More Accuracy with Less Precision.

Recognising the crucial role of uncertainty for developing high-resolution NWP systems

Tim Palmer

With thanks to

Judith Berner, Glenn Carver, Hannah Christensen, Andrew Dawson, Peter Düben, Martin Leutbecher, David Richardson, Filip Vana, Antje Weisheimer, Krishna Palem

"Forecasts possess no intrinsic value. They acquire value through their ability to influence the decisions made by users of the forecasts."

Allan Murphy (Wea. Forecasting, **8**, 281-293, 1993)

Potential Economic Value

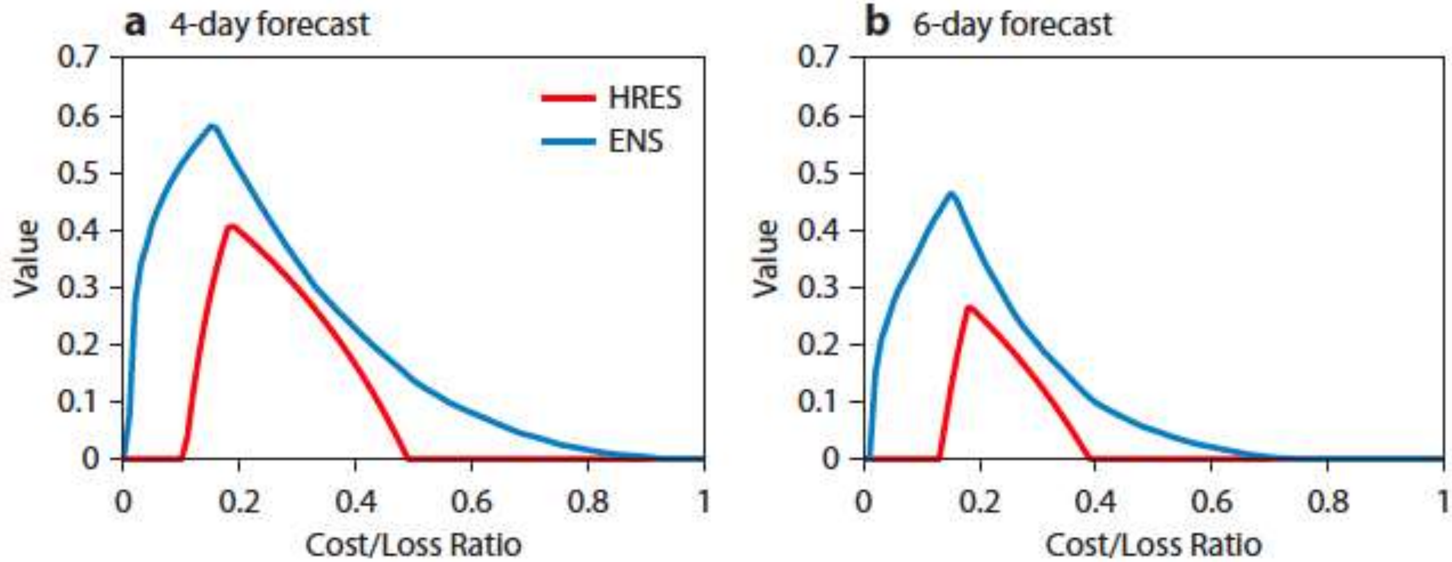
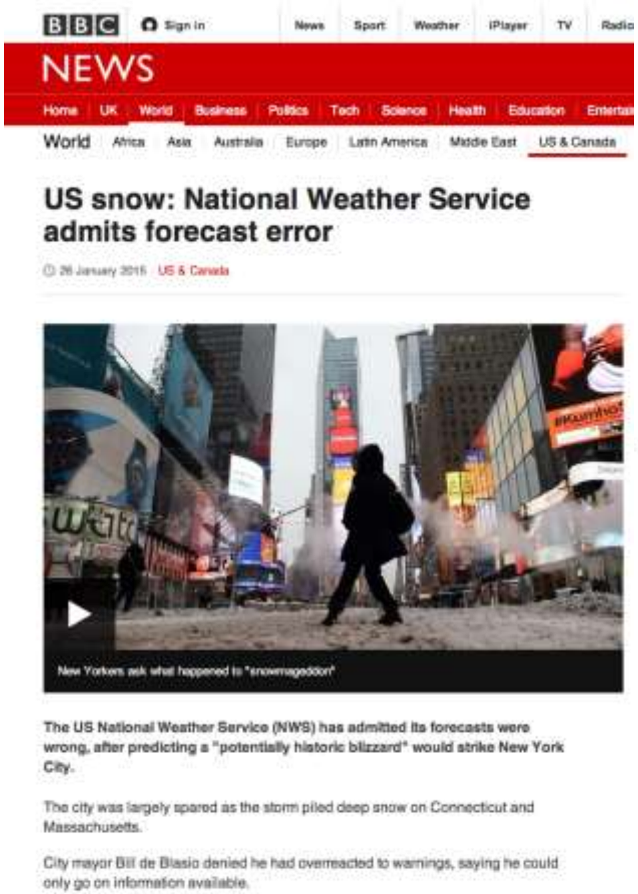


Figure 1 Potential economic value of the high-resolution forecast (HRES) and the ensemble forecast (ENS) in predicting 24-hour precipitation amounts greater than 5 mm over Europe for the summer (June to August) 2014: (a) four-day forecast and (b) six-day forecast.

$$\mathrm{CRPSS} = \langle \mathrm{Value} \rangle_{\mathrm{users,\ events}}$$

(Murphy, 1966)

A Primary strategic score for ECMWF.



BBC | Sign in | News | Sport | Weather | iPlayer | TV | Radio

NEWS

Home | UK | World | Business | Politics | Tech | Science | Health | Education | Entertai

World | Africa | Asia | Australia | Europe | Latin America | Middle East | US & Canada

US snow: National Weather Service admits forecast error

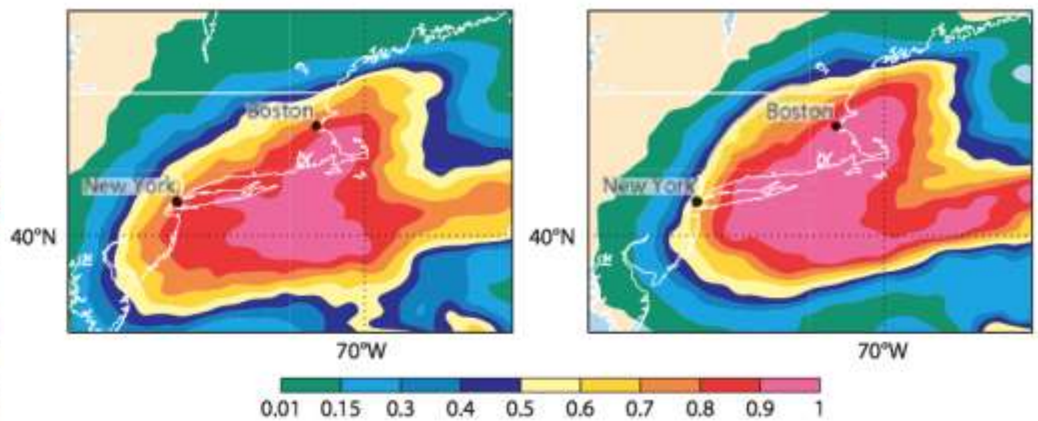
26 January 2015 | US & Canada

New Yorkers ask what happened to "snowmageddon"

The US National Weather Service (NWS) has admitted its forecasts were wrong, after predicting a "potentially historic blizzard" would strike New York City.

The city was largely spared as the storm piled deep snow on Connecticut and Massachusetts.

City mayor Bill de Blasio denied he had overreacted to warnings, saying he could only go on information available.



Probability for more than 30 mm of precipitation on 27 January in ENS forecasts from 00 UTC on 25 January (left) and 00 UTC on 26 January (right) .

ECMWF Newsletter: Richardson et al, 2015

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 6, 2009

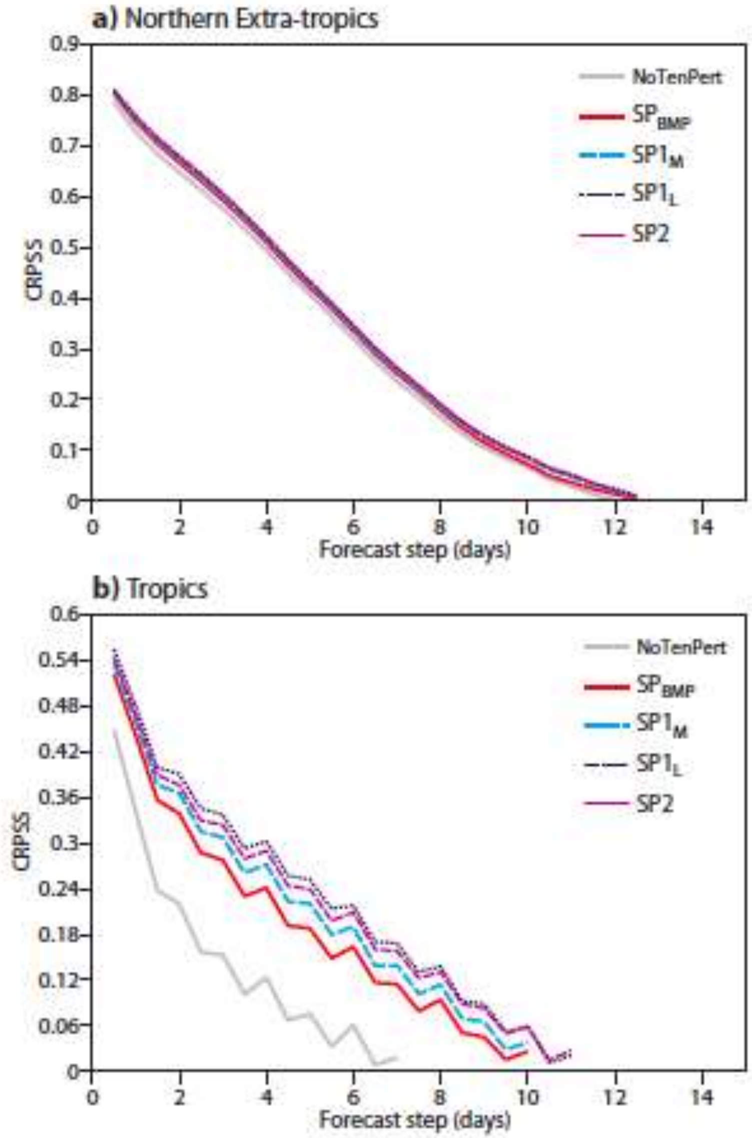
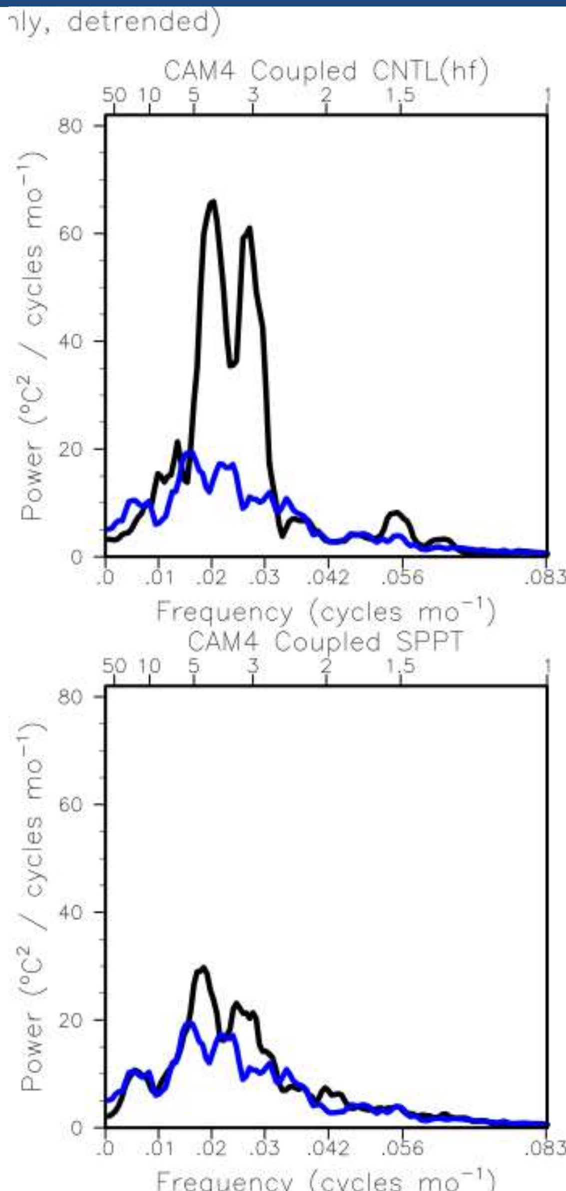


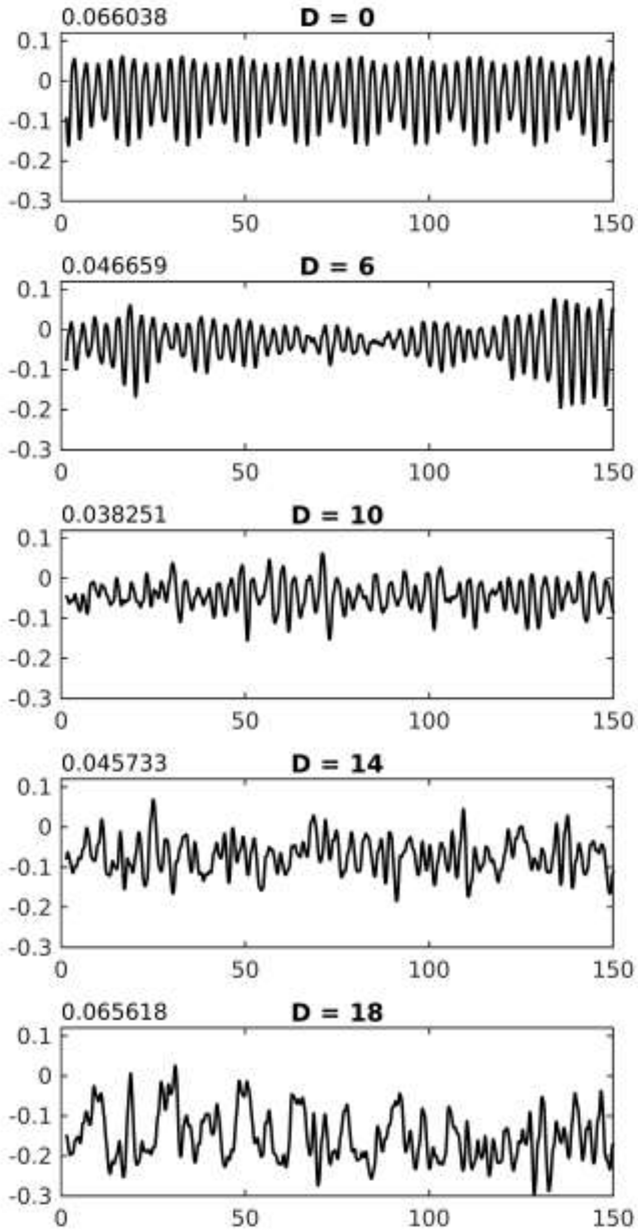
Figure 3: Continuous Ranked Probability Skill Score for 850 hPa temperature.

Tropical variability: SPPT results in an improvement



- SPPT -> better variability in SST, T2m and precipitation in tropical pacific ocean in 1° CAM4 coupled runs
  - Begin to see improvements in power spectra on N3.4 timeseries

H. M. Christensen and J. Berner



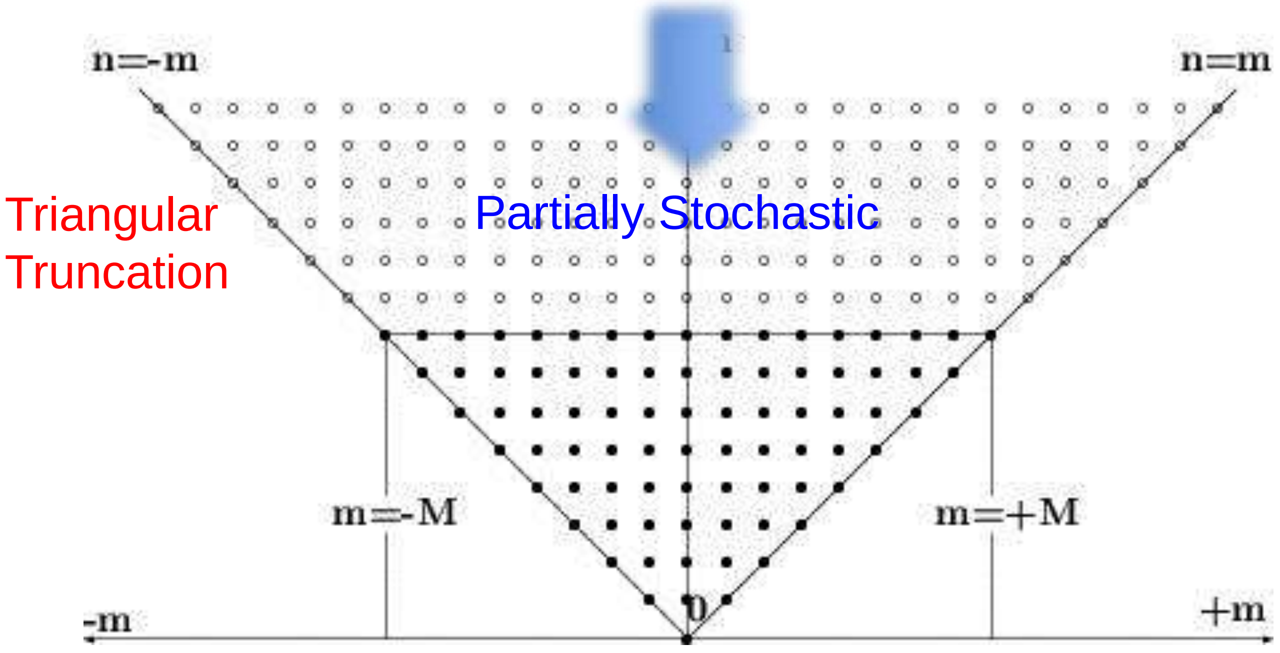
Delayed oscillator for ENSO (Stone, Sarapin, Huppert, Price, GRL, 1998 – simplified version of Muennich, Cane Zebiak).

D is amplitude of uncorrelated noise associated with a multiplicative noise perturbation to the wind stress forcing.

No such effect with additive noise.

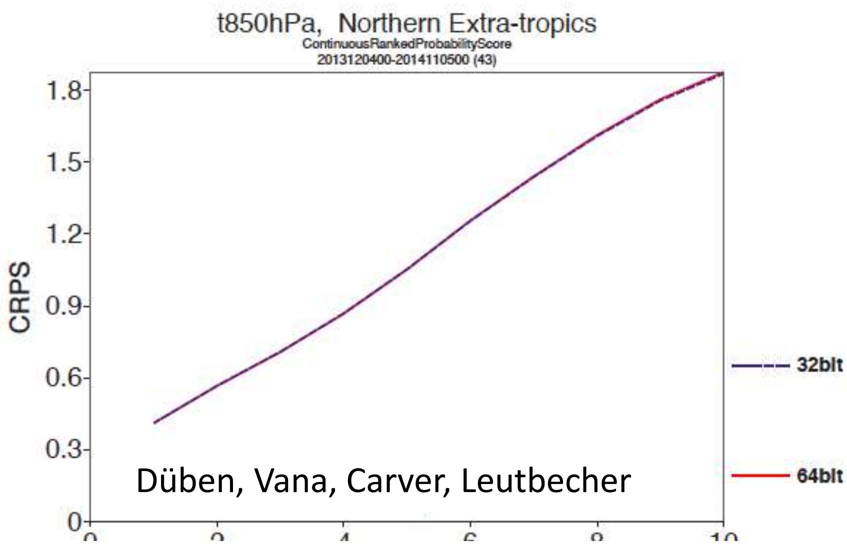
Hannah Christensen

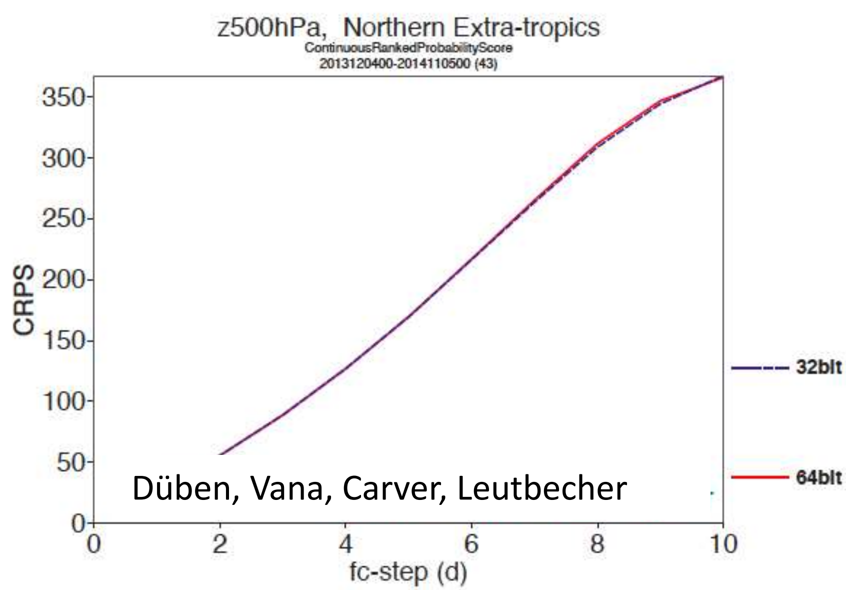
Stochastic Parametrisation



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?



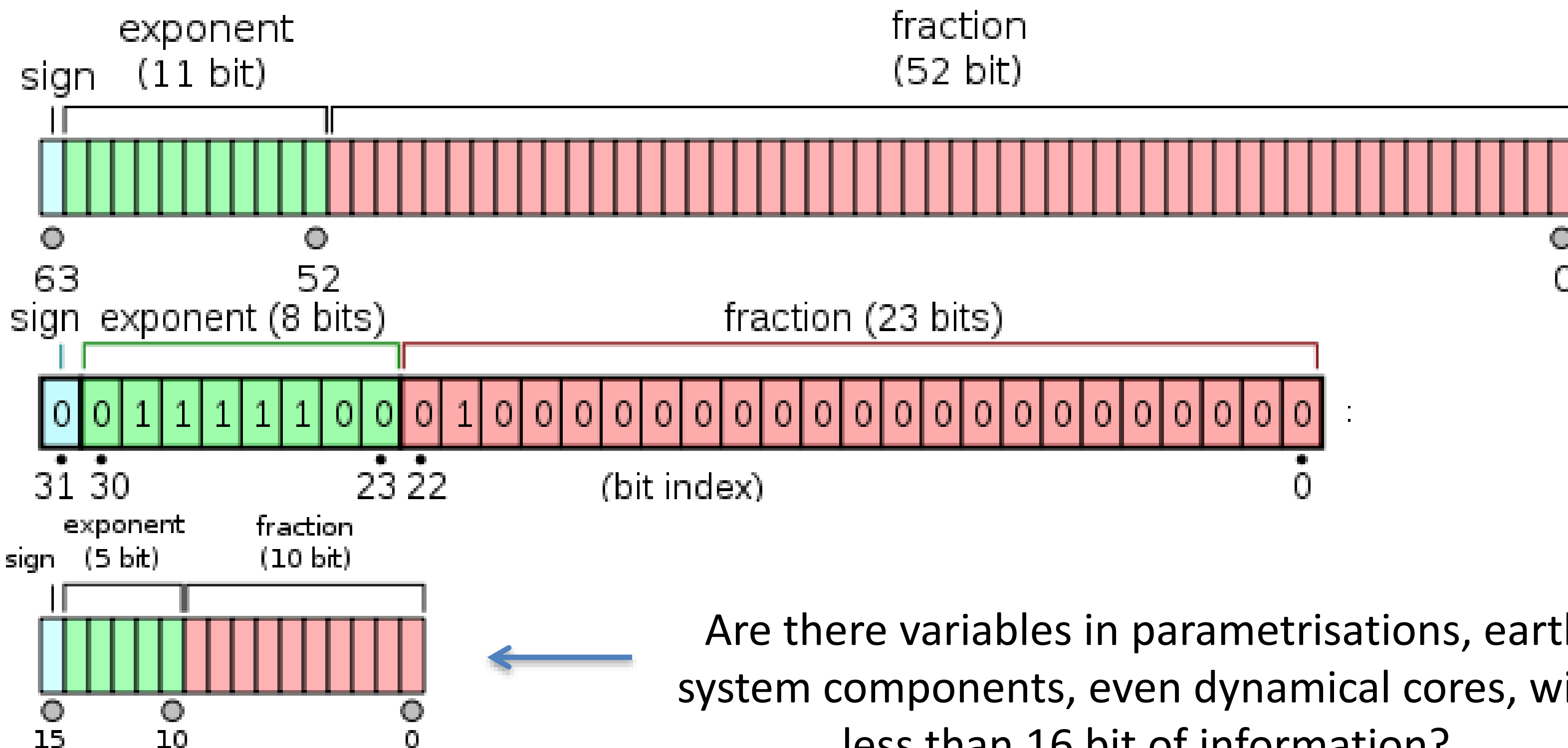


IFS: Single vs Double Precision

T399 20 member IFS

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

Reduced precision allows computations to proceed more quickly, and data to be moved around with less energy overhead.

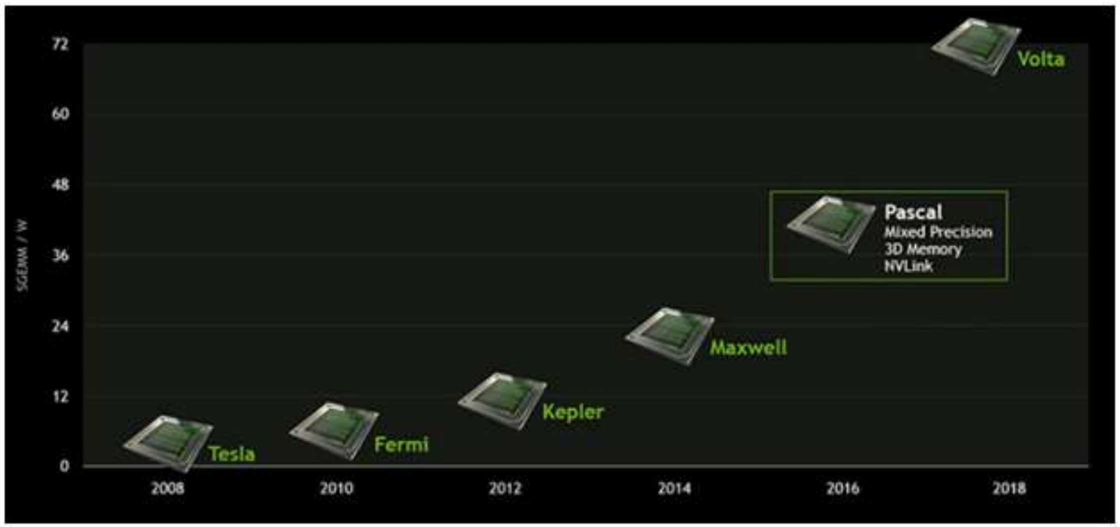


Are there variables in parametrisations, earth system components, even dynamical cores, with less than 16 bit of information?

Nvidia's Pascal to feature mixed-precision mode, up to 32GB of RAM

by Jeff Kampman — 5:55 PM on March 17, 2015

We learned some details about Nvidia's next-gen Pascal GPU architecture at last year's GTC. After today's GTC keynote, we now know a little more about how Pascal will operate. We also have new info about its memory capacity and bandwidth.



Perhaps the biggest feature of Pascal revealed today is its ability to operate in a mixed-precision mode, similar to mobile GPUs like ARM's Mali. Nvidia's current-gen GPU architecture, Maxwell, is limited to fp32 operation, meaning that int8, fp16, and fp32 operations are all processed internally at the same rate. The Maxwell GPU in the Tegra X1 SoC adds the abillity to operate in fp16 mode, which can effectively double its throughput.

Emulated reduced precision using overloaded operators

“Overloaded Operators” allows one to emulate reduced precision and specific inexact hardware in comprehensive models (such as IFS) with no need for extensive changes of model code.

Method:

1. Floating point declarations are replaced by predefined types.
2. Numerical operations are overloaded (redefined).

Example:

Emulated 5 bit significand with overloaded operation “+”

Standard Fortran:

```
REAL*8 :: a,b,c
a = 1.442221
b = 2.136601
c = a+b
```

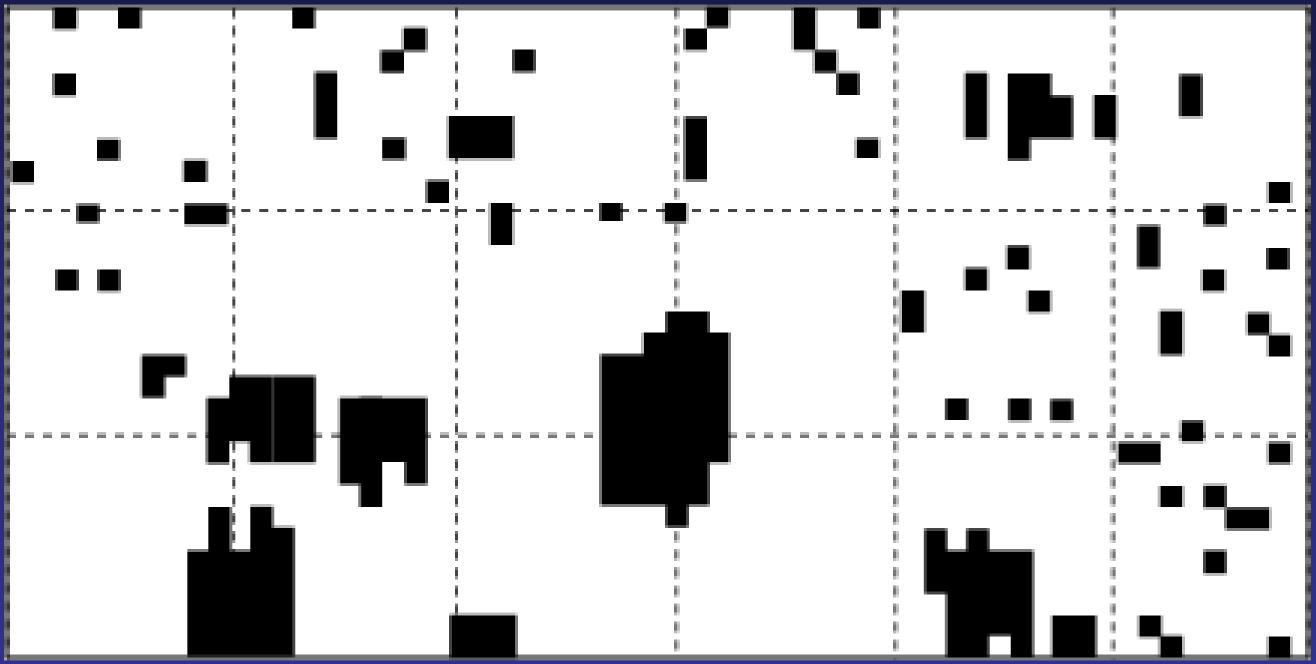
→ c=3.578822

Overloaded Operators:

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

→ c=3.562500

Nb – 1/32nd Precision = 1 bit: Cellular Automata!



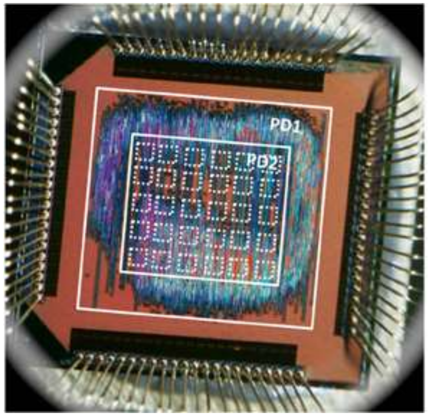
EG for representing convective organisation

(Palmer; 1997, 2001)

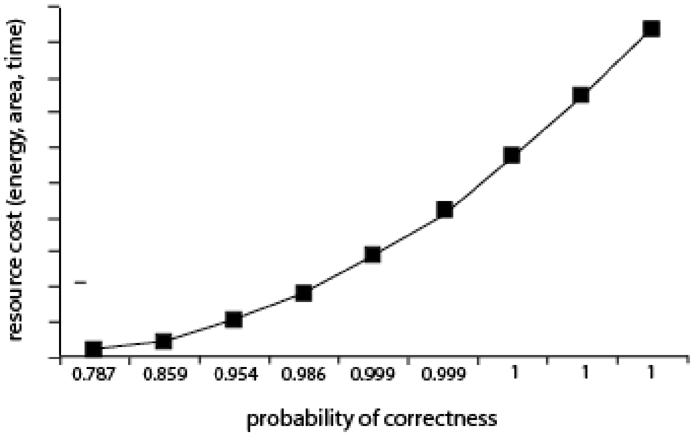
Superefficient inexact chips

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

Prototype
Probabilistic
CMOS
Chip



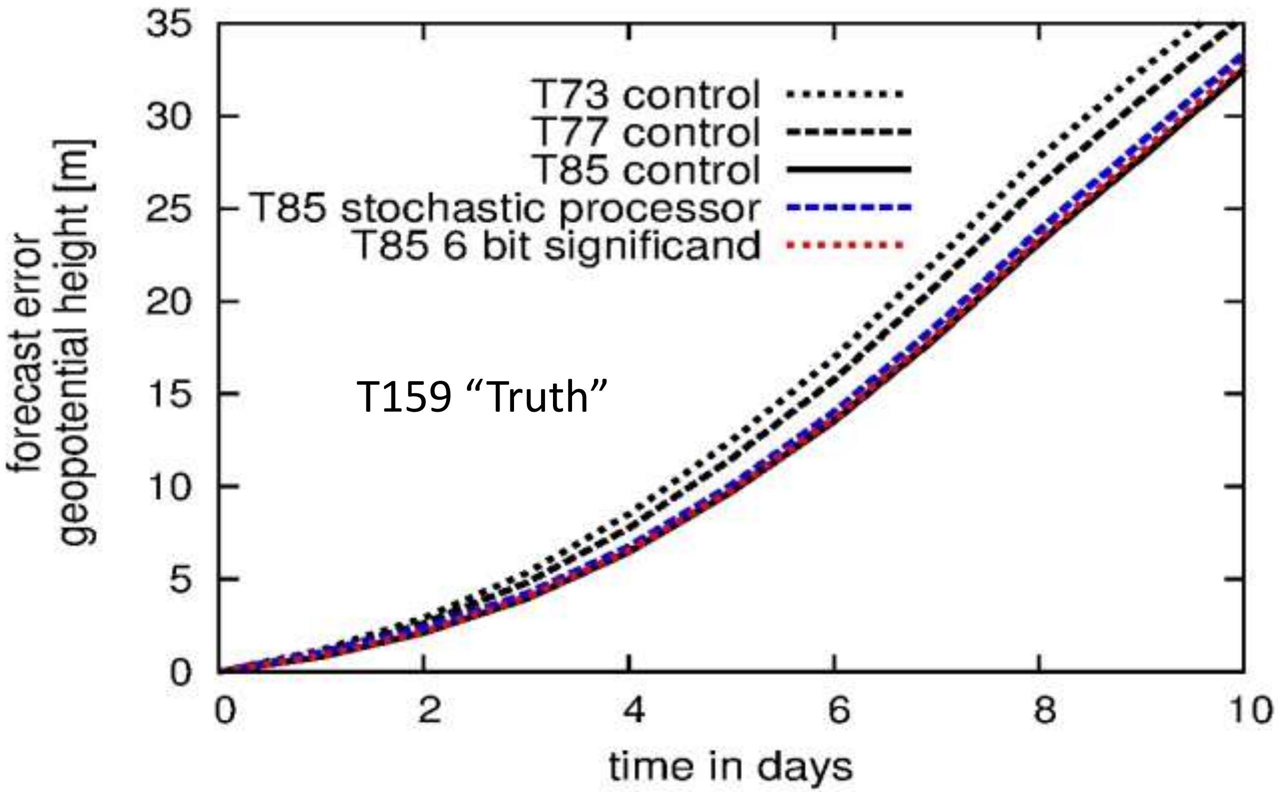
Krishna Palem.
Rice University



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).

More accurate "weather forecasts" with less precision
Reading IGCM

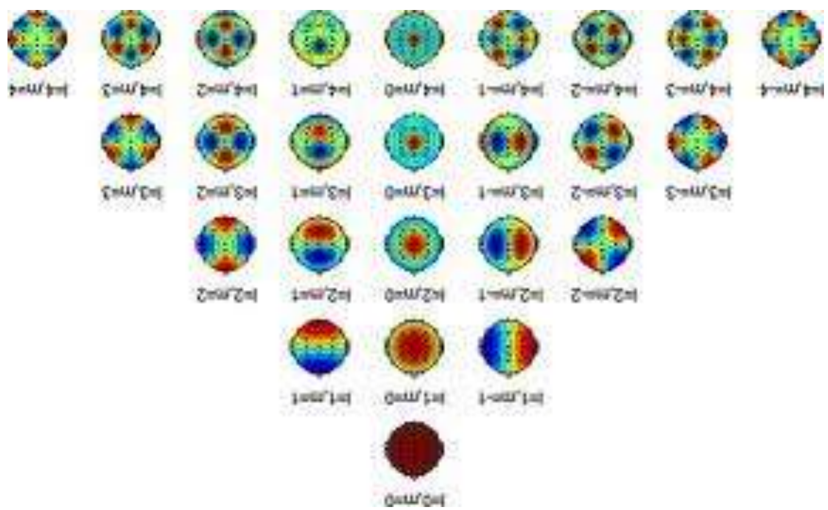
Düben and Palmer, 2014. MWR.



- 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
- Scale dependent precision easy to code in a spectral model (potential advantage for the future?)

Greater Accuracy with Less Precision?

Use freed-up computing resource to extend simulator to cloud resolved scales?



Decreasing precision, and determinism

The underpinning theoretical question

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

How to quantify the information content in the N-S variables as a function of scale (in the light of scale-dependent chaotic variability)?

Conclusions

- Forecasts acquire value through their ability to influence decisions.
- Stochastic parametrisation increases the value of forecasts.
- Because of inherent stochasticity in the sub-grid closure schemes, we are likely wasting valuable resources by computing precisely and deterministically.
- Taking account of the information content in NWP and climate model variables, we should be able to use freed-up computing resources to develop higher resolution models.
- This may be an important element in allowing ECMWF to achieve its strategic goal of developing a high-res (5km) ensemble system in the coming years.