More accuracy with less precision - assessing information content for reliable weather and climate prediction

Tim Palmer

University of Oxford

Annual Seminar 2017

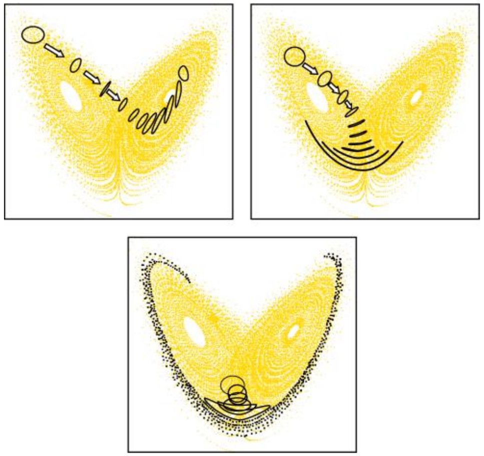
Ensemble prediction: past, present and future

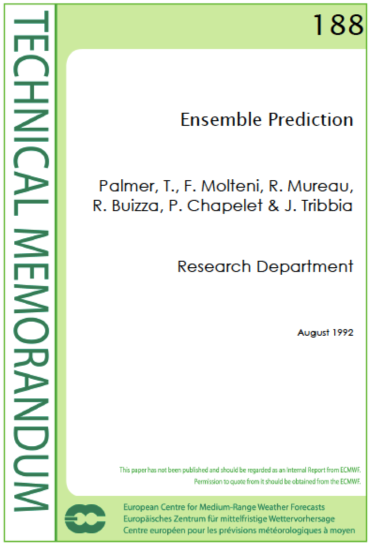
11 – 14 September

Welcome

ECMWF

#AS2017





TECHNICAL MEMORANDUM

188

Ensemble Prediction

Palmer, T., F. Molteni, R. Mureau,
R. Buizza, P. Chapelet & J. Tribbia

Research Department

August 1992

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

1992: Operational Ensemble Forecasts at ECMWF

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

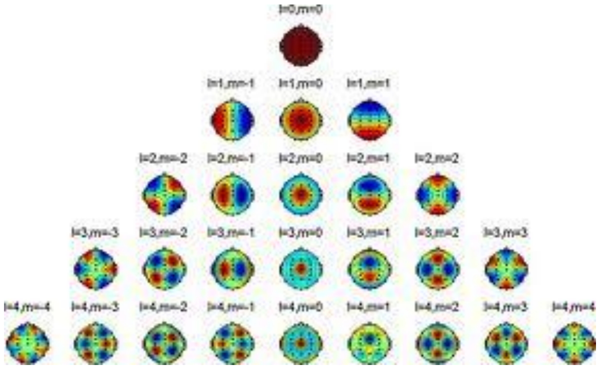
The Canonical Numerical Ansatz

Resolved scales ←

→ Unresolved scales

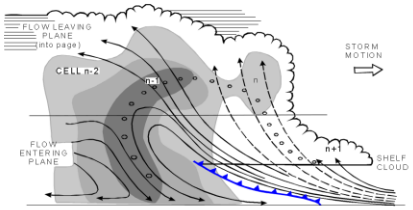
Dynamical Core

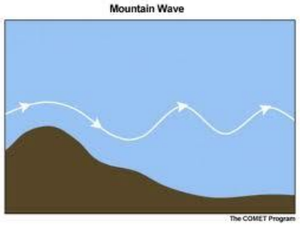
$$z = \overset{¥}{\underset{m\ l}{\text{å}}} z_{ml}e^{iml}P_{l}^{m}(f)$$



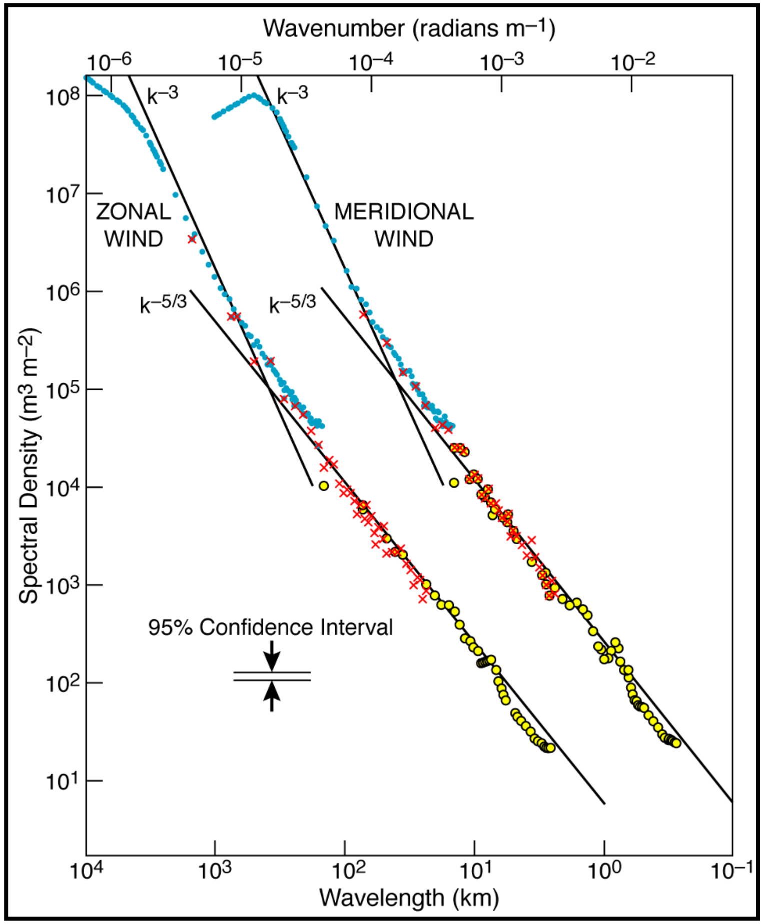
Parametrisations

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

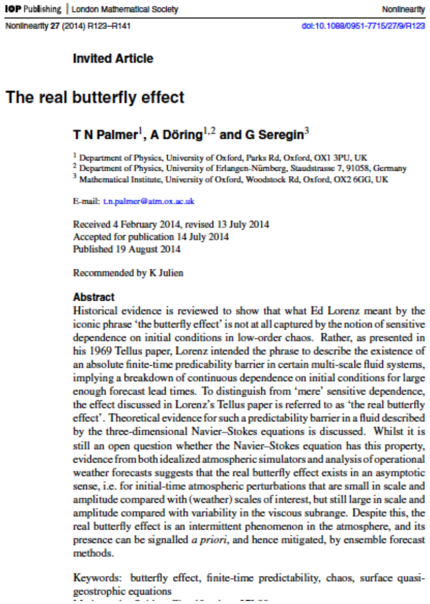




$$D = P$$



(Nastrom and Gage, 1985)



IOP Publishing | London Mathematical Society Nonlinearity

Nonlinearity **27** (2014) R123–R141 doi:10.1088/0951-7715/27/9/R123

Invited Article

The real butterfly effect

T N Palmer[1], **A Döring**[1,2] **and G Seregin**[3]

[1] Department of Physics, University of Oxford, Parks Rd, Oxford, OX1 3PU, UK
[2] Department of Physics, University of Erlangen-Nürnberg, Staudstrasse 7, 91058, Germany
[3] Mathematical Institute, University of Oxford, Woodstock Rd, Oxford, OX2 6GG, UK

E-mail: t.n.palmer@atm.ox.ac.uk

Received 4 February 2014, revised 13 July 2014
Accepted for publication 14 July 2014
Published 19 August 2014

Recommended by K Julien

Abstract

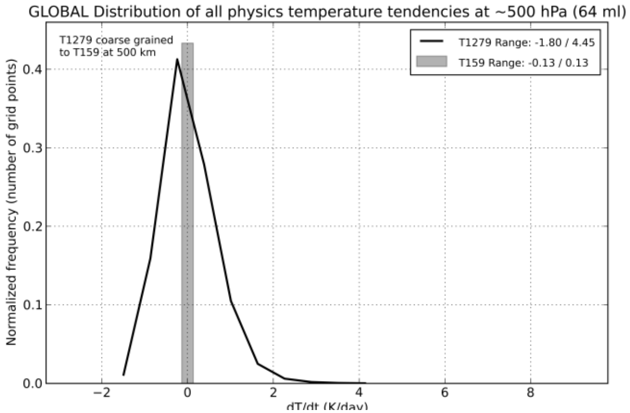
Historical evidence is reviewed to show that what Ed Lorenz meant by the iconic phrase 'the butterfly effect' is not at all captured by the notion of sensitive dependence on initial conditions in low-order chaos. Rather, as presented in his 1969 Tellus paper, Lorenz intended the phrase to describe the existence of an absolute finite-time predicability barrier in certain multi-scale fluid systems, implying a breakdown of continuous dependence on initial conditions for large enough forecast lead times. To distinguish from 'mere' sensitive dependence, the effect discussed in Lorenz's Tellus paper is referred to as 'the real butterfly effect'. Theoretical evidence for such a predictability barrier in a fluid described by the three-dimensional Navier–Stokes equations is discussed. Whilst it is still an open question whether the Navier–Stokes equation has this property, evidence from both idealized atmospheric simulators and analysis of operational weather forecasts suggests that the real butterfly effect exists in an asymptotic sense, i.e. for initial-time atmospheric perturbations that are small in scale and amplitude compared with (weather) scales of interest, but still large in scale and amplitude compared with variability in the viscous subrange. Despite this, the real butterfly effect is an intermittent phenomenon in the atmosphere, and its presence can be signalled *a priori*, and hence mitigated, by ensemble forecast methods.

Keywords: butterfly effect, finite-time predictability, chaos, surface quasi-geostrophic equations

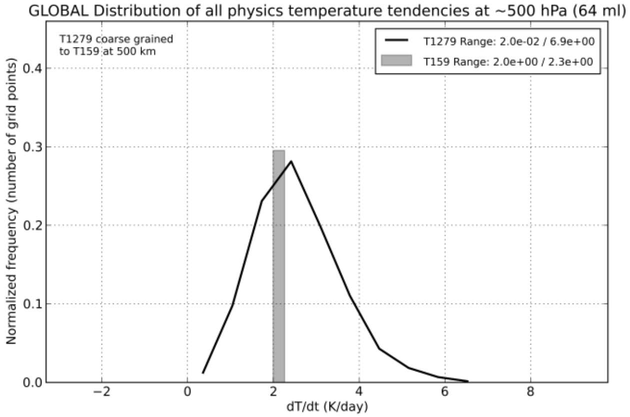
Coarse-graining

(Shutts and Palmer, 2007)

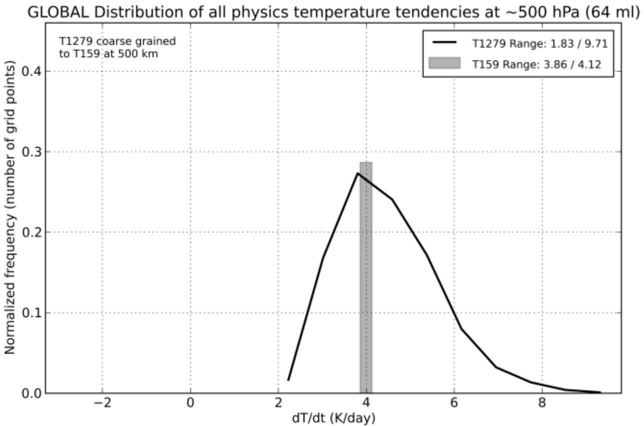
Small tendency



Medium tendency



Large tendency



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.

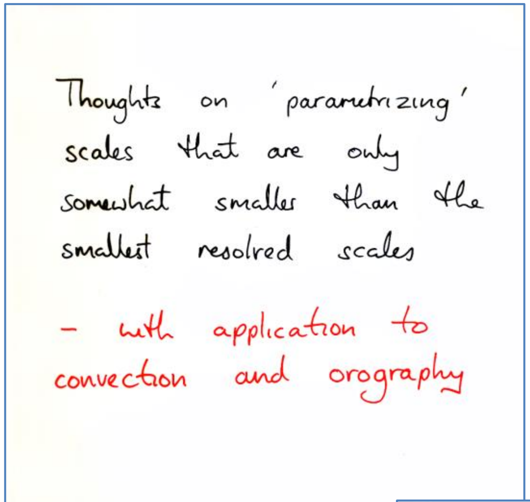
Ie when the parametrisations think the sub-grid pdf is a thin hat function, the reality is a much broader pdf.

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

Callado-Palarès and Shutts,Phil Trans 2014

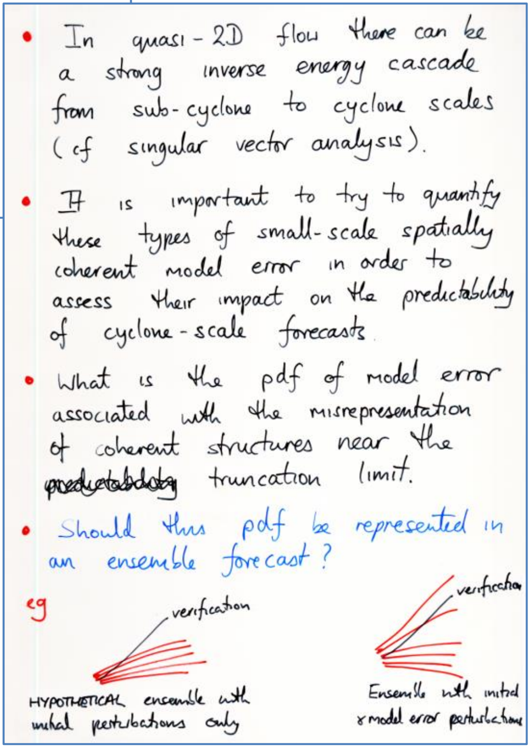
Workshop on New Insights and Approaches to Convective Parametrization

4-7 November 1996



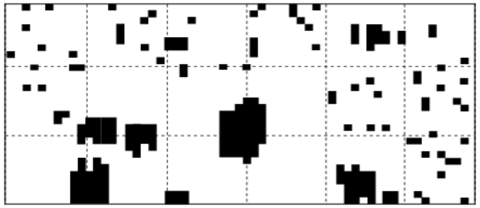
Thoughts on 'parametrizing' scales that are only somewhat smaller than the smallest resolved scales

– with application to convection and orography



- In quasi-2D flow there can be a strong inverse energy cascade from sub-cyclone to cyclone scales (cf singular vector analysis).
- It is important to try to quantify these types of small-scale spatially coherent model error in order to assess their impact on the predictability of cyclone-scale forecasts.
- What is the pdf of model error associated with the misrepresentation of coherent structures near the ~~predictability~~ truncation limit.
- Should this pdf be represented in an ensemble forecast?

eg



Probability of an "on"cell proportional to CAPE and number of adjacent "on" cells – "on" cells feedback to the resolved flow

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

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

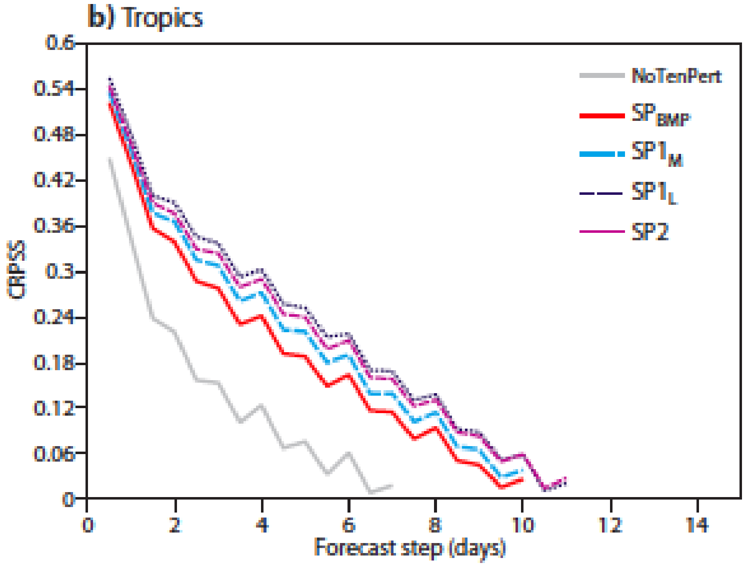
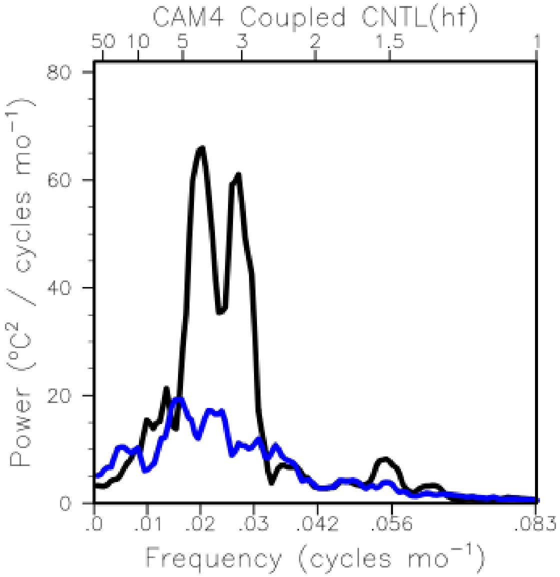


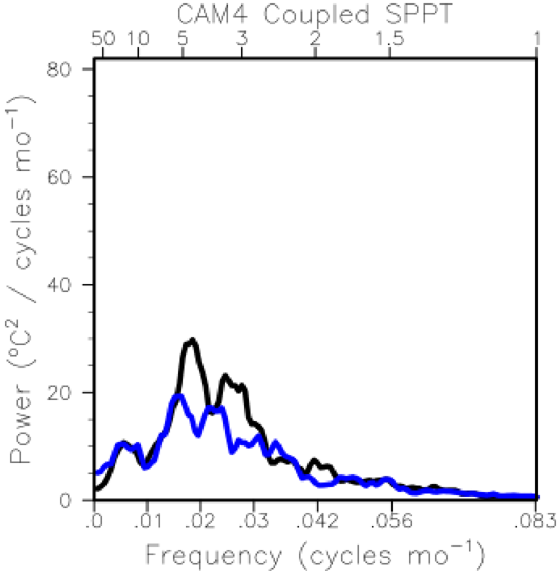
Figure 3: Continuous Ranked Probability Skill Score for 850 hPa temperature.

Hannah Christensen

Stochastic parametrisation can also improve NCAR climate model El Nino climatology



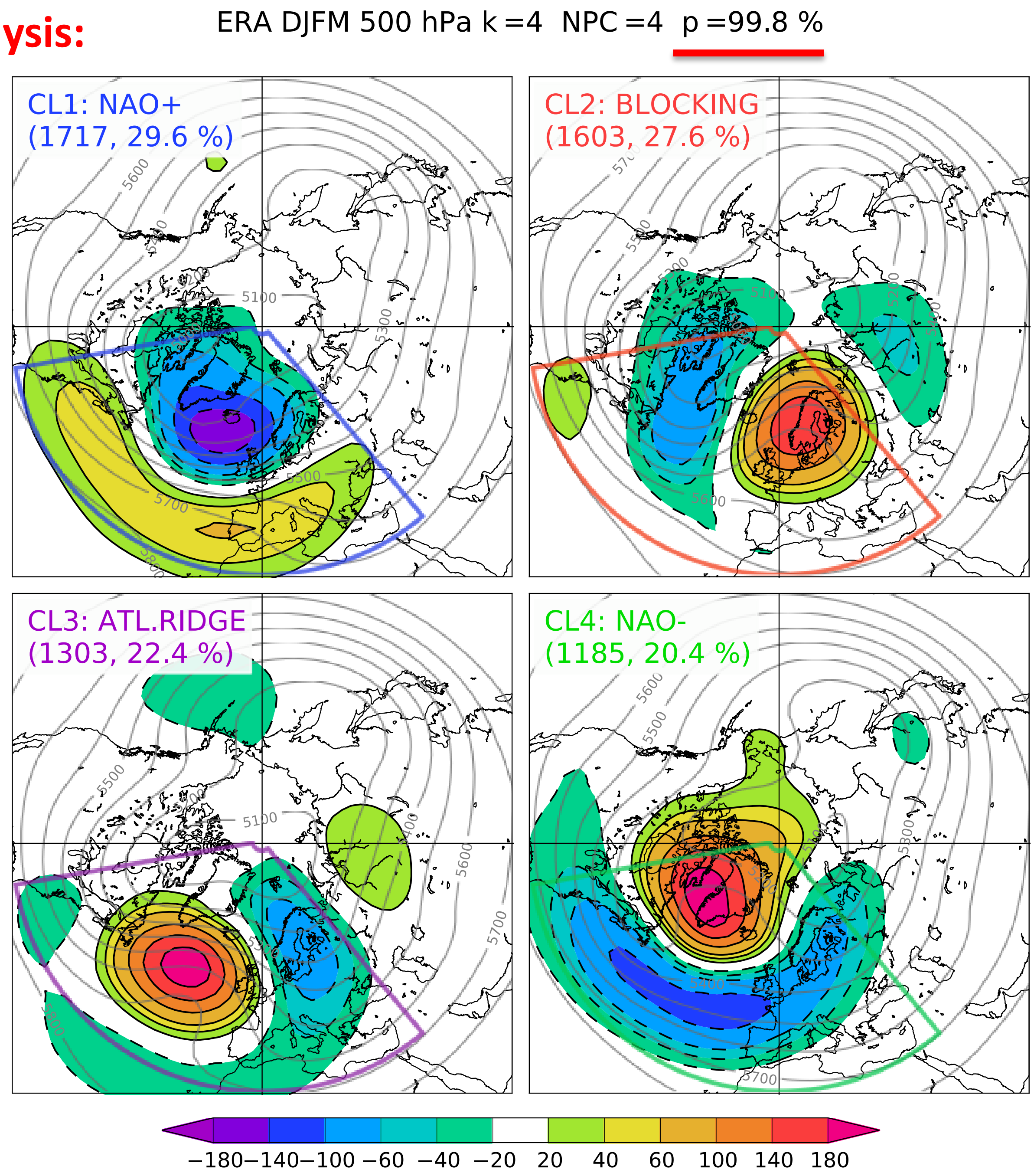
Nino3 variability without SPPT



Nino3 variability with SPPT

Christensen et al 2016

Regime Analysis: k-means clustering



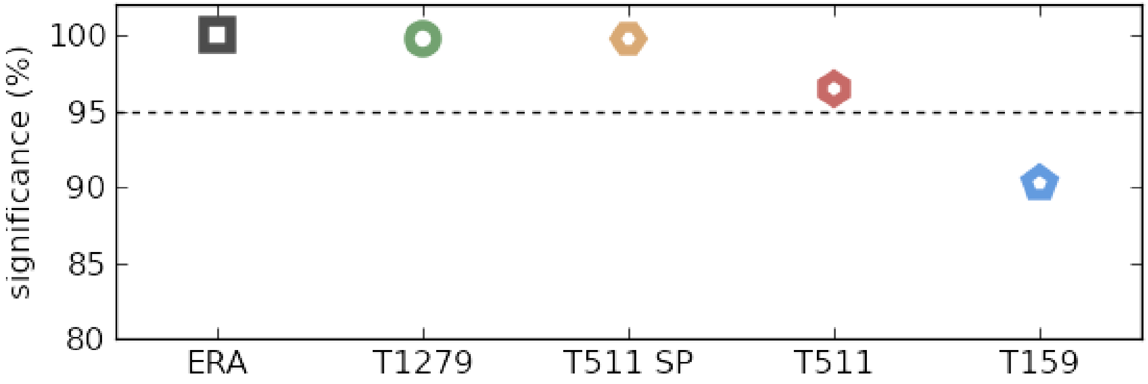
High Zonal Index

Low Zonal Index

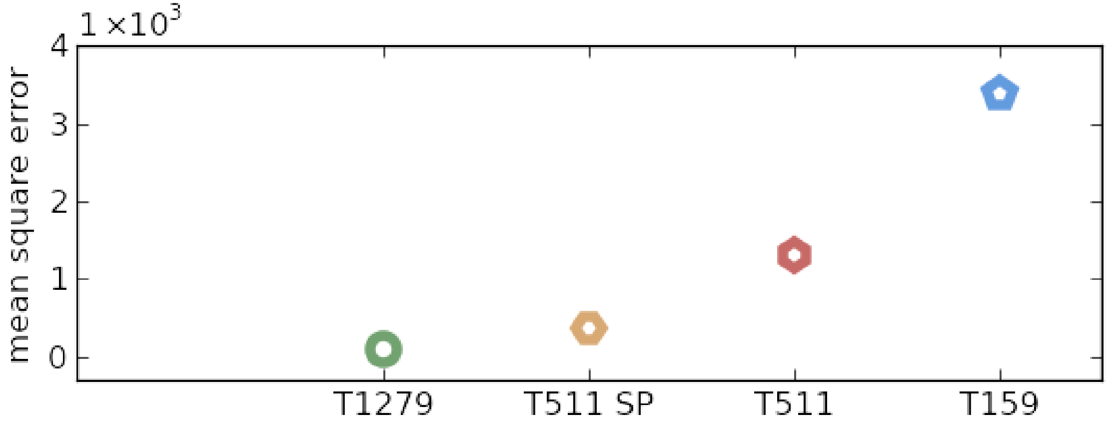
Low Zonal Index

Low Zonal Index

Athena: AMIP runs



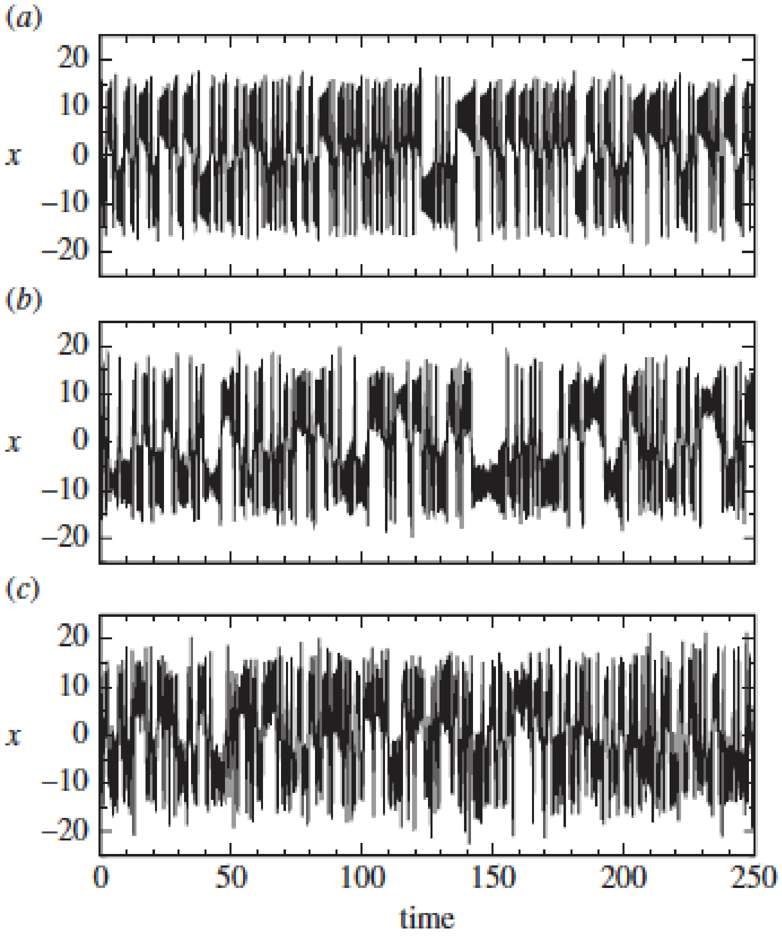
Probability that clusters are not produced from a chance sampling of a gaussian



RMS error of simulated clusters against ERA

Dawson,
Corti Palmer,
GRL 2012

$$\dot{X} = -\sigma X + \sigma Y + s\eta_1$$
$$\dot{Y} = -XZ + rX - Y + s\eta_2$$
$$\dot{Z} = XY - bZ + s\eta_3$$



s=0

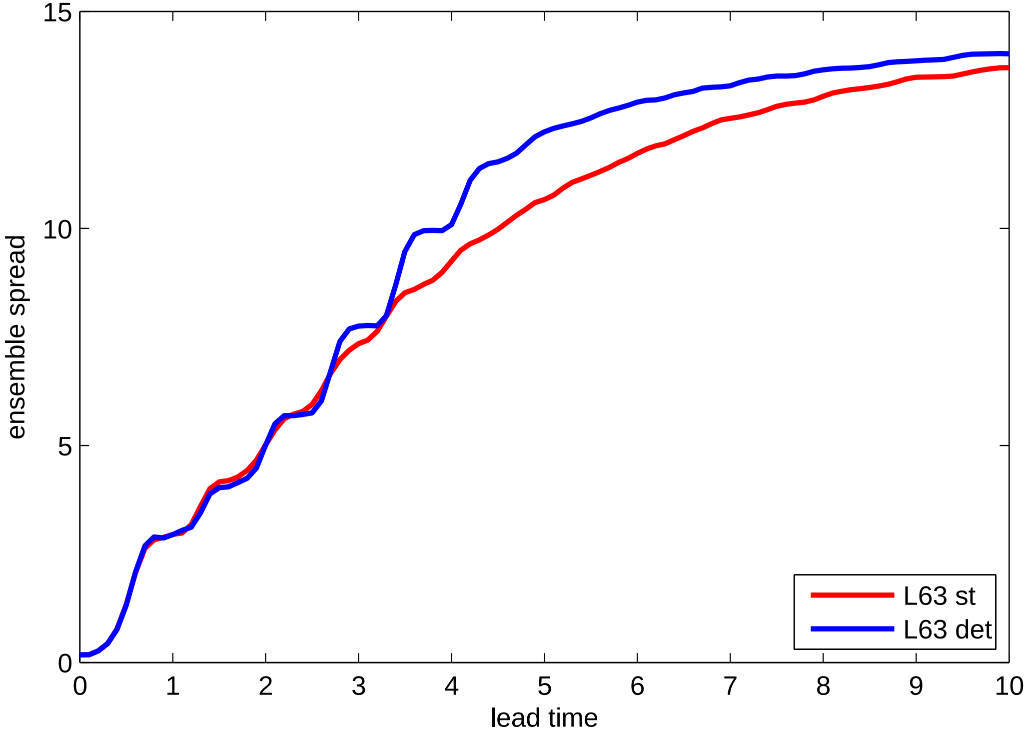
s=3.2

s=8

Kwasniok, 2014

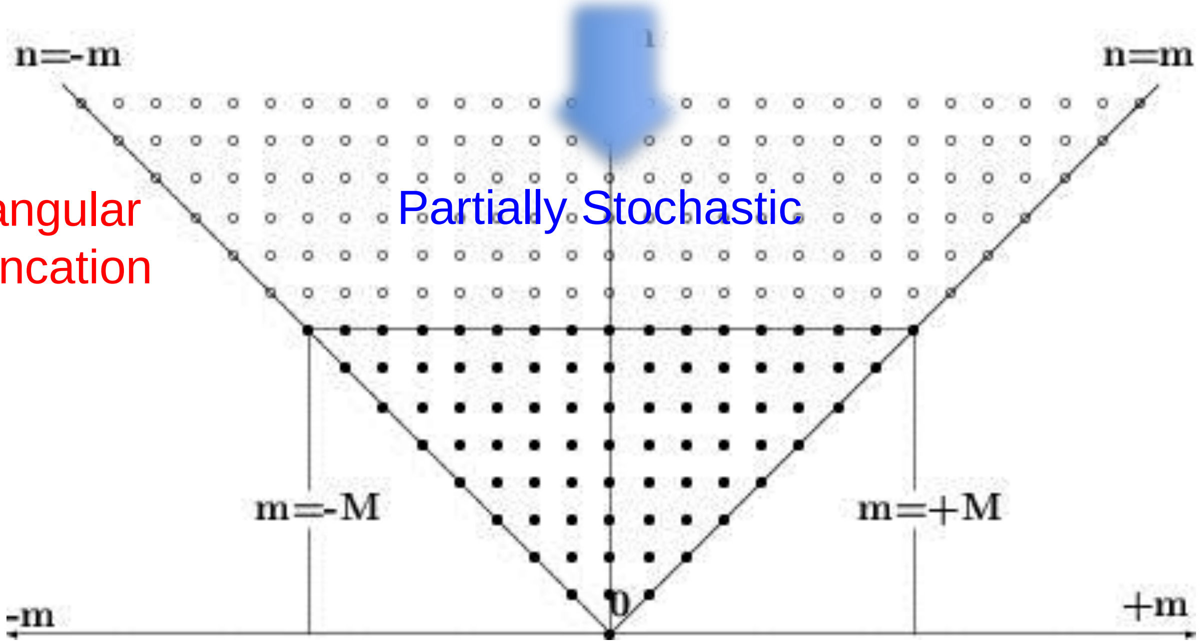
Figure 4. Lorenz '63 system: sample time series of *x* for noise levels (*a*) $\sigma = 0$, (*b*) $\sigma = 3.2$ and (*c*) $\sigma = 8$. At intermediate noise level, the distribution of regime residence times is shifted to larger values.

Spread L63 and L63 stoch.



Spread decreases (not increases!) with stochastic noise

Stochastic Parametrisation



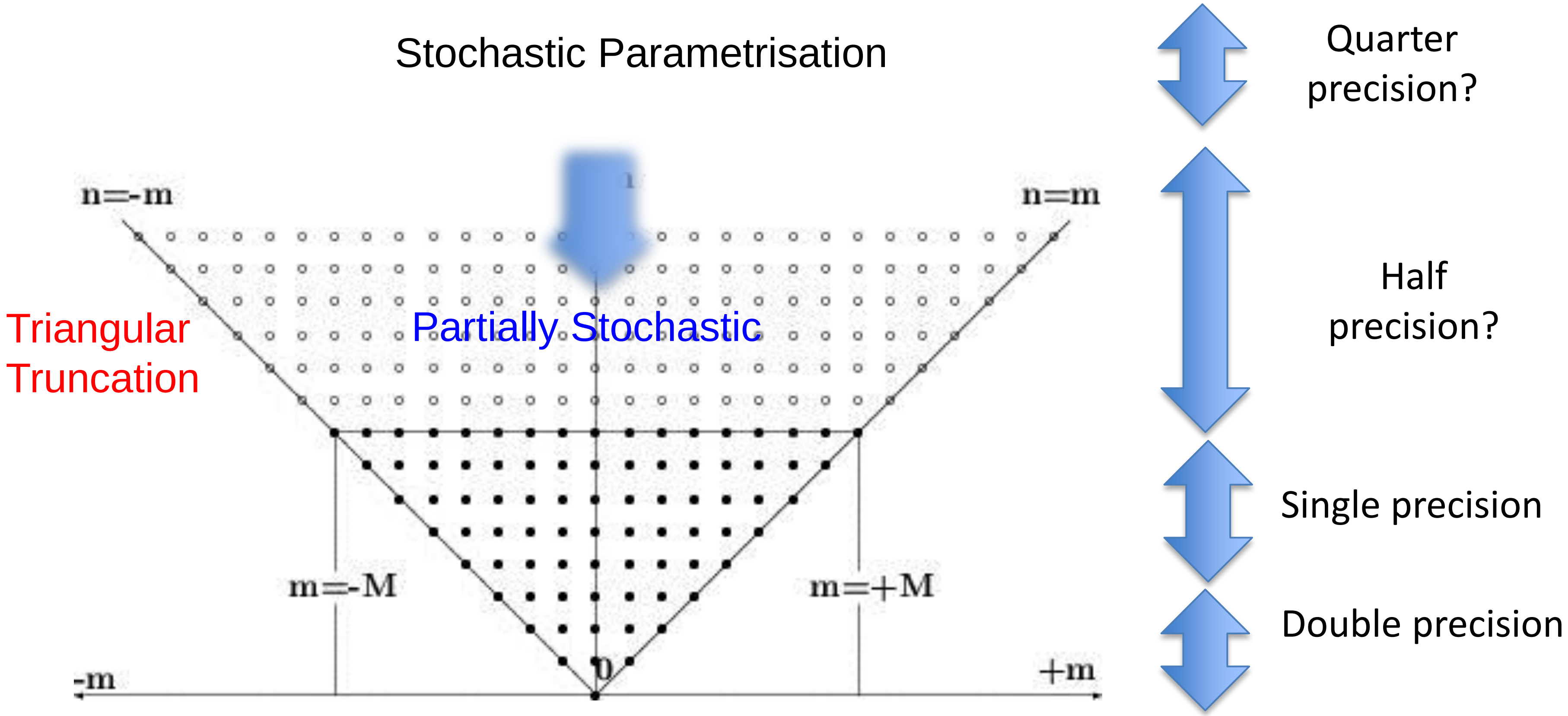
Triangular
Truncation

Partially Stochastic

If parametrisation is partially stochastic, are we “over-engineering” our dynamical cores by using double precision bit-reproducible computations for wavenumbers near the truncation scale?

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

State-dependent precision….



Stochastic Parametrisation

Triangular Truncation

Partially Stochastic

Quarter precision?

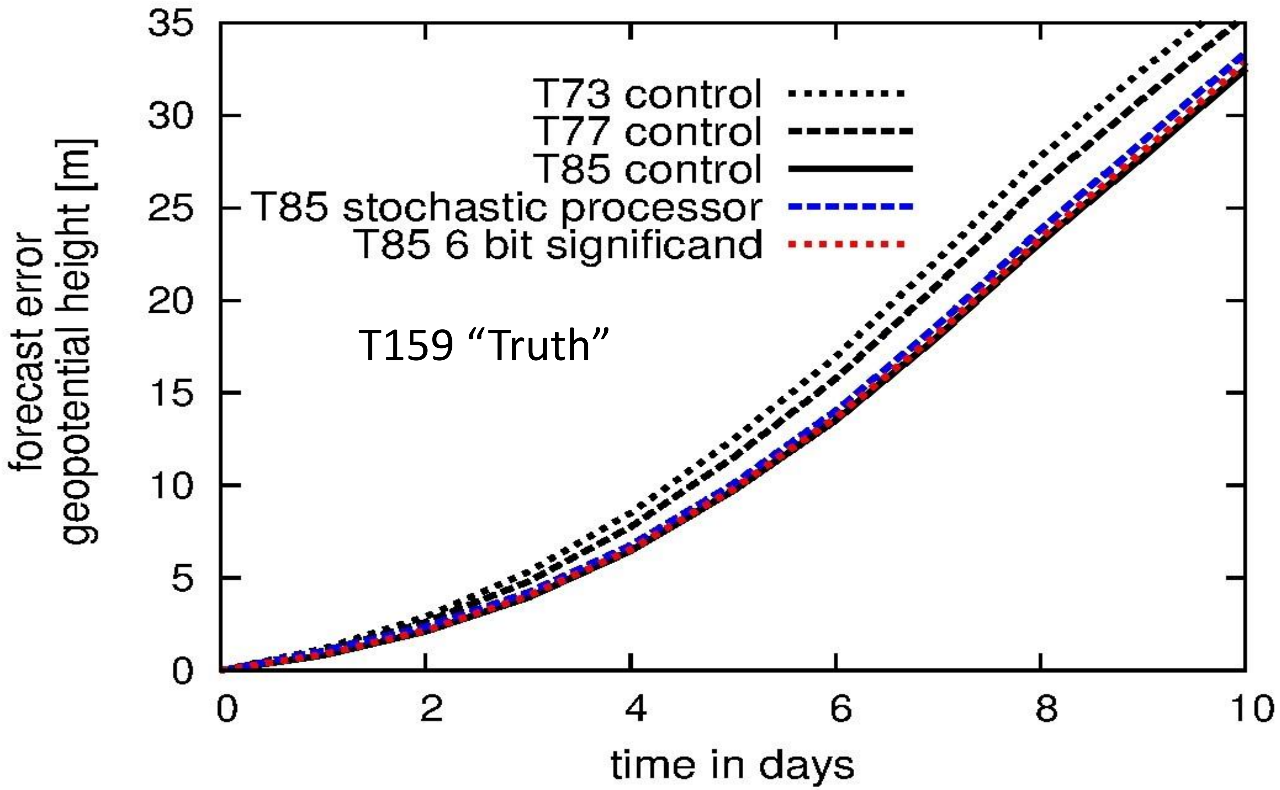
Half precision?

Single precision

Double precision

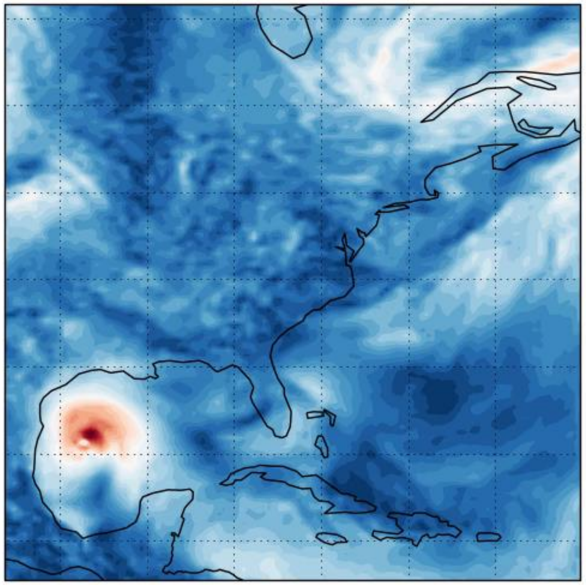
More accurate “weather forecasts“ with less precision
Reading Spectral Model

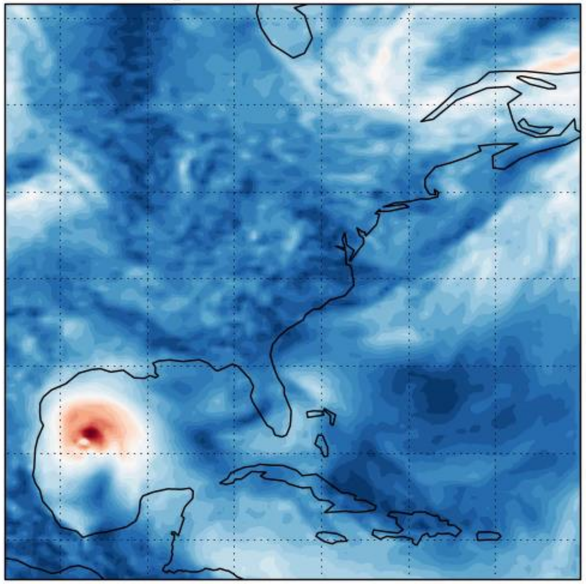
Düben and Palmer, 2014. Monthly Weather Review

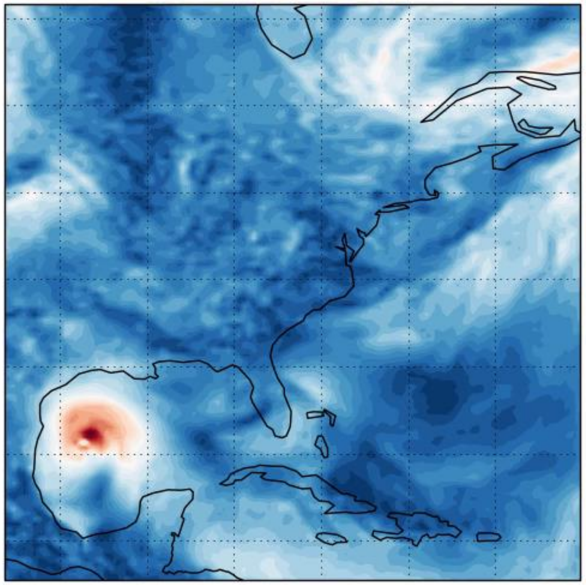


The stochastic chip / reduced precision emulator is used on 50% of numerical workload:
All floating point operations in grid point space
All floating point operations in the Legendre transforms between wavenumbers 31 and 85.
T85 cost approx that of T73

Hurricane Harvey


Double - 52 SBITS 0.00 h

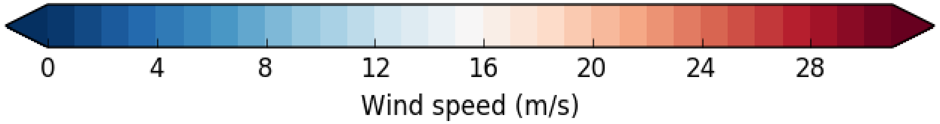

Single - 23 SBITS 0.00 h


Scale selective 0.00 h

$n = 0$	52-bit significand
0 < n £ 160	11-bit significand
160<n £ 320	9-bit significand
320<n £ 511	7-bit significand

25/08/17 00:00
850hP wind speed
T511L91

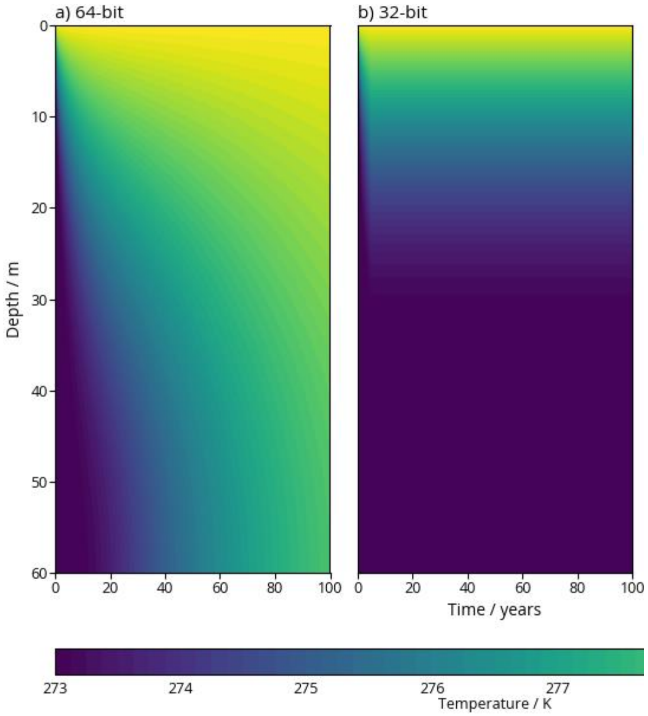
Matthew Chantry, Oxford – Peter Düben, ECMWF.



$$\frac{\partial T}{\partial t} = D\frac{\partial^2 T}{\partial z^2}$$

Andrew Dawson, Peter Düben

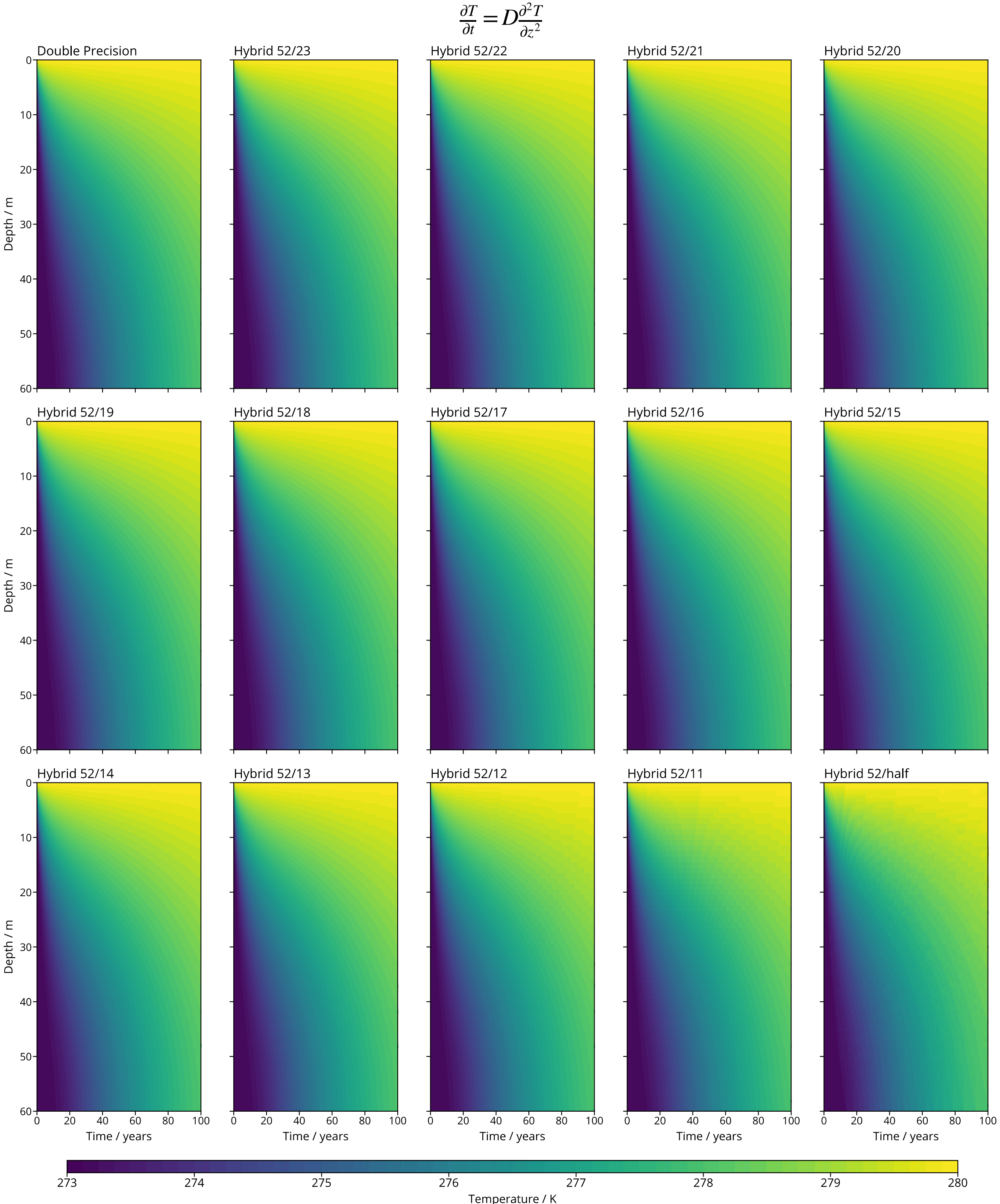
$$T_j^{n+1} = T_j^n + \Delta t\ D\ \frac{(T_{j+1}^n - 2T_j^n + T_{j-1}^n)}{(\Delta z)^2}$$



$$\frac{\partial y}{\partial t} = S(y, \overset{\substack{\text{Highly uncertain}\\ \swarrow}}{f}, t)$$

$$\underbrace{y(t^{n+1}) = y(t^{n})}_{\text{Represent at high precision}} + \underbrace{\Delta t \cdot S(y(t^{n}), f(t^{n}), t^{n}}_{\substack{\text{Compute (and retrieve}\\ \text{fields from memory) at low}\\ \text{precision}}})$$

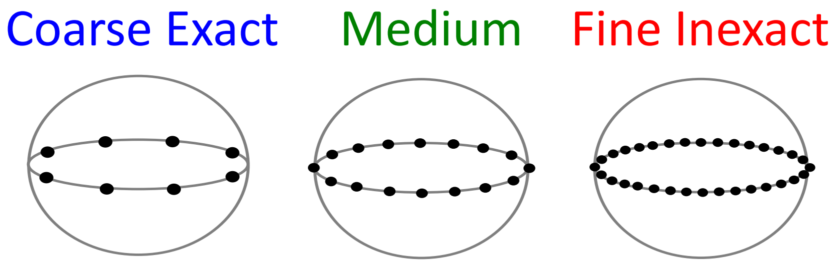
Andrew Dawson



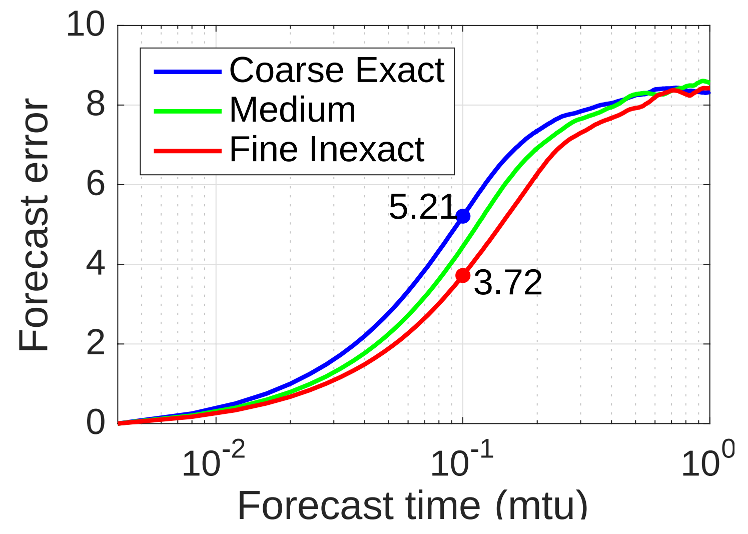
Using FPGA resources for resolution rather than precision yields 28.9% forecast improvement and 10x energy reduction (Stephen Jeffress)

Lorenz '96:

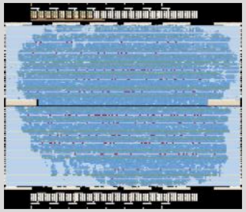
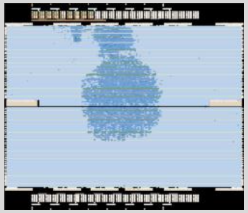
$$\dot{x}_k = x_{k-1}\left(x_{k+1} - x_{k-2}\right) - x_k + F$$



Forecast Accuracy



FPGA Resources

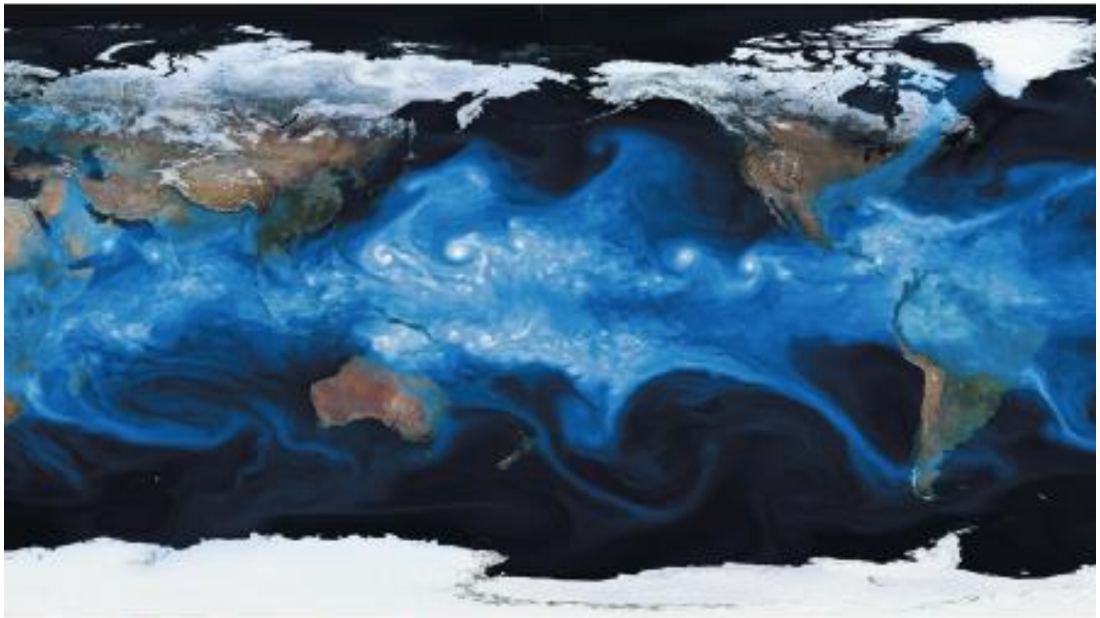
Chip area			
Gridpoints	8	16	32
Precision	Double	Single	Int16
Clock Cycles	355	355	34
Energy (μJ)	**8.2**	**15.1**	**0.9**

Forecast error is with respect to a 64 gridpoint double precision model. The FPGA for each model calculates one 4th order Runge Kutta time step. Energy is per model time unit. FPGAs (Altera Stratix V) designed with tools from Maxeler Technologies.

Jeffress et al (2017) Proc. Roy. Soc.

What is the real information content in each of the billions of bits that represent variables in a weather/climate model?

Only communicate (on and off chip) the bits that contain real information.



A simulation of Earth's atmosphere generated by the Community Atmosphere Model.

Build imprecise supercomputers

Energy-optimized hybrid computers with a range of processor accuracies will advance modelling in fields from climate change to neuroscience, says **Tim Palmer**.

Today's supercomputers lack the power to model accurately many aspects of the real world, from the impact of cloud systems on Earth's climate to the processing ability of the human brain. Rather than wait decades for sufficiently powerful supercomputers — with their potentially unsustainable energy demands — it is time for researchers to reconsider the basic concept of the computer. We must move beyond the idea of a computer as a fast but otherwise traditional "Turing machine", churning through calculations bit by bit in a sequential, precise and reproducible manner.

In particular, we should question whether all scientific computations need to be performed deterministically — that is, always producing the same output given the same input — and with the same high level of precision. I argue that for many applications they do not.

Energy-efficient hybrid supercomputers with a range of processor accuracies need to be developed. These would combine conventional energy-intensive processors with low-energy, non-deterministic processors, able to analyse data at variable levels of precision. The demand for such machines could be substantial, across diverse sectors of the scientific community.

MORE WITH LESS

Take climate change, for example. Estimates of Earth's future climate are based on solving nonlinear (partial differential) equations for fluid flow in the atmosphere and oceans. Current climate simulators — typically with grid cells of 100 kilometres in width — can resolve the large, low-pressure weather systems typical of mid-latitudes, but not individual clouds. Yet modelling cloud systems accurately is crucial for reliable estimates of the impact of anthropogenic emissions on global temperature[1].

The resolution of this computational grid is determined by the available computing power. Current petaflop computers can perform up to 10^{15} additions or multiplications — floating-point operations — per second (flops). By the early 2020s, next-generation exaflop supercomputers, capable of 10^{18} operations per second, will be able to resolve the largest and most vigorous types of thunderstorm[2]. But cloud physics on scales smaller than a grid cell will still have to be approximated, or