Greater Accuracy with Less Precision.

A new paradigm for weather and climate prediction.

Tim Palmer

University of Oxford

Oxford Physics

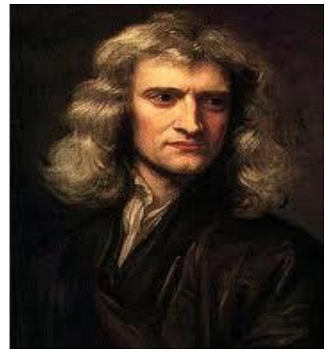
OXFORD MARTIN SCHOOL

Comprehensive weather/climate models play an important role in modern society

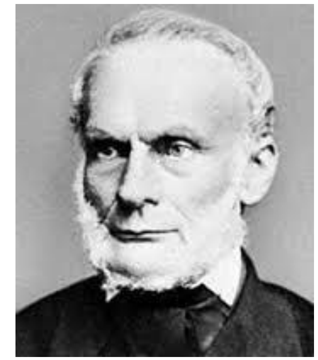
- Forecasting extreme weather events (e.g. for Disaster Risk Reduction)
- To provide estimates of future global climate – key scientific input on climate mitigation (decarbonising the world economy)
- To provide guidance on infrastructure investment for regional climate adaptation
- To foresee regional consequences of geoengineering proposals (“Plan B”)

Comprehensive Earth-System models are based on the laws of physics eg

$$E = \hbar W$$



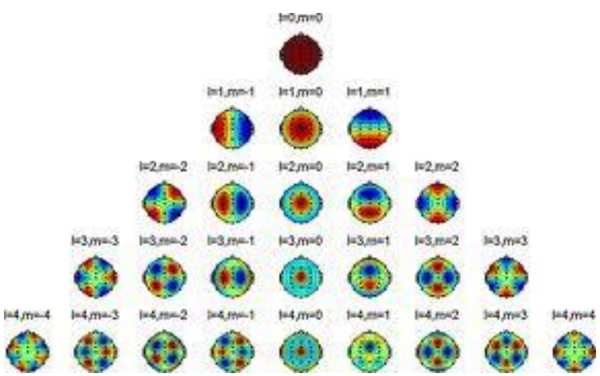
$$\mathbf{F} = m\mathbf{a}$$



$$dQ = TdS$$

$$r\left(\frac{\partial}{\partial t} + \mathbf{u}.\nabla\right)\mathbf{u} = r\mathbf{g} - \nabla p + m\nabla^2\mathbf{u}$$

$$\zeta = \sum_{m}\sum_{l}^{\infty} \zeta_{ml} e^{im\lambda} P_l^m(\phi)$$



$$\rho\left(\frac{\partial}{\partial t}+\mathbf{u}.\nabla\right)\mathbf{u} = \rho\mathbf{g} - \nabla p + \mu\nabla^2\mathbf{u}$$

↑

Unpacks into billions of individual equations, describing scales of motion from planetary scales to microscopic scales.

Even the world's biggest computers aren't big enough to represent all scales of motion in the atmosphere down to viscous scales

10,000km

10-100km

millions of equations

Simplified closure formulae to approximate processes (eg clouds) that the simulator can't resolve. Some improvements if these closure schemes are formulated stochastically.

$$\rho\left(\frac{\partial}{\partial t}+\mathbf{u}.\nabla\right)\mathbf{u}=\rho\mathbf{g}-\nabla p+\nu\ \nabla^2\mathbf{u}$$

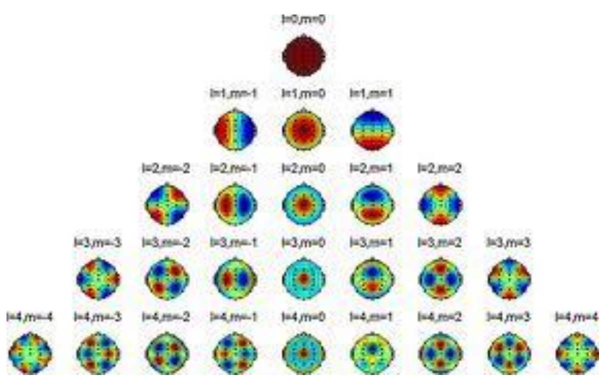
The Canonical Numerical Ansatz

Resolved scales ←

→ Unresolved scales

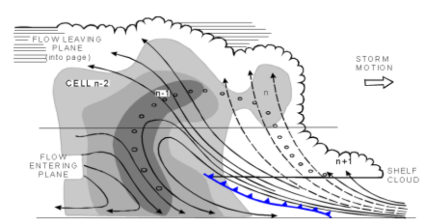
Dynamical Core

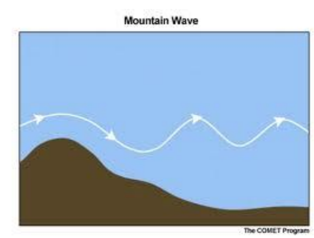
$$\zeta=\sum_{m}\sum_{l}^{\infty}\zeta_{ml}e^{im\lambda}P_l^m(\phi)$$



Parametrisations

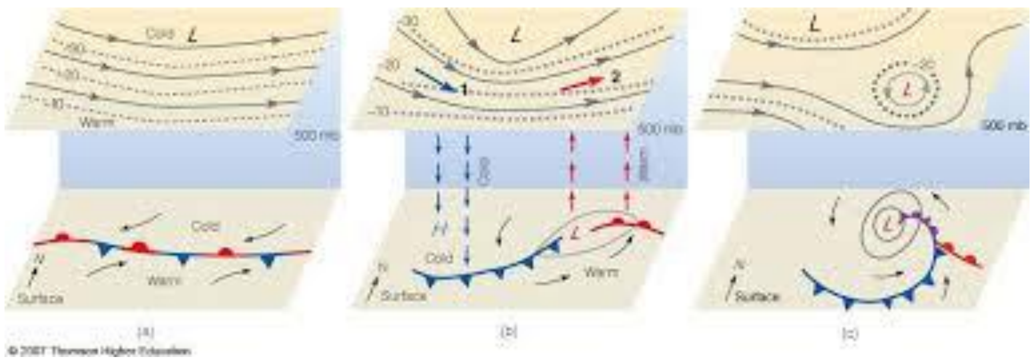
$$P\left(X_{\mathrm{Tr}};a\right)$$





$$D=P$$

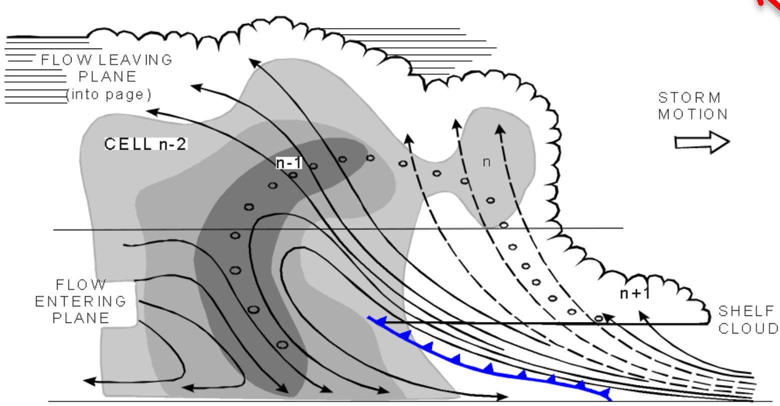
Truncation scale?



O(1000km)

O(100km) Climate

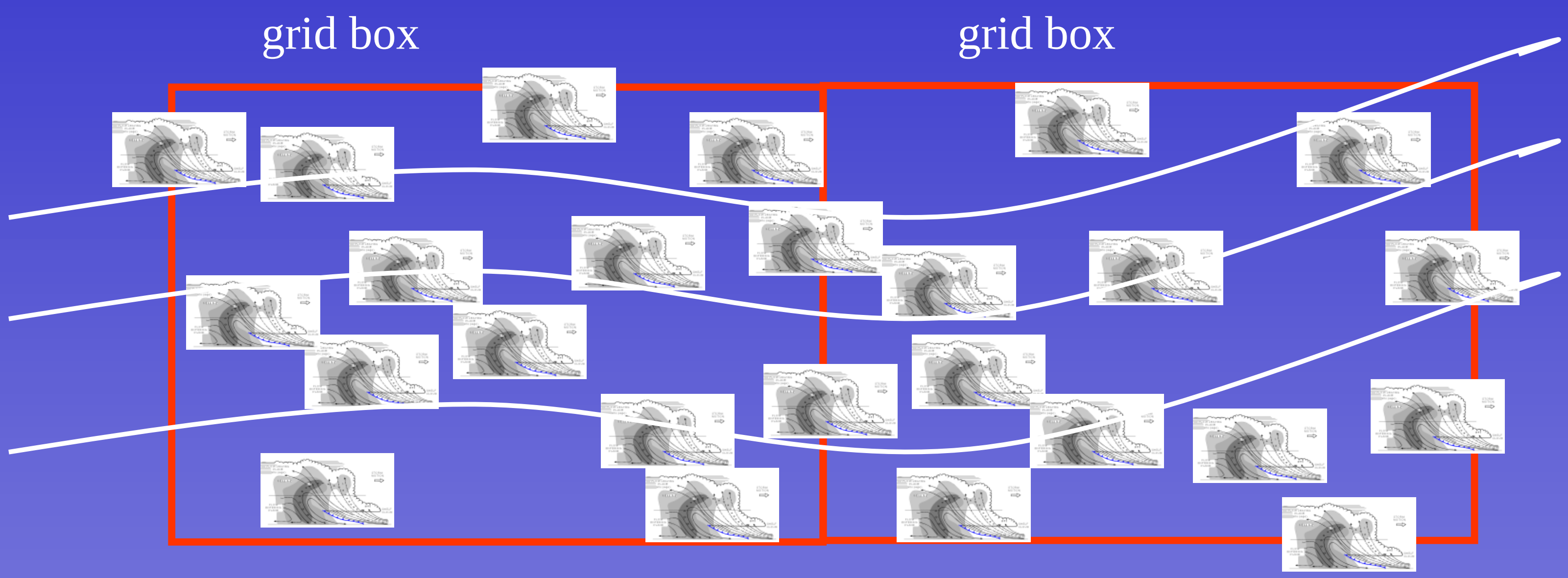
O(10km) NWP

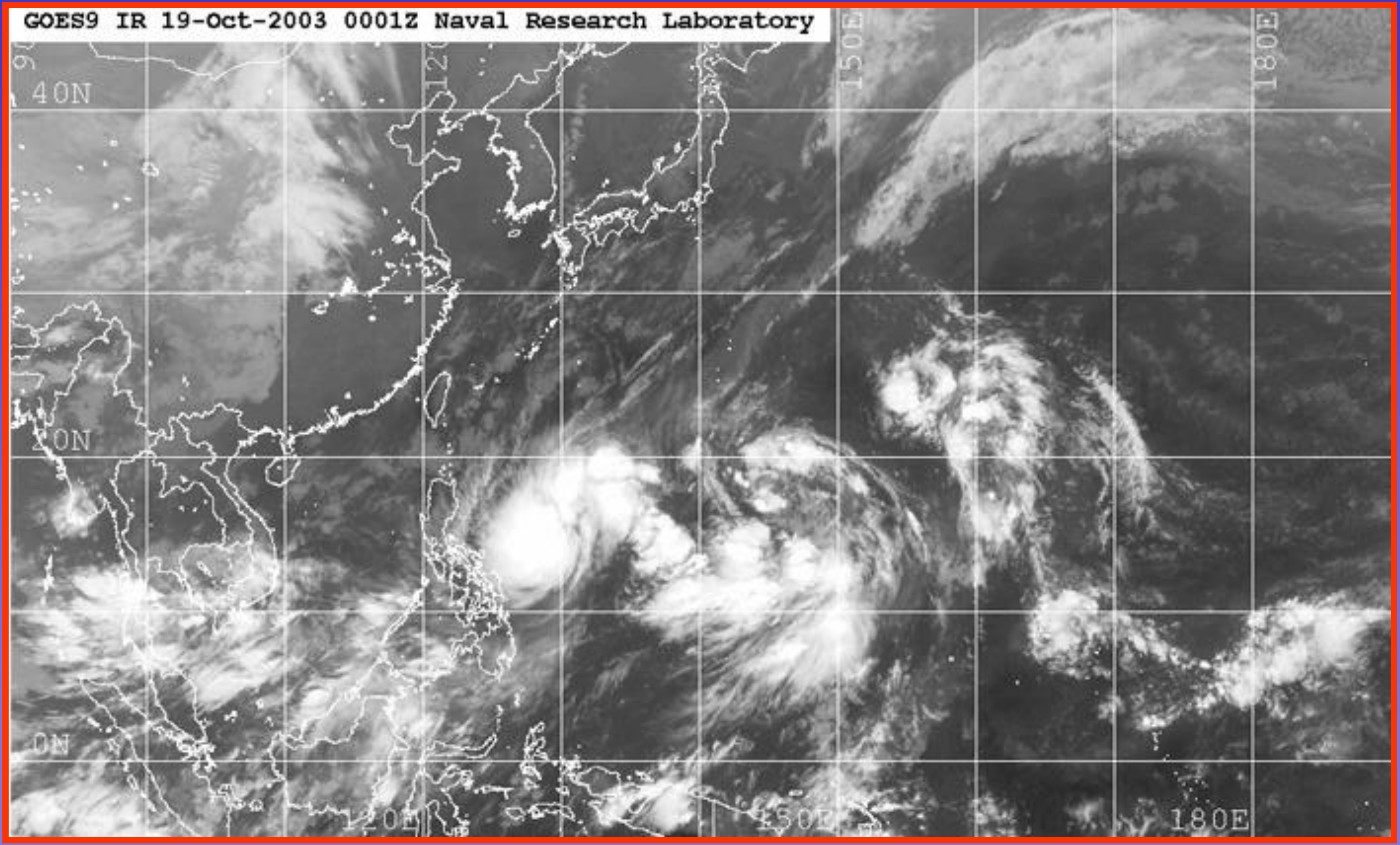


O(1km)

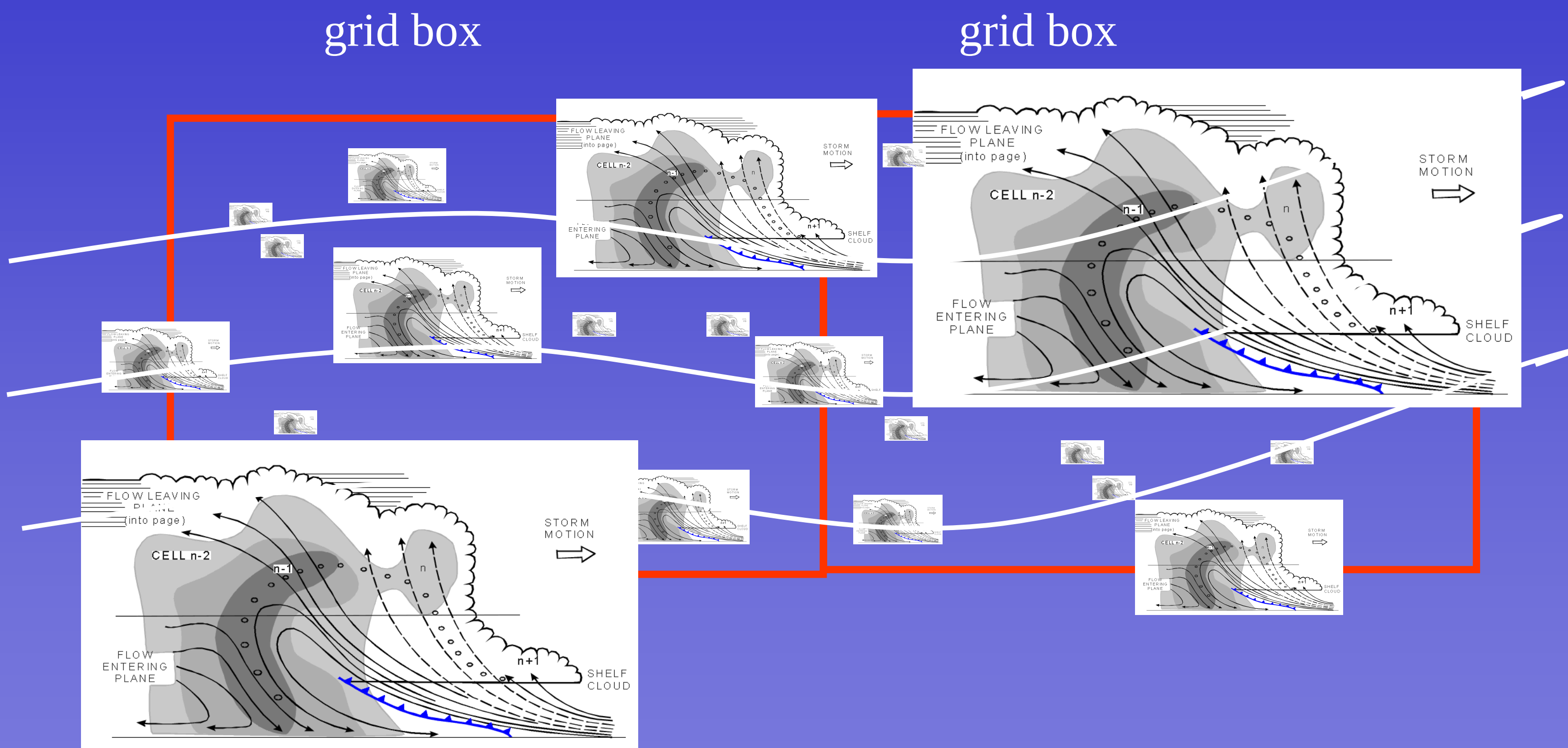
8-9 orders of magnitude above viscous scale

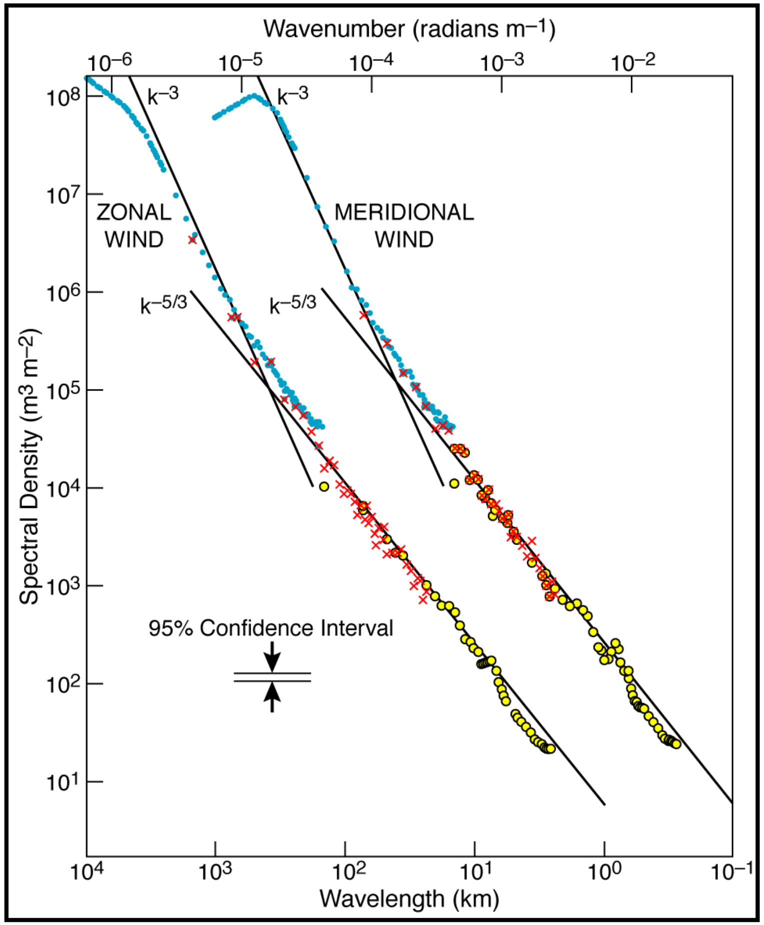
Convective parametrisation OK if the world looks like this…




GOES9 IR 19-Oct-2003 0001Z Naval Research Laboratory
40N
20N
0N
120E
150E
180E

The reality of the situation

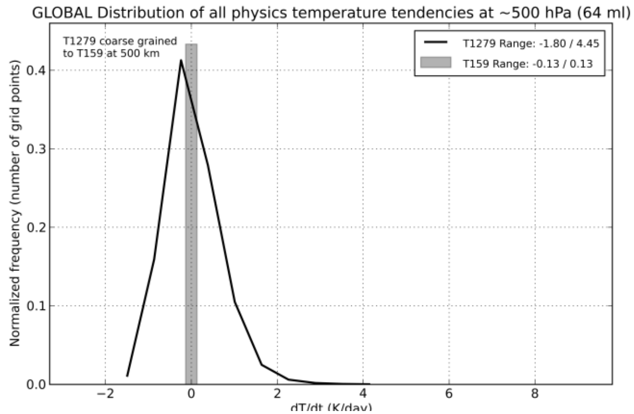




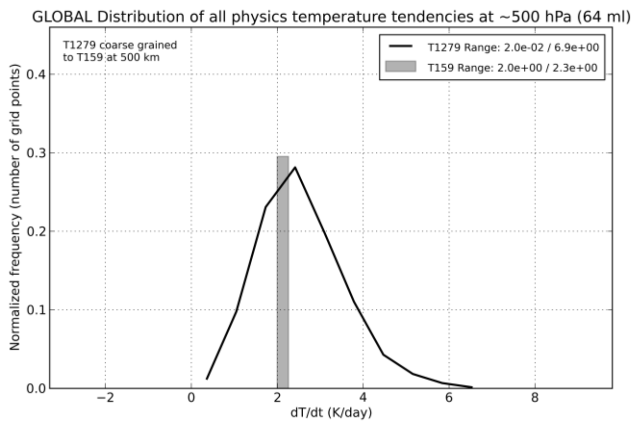
(Nastrom and Gage, 1985)

Coarse-graining
(Shutts and Palmer, 2007)

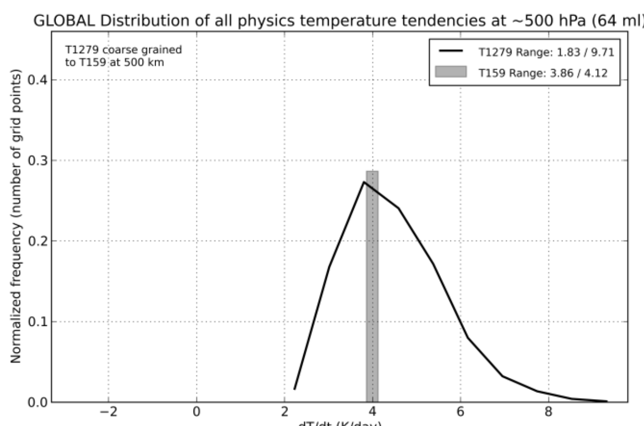
Small tendency



Medium tendency



Large tendency



Callado-Palarès and Shutts,Phil Trans 2014

Assume T1279 (16km) model = “truth”.

Assume T159 coarse-grain “model” grid.

Bar= Subset of total temperature parametrisation tendencies driven by T1279 fields coarse-grained to T159.

Curve= Corresponding “true” sub-T159-scale tendency.

Ie when the parametrisations think the sub-grid pdf is a thin hat function, the reality is a much broader pdf.

The standard deviation increases with parametrised tendency – consistent with multiplicative noise stochastic schemes.

Does it matter that we can't resolve convective cloud systems?

http://www.newscientist.com - Soaring global warming 'can't be r

File Edit View Go Bookmarks Tools Window Help

NewScientist.com

HOME | NEWS | EXPLORE BY SUBJECT | LAST WORD | SUBSCRIBE | SEARCH | ARCHIVE | RSS | JOBS

Soaring global warming 'can't be ruled out'

Click to Print

19:03 26 January 2005
NewScientist.com news service
Jenny Hogan

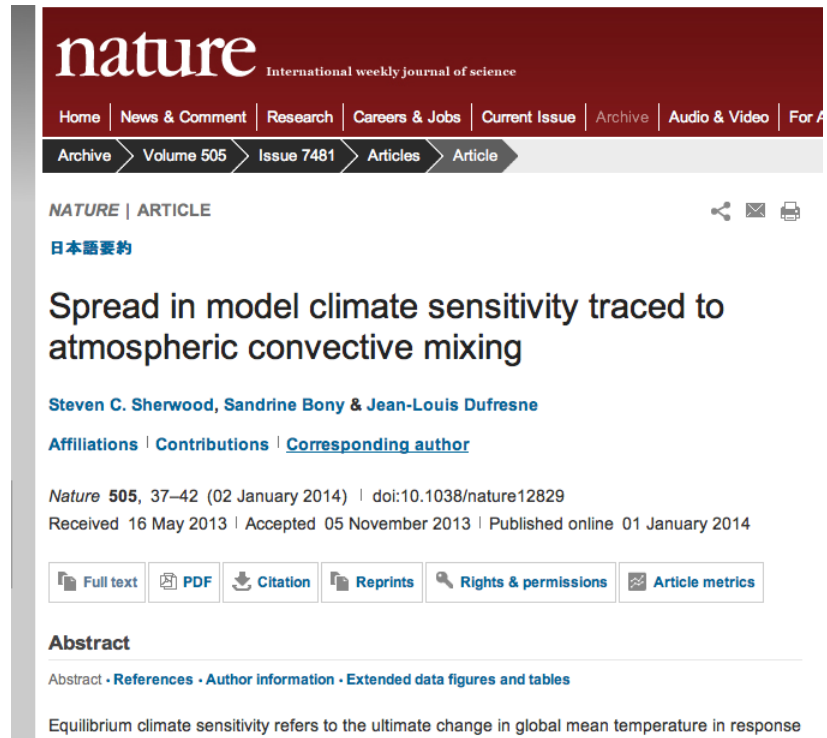
The Earth may be much more sensitive to global warming than previously thought, according to the first results from a massive distributed-computing project.

The project tested thousands of climate models and found that some produced a world that warmed by a huge 11.5°C when atmospheric carbon dioxide concentrations reached the levels expected to be seen later this century.

This extreme result is surprising because it lies far outside the 1.4°C to 4.5°C range predicted by the Intergovernmental Panel on Climate Change (IPCC) for the same CO_2-level increase - a doubling of CO_2 concentration from pre-industrial times. But it is possible the IPCC range was wrong because its estimate is based on just a handful of different computer models.

Enlarge image

The climate modelling software divides the Earth's surface into boxes hundreds of kilometres square (image: Climateprediction.net)



nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For A

Archive > Volume 505 > Issue 7481 > Articles > Article

NATURE | ARTICLE

日本語要約

Spread in model climate sensitivity traced to atmospheric convective mixing

Steven C. Sherwood, Sandrine Bony & Jean-Louis Dufresne

Affiliations | Contributions | Corresponding author

Nature **505**, 37–42 (02 January 2014) | doi:10.1038/nature12829
Received 16 May 2013 | Accepted 05 November 2013 | Published online 01 January 2014

Full text | PDF | Citation | Reprints | Rights & permissions | Article metrics

Abstract

Abstract • References • Author information • Extended data figures and tables

Equilibrium climate sensitivity refers to the ultimate change in global mean temperature in response

Yes!

How do we get to a 1km- 100m global grid?

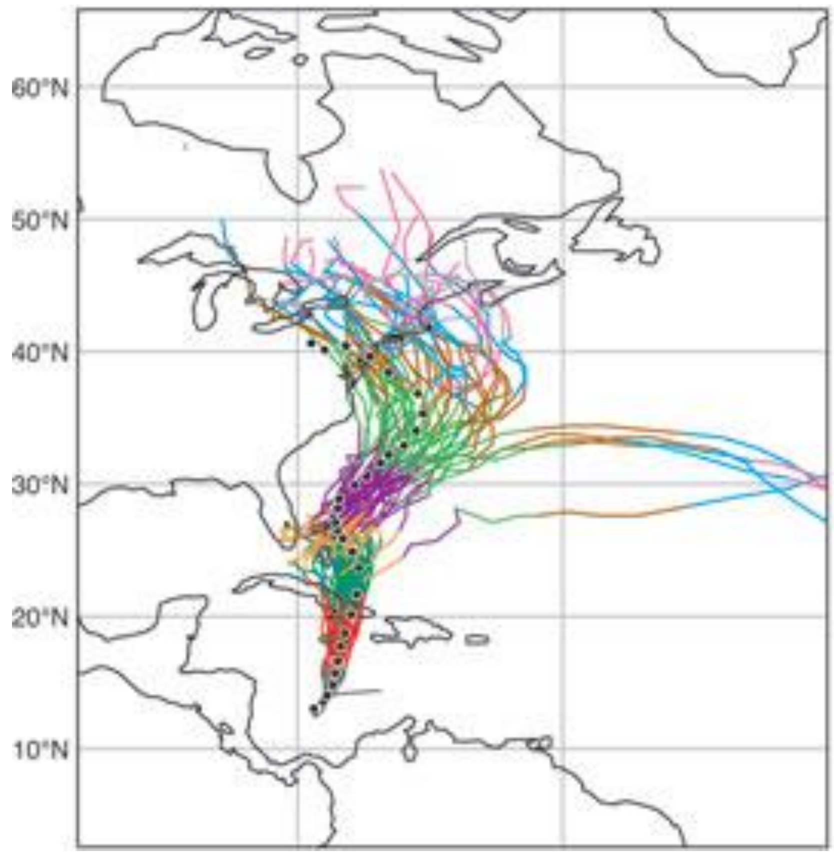
- Wait! >20yrs? Even then maybe can’t afford the power costs. Can society afford to wait that long?
- Fund a “Climate CERN” to house a prototype multi-exaflop computer dedicated to climate.
- Revisit the way we go about solving the equations numerically amd embrace the concept of approximate computing!

(Nb 2 and 3 are not mutually exclusive!)

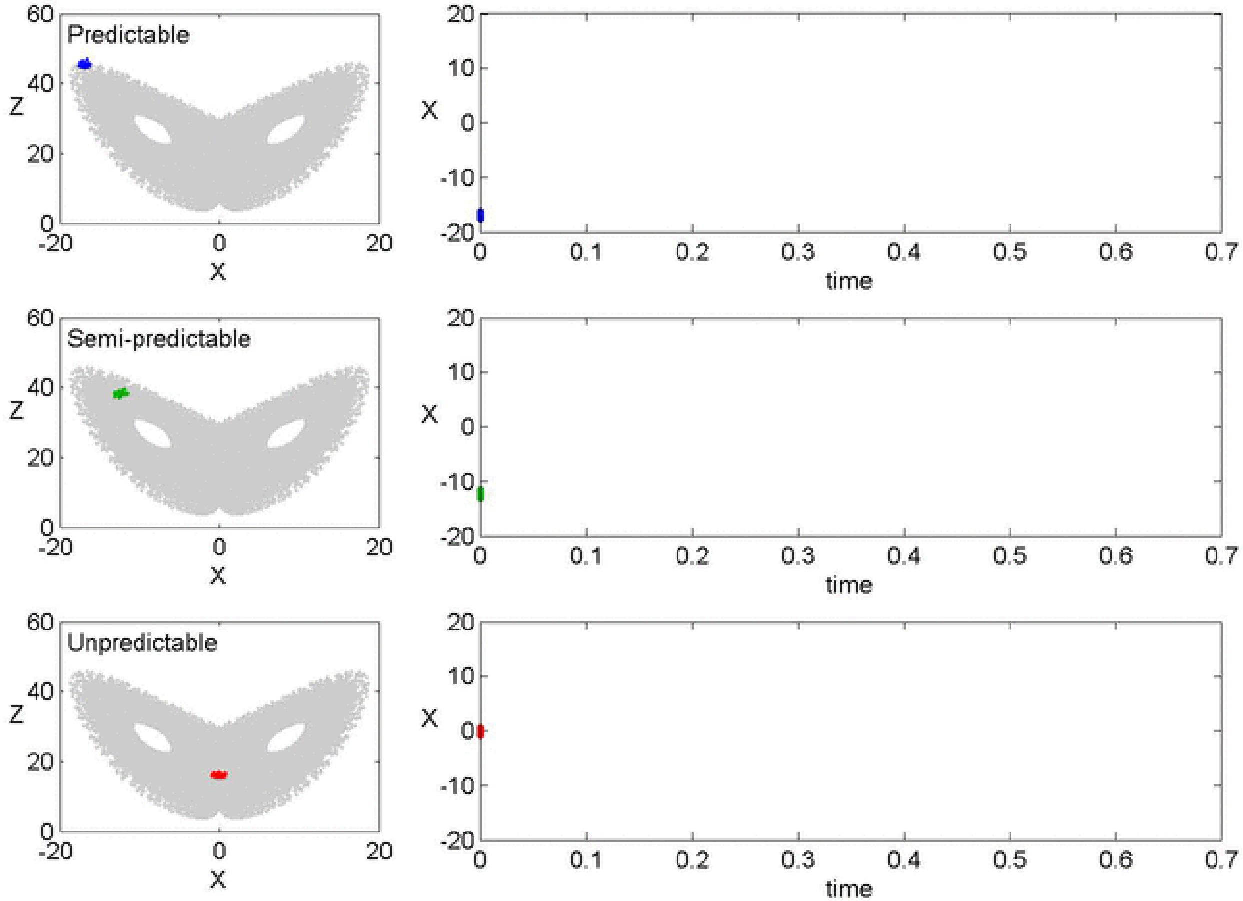

Local effects such as thunderstorms, crucial for predicting global warming, could be simulated by fine-scale global climate models.

Build high-resolution global climate models

International supercomputing centres dedicated to climate prediction are needed to reduce uncertainties in global warming, says Tim Palmer.



SANDY



$$\dot{X} = -\sigma X + \sigma Y$$

$$\dot{Y} = -XZ + rX - Y$$

$$\dot{Z} = XY - bZ$$

$$\dot{X} = \mathrm{F}[X]$$

$$d\dot{X} = \frac{\mathrm{dF}}{\mathrm{dX}} \; dX$$


TECHNICAL MEMORANDUM

598

Stochastic Parametrization and Model Uncertainty

Palmer, T.N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G.J. Shutts, M. Steinheimer, A. Weisheimer

Research Department

October 8, 2009

This paper has not been published and should be regarded as an Internal Report from ECMWF. Permission to quote from it should be obtained from the ECMWF.

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

- Multiplicative Noise

$P \rightarrow (1+\sigma)P$

- Improved reliability of probabilistic weather forecasts
- Reduced systematic errors, e.g. warm pool convection, wind stress, MJO (e.g. Weisheimer et al, 2014)

Experiments with the Lorenz '96 System

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

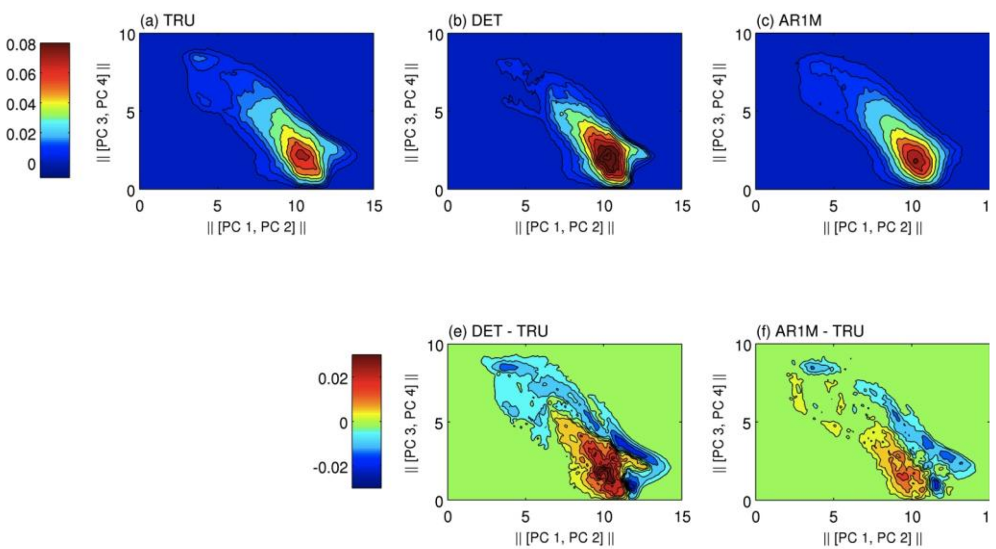
$$\frac{dY_j}{dt} = -cbY_{j+1}\left(Y_{j+2} - Y_{j-1}\right) - cY_j + \frac{hc}{b} X_{\text{int}[(j-1)/J+1]}$$

Assume Y unresolved

Approximate sub-grid tendency by U

Deterministic parametrisation of Y

Stochastic parametrisation of Y



Deterministic: $U = U_{det}$

Additive: $U = U_{det} + e_{w,r}$

Multiplicative: $U = (1+e_r)\, U_{det}$

Where:

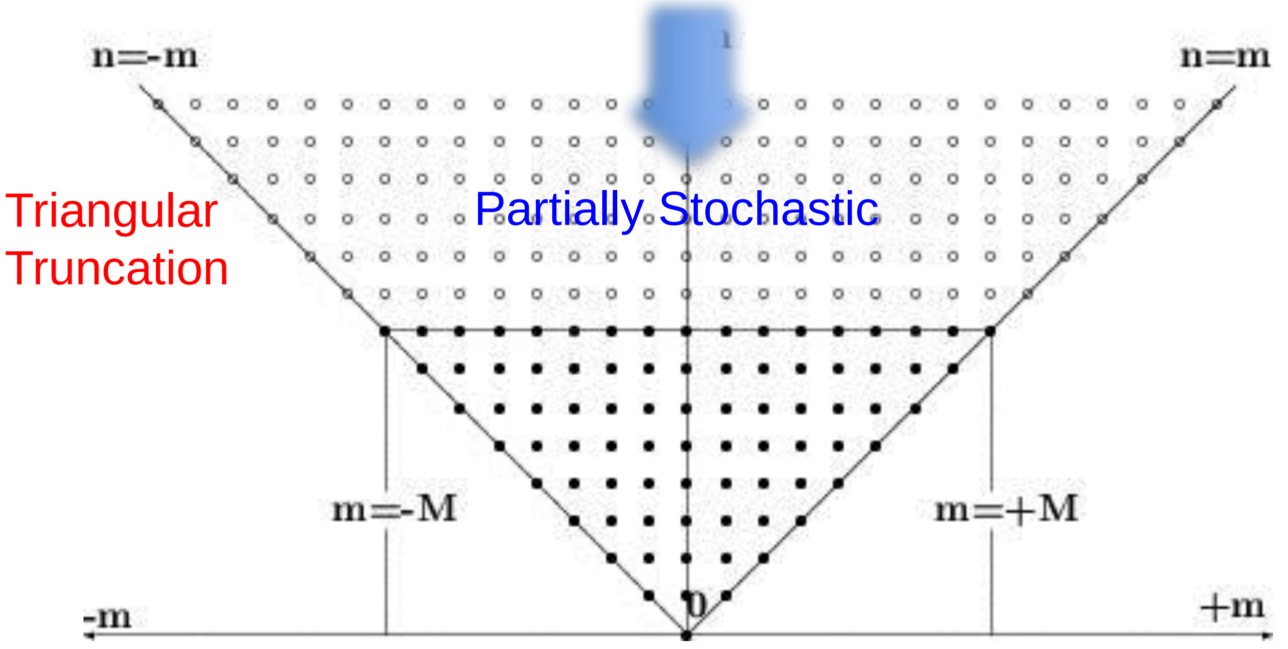
U_{det} = cubic polynomial in X

$e_{w,r}$ = white / red noise

Fit parameters from full model

Christensen et al, Climate Dynamics 2014

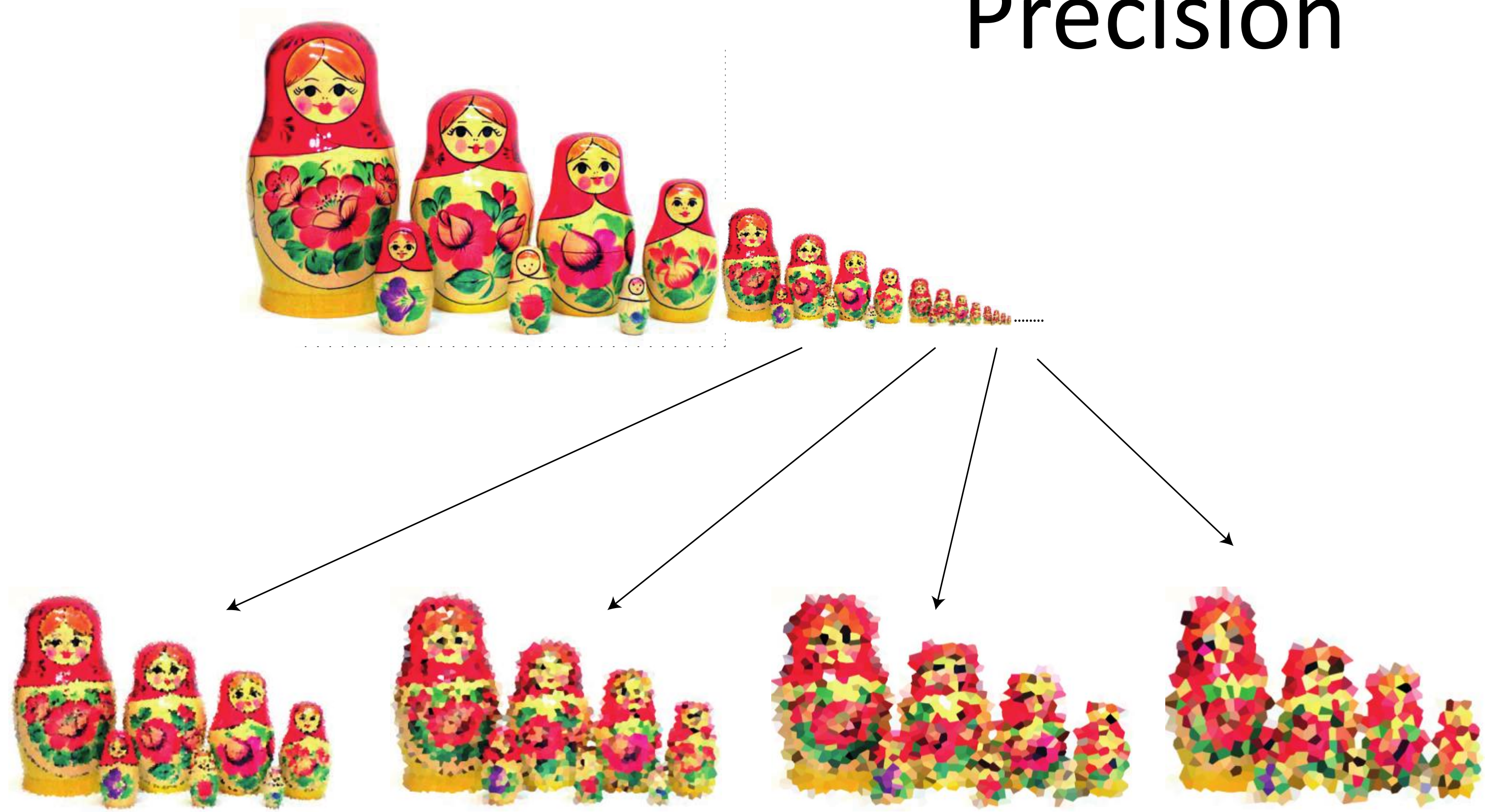
Stochastic Parametrisation

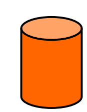


If parametrisation is partially stochastic, are we "over-engineering" our models (parametrisations, dynamical core) by using double precision bit-reproducible computations throughout?

Are we making inefficient use of computing resources that could otherwise be used to increase resolution?

Greater Accuracy with Less Precision

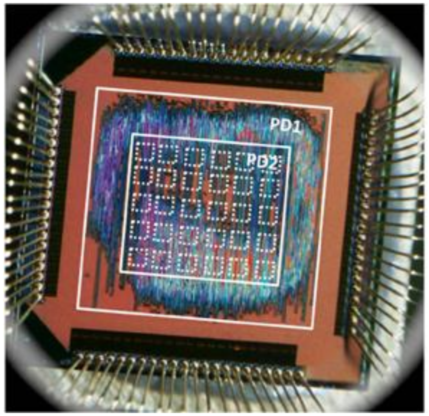


More accurate than but as computationally cheap as 

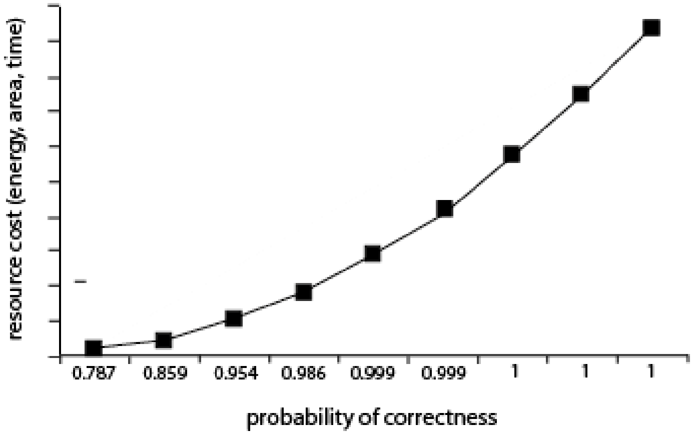
Superefficient inexact chips

http://news.rice.edu/2012/05/17/computing-experts-unveil-superefficient-inexact-chip/

Prototype
Probabilistic
CMOS
Chip



Krishna Palem.
Rice University



The chip that produced the frame with the most errors (right) is about 15 times more efficient in terms of speed, space and energy than the chip that produced the pristine image (left).

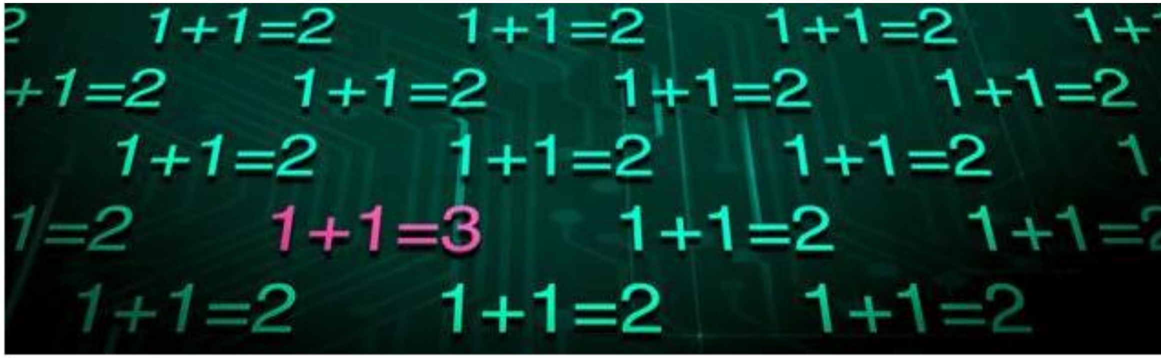


Illustration: Jose-Luis Olivares/MIT

Harnessing error-prone chips

New system would allow programmers to easily trade computational accuracy for energy savings.

Larry Hardesty | MIT News Office
October 30, 2014

▼ Press Inquiries

RELATED

Paper: "Chisel: Reliability- and accuracy-aware optimization of approximate computational kernels"

As transistors get smaller, they also grow less reliable. Increasing their operating voltage can help, but that means a corresponding increase in power consumption.

Experiments with the Lorenz '96 System

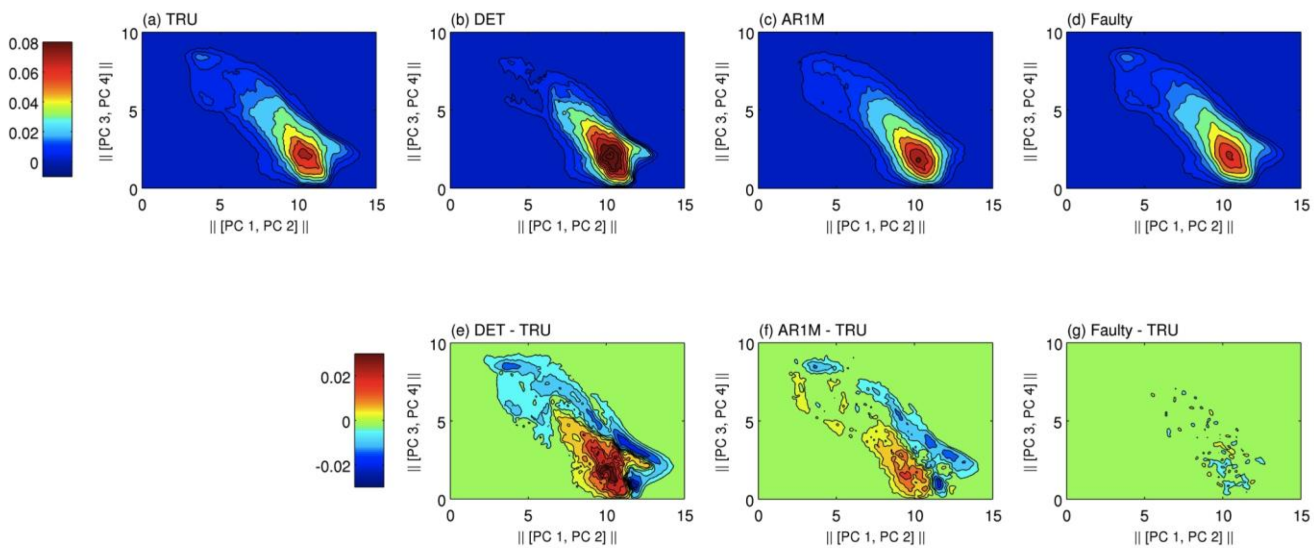
$$\frac{dX_k}{dt} = -X_{k-1}\left(X_{k-2} - X_{k+1}\right) - X_k + F - \frac{hc}{b}\sum_{j=J(k-1)+k}^{kJ} Y_j$$

$$\frac{dY_j}{dt} = -cbY_{j+1}\left(Y_{j+2} - Y_{j-1}\right) - cY_j + \frac{hc}{b} X_{\mathrm{int}[(j-1)/J+1]}$$

Deterministic parametrisation of Y

Stochastic parametrisation of Y

Stochastic Chip Emulator

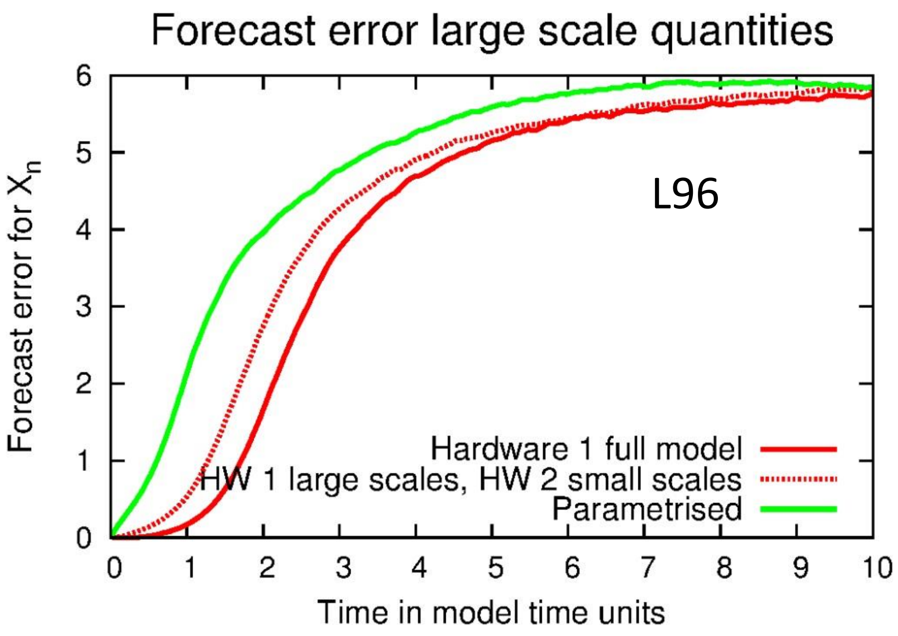


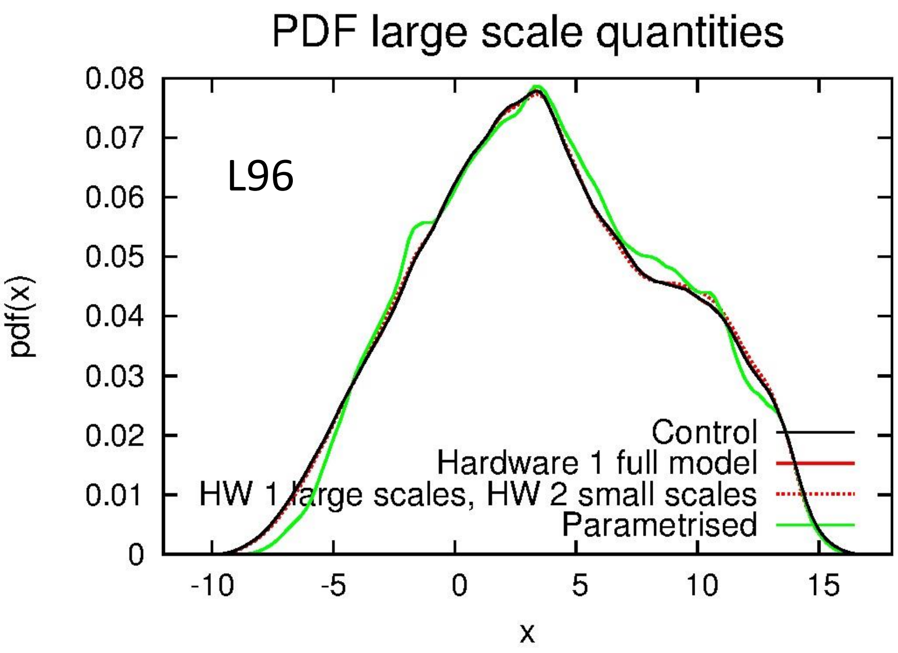
Pruned Hardware

- Parts of the floating-point unit that are hardly used or do not have a strong influence on significant bits are physically removed to obtain an increase in performance and a reduction in power consumption.
- We are collaborating with Krishna Palem and Co-workers to investigate pruned hardware setups in simulations of Lorenz 96 and the Reading Spectral Model.
- We design customised adder/subtractor and multiplier blocks for the floating point unit.

Power and Performance (floating-point unit only)	**Power** Double precision: 100%	**Performance** Double precision: 100%
Hardware 1 Adder/Subtractor Multiplier	 51% 25%	 119% 128%
Hardware 2 Adder/Subtractor Multiplier	 34% 8%	 135% 123%

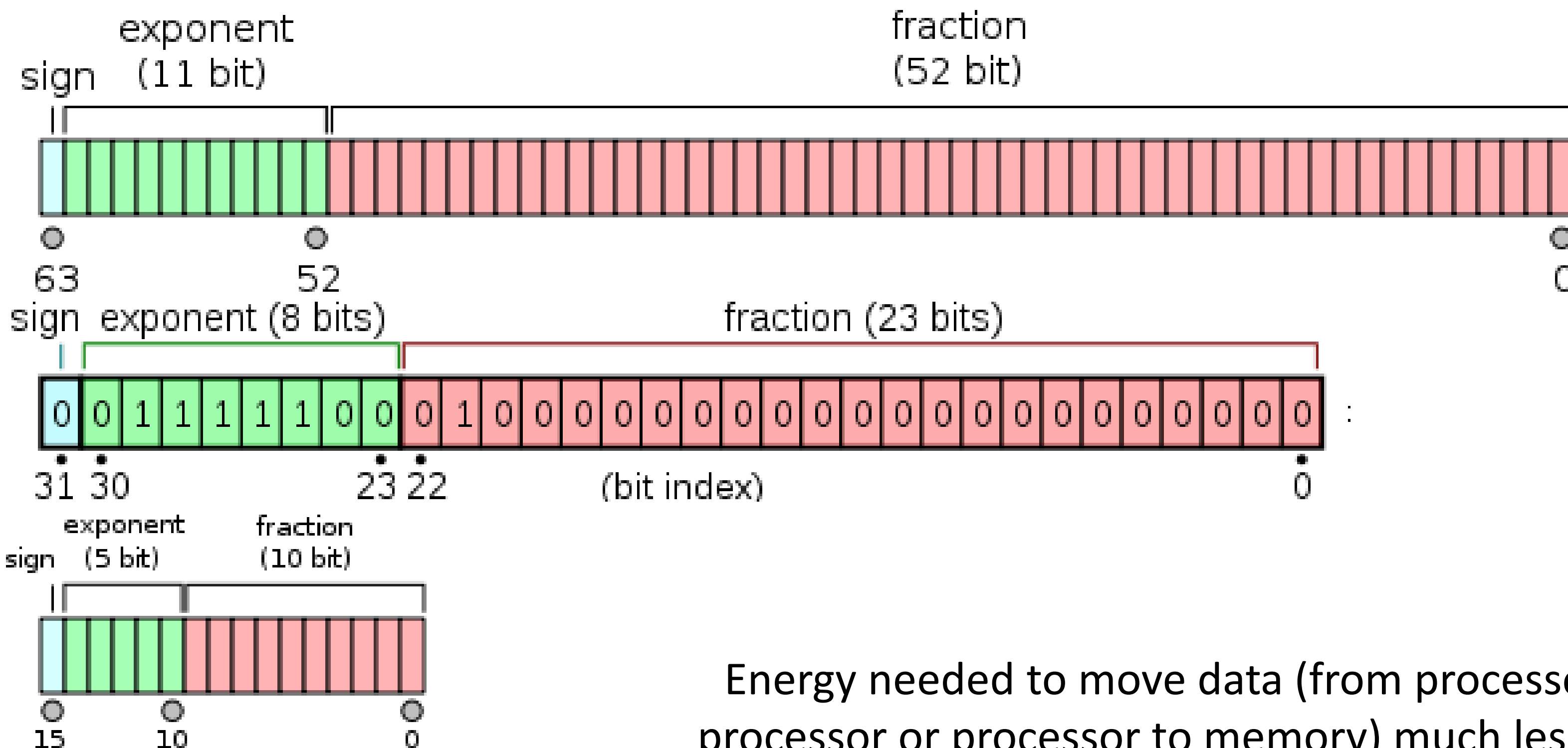
Pruned Hardware



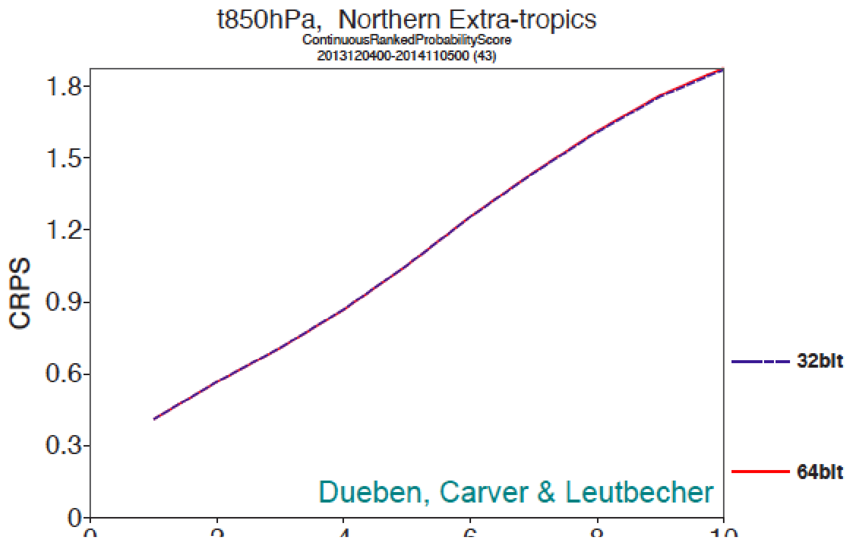


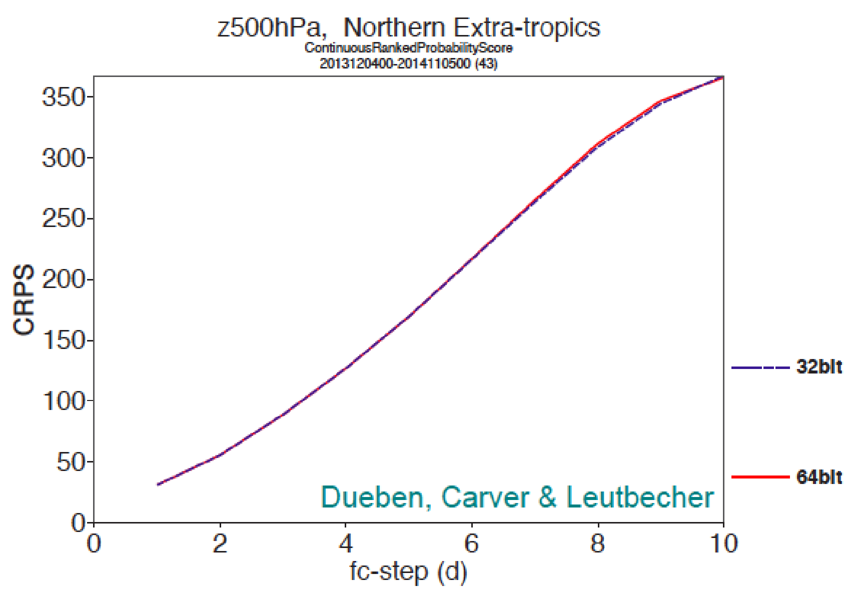
- The error due to inexact hardware is much smaller compared to the error with parametrised small scales.
- We are currently investigating the use of pruned hardware and inexact memory in a spectral dynamical core (IGCM)

Do we need to represent all variables, e.g. near the truncation scale, by double precision floating point numbers?



Energy needed to move data (from processor to processor or processor to memory) much less with low precision representations





IFS: Single vs Double Precision

T399 20 member IFS

Can run 15 day T639 at single precision for cost of 10-day T639 at double precision

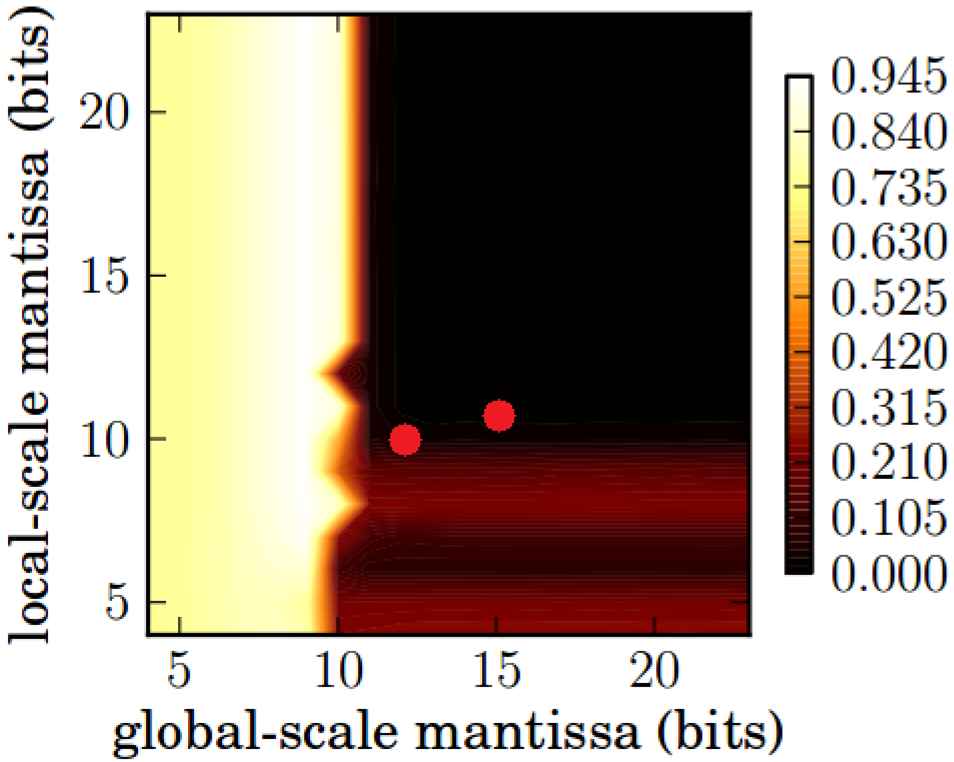
Field Programmable Gate Array (FPGA)

- FPGAs are integrated circuits that can be configured by the user (programmable hardware).
- Numerical precision can be customised to the application.
- We collaborate with Imperial College to implement Lorenz 96 on FPGAs.
- We scale the size of the Lorenz 96 setup to the size of a high performance application to obtain realistic estimates for performance (N_X = 20,000).

Field Programmable Gate Array (FPGA)

Hellinger distance for large scale quantities with reduced precision:

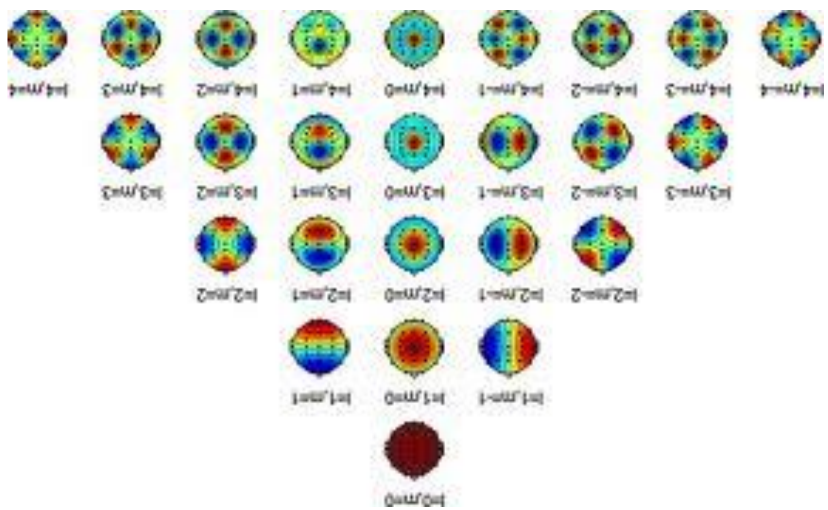
Lorenz 96



Relative speed-up:

Single precision on FPGA	Reduced precision with 15 bit significand for large scales and 11 bit for small scales	Reduced precision with 12 bit significand for large scales and 10 bit for small scales
1.0	1.9	2.5

Greater Accuracy with Less Precision?

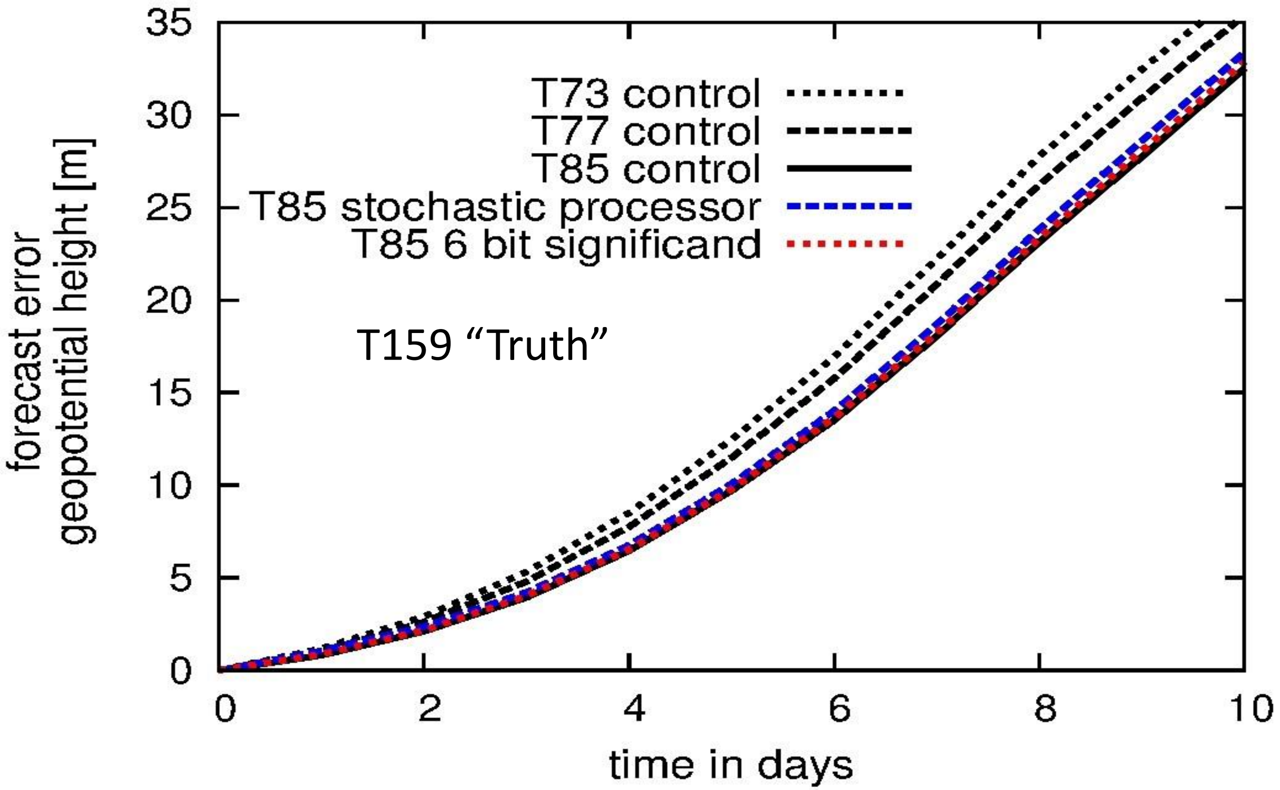


Use freed-up computing resource to extend simulator to higher resolution?

Decreasing precision, and determinism

More accurate “weather forecasts“ with less precision
Reading Spectral Model

Düben and Palmer, 2014. MWR.



- The stochastic chip / reduced precision emulator is used on 50% of numerical workload:
  All floating point operations in grid point space
  All floating point operations in the Legendre transforms between wavenumbers 31 and 85.
- Imprecise T85 cost approx that of T73

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

ISSN 1364-503X

volume 372

number 2018

Stochastic modelling and energy-efficient computing for weather and climate prediction

Papers of a Theme Issue organised and edited by Tim Palmer, Peter Düben and Hugh McNamara

The world's first science journal

Royal Society **Publishing**

28 June 2014

The important practical question

Supercomputers with variable and programmable levels of inexactness will require significant hardware redesign.

Chip manufacturers will not develop these unless they perceive there is a substantial market for them.

Could the notion of inexact computing be relevant in other areas of physics (plasma, computational fluid dynamics, astro, cosmology, human brain, etc….)?