used for machine learning, where high precision is often unnecessary.
- Support half-precision floats.

FPGA - Field programmable gate arrays

- Programming at a logic gate level (very hard).
- Configure a chip to solve only your equations (very power efficient).
- Can use arbitrary numerical precisions (not just double, single, half).
- Now available on cloud computing, e.g. Amazon, Microsoft.

Can we take advantage of these developments?

What’s been done

Single precision

- Met office - Pressure solve (operational) and large-scale precipitation.
- ECMWF - “full” forward integration model. Now used for testing and future model development.
- MeteoSwiss - most of model running operationally (60% savings over double).

Lower than single

- Reduced GCMs and simplified models at Oxford.
- Nemo ocean model in mixed precision at Barcelona Supercomputing Center.

What we'd like to do

Precision driven by the uncertainty in the model.

Dynamics accurate to the level masked by uncertainty in parameterisation schemes.

See upcoming paper by Subramanian et al.

Precision errors in parameterisations comparable to deviation from *truth* scheme.

What we've actually done so far

1. Investigation of impact of reduced precision in absence of coupling.
2. Precision in coupled models tuned for "minimal" change in output.

Emulated reduced precision

- Replace standard precision declaration with our derived types.
- Emulates arbitrary precision without large language/hardware changes (e.g. CUDA/FPGAs).
- Increases run-time, only useful for investigation.

Standard Fortran:

```
REAL :: a,b,c
a = 1.442221
b = 2.136601
c = a+b
```

$\rightarrow$ *c=3.578822*

Reduced precision declarations:

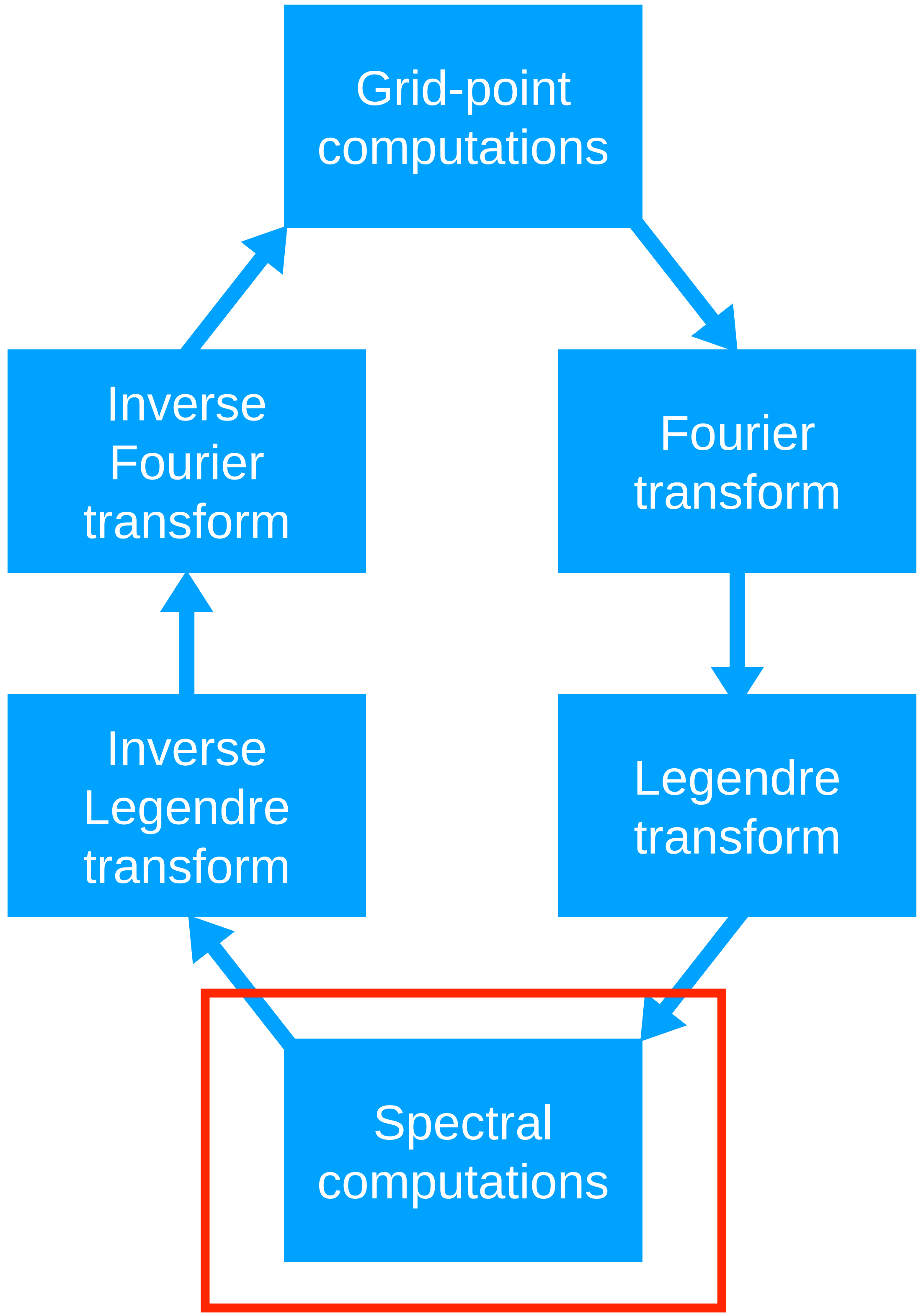
```
TYPE(reduced_precision) :: a,b,c
a = 1.442221
b = 2.136601
c = a+b
```

$\rightarrow$ *c=3.562500*

Dawson and Düben 2016

Spectral space
OpenIFS

Spectral dynamical core schematic



What we've done

- Reduced precision calculations in spectral-space only.
- Spectral transforms and grid-point calculations in double precision.

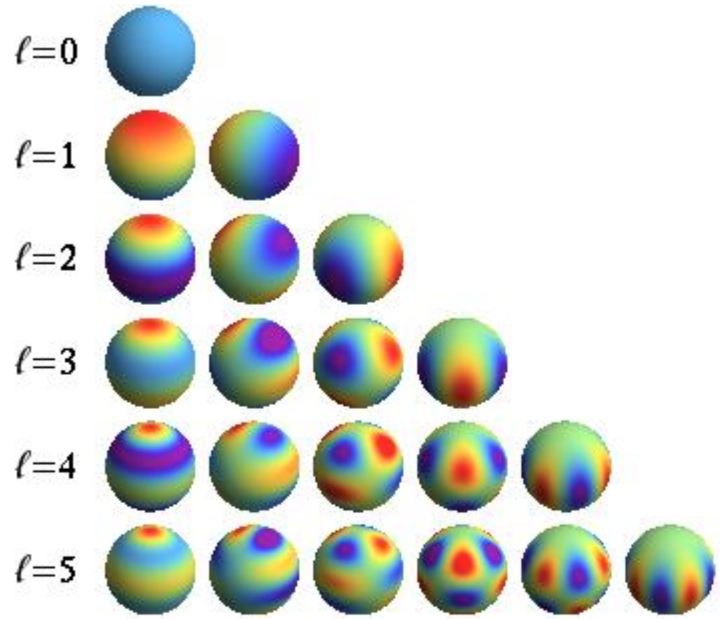
Will …

- introduce rounding errors to prognostic variables: vorticity, temperature etc.

Won't …

- cover all algorithmic error propagation

Why spectral space?

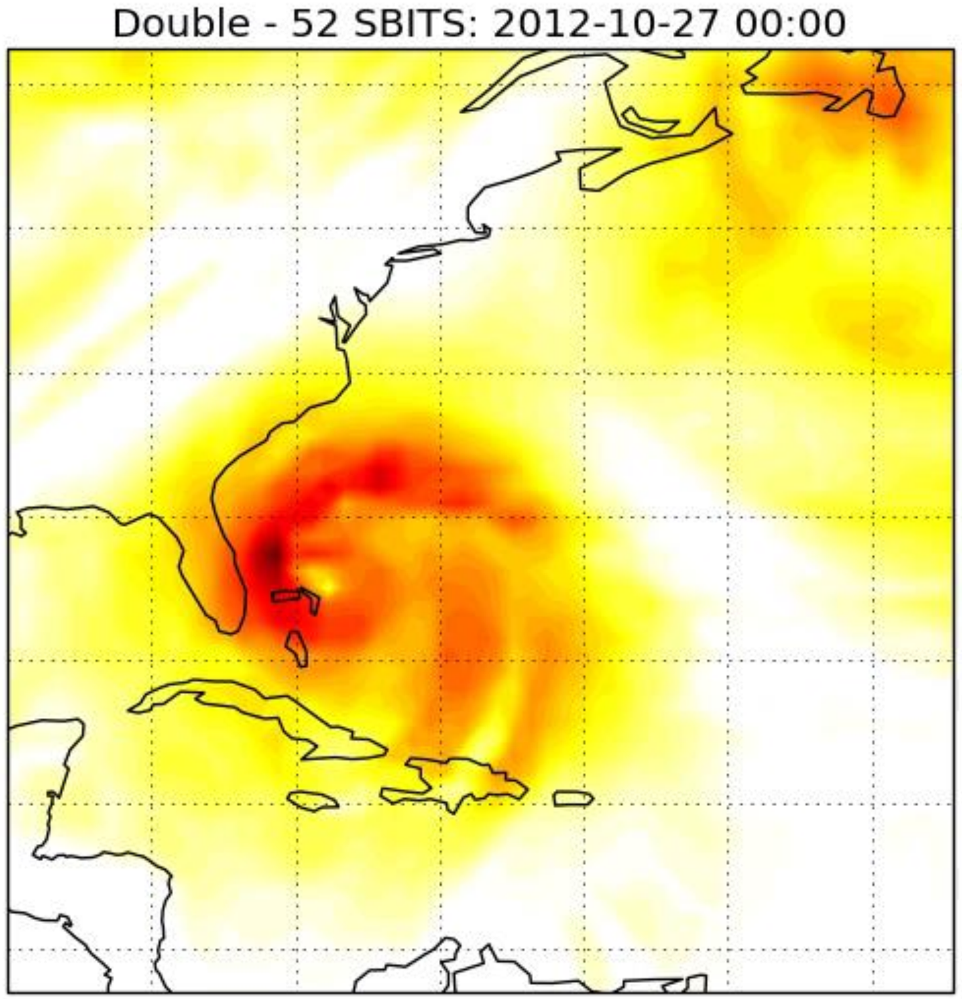


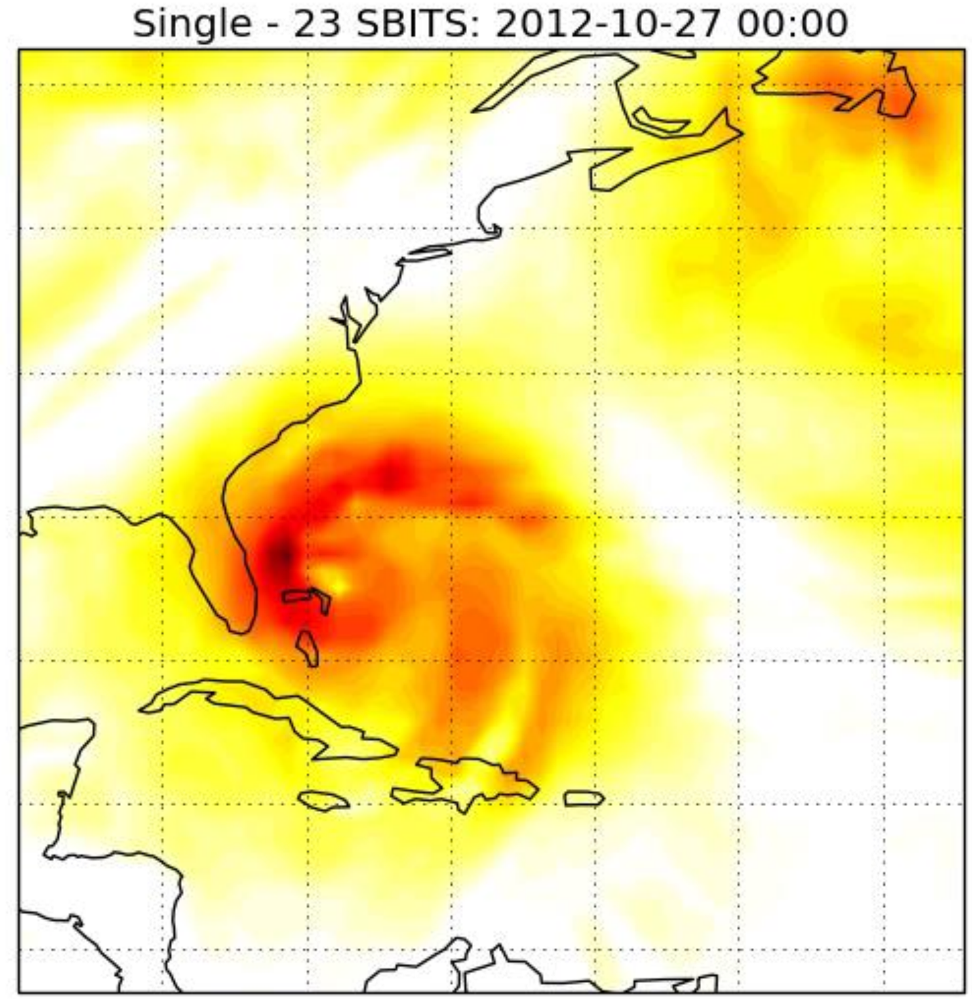
- Spectral models represent fields as a sum of modes representing different lengthscales.

- Can we reduce precision when calculating the small scales?

- This is appealing due to the high inherent uncertainty in small scale dynamics (parametrisation, viscosity, data-assimilation,…).

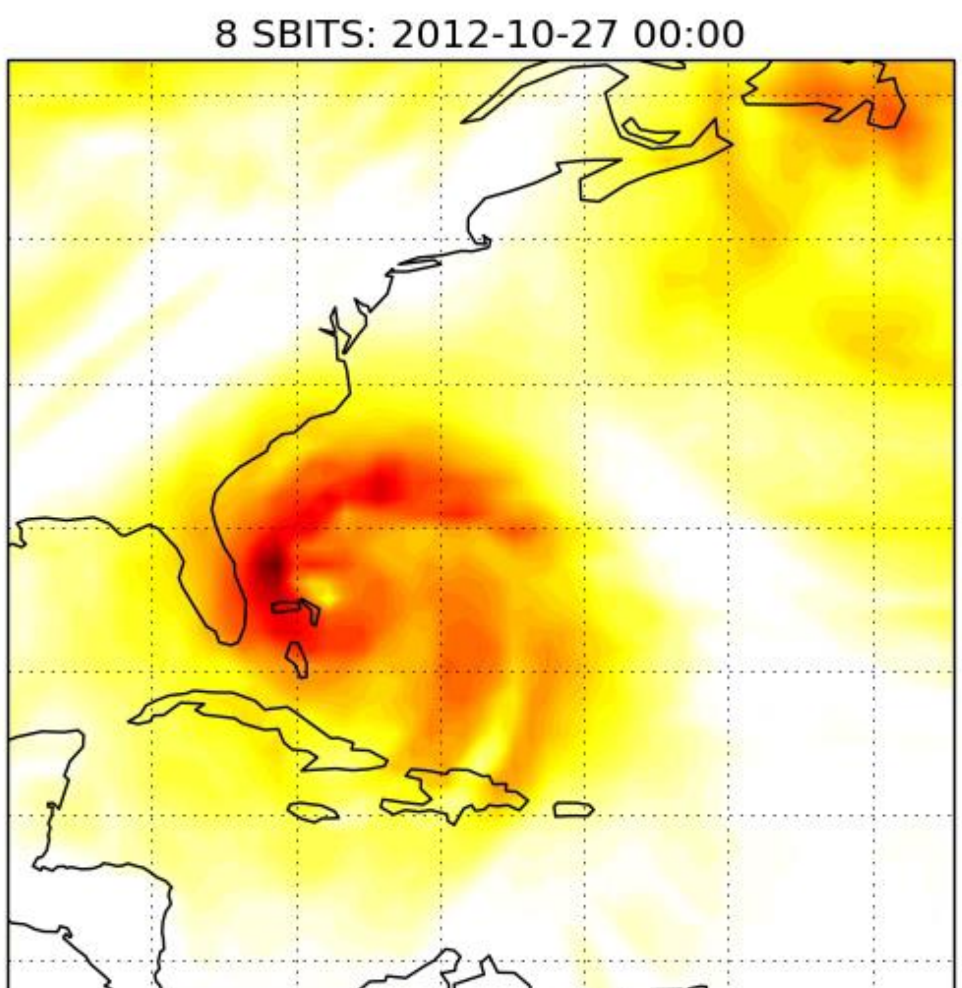
Double precision (52)





Single precision (23)

8 significand bits

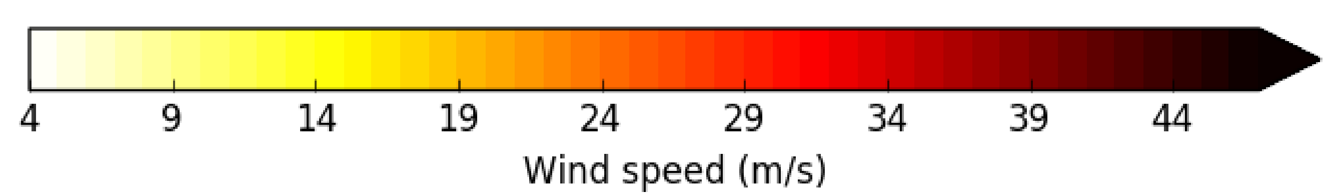


Hurricane Sandy

27/10/12 00:00

850hP wind speed

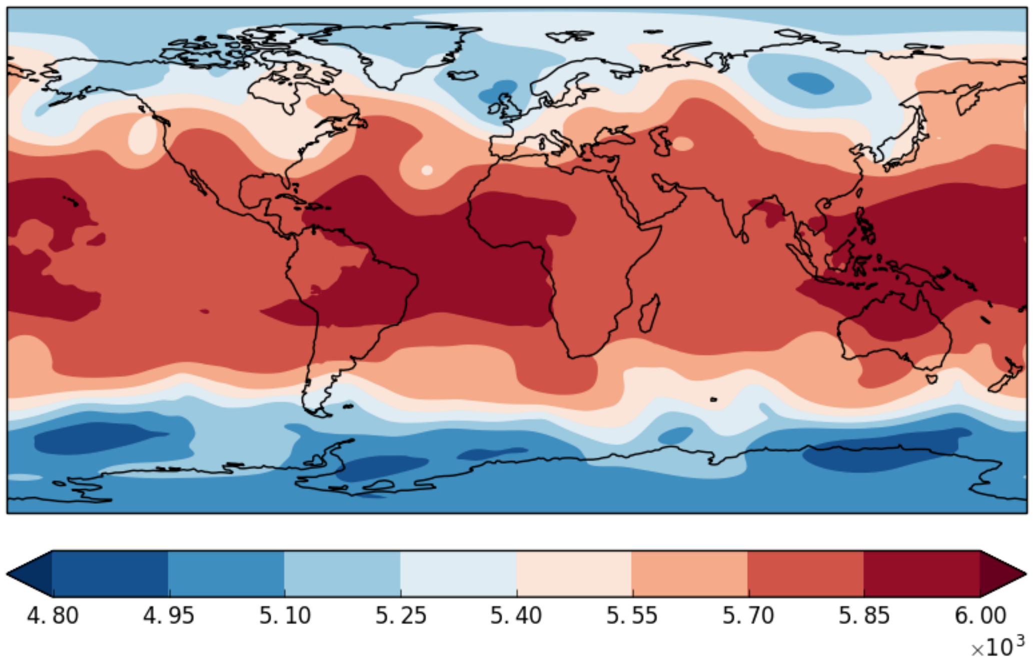
T255L91 ~ 80km



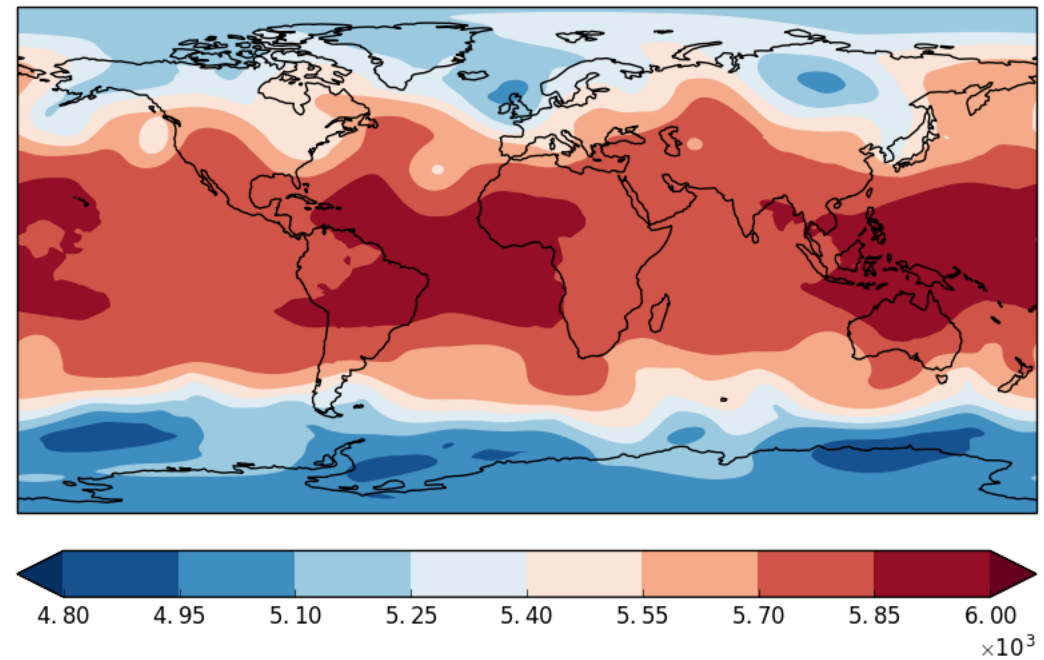
No need for scale-selectivity?

Z500hP after 120hours

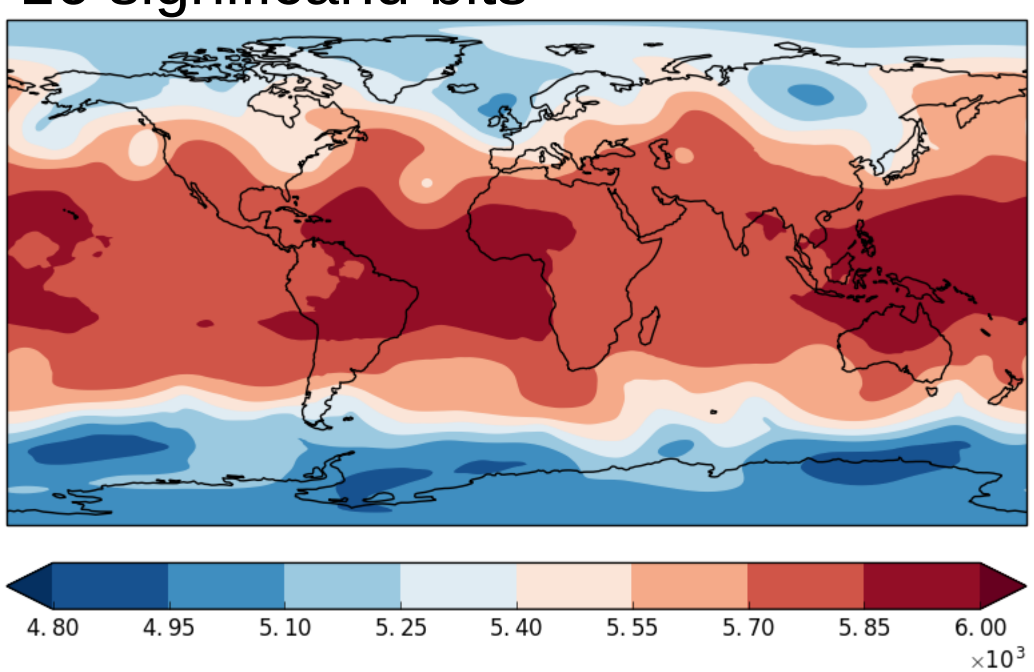
Double precision (52 sbits)



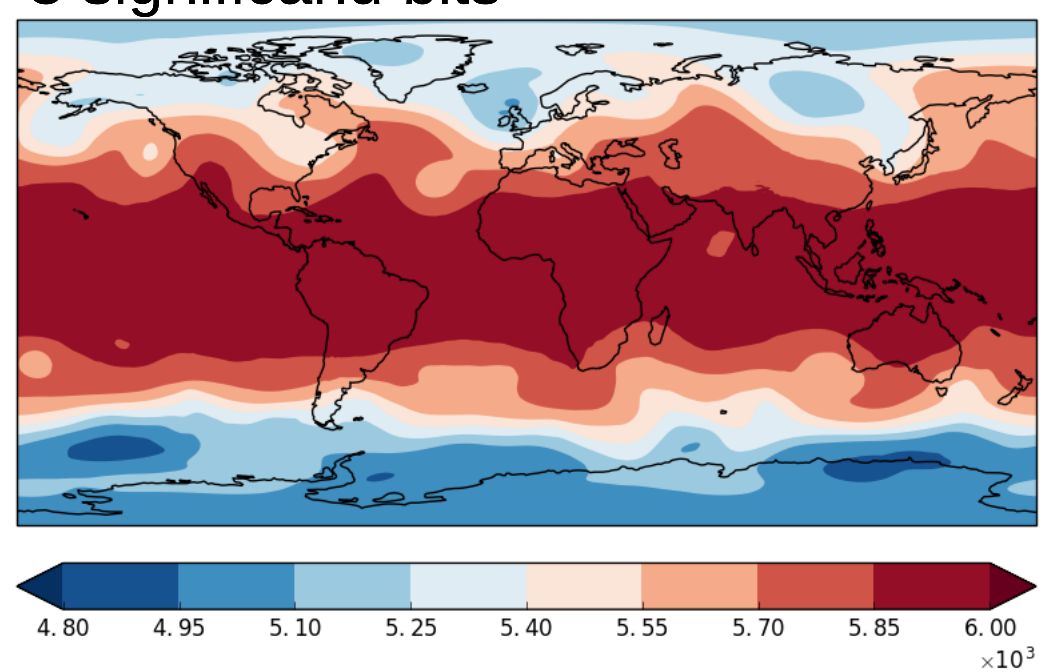
Single precision (23 sbits)



16 significand bits



8 significand bits

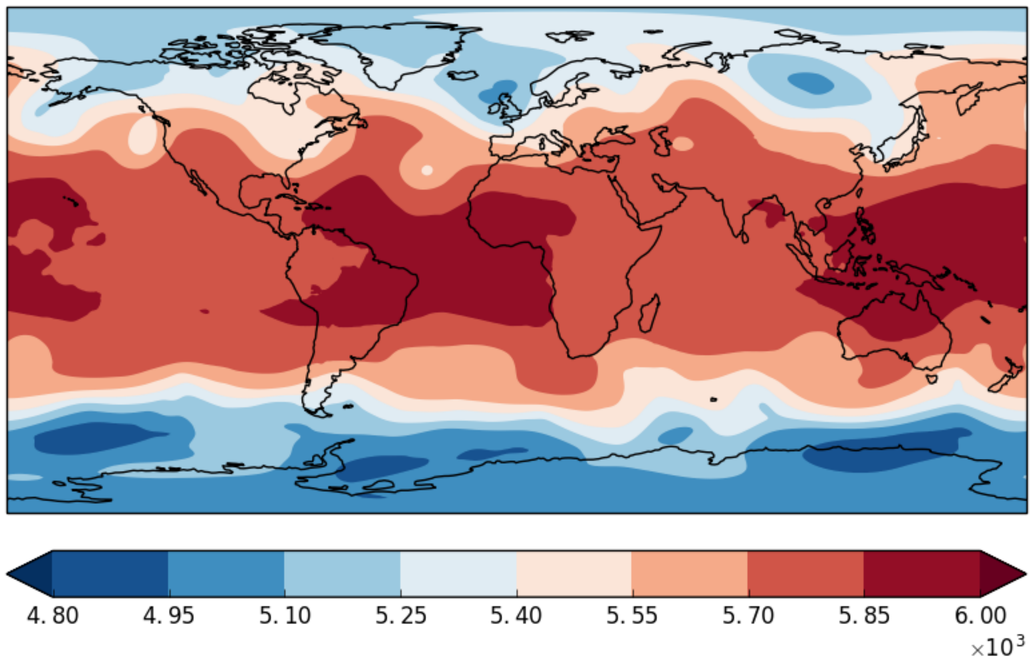


Bias in uniform 8 sbits.

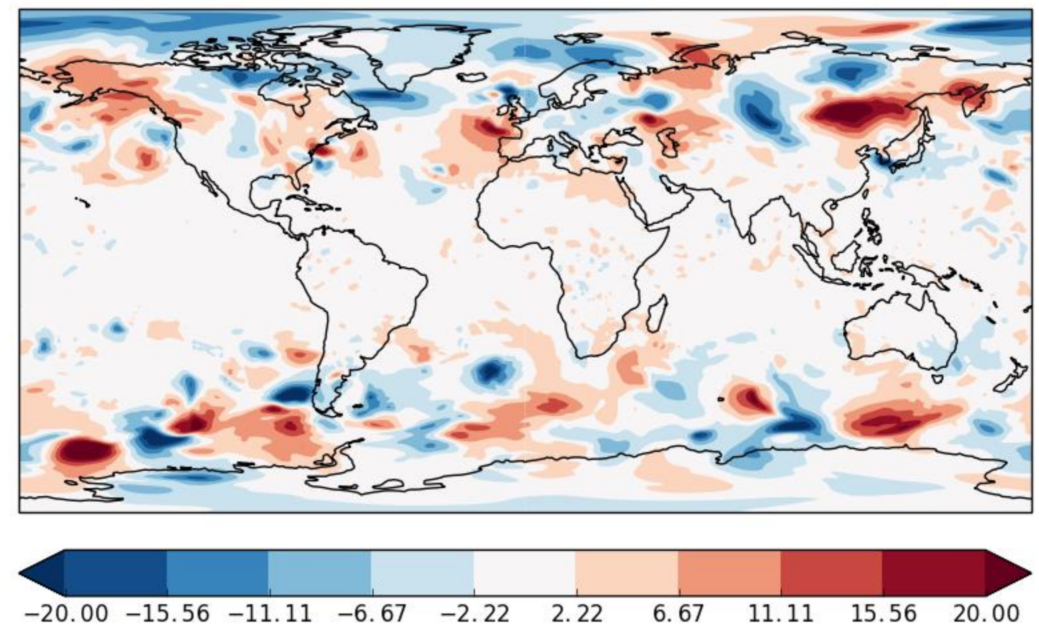
No need for scale-selectivity?

Z500hP after 120hours

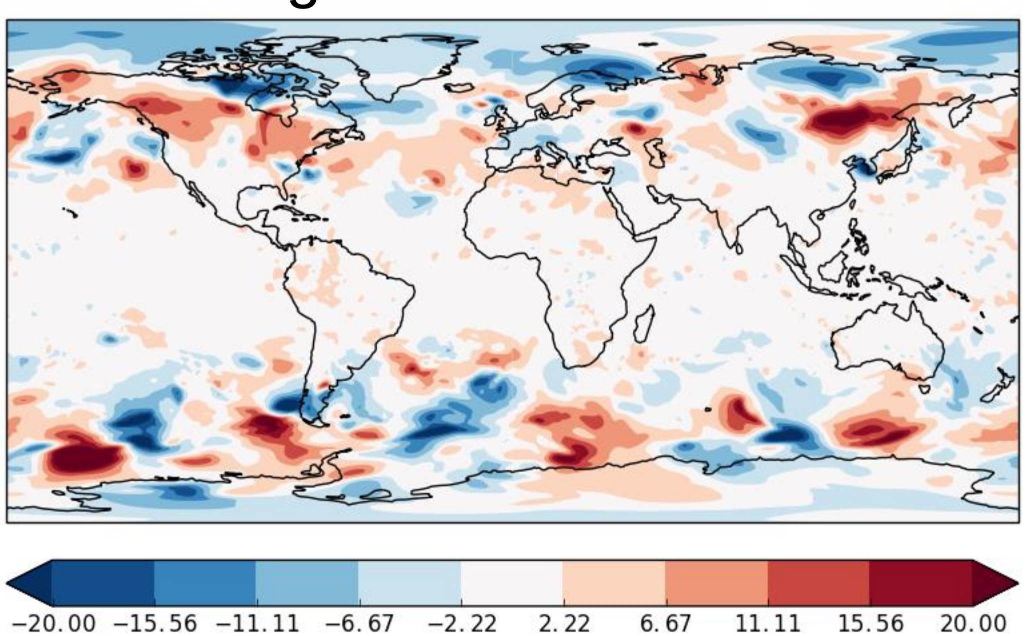
Double precision (52 sbits)



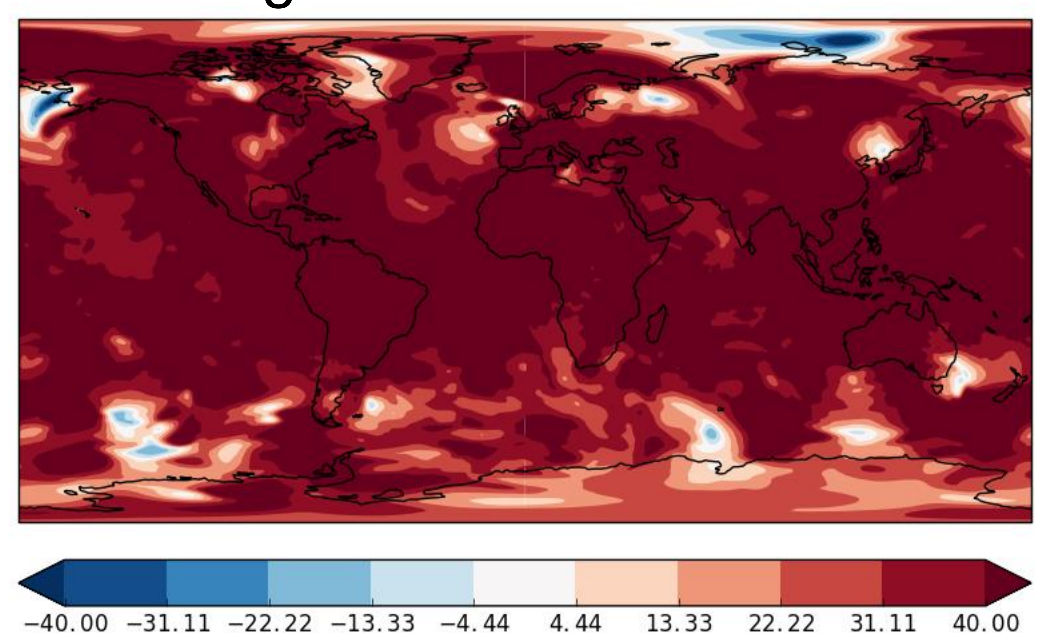
Db - Single precision



Db - 16 significand bits

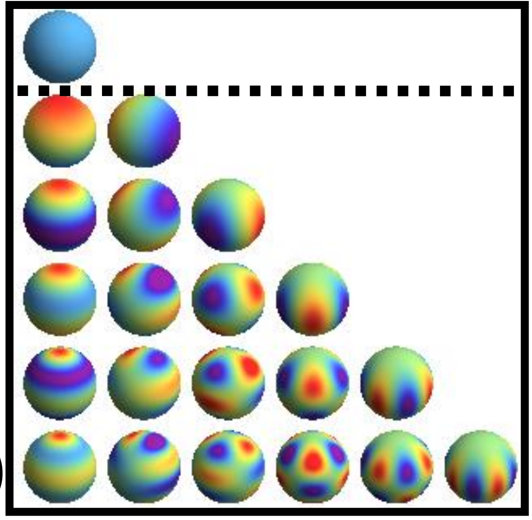


Db - 8 significand bits



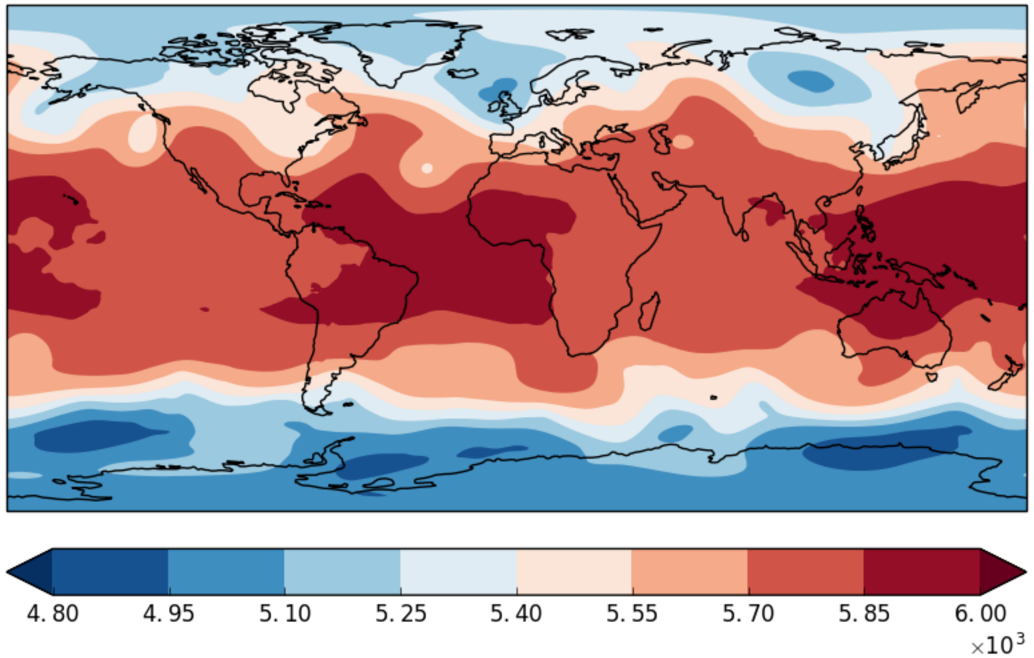
Global bias

First-order scale-selectivity

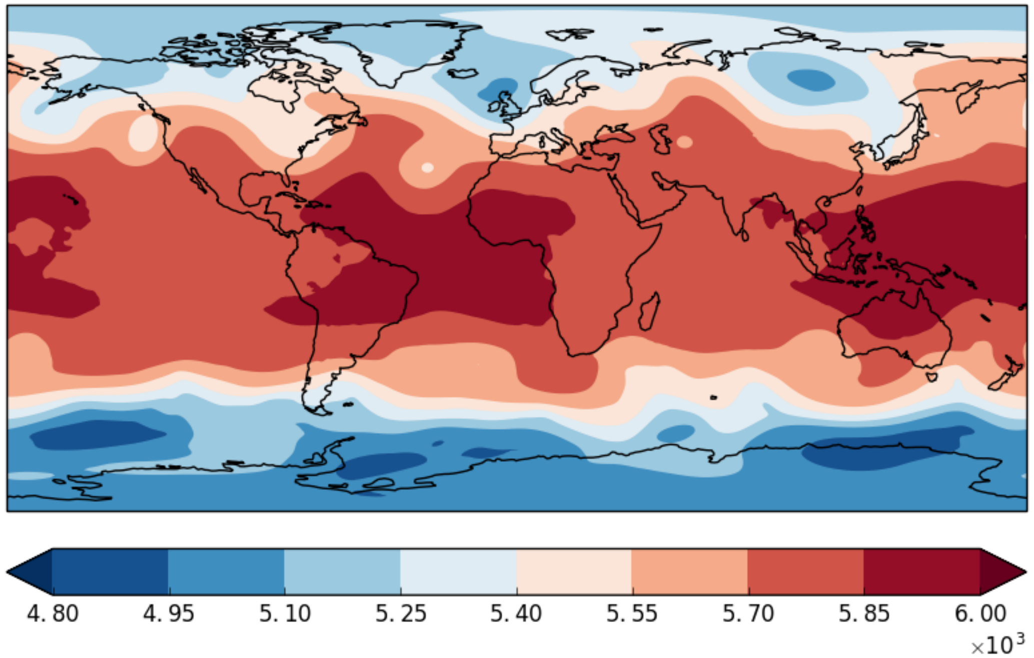


Z500hP after 120hours

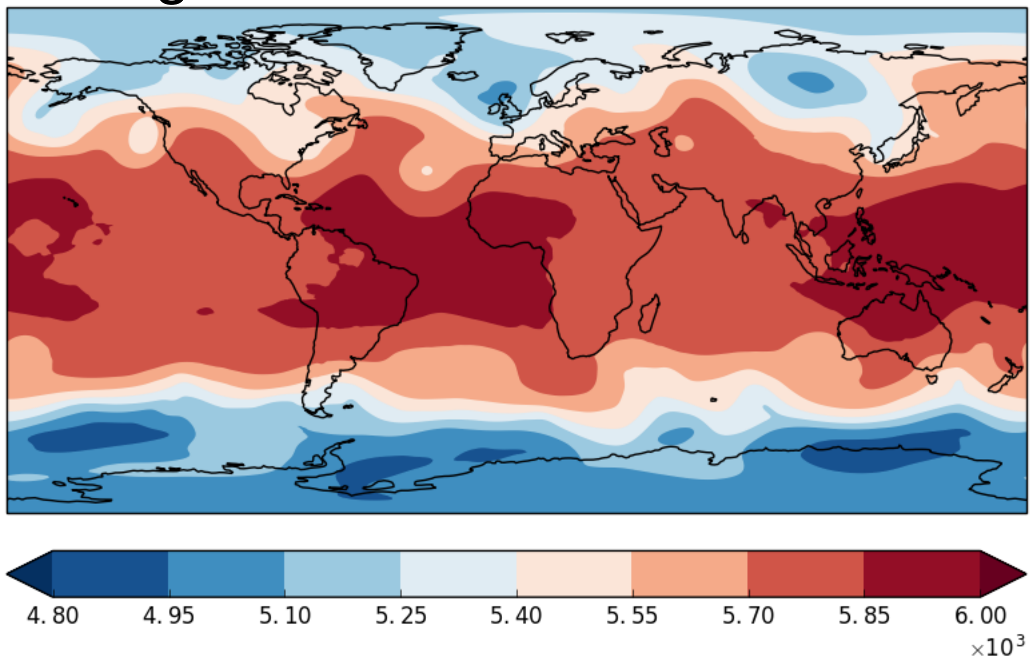
Double precision (52 sbits)



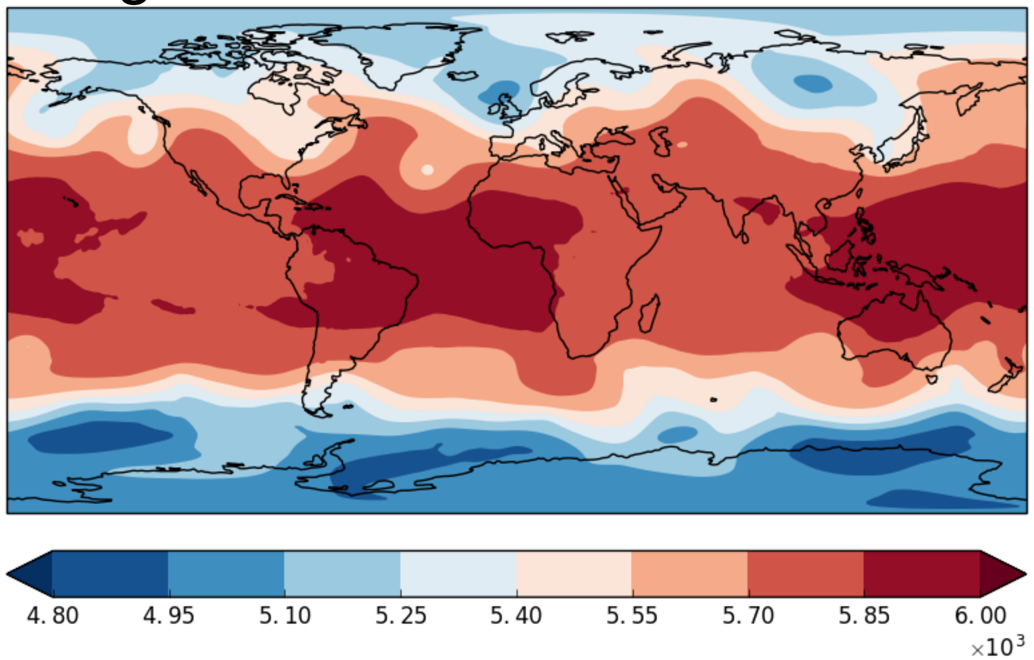
Single precision (23 sbits)



16 significand bits

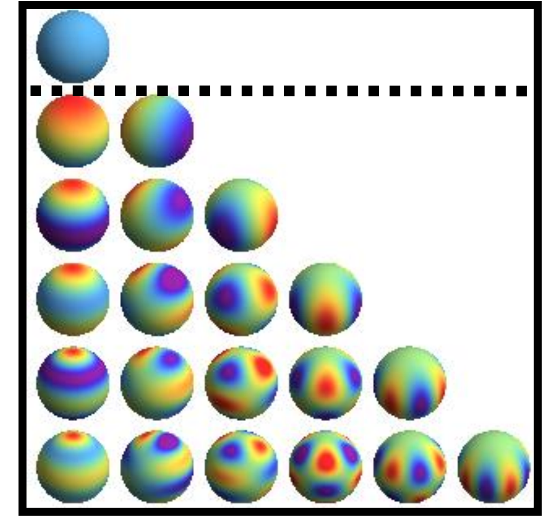


8 significand bits + …



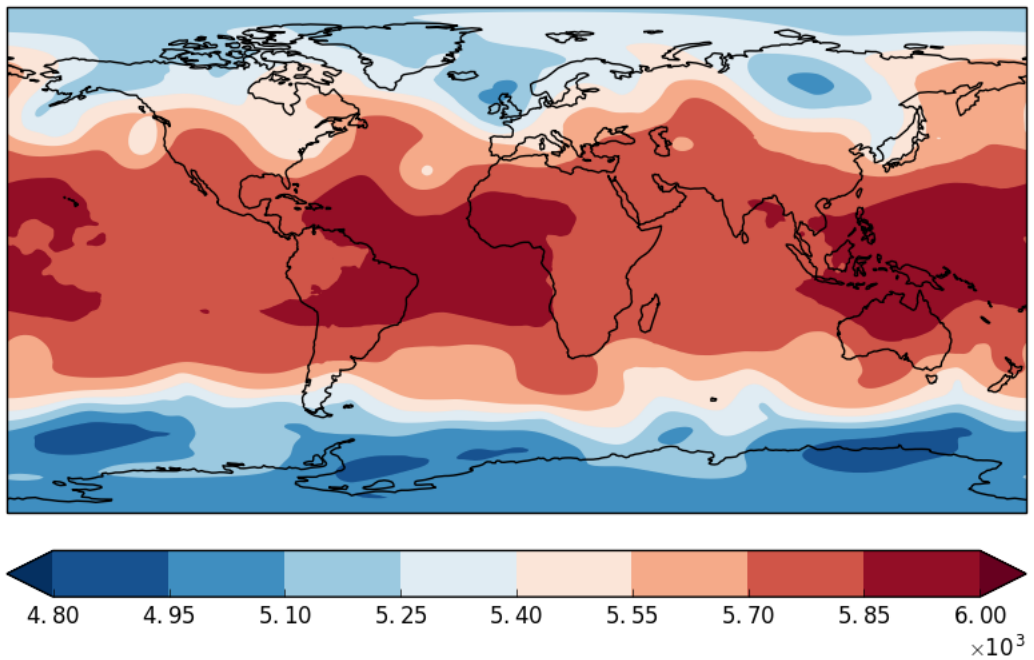
… double precision zero mode

First-order scale-selectivity

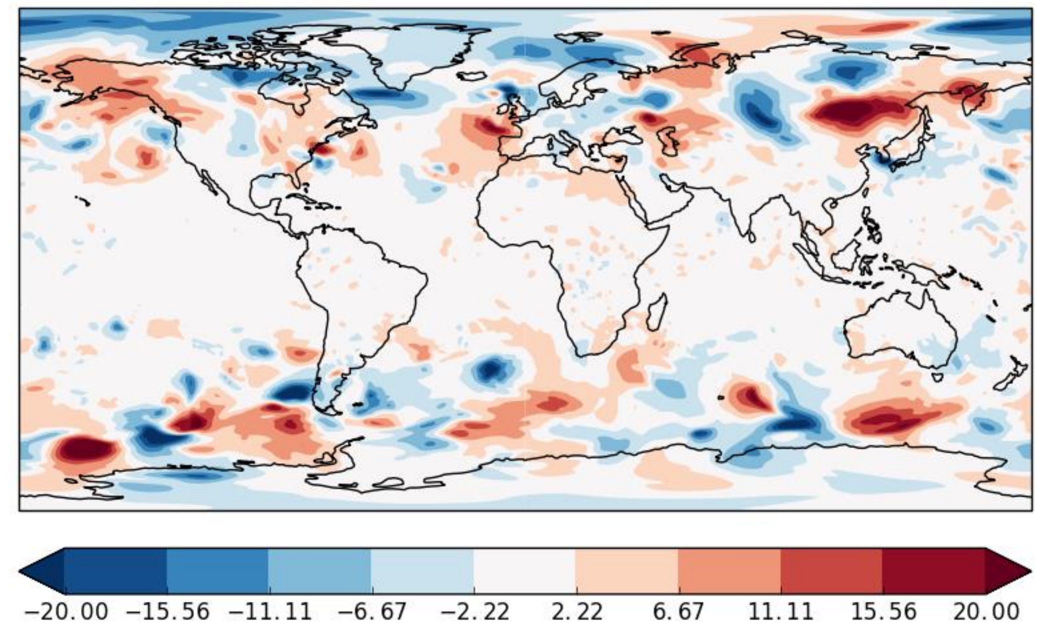


Z500hP after 120hours

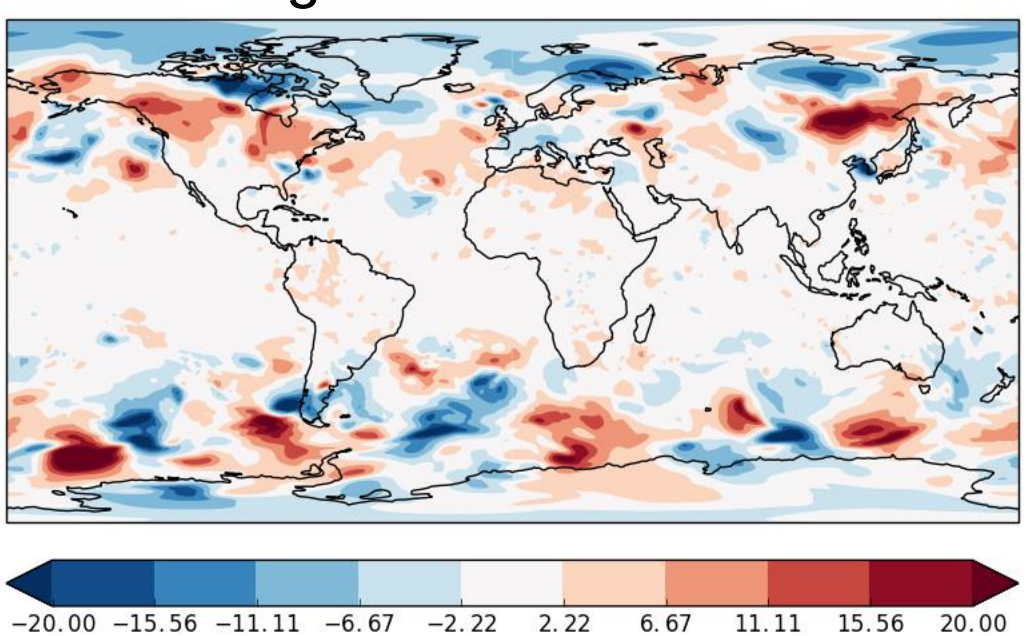
Double precision



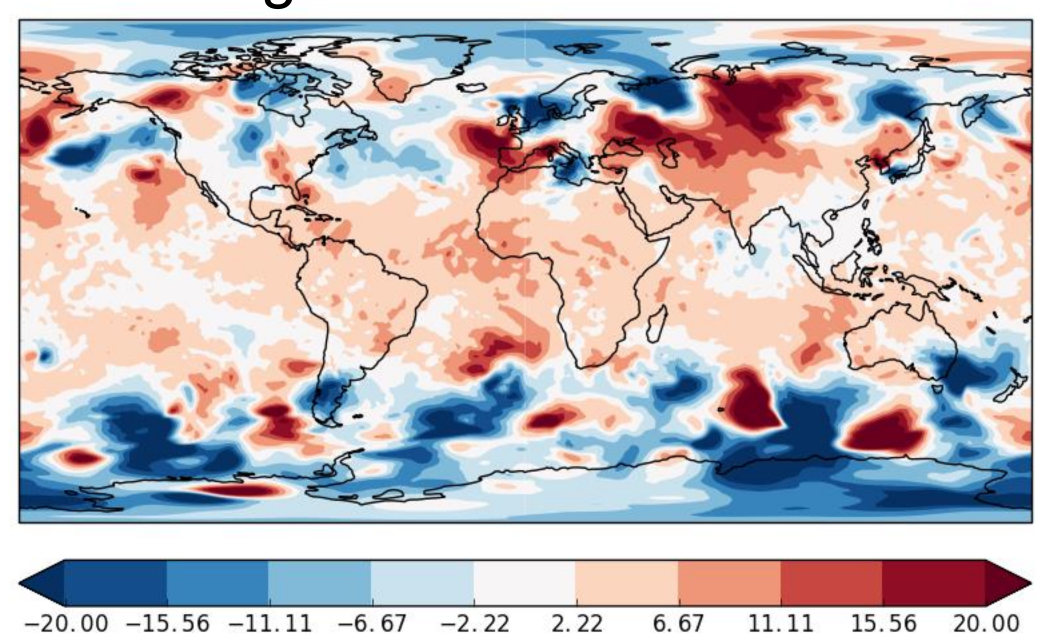
Db - Single precision



Db - 16 significand bits

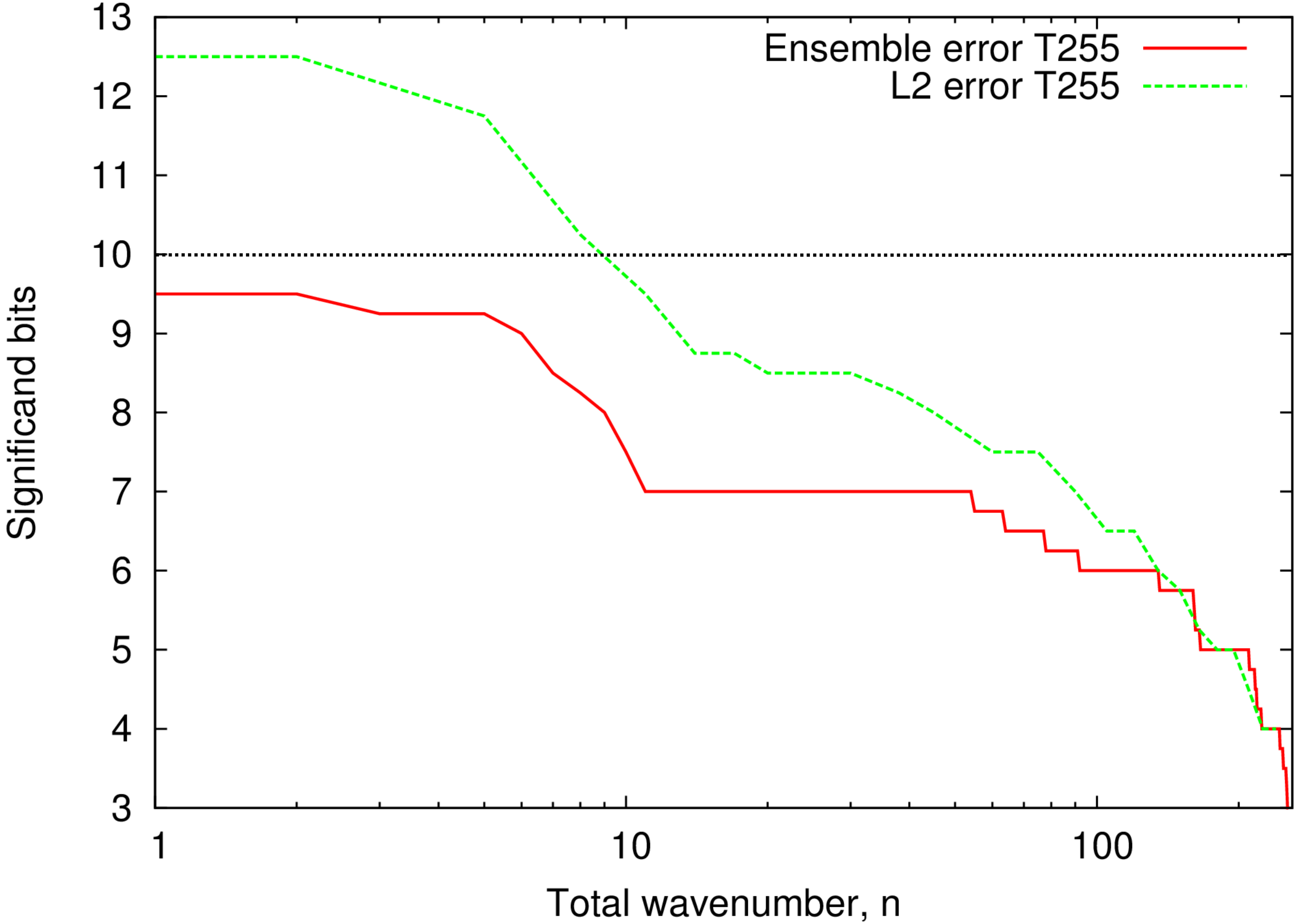


Db - 8 significand bits + …



… double precision zero mode

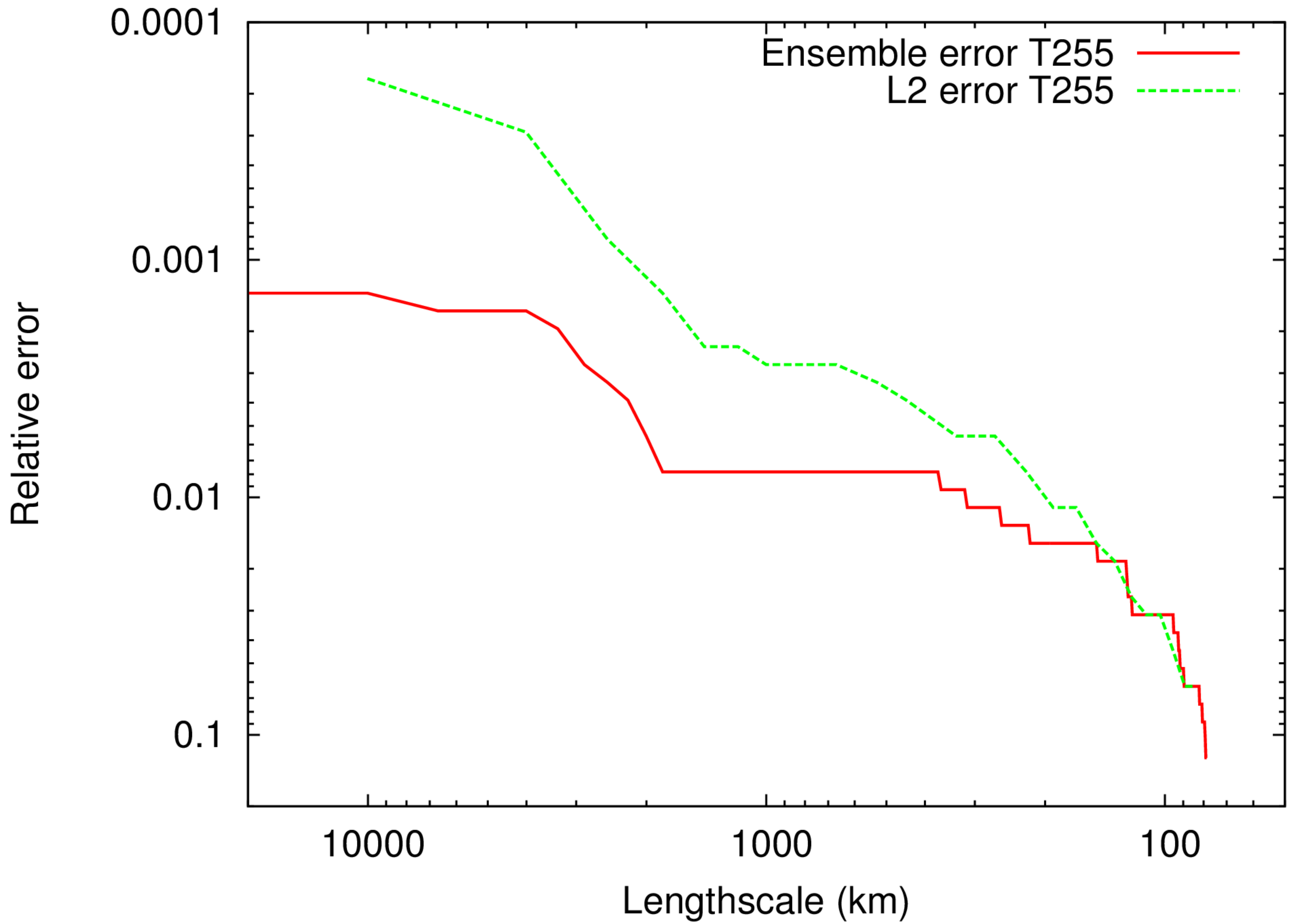
How many bits?



Half precision

~80km horizontal resolution

What error does this introduce?



~80km horizontal resolution

Climate investigation

- 11 year integration at T159L91 (~125km horizontal resolution).

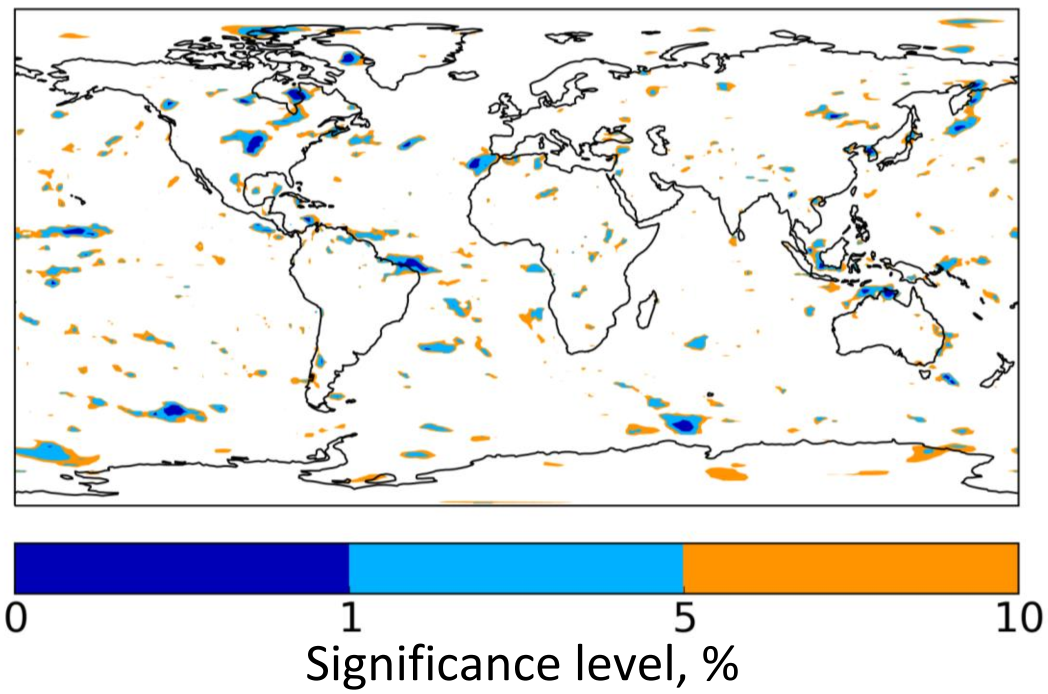
- 10 member ensembles for 2005-2015:
  double precision
  single precision
  half precision (zero mode in double precision)

- Do reduced precision errors accumulate?

T-test difference from double precision

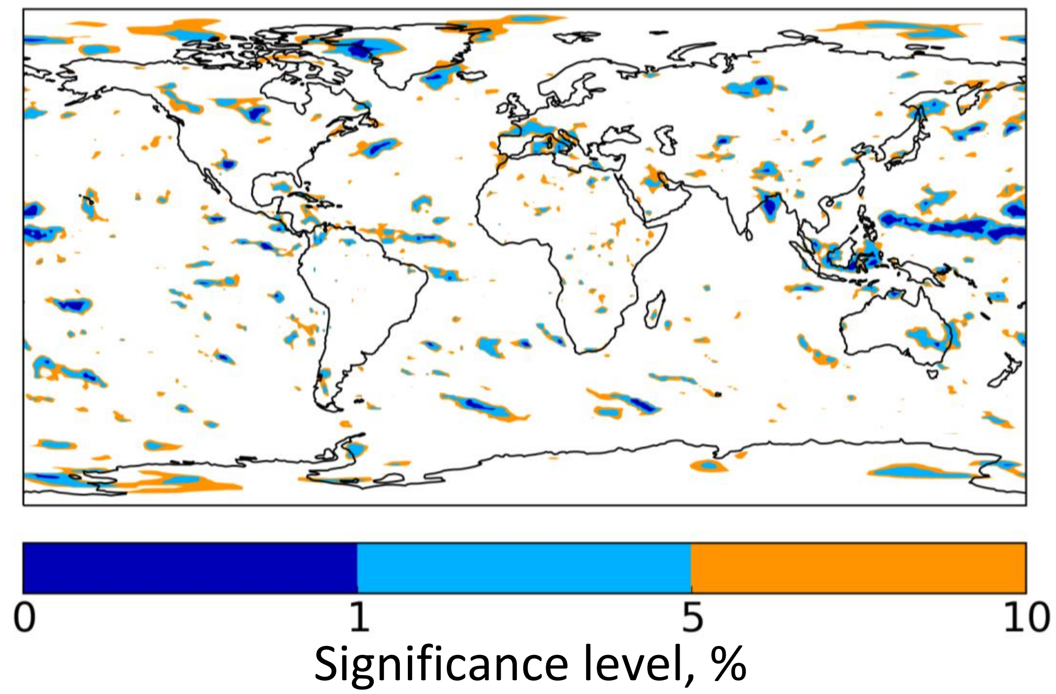
Accumulated precipitation

Single precision



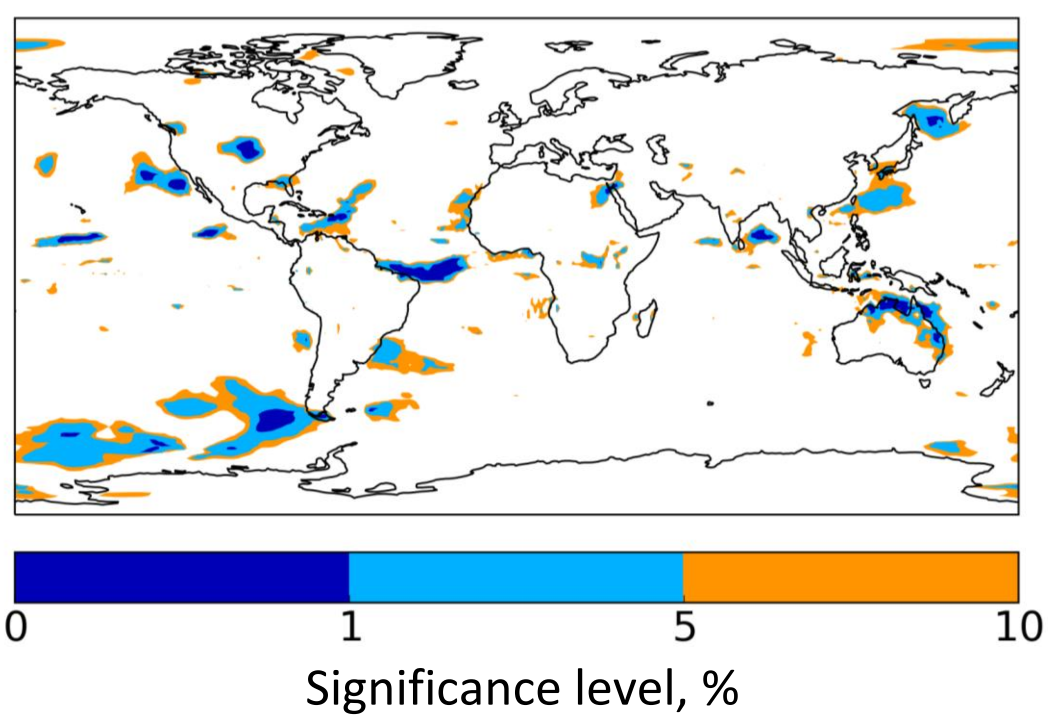
Proportion of significant grid points 3.2%

Half precision (zm)

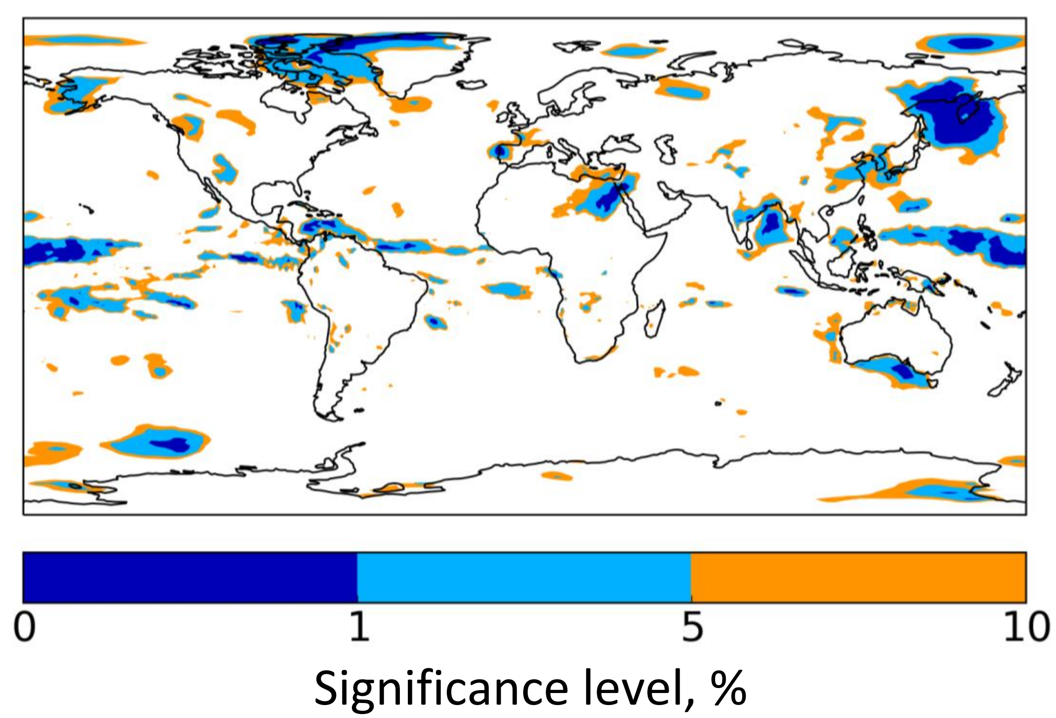


Proportion of significant grid points 5.5%

Mean 2m temperature



Proportion of significant grid points 4.6%



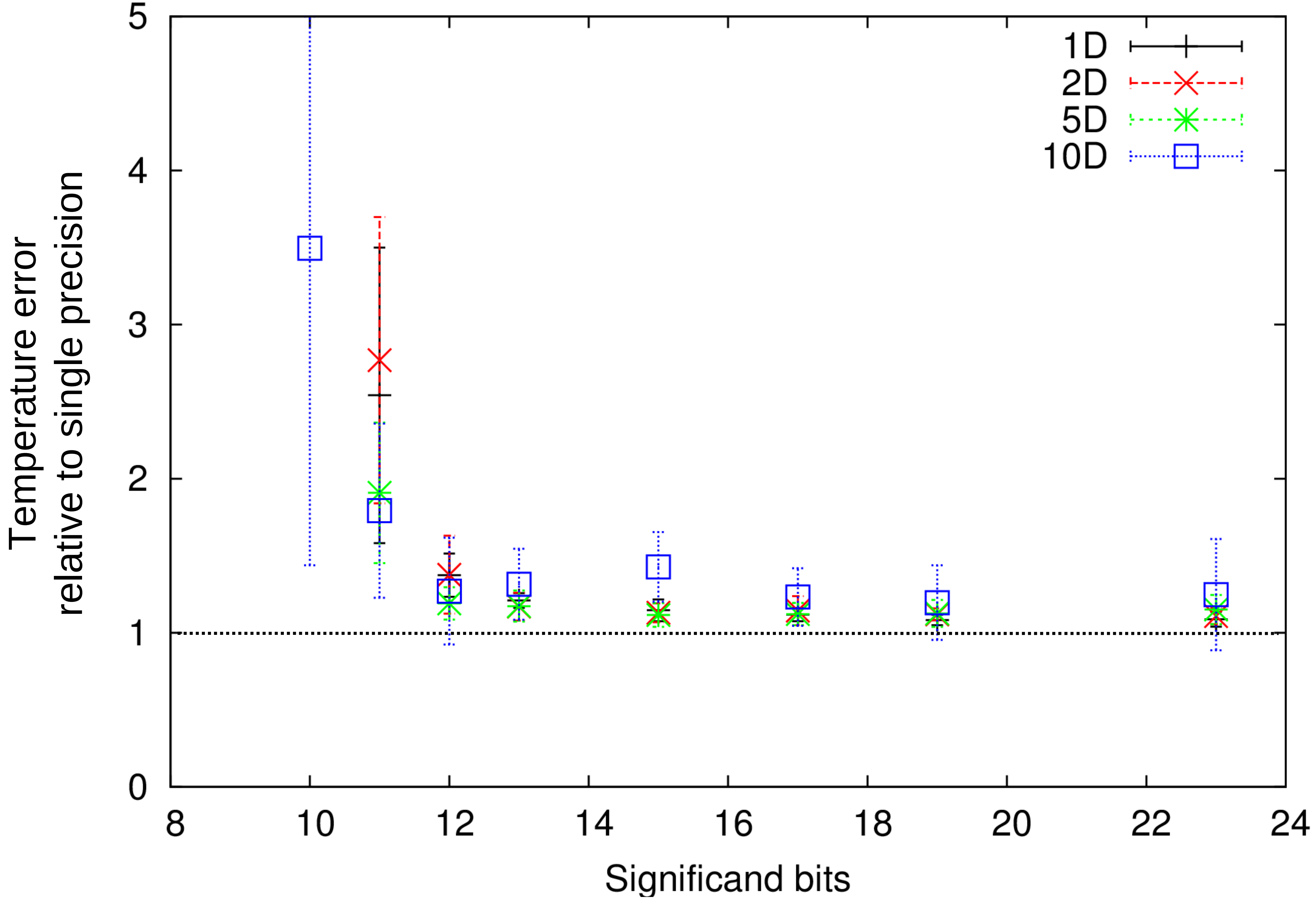
Proportion of significant grid points 7.7%

High resolution tests

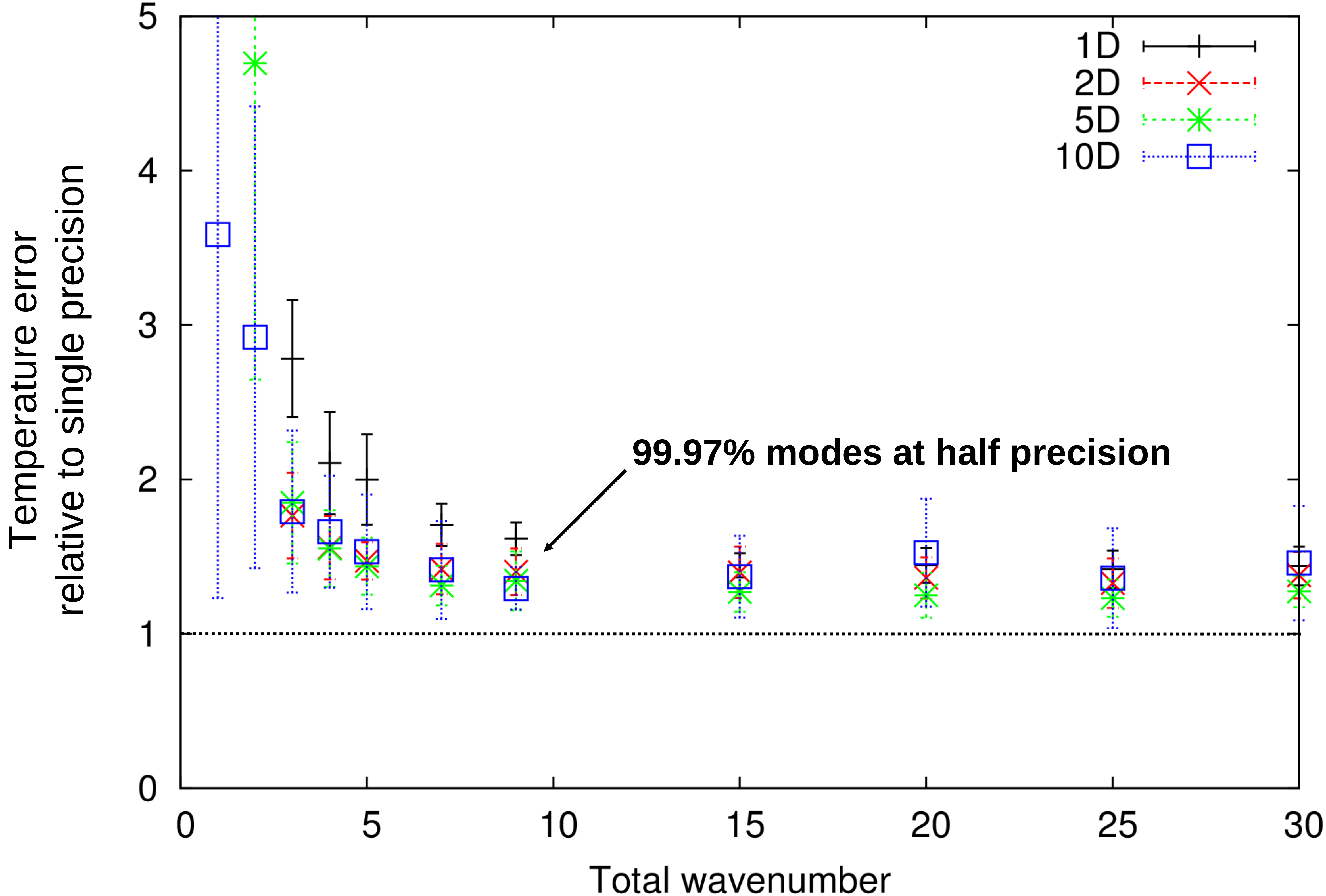
- T511: ~40km grid spacing (compare with ~20km UKMO/ECMWF ensembles).

- 10 start dates between 1999 and 2017.

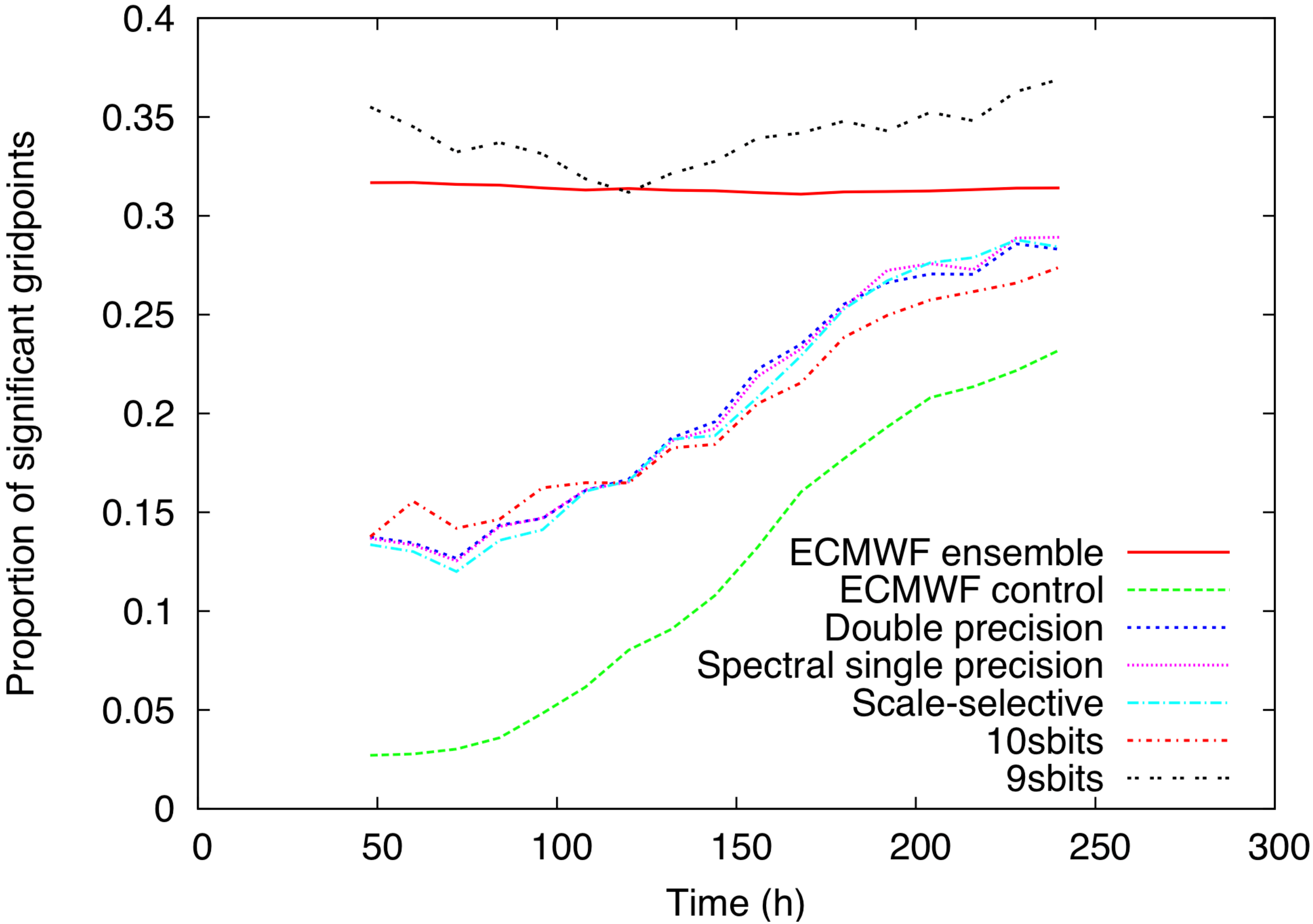
L2-norm — Global precision



Half-precision from which wavenumber?



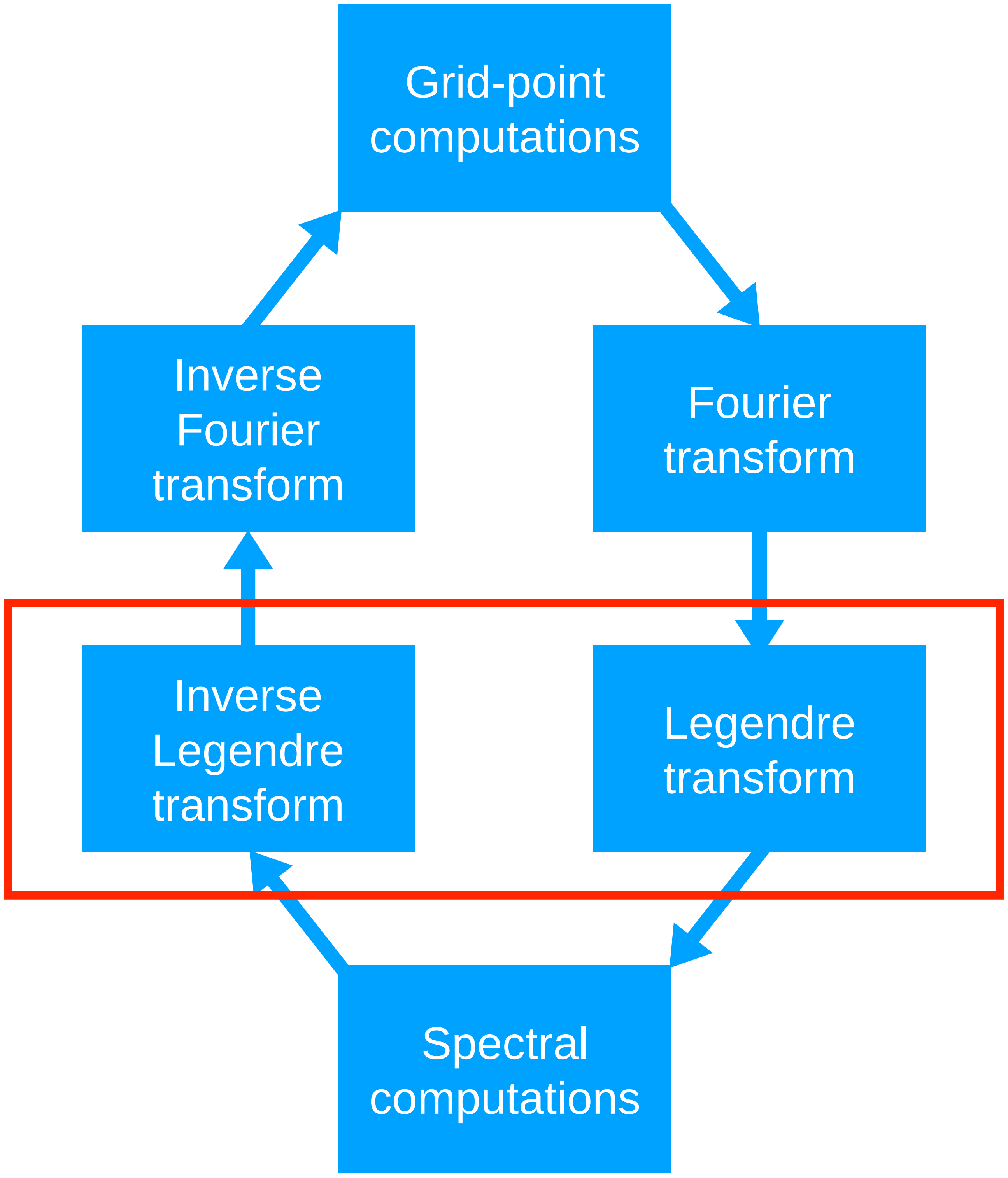
Plausible ensemble member?



Legendre transforms OpenIFS

Sam Hatfield

Spectral dynamical core schematic



Legendre transforms

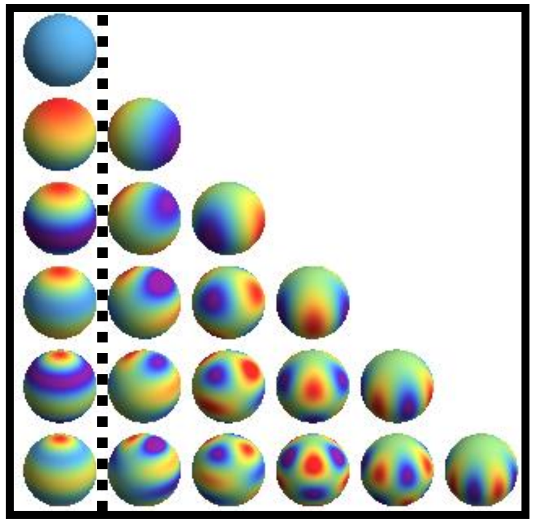
Matrix multiplication - $O(N^3)$ operations.

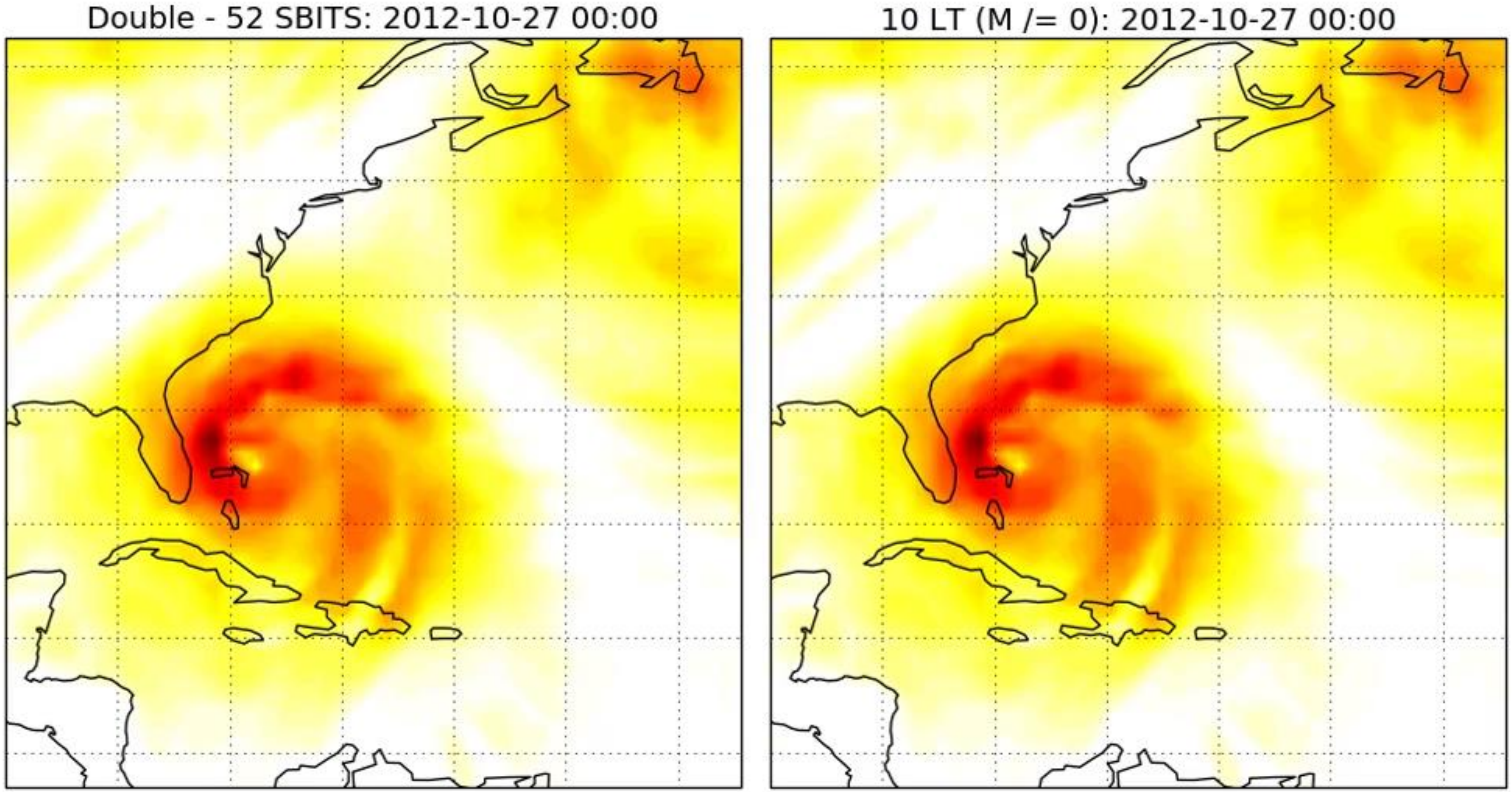
Linear, can rescale variables to fit within dynamic range of half-precision.

How large are the errors introduced?

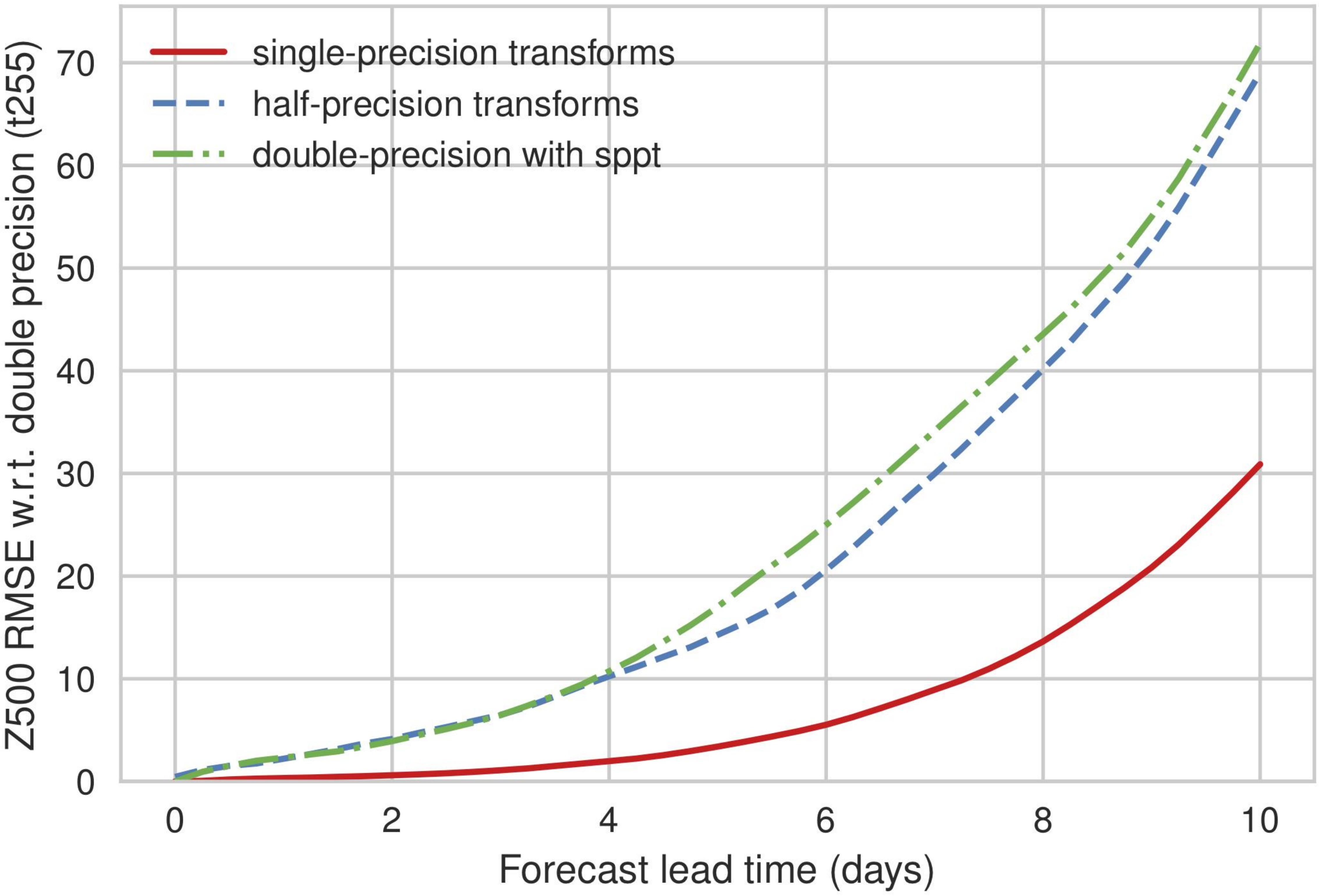
Hurricane Sandy

Double precision for m=0, half-precision for all other values of m.





Deviation from double precision



Physical parameterisations SPEEDY and OpenIFS

Leo Saffin

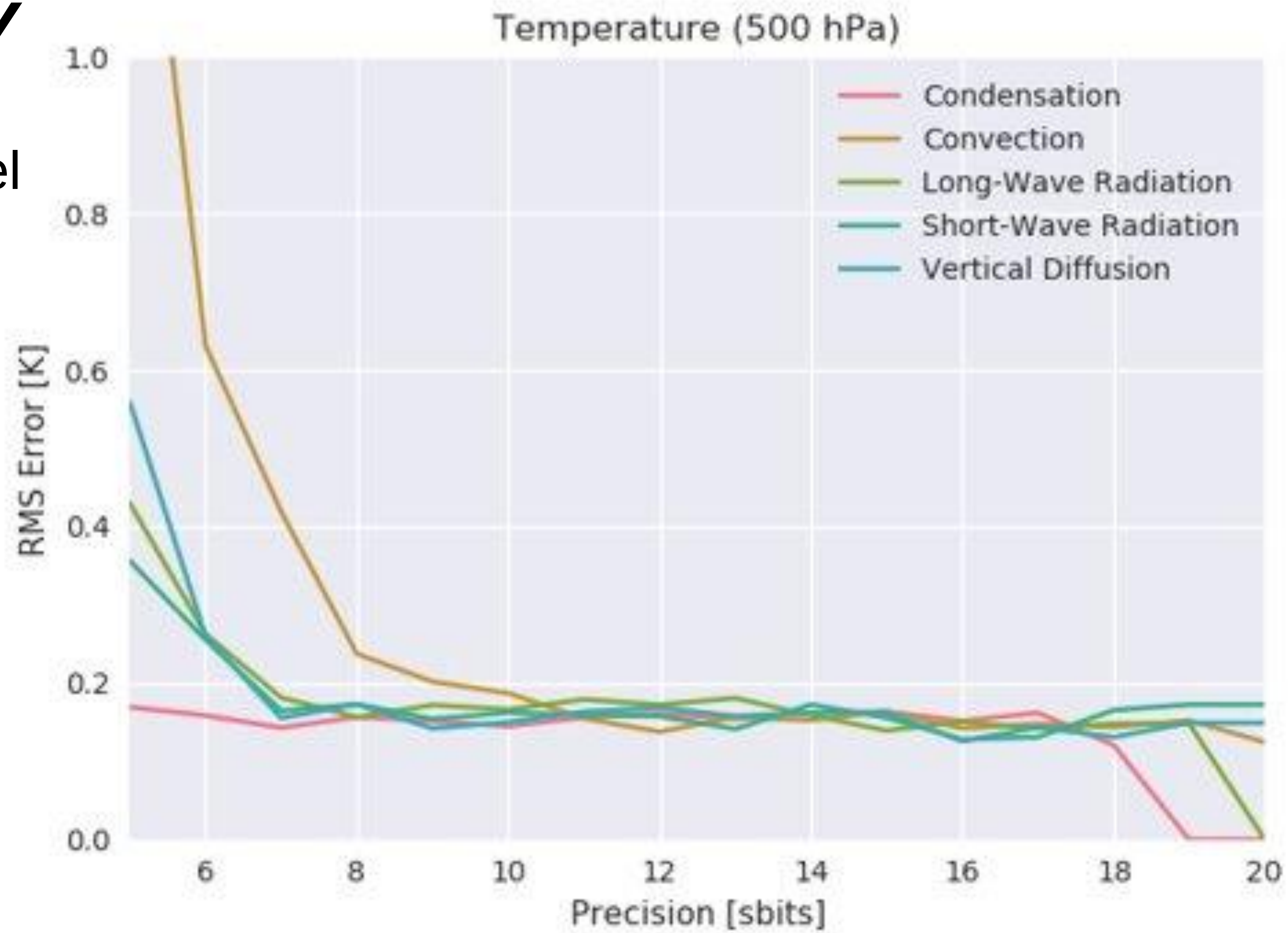
SPEEDY

T30 spectral model

8 vertical levels

Simplified physics

1 week “forecast”

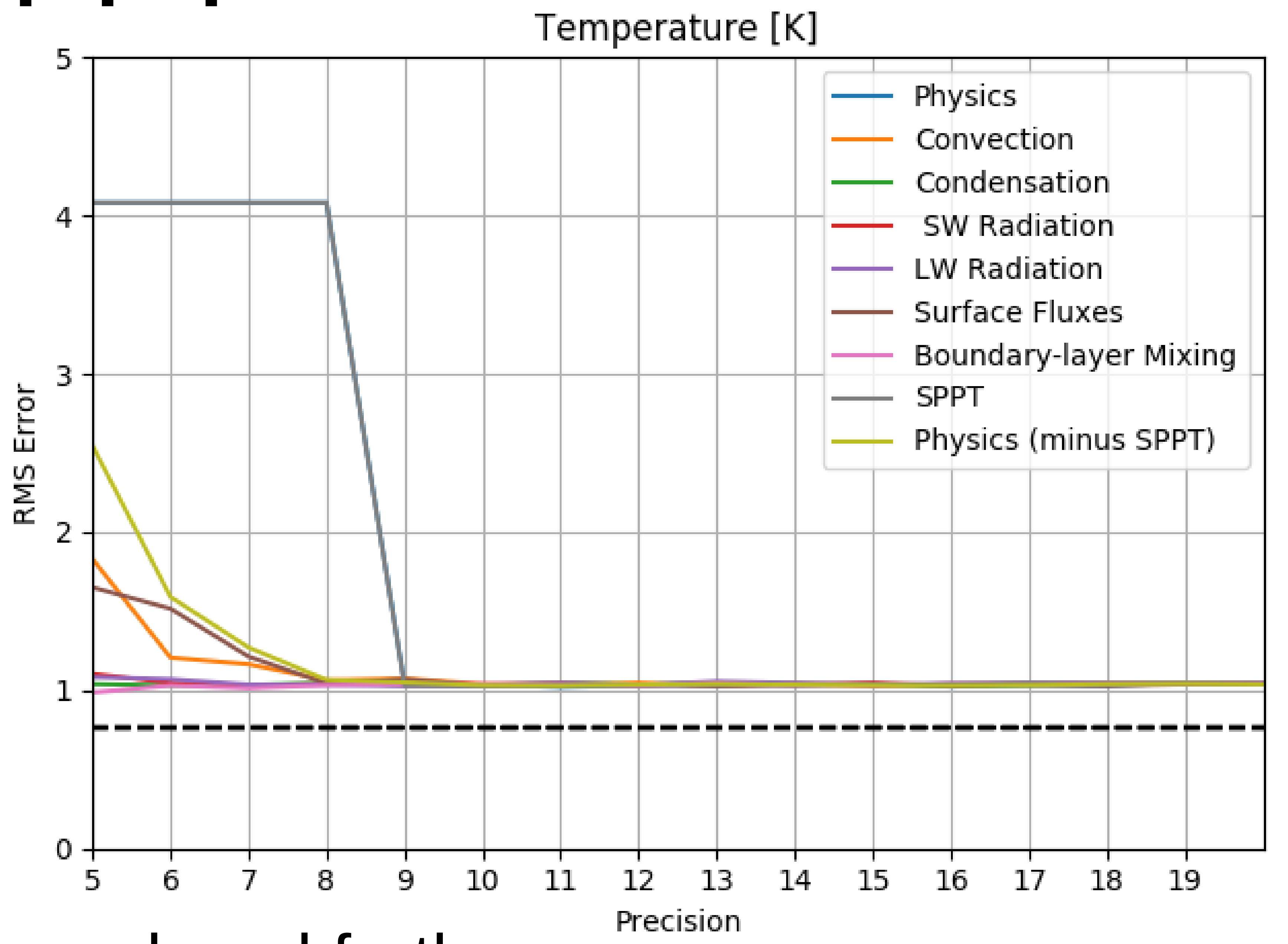


- Half-precision plausible for many schemes.

SPEEDY-SPPT

Introduce Stochastic Perturbation of Parameterisation Tendencies (SPPT) to Speedy

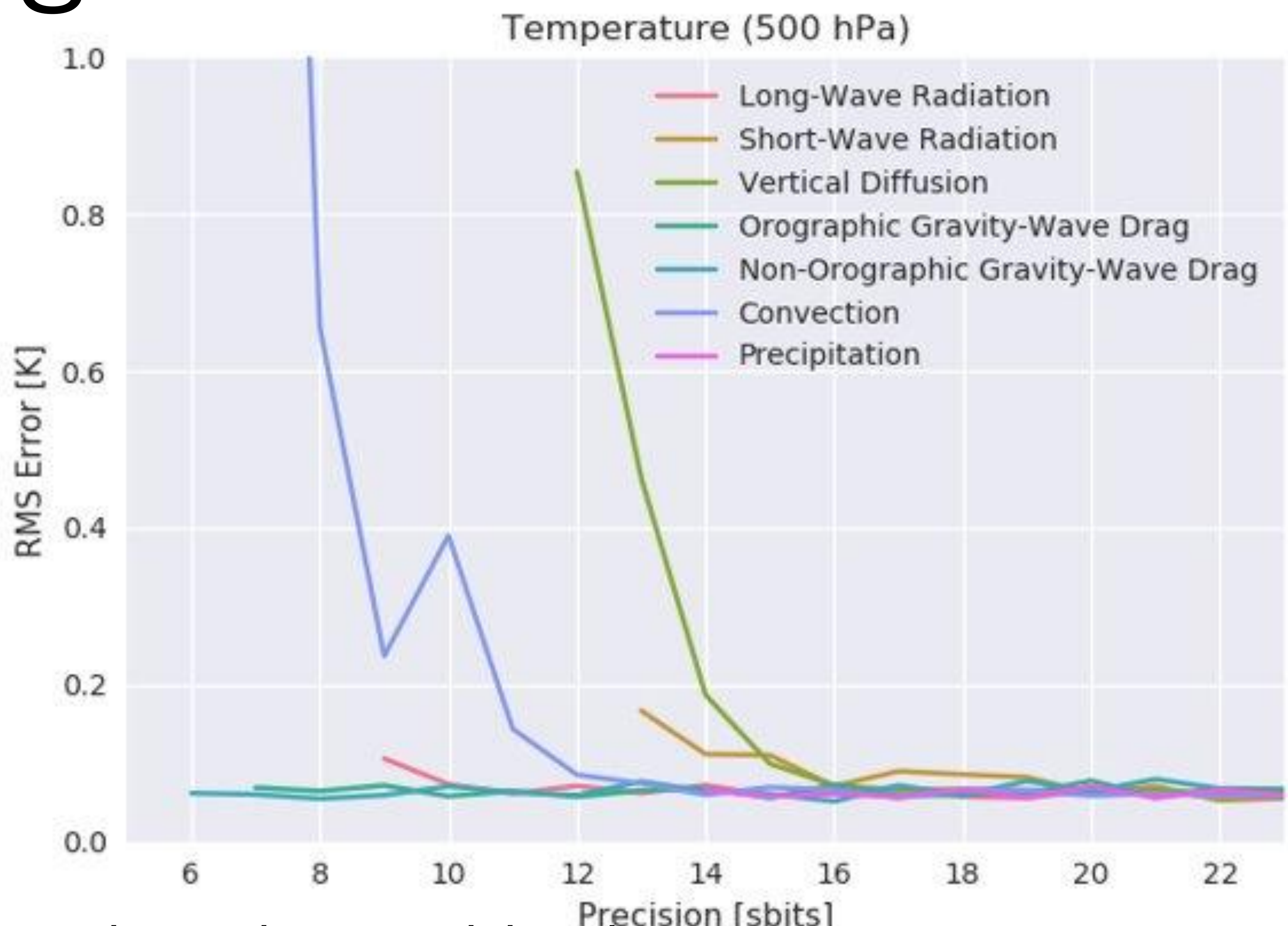
Equivalent scheme to IFS.



- Allows precision to be reduced further.
- SPPT scheme first to fail! Caused by spectral transforms involved in SPPT.

OpenIFS

T21, 60 levels
1-day forecast



- More scheme dependent precision than SPEEDY.
- Initial tests at T159 fail at similar precisions or higher.
- Now testing with SPPT, higher resolution and longer runs.

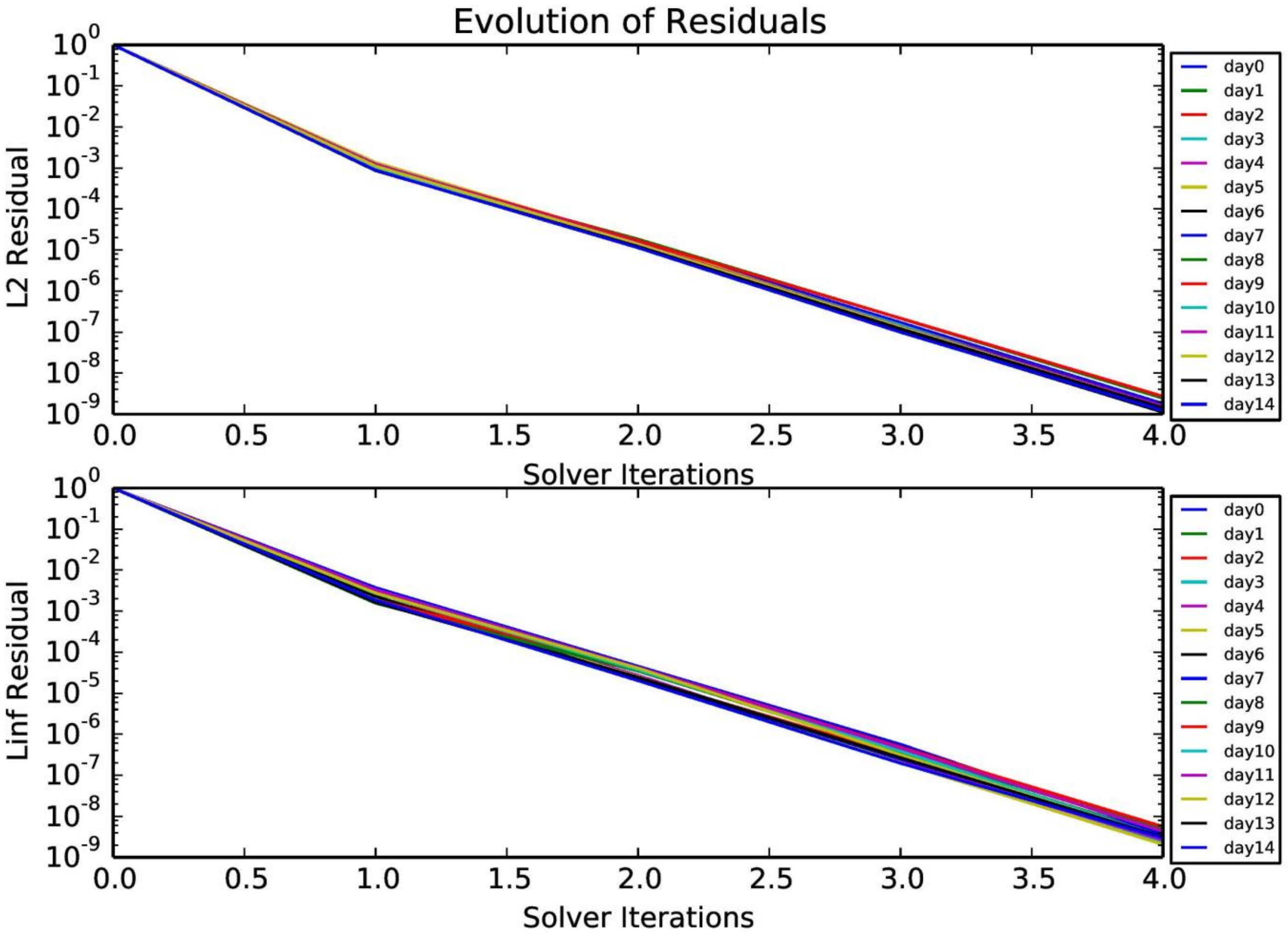
And more!

Preconditioning linear solvers

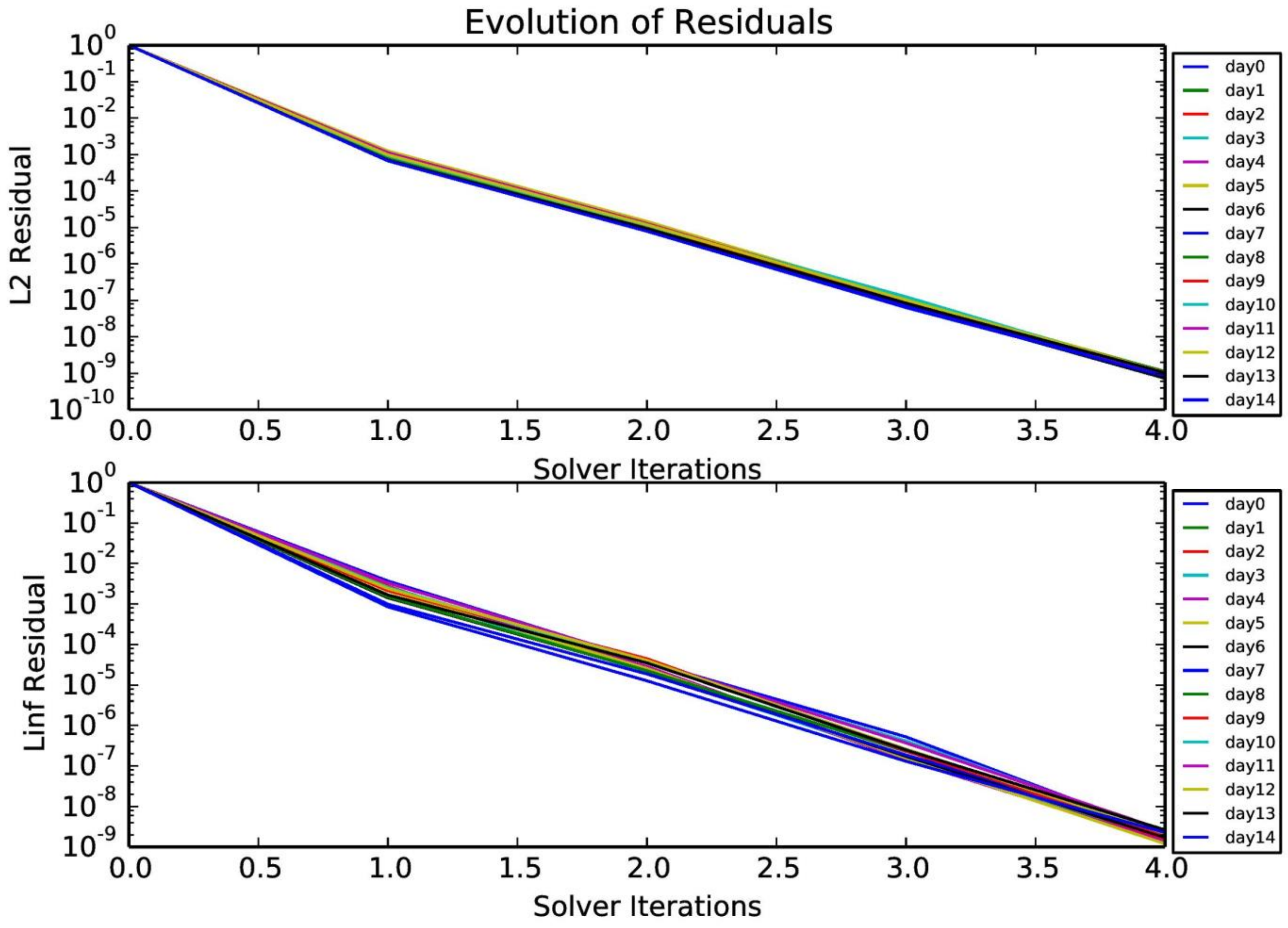
Jan Ackmann

Rossby-Haurwitz wave - MPDATA timestepping scheme

10 significant bits



Double precision

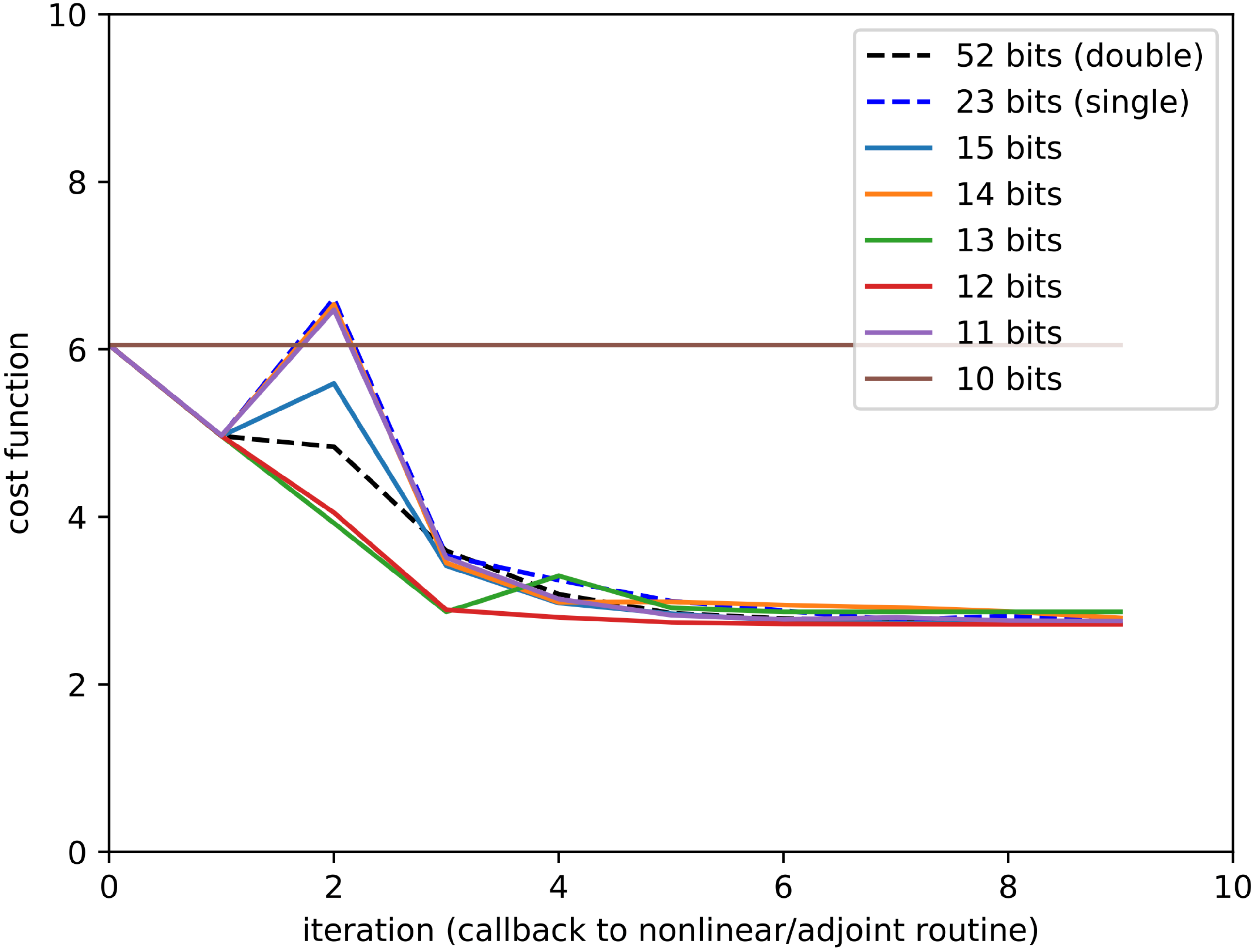


Retain high precision at the poles

Work with Piotr Smolarkiewicz

Adjoint-based minimisation in MITGCM

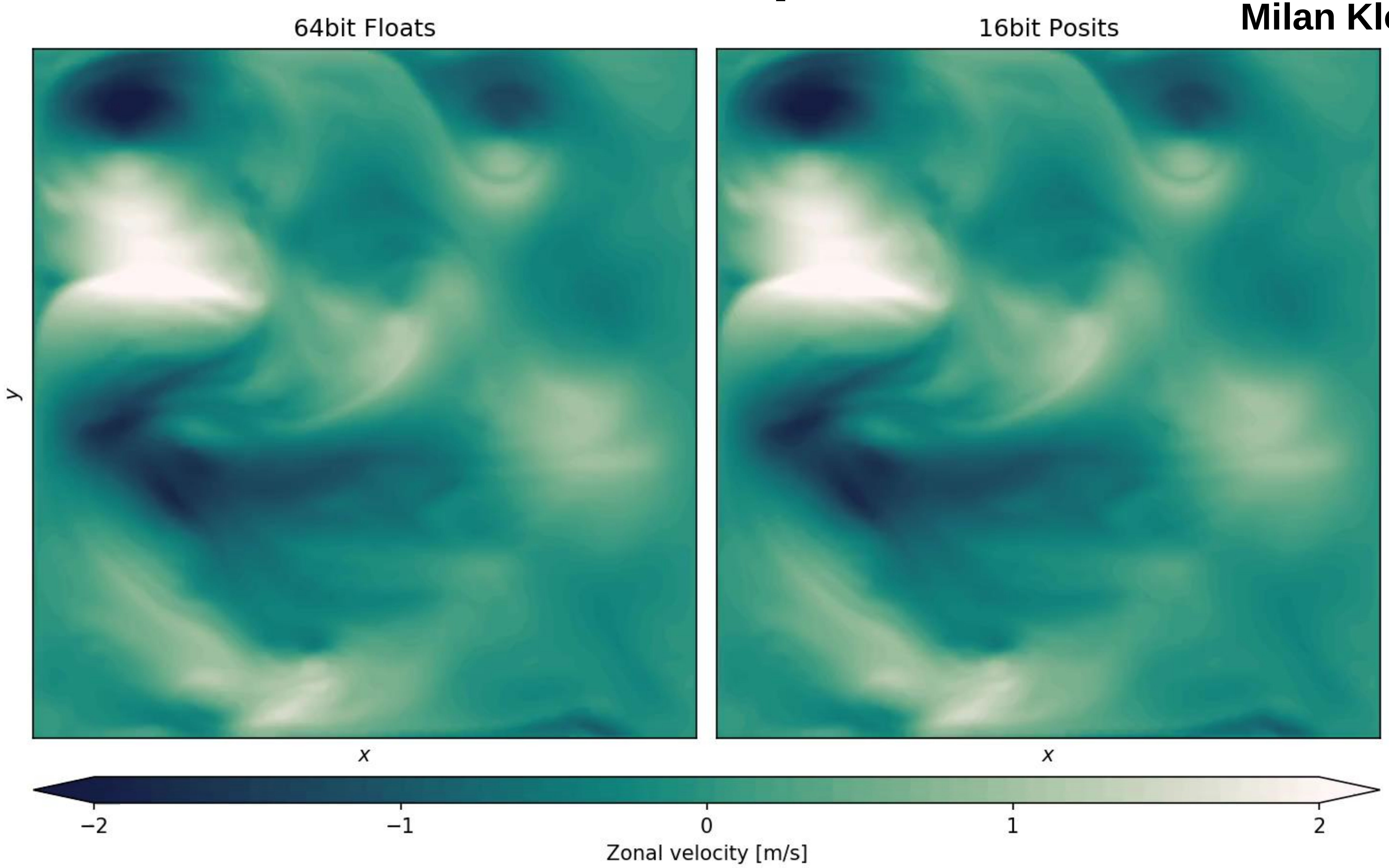
Andrew McRae



Shallow water simulation using floats and half-precision posits.

Alternate number types: better for low precision?

Milan Kloever



Shallow water simulation using floats and half-precision posits.

Take away?

- Precision can generally be reduced below single…
- …but not all fields/computations to half-precision.
- Can we do more to be guided by knowledge of uncertainty?
- Doing this job completely will take man-hours, but save computer costs (or allow higher resolutions). Overall savings?