An overview of the Isaac Newton Institute programme, “Mathematical and statistical approaches to climate modelling and prediction”

Kevin Horsburgh (with thanks to my co-organisers and all participants of CLP

Liverpool Marine Symposium. 17 January 2011

- Aims/goals of the programme
- What did we do?
- Emerging themes and findings
    - stochastic models
    - statistical frameworks

Isaac Newton Institute for Mathematical Sciences

Mathematical and Statistical Approaches to Climate Modelling and Prediction

11 August - 22 December 2010

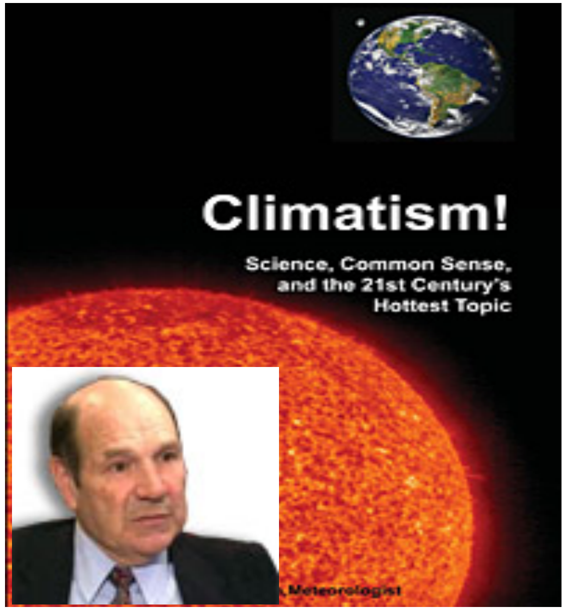
Organisers: Professor R Chandler *(University College London)*, Dr M Collins *(Met Office)*, Professor P Cox *(University of Exeter)*, Dr K Horsburgh *(Proudman Oceanographic Laboratory)*, Professor JM Huthnance *(Proudman Oceanographic Laboratory)*, Dr JC Rougier *(University of Bristol)*, Professor DB Stephenson *(University of Exeter)* and Professor J Thurburn *(University of Exeter)*

- The programme will complement NERC endeavours in climate research

- It is intended to bring together world-leading researchers in climate modelling, mathematics and statistics in order to make progress in solving some of the major issues facing climate prediction

Extreme positions on both sides of the climate debate arise because of the high degree of uncertainty surrounding the amount of future warming and the lack of detailed predictions at the regional level

"*The Northern Hemisphere winter is already proving once again that global warming is another undelivered government promise*". Dr Tim Ball, Dec 2010



Coldest Winter in 1000 Years Cometh – not.

A major goal for climate scientists is to provide credible and transparent assessments of model accuracy and other uncertainties. This requires the close collaboration of mathematicians, statisticians and climate scientists

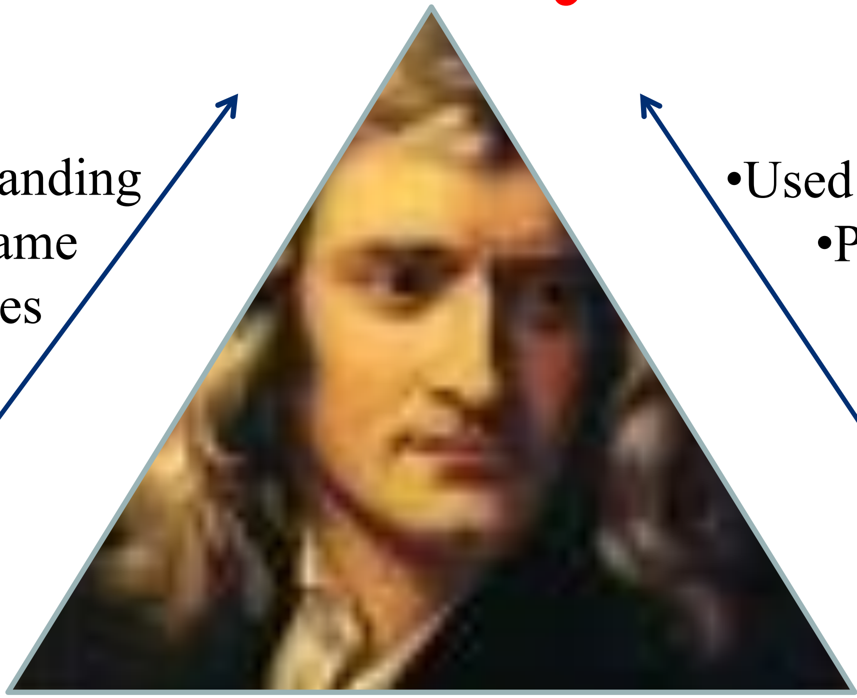
Overarching goals

- To bring together world-leading researchers in climate modelling, mathematics & statistics to face the challenges of climate prediction
- Provide a credible, systematically-derived, transparent assessment of uncertainty…..
- ….thus engage in an honest dialogue with policy makers and the public (by answering the question asked; and by pointing out when the question is unanswerable)

Reality

•Aids understanding
•Used to frame hypotheses

•Used to test hypotheses
•Policy relevant



Idealised mathematical models

Comprehensive mathematical models

INI has brought these various communities together

Specific aims of our programme

- To stimulate improved stochastic sub-grid-scale physics models
  - to improve the variability of ensemble climate simulations
  - improve the connection between deterministic models and statistical tools
  - establish frameworks for relating models of different resolutions
- Develop statistical techniques to provide a sound basis for probabilistic climate prediction
  - Synthesis of vast amounts of model data
  - quantitative answers to questions of interpreting probabilistic output
  - how best to use measurements to assess quality of climate model predictions (and initialisation – D/A)

The CLP programme in numbers

- 157 – number of participants
- 19 week programme
- 95 days of scientific interaction
- 150 – hours of formal presentations
- 2-5 new collaborations per participant (as indicated by individual final reports)
- Many potential new papers, grant applications
- In short, an unrivalled opportunity for cross discipline research on a large scale

What happened – workshops and outputs

- **August 23-27 workshop in Cambridge**
  - **“Stochastic methods in climate modelling”**
  - **Programme, abstracts, videos at**
  http://www.newton.ac.uk/programmes/CLP/clpw01p.html
  - **Some papers in *Phil Trans. Roy. Soc. Lond. A* special issue “Climate predictions: the influence of nonlinearity and randomness”**
- **September 20-23 workshop at University of Exeter**
  - **“Probablistic Climate Prediction”**
  - **Programme, abstracts, videos at**
  http://www.newton.ac.uk/programmes/CLP/clpw02p.html
  - **Emerging research questions (10) and discussion on Wiki**
- **November 24, Willis Research Network, London**
  - **“Climate Change Question Time” Open for Business event**
- **December 6-10 2010 – final workshop**
  - **Uncertainty in climate prediction: models, methods and decision support**

What happened – many discussion sessions

- Emulators and “particle filtering” September 2
- Palaeo-climate reconstruction and SUPRAnet September 15
- Tipping points October 18-21
- Allied RSS meeting October 22

“Complexity and statistics: tipping points and crashes”

- Multi-model ensembles / probabilistic climate projection October 26

Statistical aspects of the programme

- Role of phenomenological models in studying complex systems (Michael Ghil, Didier Paillard, Frank Kwasniok, Hank Dijkstra, Michel Crucifix, Arthur Dempster)
- Data assimilation with uncertain static parameters: statistical approach using Particle-MCMC (Jonty Rougier, Michel Crucifix).
- Approximations of the above for large problems (John Haslett, Vincent Garreta, Andrew Parnell, Nathan Urban, Richard Wilkinson, Neil Edwards)
- General Bayesian methods for palaeoclimate reconstruction (Caitlin Buck, John Haslett, Vincent Garreta, Andrew Parnell, Michel Crucifix, Jonty Rougier)
- Statistical models for multi-model ensembles (James Annan, Julia Hargreaves, David Stephenson, Bryson Bates, Richard Chandler, Jonty Rougier, Michael Goldstein)

Emerging themes and findings

- Statistical framework for handling uncertainty
- Probabilistic Projections
- Emulators, Accelerated Bayesian Computation
- Paeleo-climate and proxy data
- Tipping points
- Stochastic parameterisations and turbulence spectrum in GCMs
- Data assimilation for decadal climate prediction
- Maximum Entropy Production

From UKCP09

UK CLIMATE PROJECTIONS

"The mechanisms for atmospheric blocking are only partially understood, but it is clear that there are complex motions, involving meso-scale atmospheric turbulence, and interactions that climate-resolution models may not be able to represent fully."

"In developing the UKCP09 projections it was decided not to include probabilistic projections for future wind due to the high level of associated uncertainty."

Eg $\quad + \quad p \quad ^{2}$

$$X_1 \quad X_2 \quad X_3 \quad \ldots \qquad\qquad \ldots \quad X_n$$

Increasing scale

e.g. momentum transport by:

- Turbulent eddies in boundary layer
- Orographic gravity wave drag.
- Convective clouds

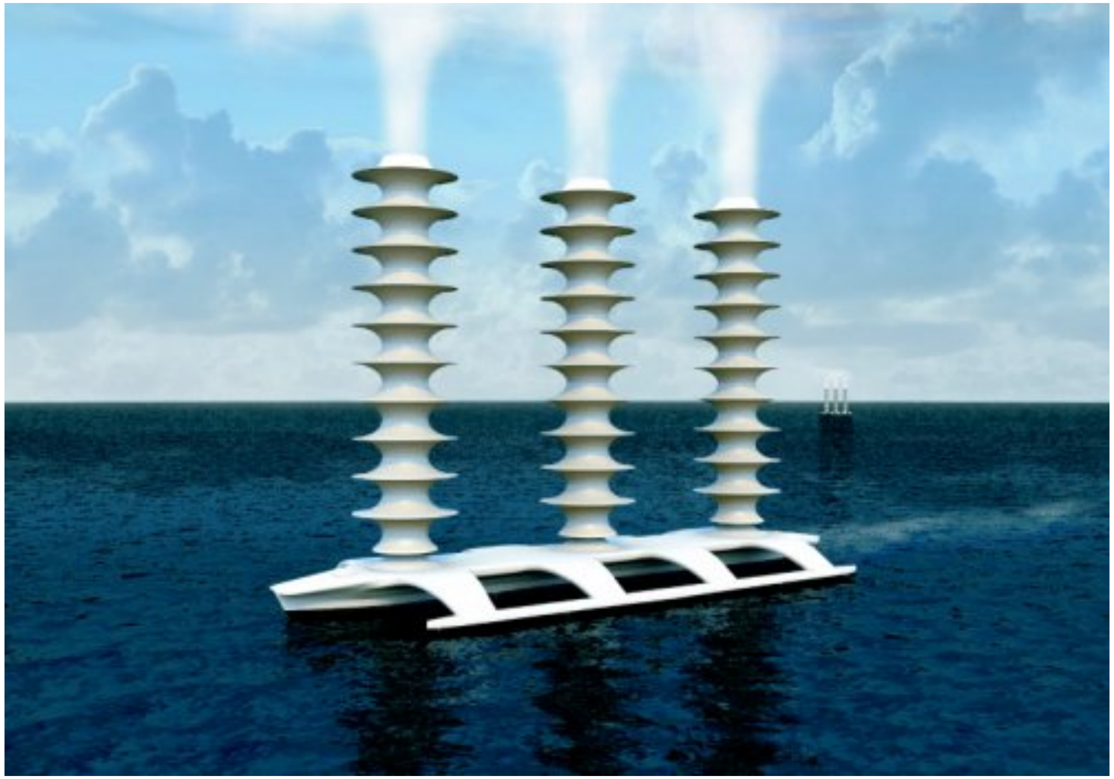
$$P(X_n; \in)$$

Local bulk-formula parameterisation

Will future UK offshore winds be reliably strong enough to provide projected energy needs from renewables?

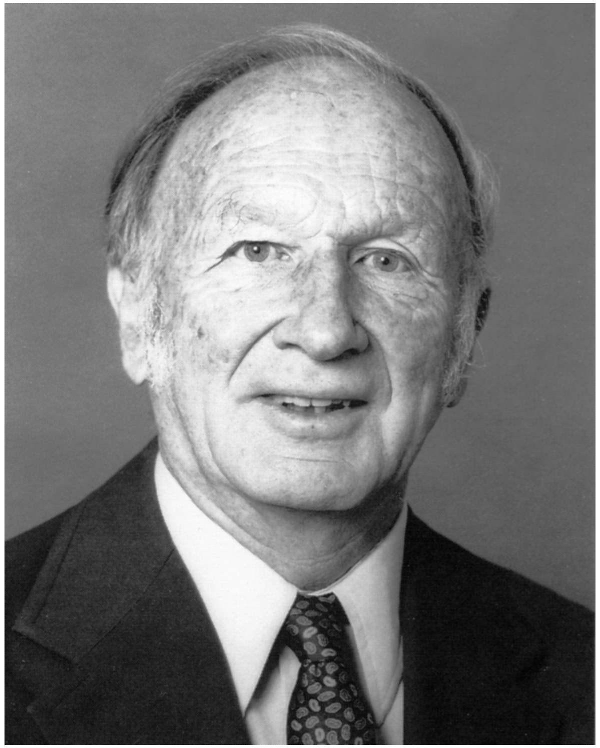
Primitive equation models are required to resolve the variability of planetary waves and the processes leading to blocking anticyclones

…and also to evaluate consequences of geoengineering proposals



Permanent El Nino, alteration of monsoon patterns?

"*I believe that the ultimate climate models..will be stochastic, ie random numbers will appear somewhere in the time derivatives*" Lorenz 1975.



Stochastic-dynamic approaches to probabilistic Earth-system modelling

Computationally-cheap nonlinear stochastic-dynamic models (potentially on a secondary grid) providing specific realisations of sub-grid motions rather than sub-grid bulk effects.

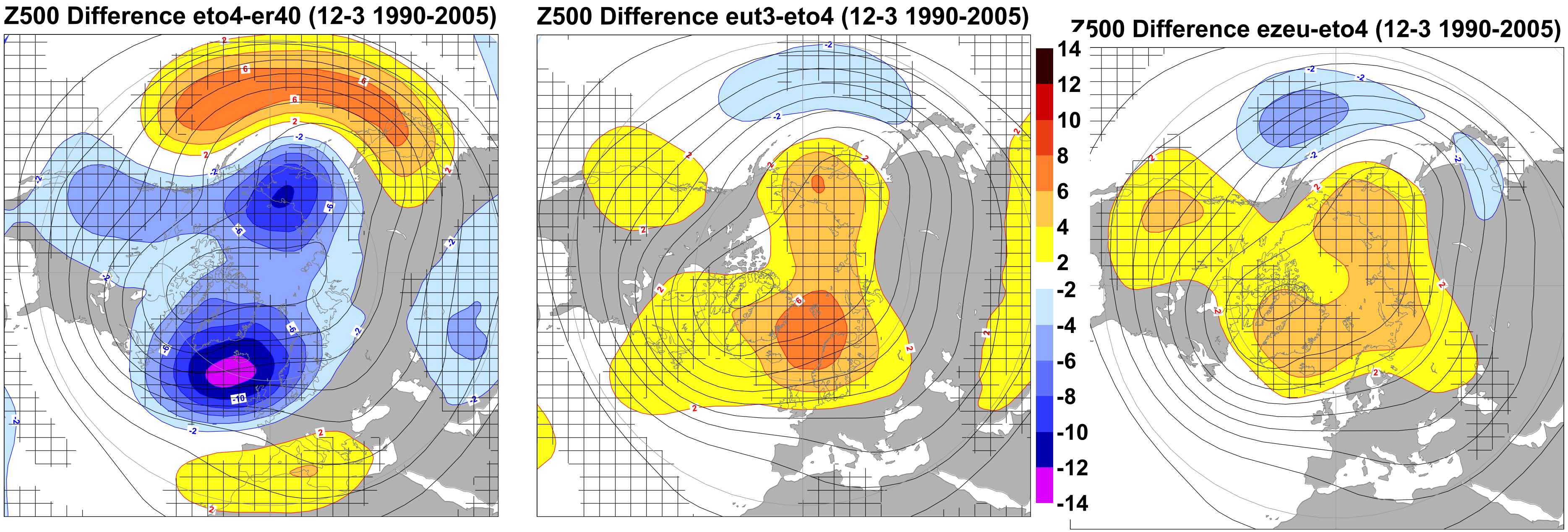
Potentially coupled over a range of scales

(Palmer, 1997; 2001)

Examples :

- Multiplicative Noise (Stochastically Perturbed Parametriatsion Tendencies; SPPT - Buizza et al, 1999)
- Stochastic Backscatter (Stochastic Spectral Backscatter Scheme; SPBS, Shutts, 2005, Berner et al 2010)
- Cellular Automata (Palmer 1997, Berner et al 2010)
- Stochastic lattice models (Majda et al, 2010)
- Dual grid, stochastic mode reduction (Majda et al, 2010: Allen et al, 2010)
- Statistical mechanics of finite sized cloud ensembles (Plant and Craig 2008)

CNT_{T95}-ERA40 | CNT_{T511}-CNT_{T95} | $SPBS_{T95}$-CNT_{T95}



T95 | T511 | T95+Stochastic parametrisation

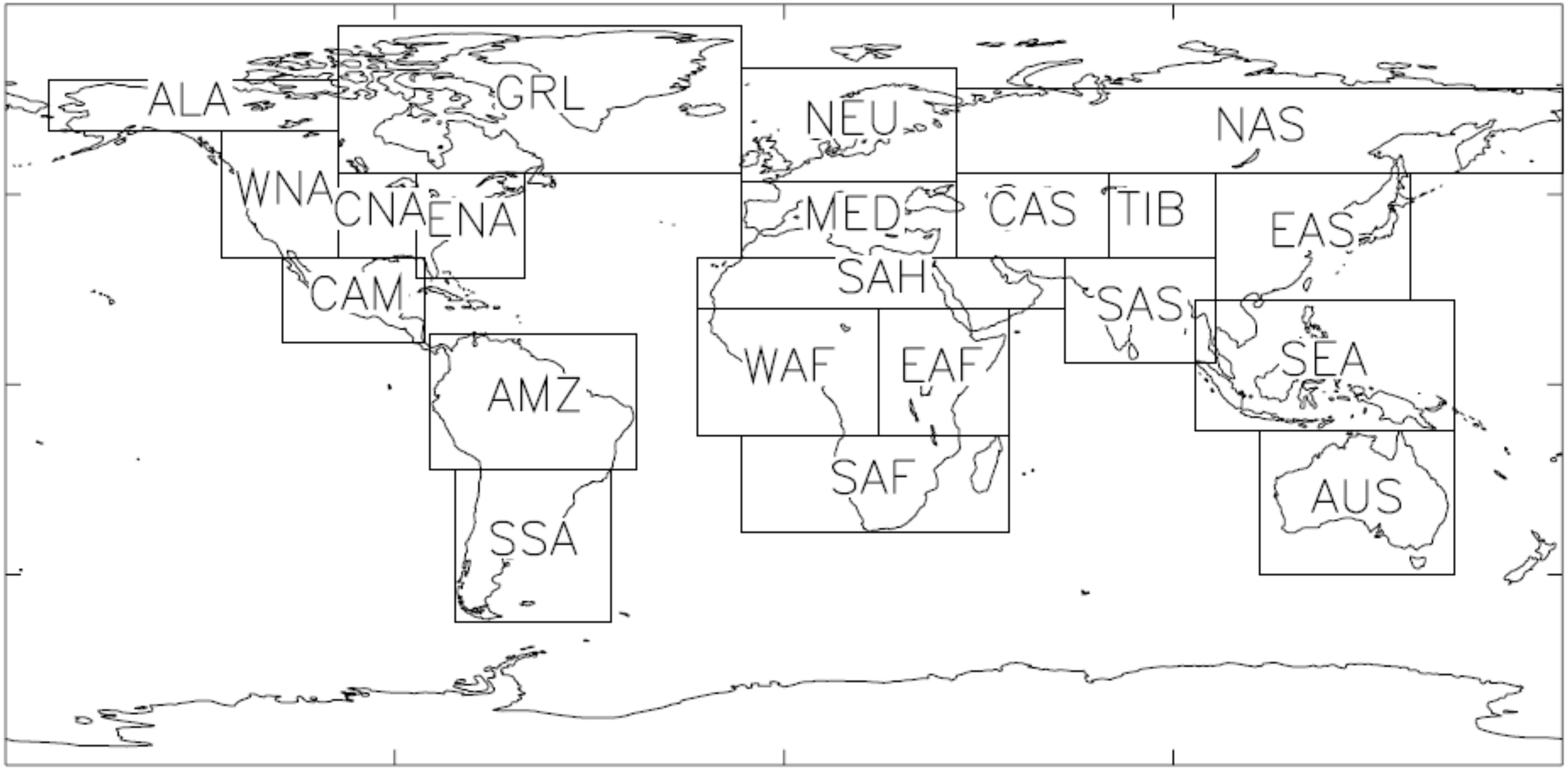
- Experiments with Berner et al (JAS 2009) stochastic backscatter scheme (partially conducted during CLP programme)
- Winters (Dec-Mar) of the period 1990-2005

Comparison of the BSS(∞) for precipitation over land regions: ENSEMBLES multi-model ensemble (MM), perturbed parameter ensemble (PP), ECMWF stochastic physics ensemble (SP) and ECMWF control ensemble (noSP)

	precipitation			
	JJA		DJF	
	dry	wet	dry	wet
Australia	1	2	2	3
Amazon Basin	3	0	6	6
Southern South America	1	1	1	1
Central America	2	3	3	2
Western North America	3	3	6	3
Central North America	6	1	3	3
Eastern North America	1	1	2	3
Alaska	3	1	2	3
Greenland	1	6	2	3
Mediterranean	3	3	3	3
Northern Europe	2	2	3	3
Western Africa	3	1	2	3
Eastern Africa	3	3	2	2
Southern Africa	3	3	3	2
Sahel	1	3	2	6
South East Asia	1	1	1	1
East Asia	3	3	3	3
South Asia	3	3	3	3
Central Asia	2	3	2	2
Tibet	1	1	6	6
North Asia	3	2	1	1

MM best	20	24%
PP best	18	21%
SP best	38	45%
no SP bes	8	10%
	84	

M Weisheimer (ECMWF), Work in progress

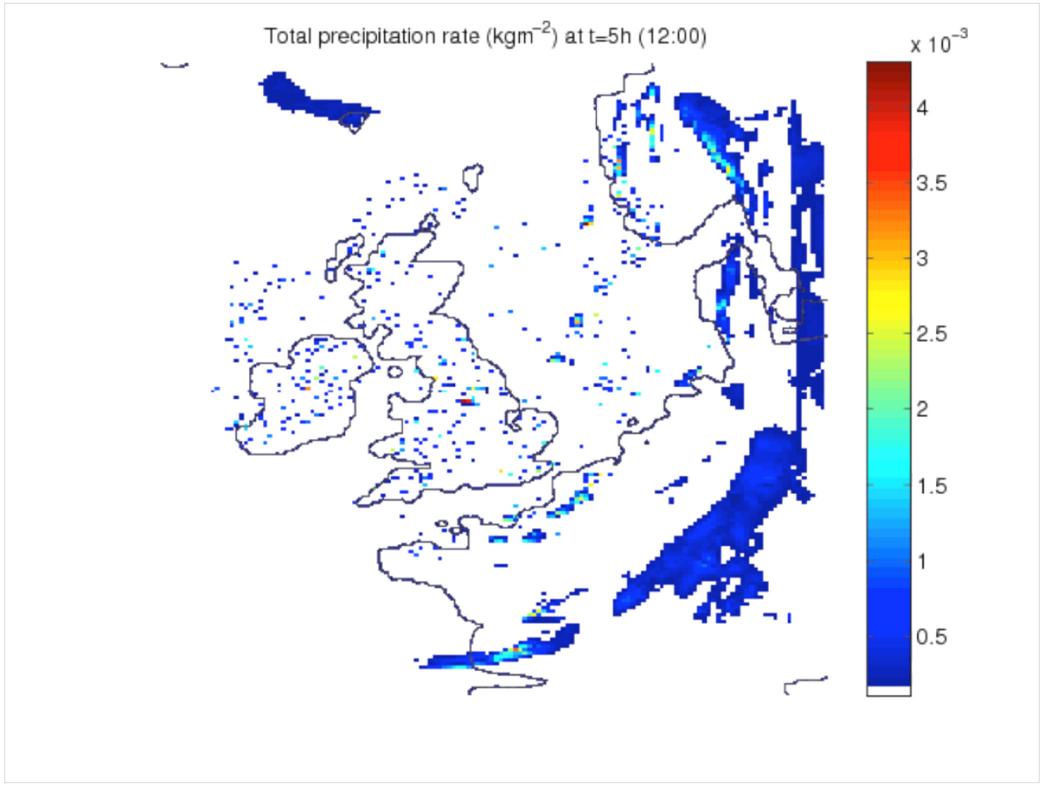


Regions from Giorgi and Francisco, 2000

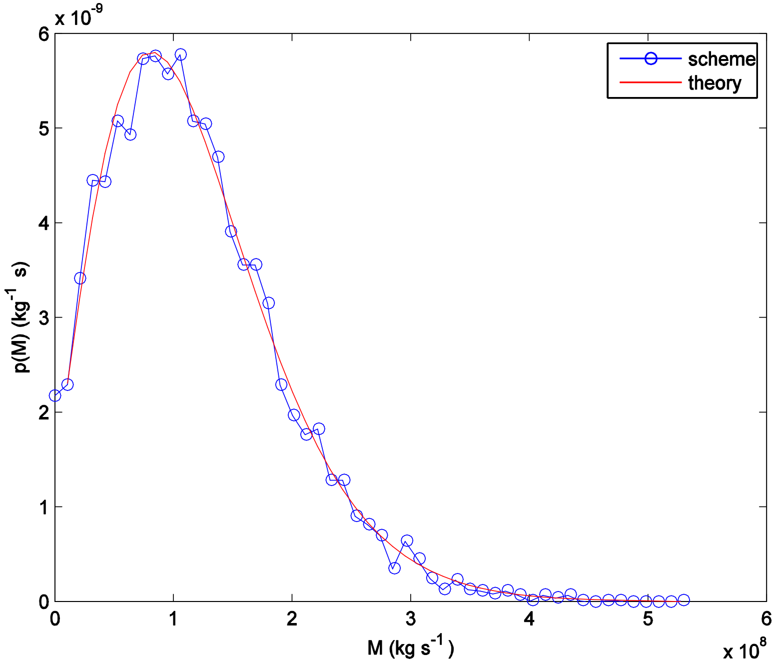
Comparison of the BSS(∞) for temperature over land regions: ENSEMBLES multi-model ensemble (MM), perturbed parameter ensemble (PP), ECMWF stochastic physics ensemble (SP) and ECMWF control ensemble (noSP)

	temperature							
	JJA		DJF					
	cold	warm	cold	warm				
Australia	3	3	3	1		MM best	32	38%
Amazon Basin	3	1	1	1		PP best	19	23%
Southern South America	1	1	3	2		SP best	25	30%
Central America	3	1	3	1		no SP bes	8	10%
Western North America	1	1	6	6			84	
Central North America	1	1	2	2				
Eastern North America	3	3	2	3				
Alaska	3	3	2	2				
Greenland	1	1	2	2				
Mediterranean	3	6	1	6				
Northern Europe	2	2	3	3				
Western Africa	1	1	3	3				
Eastern Africa	1	1	2	3				
Southern Africa	1	2	1	1				
Sahel	1	2	6	6				
South East Asia	1	1	2	1				
East Asia	3	2	1	6				
South Asia	3	1	3	3				
Central Asia	1	2	3	2				
Tibet	3	1	3	6				
North Asia	1	2	1	2				

Stochastic parameterisation for convection



Example snapshot of precipitation rate, Met Office model, with scattered convection (randomly distributed)



Pdf of convective mass flux, shows range of convective responses over an area (64km)^2

Plant-Craig scheme behaviour c.f. distribution from equilibrium theory

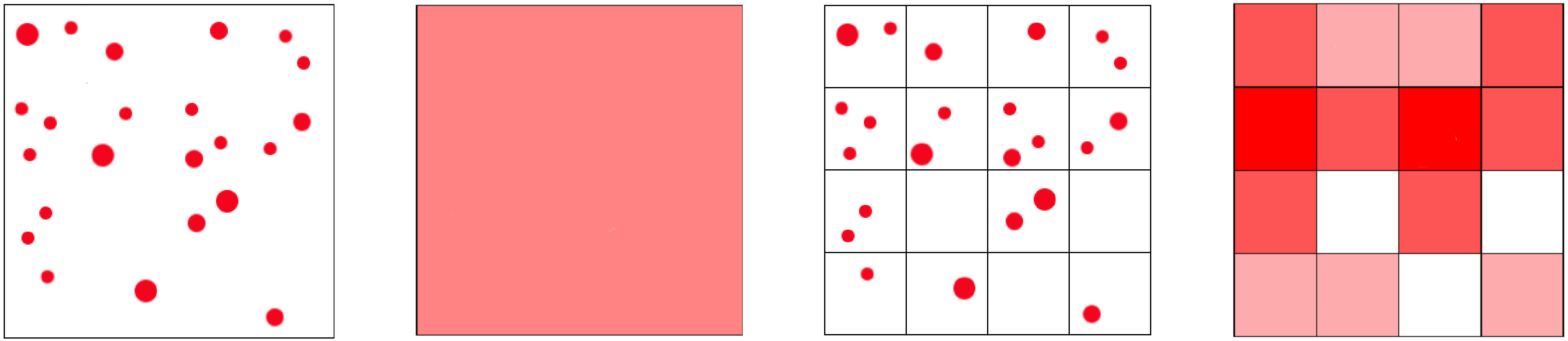
Stochastic parameterisation for convection [Bob Plant]

<u>Background</u>

Few clouds per grid box (depends on size and forcing), maybe 1 or 2
=> actual no. of clouds & convection in box unpredictable

Convective equilibrium: theory for mean strength & fluctuations

Plant-Craig parameterisation enacts :
stochastic: models <u>fluctuations</u> (and mean response).
(hitherto in single-column models & ensemble weather forecasting)

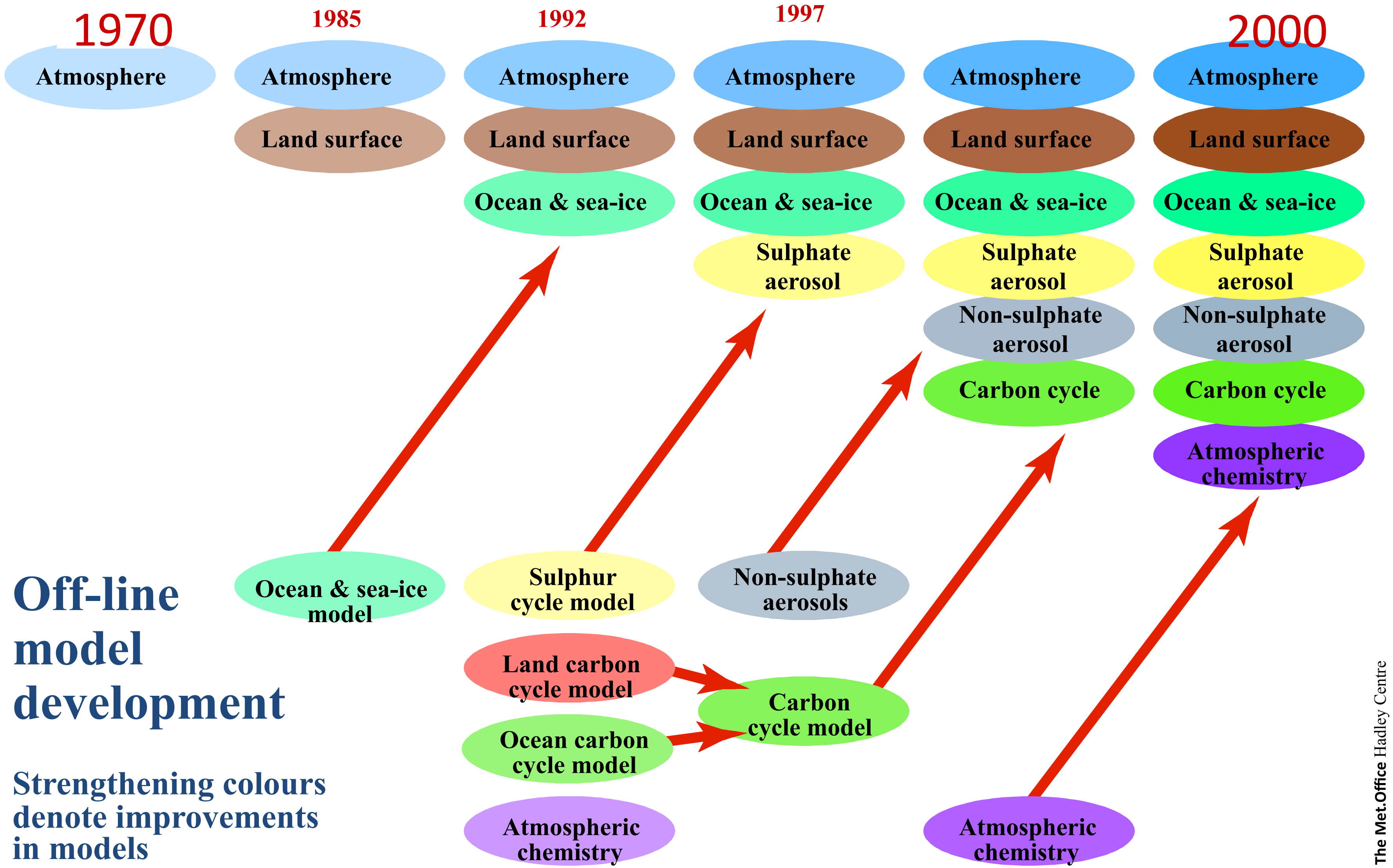


clouds vary; no effect if coarse GCM grid; finer grid needs stochastic parameterisation

Emulation as parametrisation [Peter Challenor, NOC]

- We have a GCM that cannot resolve sub-grid scale processes
- But we have a process model for these processes
- This is too slow to embed in the GCM
- Build an emulator for the process model
- Embed this in the GCM (in a deterministic or a stochastic way)

Towards Comprehensive Earth System Models



A community-wide approach to the Climate Model development?

Standard argument against the “Airbus paradigm”

“We need model diversity in order to be able to estimate prediction uncertainty”

However, development of a skilful Probabilistic Climate Model weakens this argument, opening the door to greater integration of climate model development, and to much more efficient use of the enormous human and computational resources needed to develop reliable climate prediction models.

Climate and its interpretation

- The term “climate” has undergone a number of meanings, in different disciplines and at different times
- Alexander von Humbolt (1845) described climate in holistic terms that included the impact of the total environment on peoples and their culture.
- In the late 19th century climate, was increasingly characterised as the average condition of the atmosphere at any point. Climate was the statistics of weather calculated over long periods of time, and usually for larger geographic areas. Climate research was classification of regional averages. In this era, global climate was broadly the sum of all regional climates.

- As observing technologies improved throughout the 20th century, climatology emerged as the science of climatological processes and, as a discipline, the balance fell away from geography and towards physics. Regional climate was now more correctly thought of as interaction of the global climate with regional physical geography

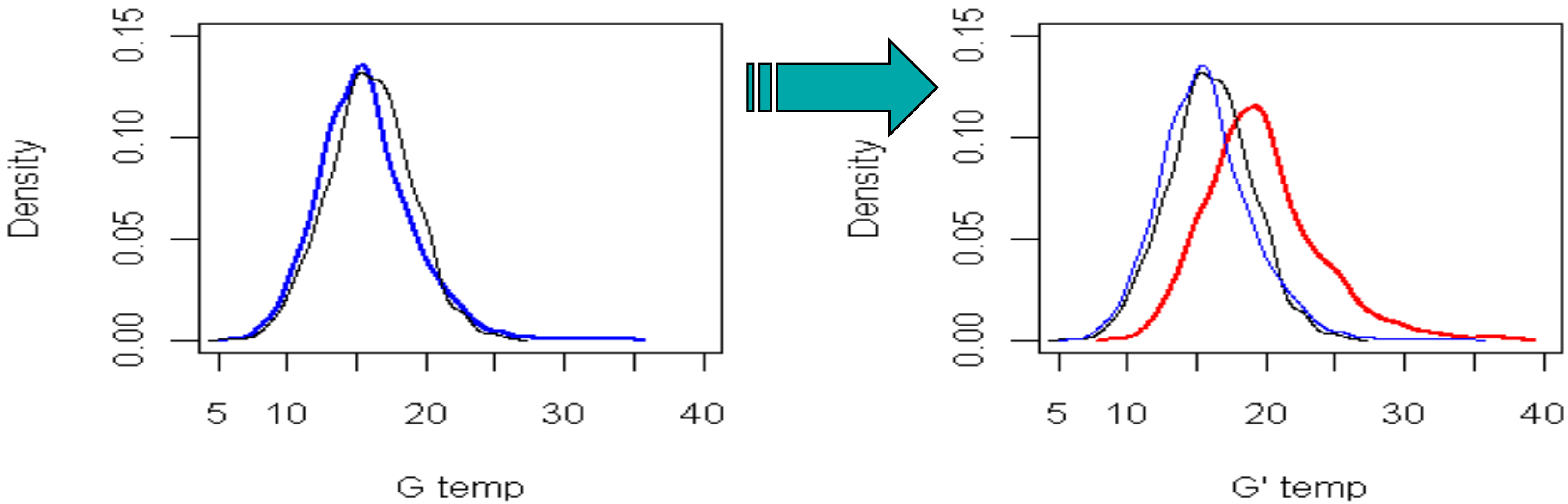
- Most recently we think of the climate system as the physical-biogeochemical system (atmosphere, ocean, cryosphere, vegetation, biogeochemical cycles) which generates the time-variable statistics of the Earth system. Only in this latter phase have dynamical models aspired to explicitly represent all of the aforementioned components. This has allowed the models to become tools for political and environmental debate.

What is climate?

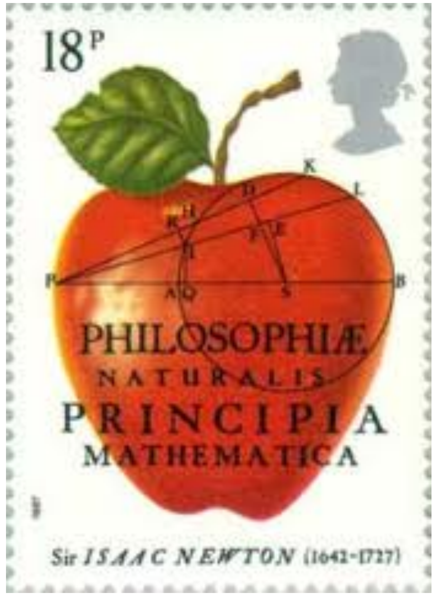
- We cannot observe the climate; we merely observe the states of the various components of the climate system at various instants in time
- Climate is:
  - the invariant measure in a hypothetical thought experiment in which the forcing is held constant indefinitely
  - what is obtained by taking the joint PDF for the state and the parameters and then integrating it over the parameters
  - the PDF of the current system state given all the historical observations
- Current GCMs are not configured to produce PDFs, and hence additional statistical assumptions are required to compute climate as defined in these terms

- Given a forward model of the climate system, a prior on the initial variables, a prior on the parameters (e.g. expert knowledge and measurements: the climate is could be defined as the joint *posterior density of the variables and parameters conditioned on the model and the available measurements.*

Black line = pdf of obs data 1970-1999
Blue line = pdf of climate data 1970-1999
Red line = pdf of climate data 2070-2099



The Multi-Model Ensemble



A “fruit bowl of opportunity” $\{X_1, X_2, \ldots, X_m\}$
Note: Not a random sample from one homogeneous population
(and it does not include all possible fruit!)

What does reality look like?

An inconvenient truth

true climate Y – inferred from observations Z
It could not have been drawn out of the fruit bowl

How can we infer properties of this from the fruit in the fruitbowl?

Smoothies (multi-model means)

$\neq$

A smoothie is a weighted average of fruits.
- **It is not an item of real fruit!**

(important information has been lost by averaging)

- **Non-unique choice of weights for making smoothies.**

→We require modelling frameworks for obtaining samples of real fruit from the posterior distribution p(Y|X) (not smoothies E (X) or E(X|Y)).

Homogeous sub-samples

How to relate Y to X?

- Are the $\{X_i\}$ independent draws from a distribution centred on Y?
- Are the $\{X_i\}$ second-order exchangeable with each other and Y?
- How best to model model discrepancy $Y-X_i$?

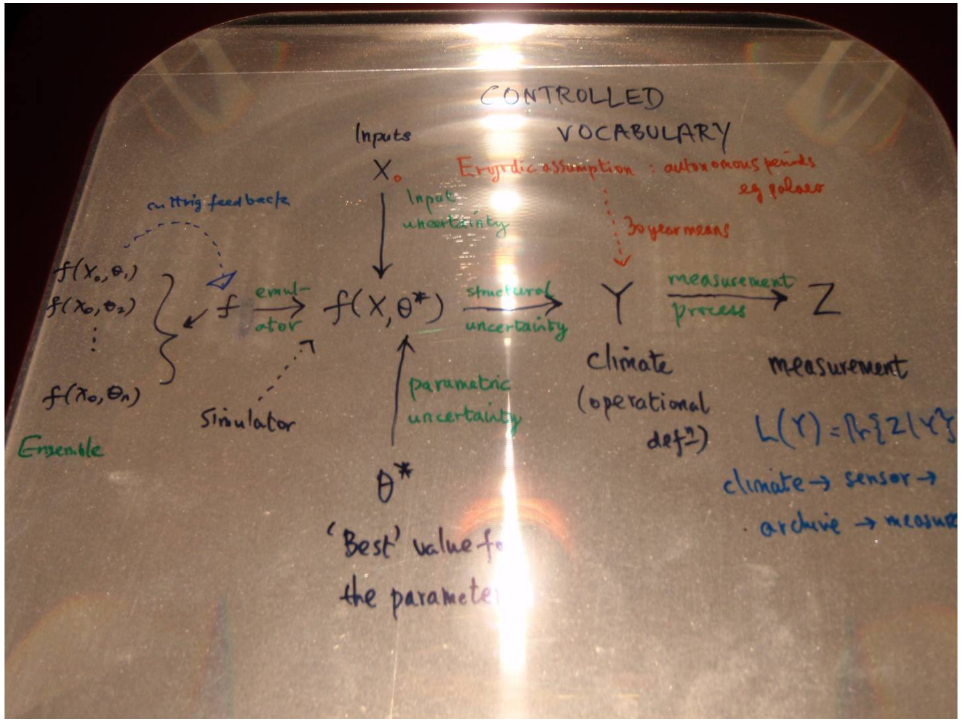
"All fruit are equal, but some are are more equal than others" – Granny Orwell

Statistical frameworks

We need more than weights – we need a credible statistical model for the distribution of real climate Y given information from a multi-model ensemble

$$X' = \sum_{i=1}^{m} w_i X_i$$

- Incorrect to interpret weights w as probabilities for each model
- Need to know dependencies in X_i to find variance of X'
- X' does not equal real climate Y
- require a statistical model for the distribution Y|X'



Y = real climate
X_i = climate model run
Z = observation of climate

Uncertainty within statistical frameworks

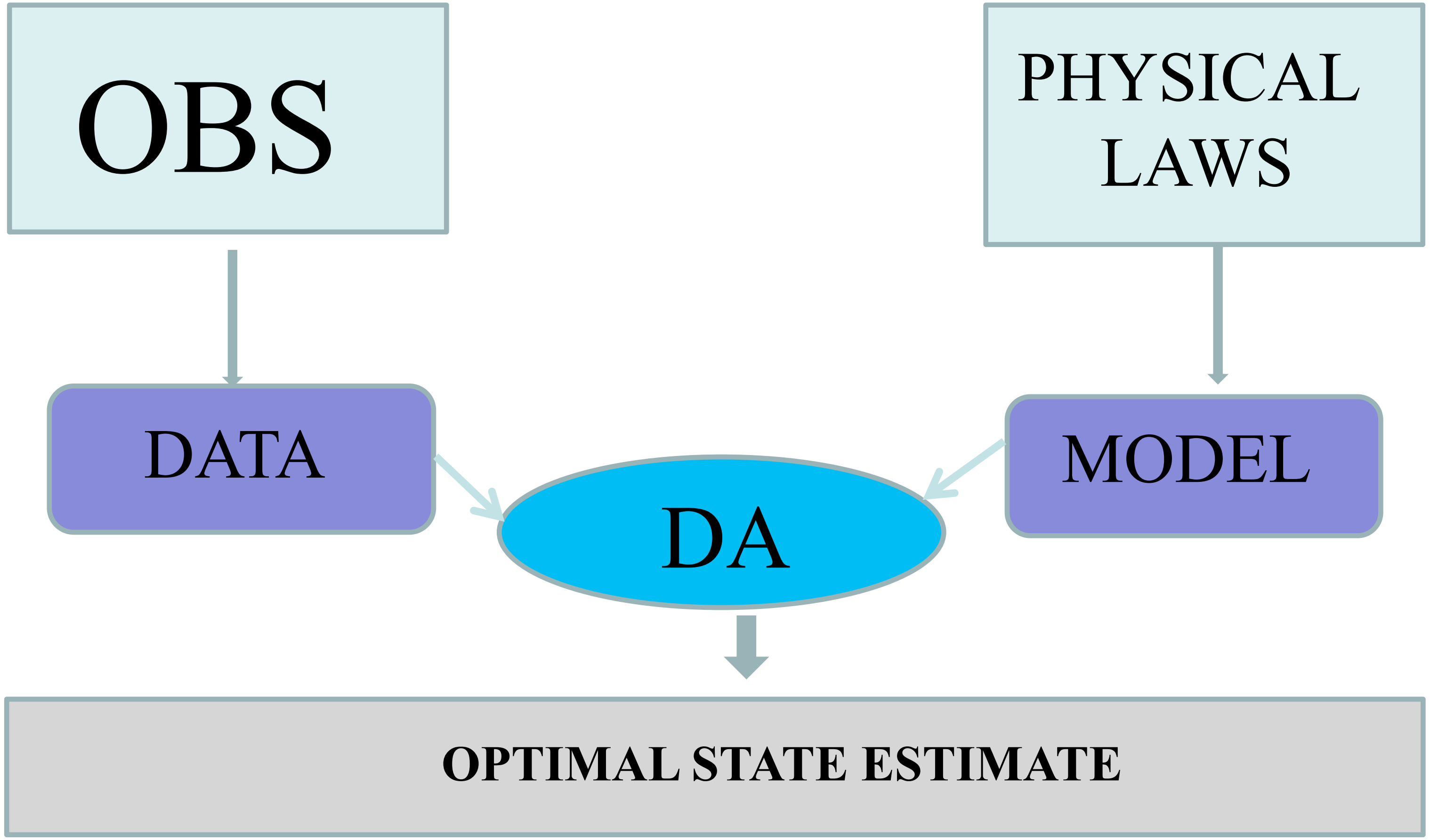
- What are relevant / dominant sources of uncertainty in any application?
- Which climate simulators to include in any study?
- How to handle fundamental simulator deficiencies?
- How to weight information from multiple simulators?
- Different conceptual frameworks e.g. “population” of simulators versus exchangeable set (operationally probably similar however)
- How to ensure decision-relevant probabilistic interpretation of information?
- How to provide uncertainty assessments in form suitable for downstream / stakeholder use?

Models of complex systems

- Using complex simulators to understand physical systems raises many technical and conceptual questions about the uncertainty introduced by moving from the model to the system
- Methodologies to deal with these difficulties have been devised under the Managing Uncertainties in Complex Models project (http://www.mucm.ac.uk)
- Understanding climate inevitably requires “best expert judgement” in a probabilistic framework
  - Probabilistic estimates from an individual (subjective Bayes analysis)
  - A probabilistic analysis so compelling it would demand agreement from all experts (objective bayes analysis)
- Advanced statistics provides the framework for simulator analysis, inverse modelling and expert judgement

Data Assimilation (DA)

Aim: to combine limited observations with our Knowledge of the laws of physics for optimal state estimation



DA and Climate

- ❑ DA developed in context of Numerical Weather Prediction
- ❑ Starting to be used in paleoclimate studies and in oceanography.
- ❑ Challenges to full implementation in Earth System Models:

Find the state x given observations y_i by minimising

$$J(x_0) = \frac{1}{2}(x_0 - x_b)^T B^{-1}(x_0 - x_b)$$
$$+ \frac{1}{2}\sum_{i=0}^{n}\left[H(x_i) - y_i\right]^T R_i^{-1}\left[H(x_i) - y_i\right]$$
$$+ \frac{1}{2}\sum_{i=0}^{n}\left[x_i - M(x_{i-1})\right]^T Q_i^{-1}\left[x_i - M(x_{i-1})\right]$$

- Multiple time scales
- High dimensions
- Sparsity of data
- Nonlinearity of climate processes
- Model Uncertainty

BENEFITS

1. Reduction in prediction uncertainty due to improved estimation of initial state, especially for seasonal and decadal prediction
2. Sophisticated diagnostics to reduce model errors- can isolate errors in physical processes
3. Sophisticated diagnostics to reduce key uncertainties in climate change e.g. due to cloud or ice feedbacks

Final remarks

- The programme has led to
  - (i) a common understanding of need for a joint effort of statisticians and climate modellers to form climate projections with uncertainties and usable by stakeholders
  - (ii) better agreement over concepts
  - (iii) a common sense of direction.
- Acceptance and a raised profile of modelling uncertainty in climate science, the need to test models, ideas to deal with uncertainty, tightening vocabulary, clarifying issues
- Impetus to think about fundamentals of climate modelling and combining state-of-the-art statistics and modelling (projects instigated).
- We see the combination of state-of-the-art statistics and modelling as the way forward for climate science and its use for policy and investment decisions.

Developing the CLP approach

- More networking programmes such as this one
- Better recognition of the important role of statistical and stochastic modelling by climate centres – it is a fundamental part of the prediction system
- More incentivisation for statisticians to be involved in climate science – there's a skill shortage of statisticians in climate science that needs to be addressed;
- Improved funding mechanisms for joint climate-statistical research e.g. coordinated joint funding by NERC and EPSRC. For example, a funding call for complexity mathematics in climate science would be ground-breaking

Emulators

- An emulator is a statistical approximation to the original simulator (Gaussian Process, or Bayesian NN)
- It gives us an estimate of what the simulator would give plus a measure of uncertainty
- It is very fast

1. Run a designed experiment with the simulator
2. Build the emulator (incl diagnostic checks)
3. Use the emulator for statistical inference on the simulator outputs