UNIVERSITY OF EXETER

Dealing with Model Uncertainties in Climate Change

Mat Collins, Exeter Climate Systems, College of Engineering, Mathematics and Physical Sciences, University of Exeter, with input from many colleagues

Models, models, models

• The Coupled Model Intercomparison Project (CMIP) collects output from the climate models from all over the world and provide web access
• This has become the "gold standard" in assessing uncertainties in projections
• CMIP5 will produce 2.3 Pb data*
• However, should all models be treated as equally likely?
• Is the sample somehow representative of the "true" uncertainty?
• Could there be surprises and unknown unknowns?

UNIVERSITY OF EXETER

* 25 years to download on home broadband link, Internet archive = 3Pb, LHC will produce 15 Pb of data, World of Warcraft = 1.3 Pb (source Wikipedia).

Uncertainties in Models and Projections

- Models have "errors" i.e. when simulating present-day climate and climate change, there is a mismatch between the model and the observations
- Differences in model formulation can lead to differences in climate change feedbacks
- Cannot post-process projections to correct errors

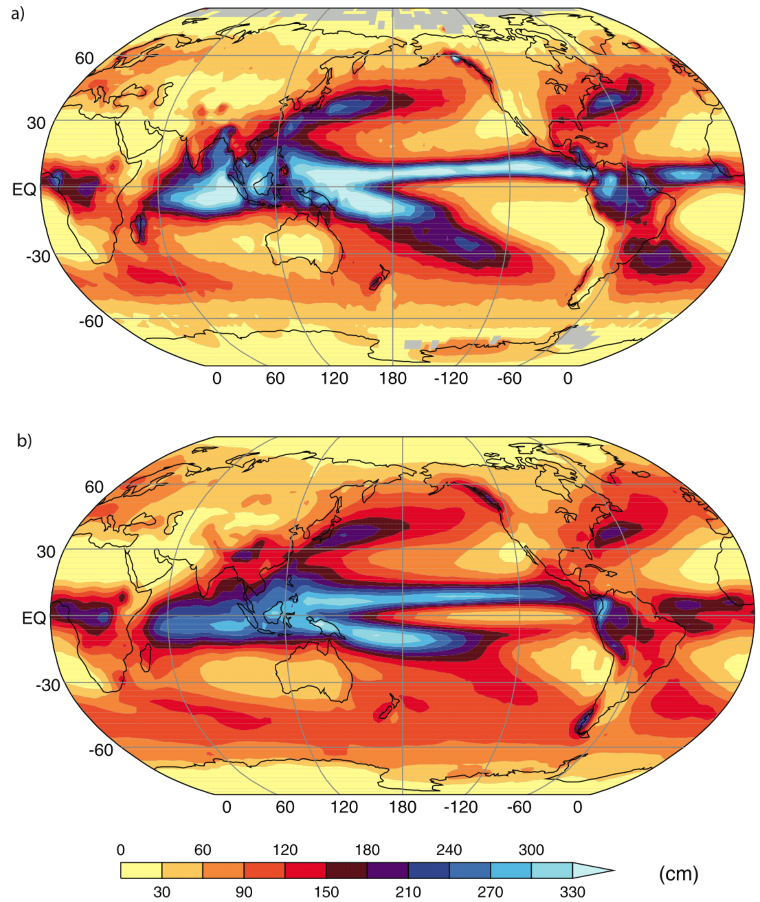


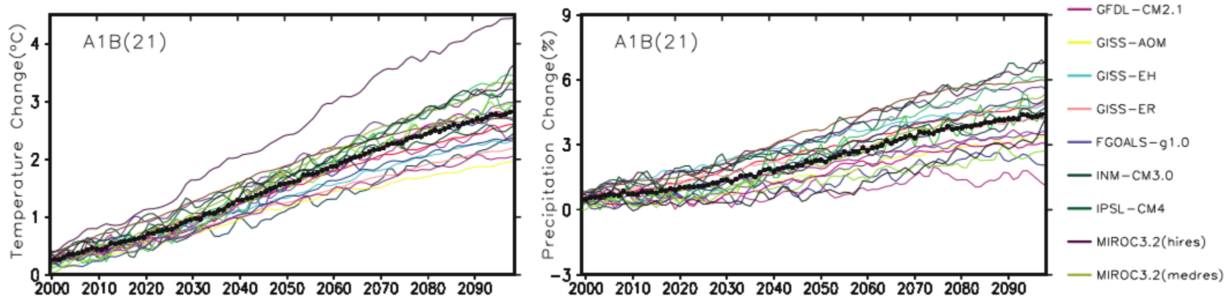
Figure 8.5

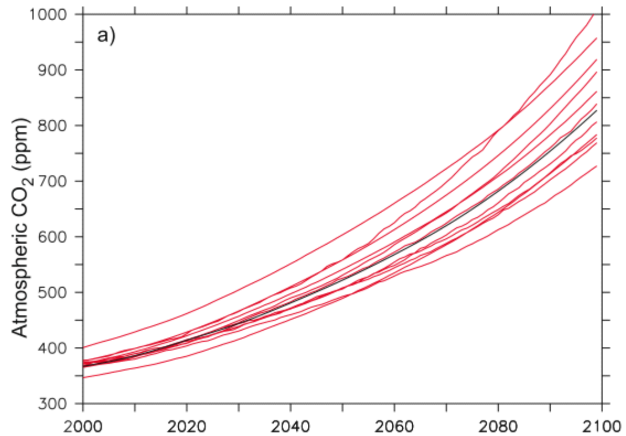
UNIVERSITY OF EXETER

Source: IPCC Fourth Assessment Report

Uncertainties in Models and Projections

Global mean projections from different models usi the same GHG concentrations are different



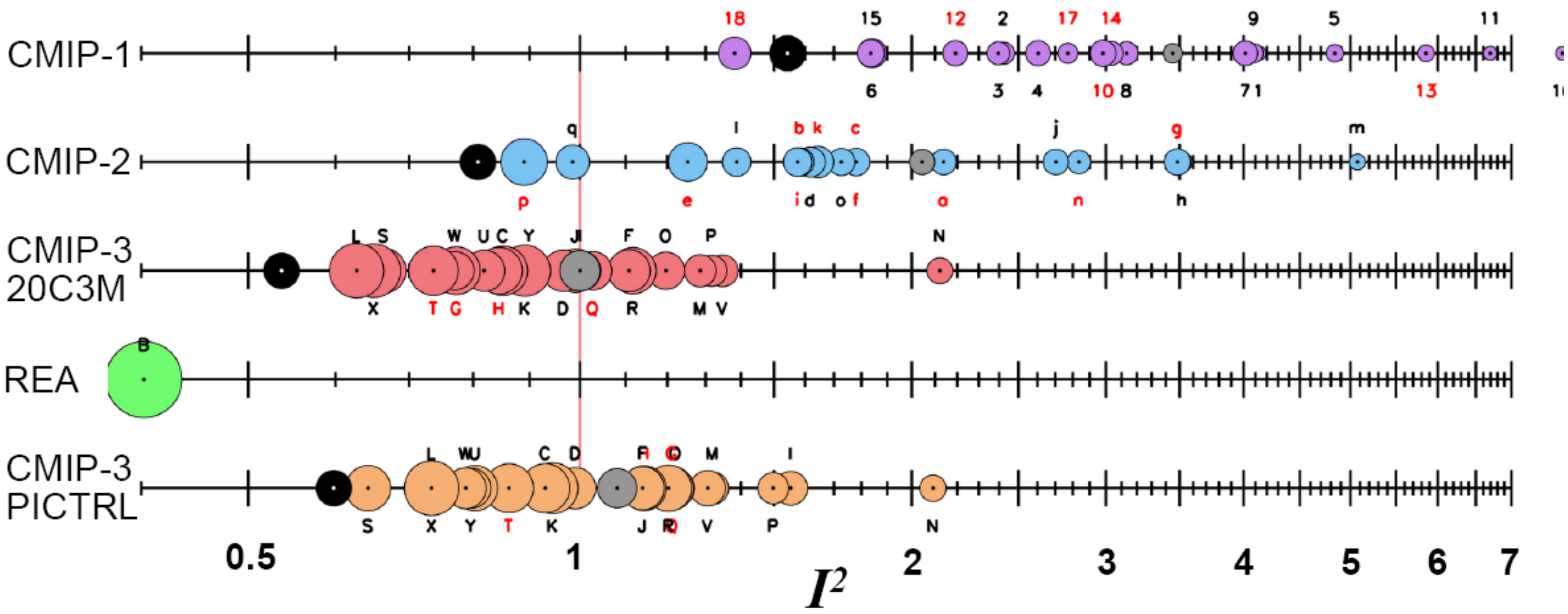


Global mean carbon cycle feedbacks from different models using the same GHG emissions are different

UNIVERSITY OF EXETER

Source: IPCC Fourth Assessment Report

Metrics, Metrics, Metrics

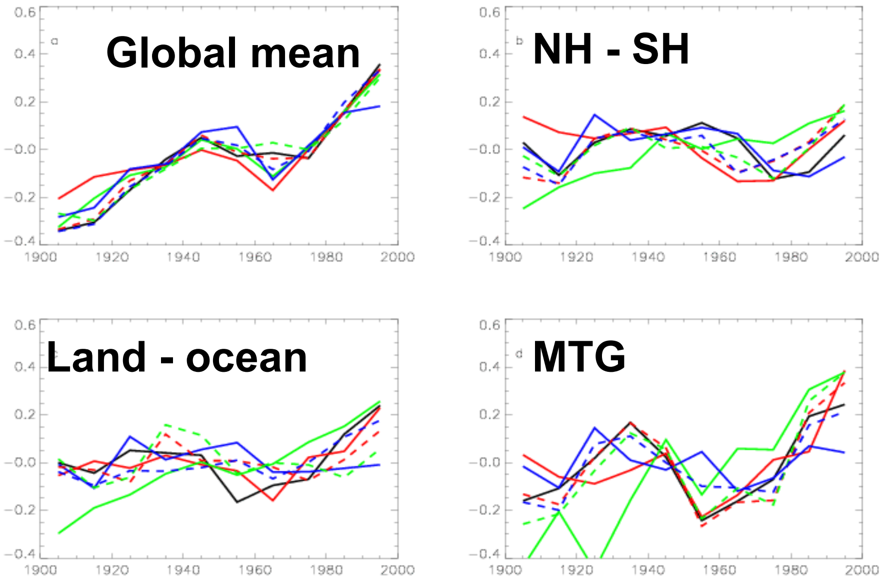


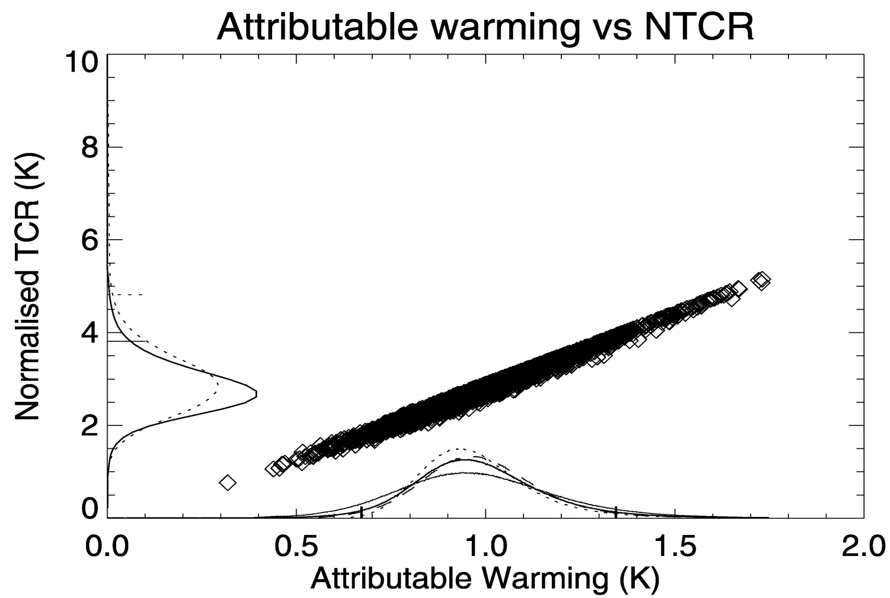
Reichler and Kim, 2008

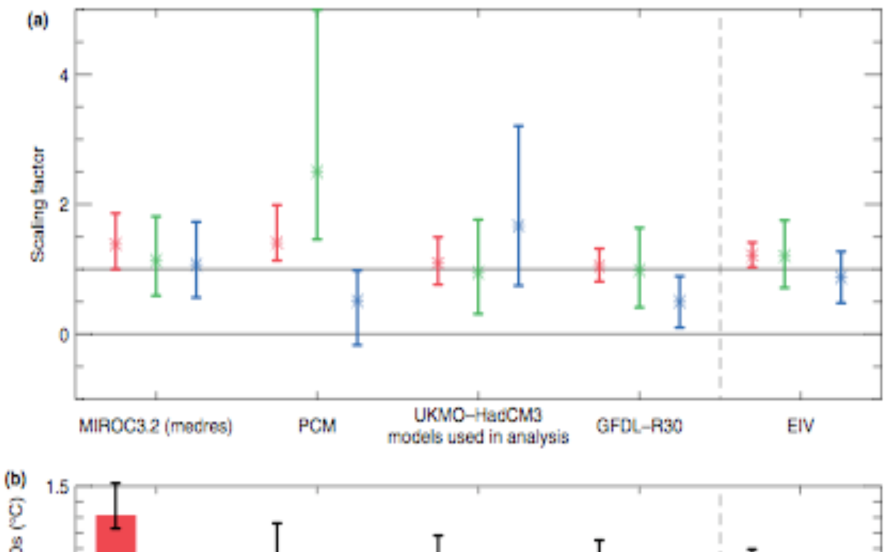
Techniques for Projections
(projection = dependent on emissions scenario)

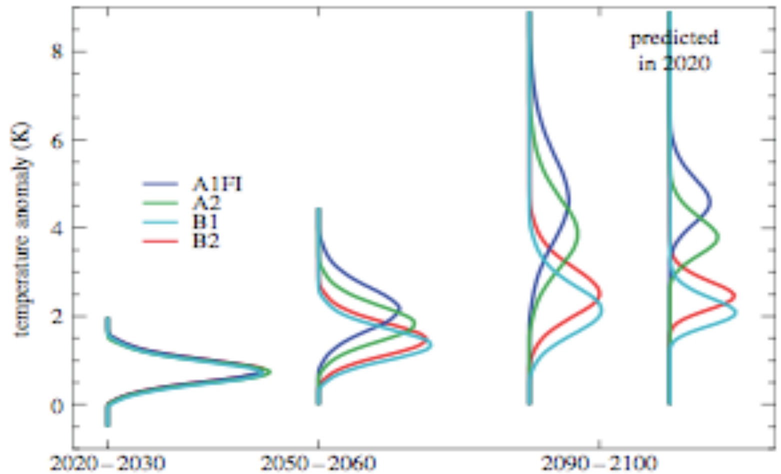
- Extrapolation of signals (e.g. ASK)
- The meaning of simple ensemble averaging
- Emergent constraints
- Single-model Bayesian approaches
- Strengths and weaknesses of different approaches
- Challenges

Robust quantification of contributors to past temperature change enables quantification of likely future rates of warming (ASK)

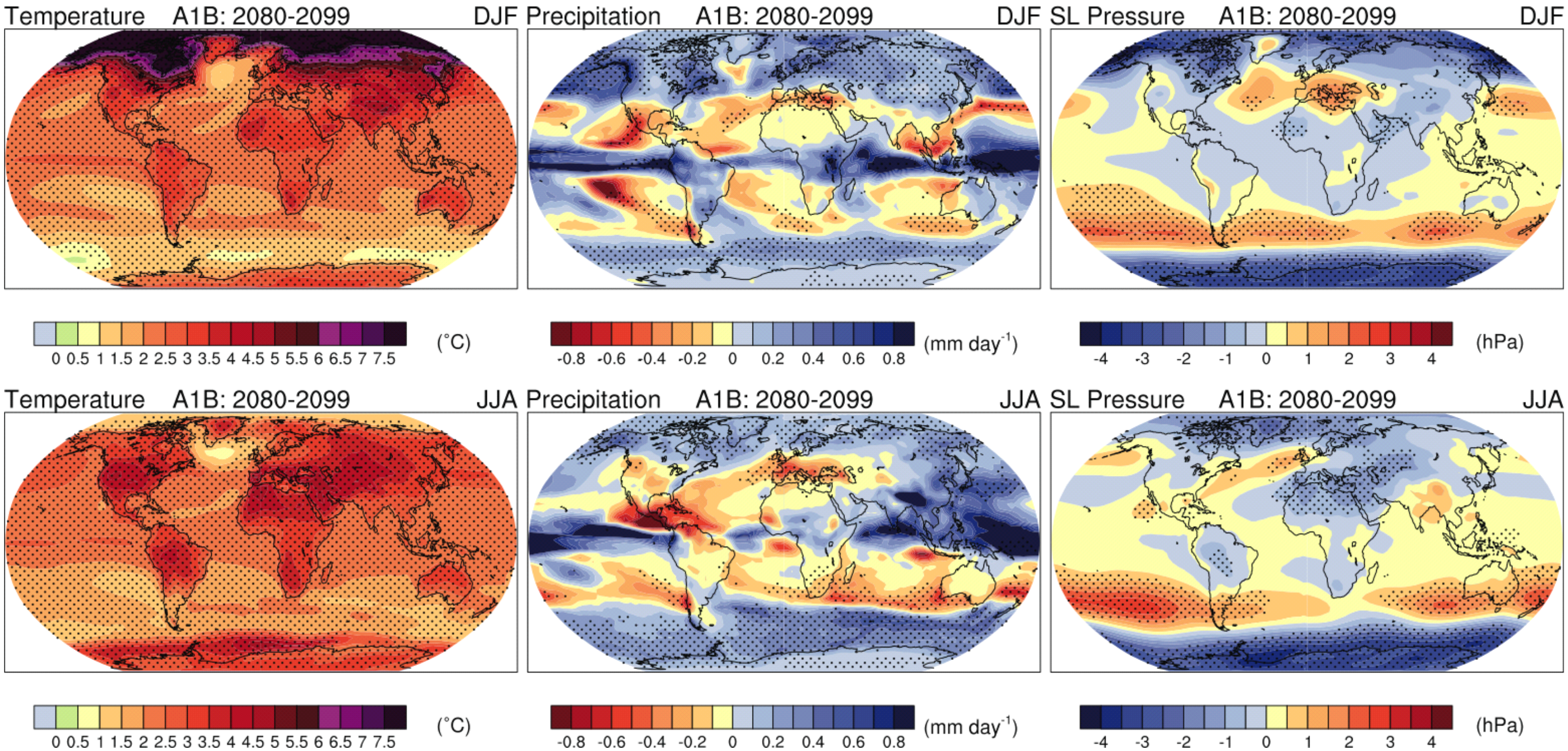






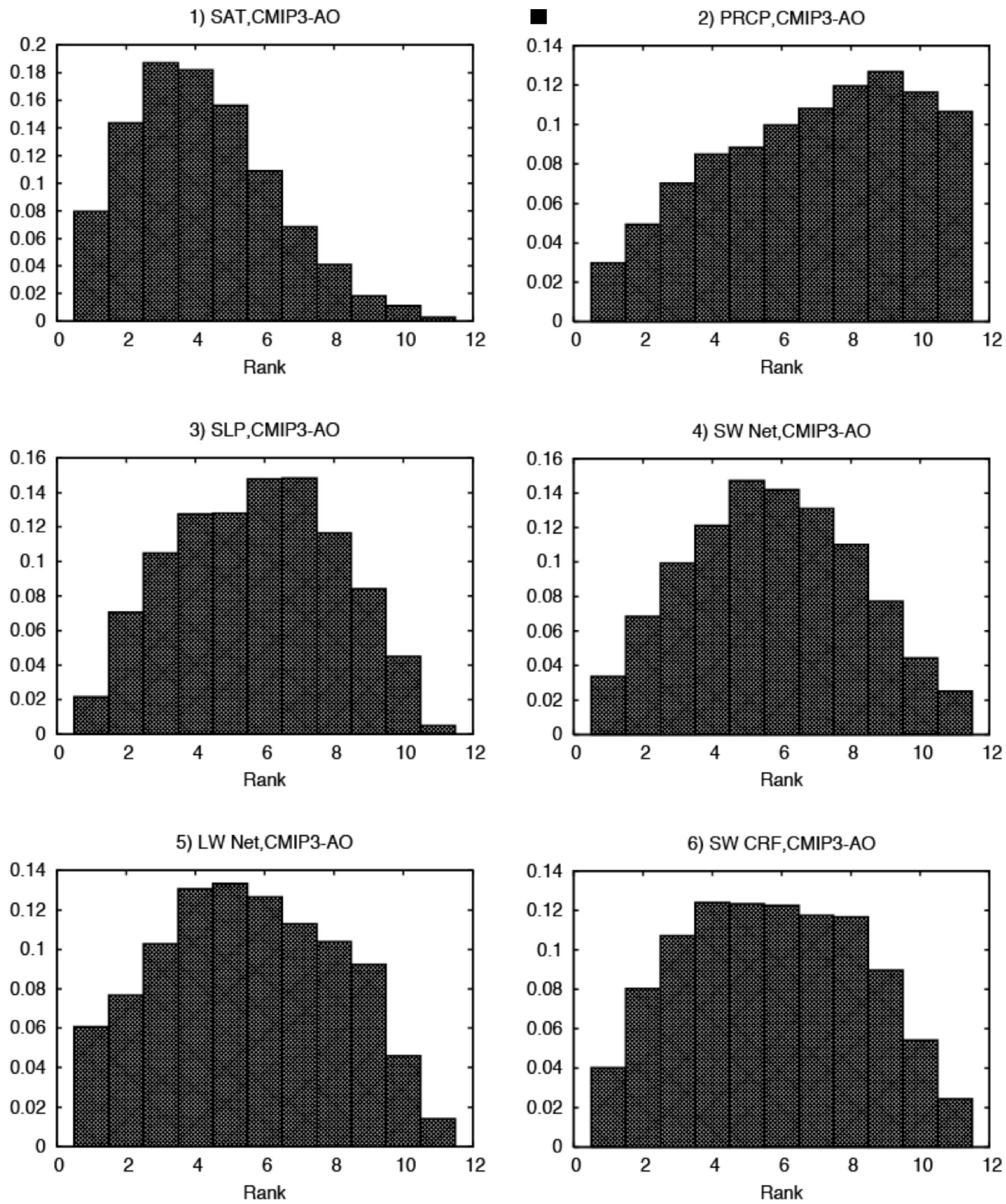


What about regional information?



How do we interpret the ensemble mean and spread?

Source: IPCC Fourth Assessment Report



Rank histograms from CMIP3 output

Loop over grid points and rank the observation w.r.t. ensemble members, compute histogram

Uniform histogram is desirable

U-shaped = too narrow

Domed = too wide

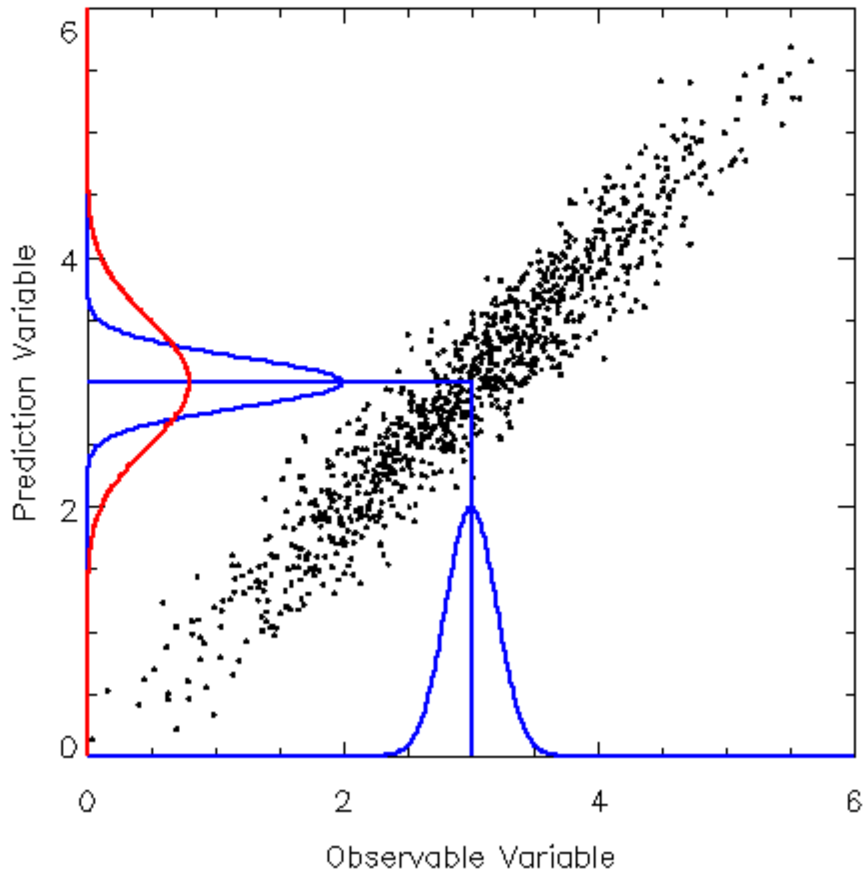
Tokuta Yokohata, James D. Annan, Julia C. Hargreaves, Charles S. Jackson, Michael Tobis, Mark J. Webb, David Sexton, Mat Collins

Emergent Constraints: Schematic

- Find relationship
- Find observed value
- Read off prediction
- Find observational uncertainty
- Add onto prediction
- Add “statistical” uncertainty from scatter

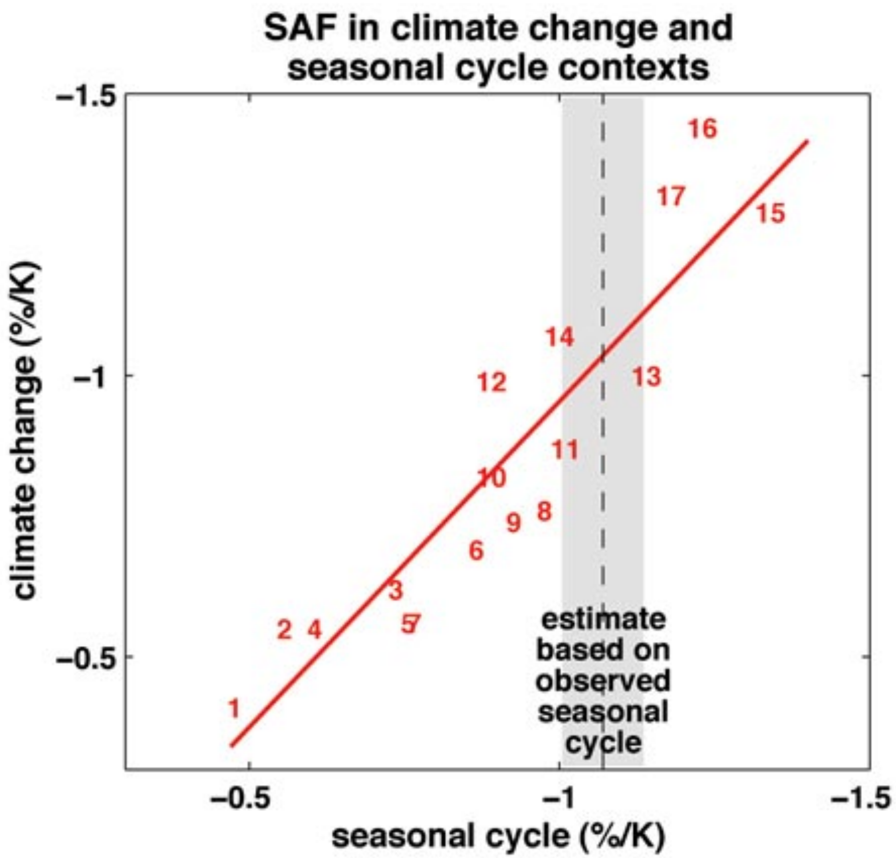
Reduce uncertainty

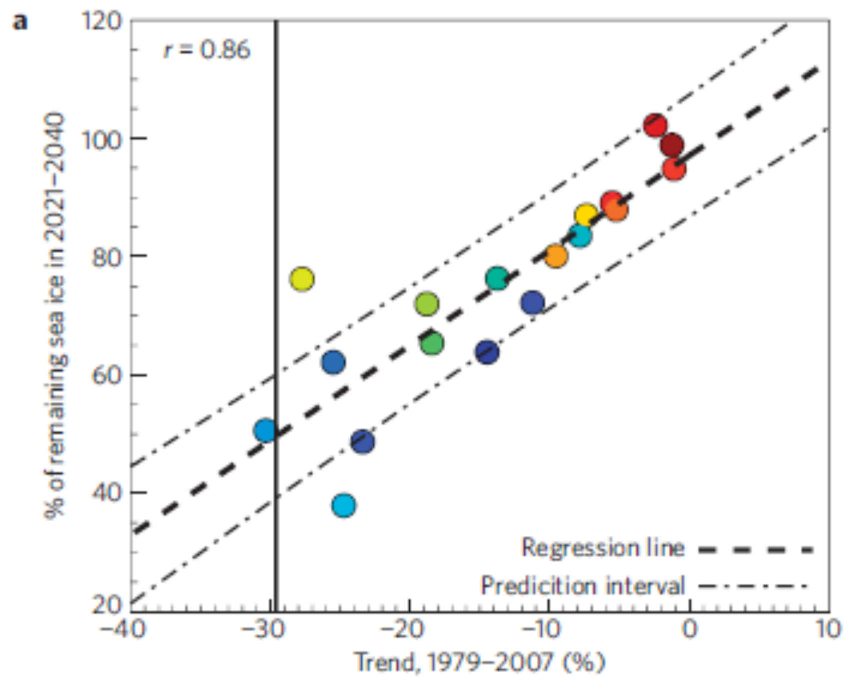
- Use better observations
- Find a better relationship (constraint)



/metric

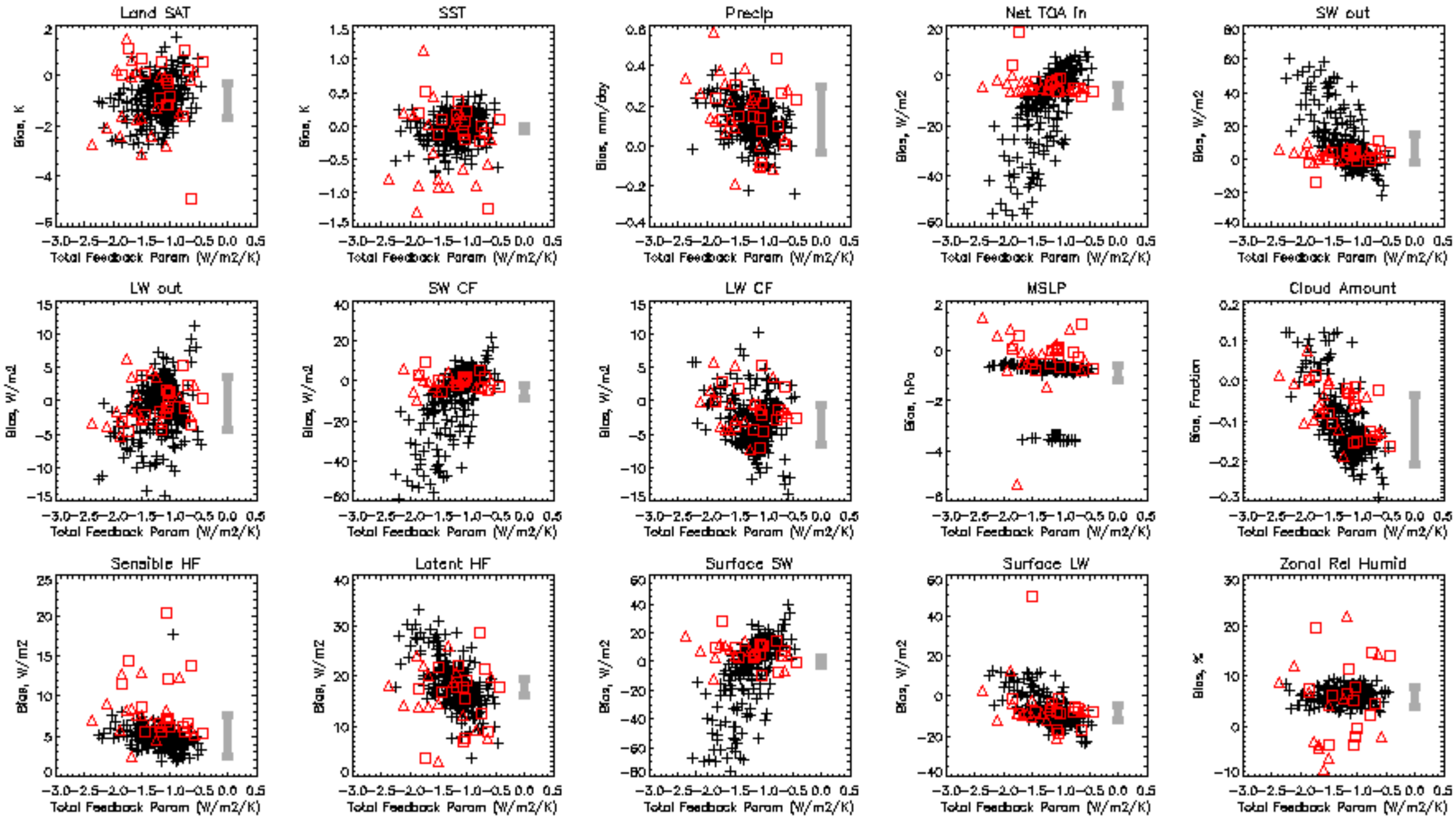
Emergent Constraints





- Boé J, Hall A, Qu X (2009), September Sea-Ice Cover in the Arctic Ocean Projected to Vanish by 2100, Nature Geosci, 2: 341-343
- Hall A, Qu X (2006) Using the current seasonal cycle to constrain snow albedo feedback in future climate change. Geophys. Res. Lett., 33, L03502

Feedback parameter (climate sensitivity) vs different metrics



Single Model Bayesian Approaches: The Perturbed Physics Ensemble

- Take one model structure and perturb uncertain parameters and possible switch in/out different subroutines
- Can control experimental design, systematically explore and isolate uncertainties from different components
- Potential for many more ensemble members
- Unable to fully explore “structural” uncertainties
- HadCM3 widely used (MOHC and climateprediction.net) but other modelling groups are building capability

Some Notation

$$y = f(x)$$

$\mathbf{y} = \{\mathbf{y}_h,\mathbf{y}_f\}$ historical and future climate variables (many)

$\mathbf{f}$ = model (complex)

$\mathbf{x}$ = uncertain model input parameters (many)

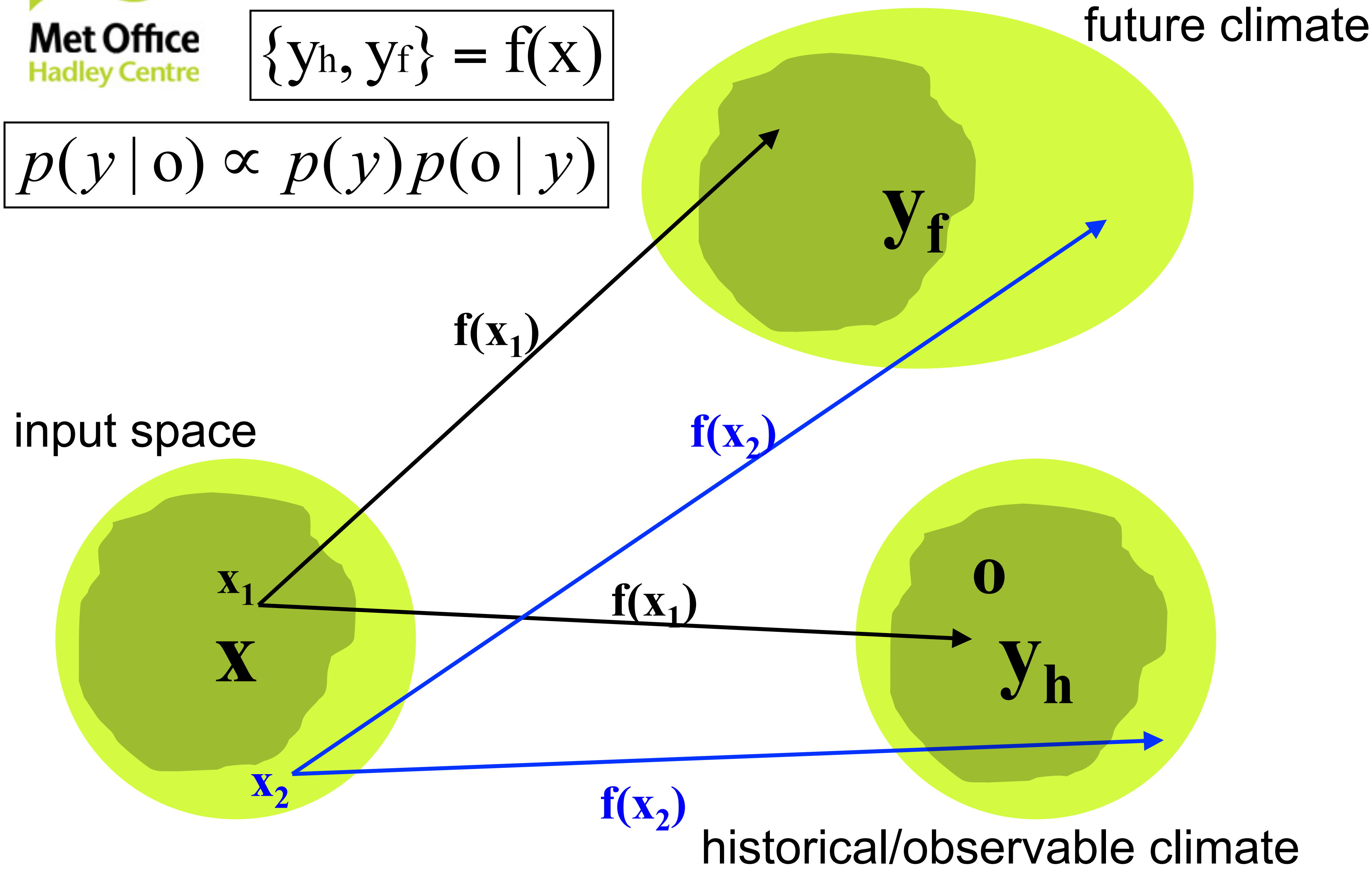
$\mathbf{o}$ = observations (many, incomplete)

- Our task is to explore $\mathbf{f(x)}$ in order to find $\mathbf{y}$ which will be closest to what will be observed in the past *and* the future (conditional on some assumptions)
- Provide probabilities which measure how strongly different outcomes for climate change are supported by current evidence; models, observations and understanding

Probabilistic Approach

$$\boxed{\{y_h, y_f\} = f(x)}$$

$$\boxed{p(y \mid o) \propto p(y)\,p(o \mid y)}$$



Met Office
Hadley Centre

Estimating Likelihood

$$\log L_0(\mathrm{x}) = -c - \frac{n}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathrm{y_h} - \mathrm{o})^T \mathbf{V}^{-1}(\mathrm{y_h} - \mathrm{o})$$

$\mathbf{V}$ = obs uncertainty + emulator error + discrepancy

$$\log L_0(\mathrm{x}) \sim (\mathbf{y_h}\text{-}\mathbf{o}) \quad \mathbf{V}^{-1} \quad (\mathbf{y_h}\text{-}\mathbf{o})$$

$\mathbf{V}$ is calculated from the perturbed physics and multi-model ensemble
It is a very complicated metric

Enhancement of “Standard” Approach (Rougier 2007)

$$y = f(x^*) + d$$

$\mathbf{y}$ = $\{\mathbf{y}_h, \mathbf{y}_f\}$ climate variables (vector)

$\mathbf{f}$ = HadCM3

$\mathbf{x}^*$ = best point in HadCM3 parameter space – for observable and non-observable fields

$\mathbf{d}$ = discrepancy – irreducible/”structural” model error (vector)

Global Climate Sensitivity

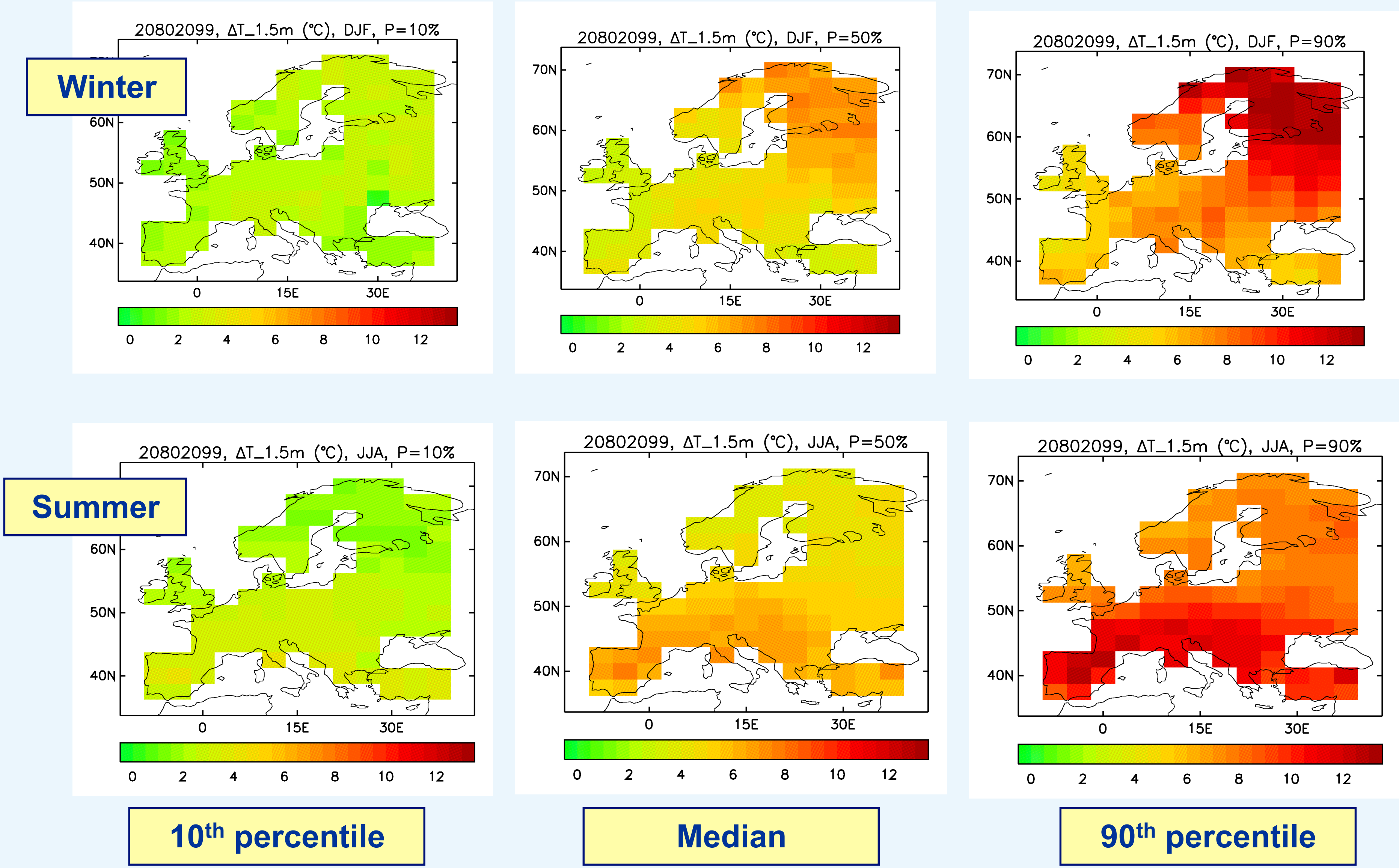
$$y = f(x_*) + d$$

Since Murphy et al. (20

- More ensemble with simultaneou perturbations to
- New emulator ba linear regressior taking into accou parameter intera
- New implementa likelihood functic
- Discrepancy cal from CMIP3/AR4 model archive

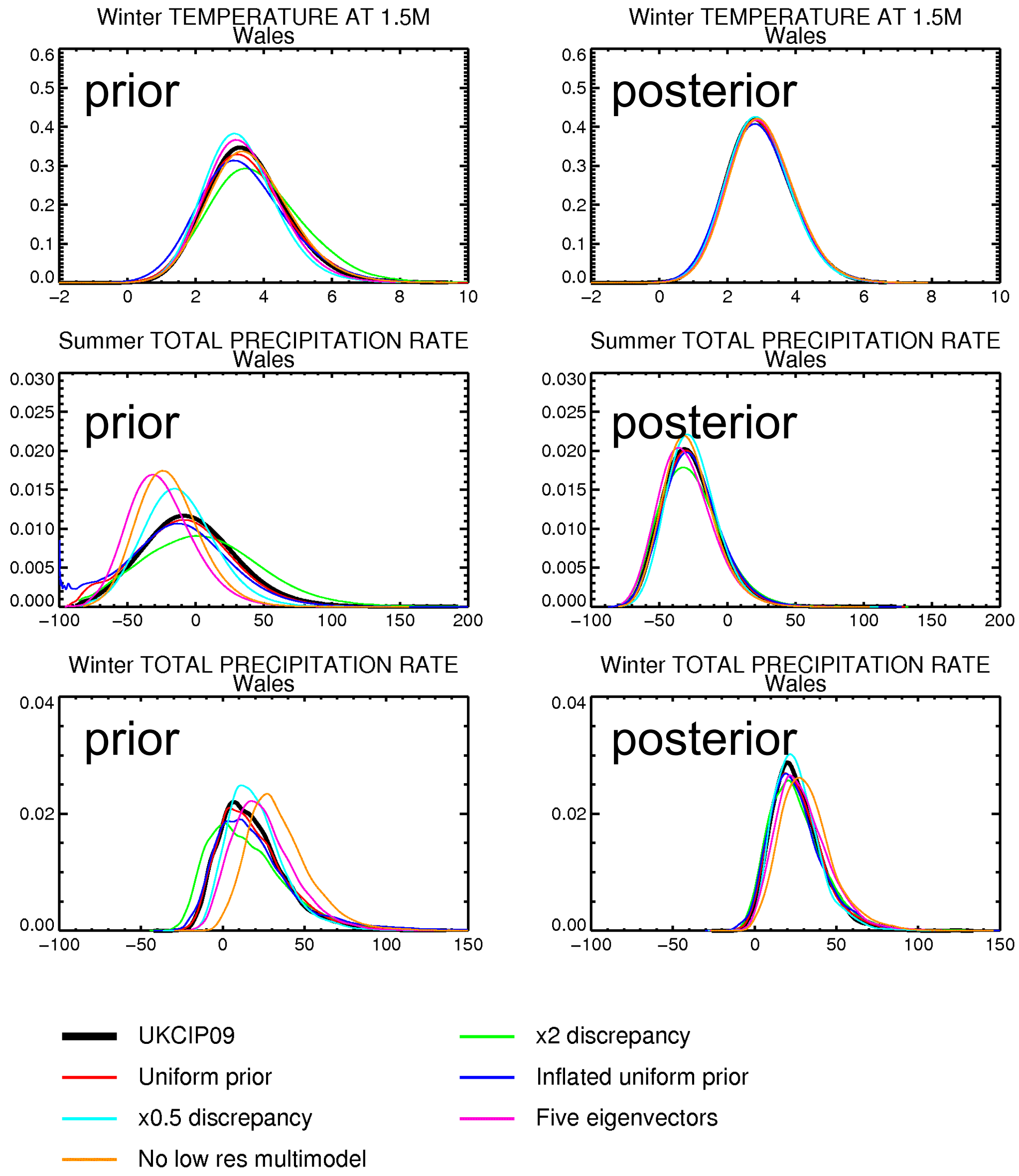
5-95% range 2.3-4.3K

Surface temperature changes for the 2080s



Sensitivity to Key Assumptions

- For pragmatic reasons that there are a number of choices and assumptions that have to be made in the implementation
- We can at least test the sensitivity to these assumptions
- Prior distributions are sensitive to assumptions
- Likelihood weighting/ discrepancy reduces that sensitivity significantly



Strengths and Weaknesses

Extrapolation of signals (e.g. ASK)

• Conceptually simple for “near-term” (linear) climate change

• Useful for global and large-scale temperature projections

• Implementation made more complex by the use of attributable warming

The meaning of simple ensemble averaging

• Consistent with current practice

• Can only be tested for historical climate variables, not future projections

• Inconsistent with the idea of errors-common-to-all-models (e.g. split ITCZ)

• Perhaps a zeroth-order test

UNIVERSITY OF EXETER

Strengths and Weaknesses

Emergent constraints

- Strength in simplicity
- Will not work for all variables (e.g. climate sensitivity)
- Consistency of projections of different variables?

Single-model Bayesian approaches

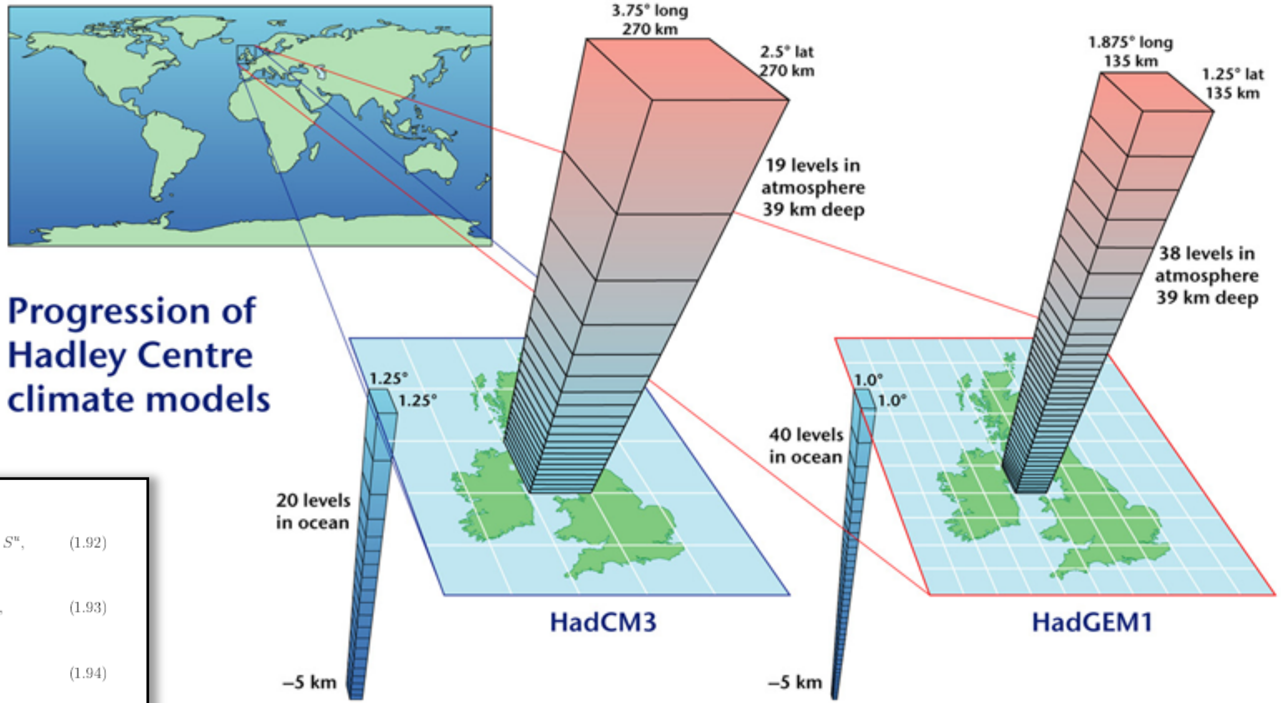
- Rigorous statistical approach
- Can be implemented for “exotic” variables
- Weak observational constrains
- Estimating discrepancy is a challenge

Challenges/Future Work

- The complexity and expense of climate models makes it hard to fit them into existing statistical frameworks (JR quote)
- (Many members of the climate modelling community are turned-off by statistics)
- Either we work hard to fit existing models into frameworks,
    - or we develop new frameworks,
    - or we develop new “probabilistic” climate models

Met Office
Hadley Centre

Progression of Hadley Centre climate models



Horizontal momentum components

$$\frac{Du}{Dt} = -\frac{uw}{r} - 2\Omega w \cos\phi + \frac{uv\tan\phi}{r} + 2\Omega v \sin\phi - \frac{c_{pd}\theta_v}{r\cos\phi}\frac{\partial \Pi}{\partial \lambda} + S^u, \tag{1.92}$$

$$\frac{Dv}{Dt} = -\frac{vw}{r} \qquad - \frac{u^2\tan\phi}{r} - 2\Omega u \sin\phi - \frac{c_{pd}\theta_v}{r}\frac{\partial \Pi}{\partial \phi} + S^v, \tag{1.93}$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \frac{u}{r\cos\phi}\frac{\partial}{\partial \lambda} + \frac{v}{r}\frac{\partial}{\partial \phi} + w\frac{\partial}{\partial r}, \tag{1.94}$$

$$\Pi = \left(\frac{p}{p_0}\right)^{\frac{R_d}{c_{pd}}}, \qquad \text{[Exner function; } p_0 = 1000hPa\text{]} \tag{1.95}$$

$$\theta_v = \frac{T}{\Pi}\left(\frac{1 + \frac{1}{\epsilon}m_v}{1 + m_v + m_{cl} + m_{cf}}\right). \quad \text{[Virtual potential temperature; } \epsilon = \frac{R_d}{R_v} \cong 0.622\text{]} \tag{1.96}$$

Vertical momentum component

$$\frac{Dw}{Dt} = \frac{(u^2 + v^2)}{r} + 2\Omega u \cos\phi \; - \; g \; - c_{pd}\theta_v \frac{\partial \Pi}{\partial r} + S^w. \tag{1.97}$$

Continuity

$$\frac{D}{Dt}\left(\rho_y r^2 \cos\phi\right) + \rho_y r^2 \cos\phi \left(\frac{\partial}{\partial \lambda}\left[\frac{u}{r\cos\phi}\right] + \frac{\partial}{\partial \phi}\left[\frac{v}{r}\right] + \frac{\partial w}{\partial r}\right) = 0, \tag{1.98}$$

where

$$\rho = \rho_y \left(1 + m_v + m_{cl} + m_{cf}\right). \tag{1.99}$$

Thermodynamics

$$\frac{D\theta}{Dt} = S^\theta = \left(\frac{\theta}{T}\right)\frac{\dot{Q}}{c_{pd}}, \tag{1.100}$$

where

$$\theta = \frac{T}{\Pi} = T\left(\frac{p_0}{p}\right)^{\frac{R_d}{c_{pd}}}. \quad \text{[Potential temperature; } p_0 = 1000hPa\text{]} \tag{1.101}$$

State

$$\Pi^{\frac{\kappa_d - 1}{\kappa_d}} \rho\theta_v = \frac{p_0}{\kappa_d c_{pd}}. \quad \left[\kappa_d \equiv \frac{R_d}{c_{pd}}\right] \tag{1.102}$$

Moisture

$$\frac{Dm_v}{Dt} = S^{m_v}, \tag{1.103}$$

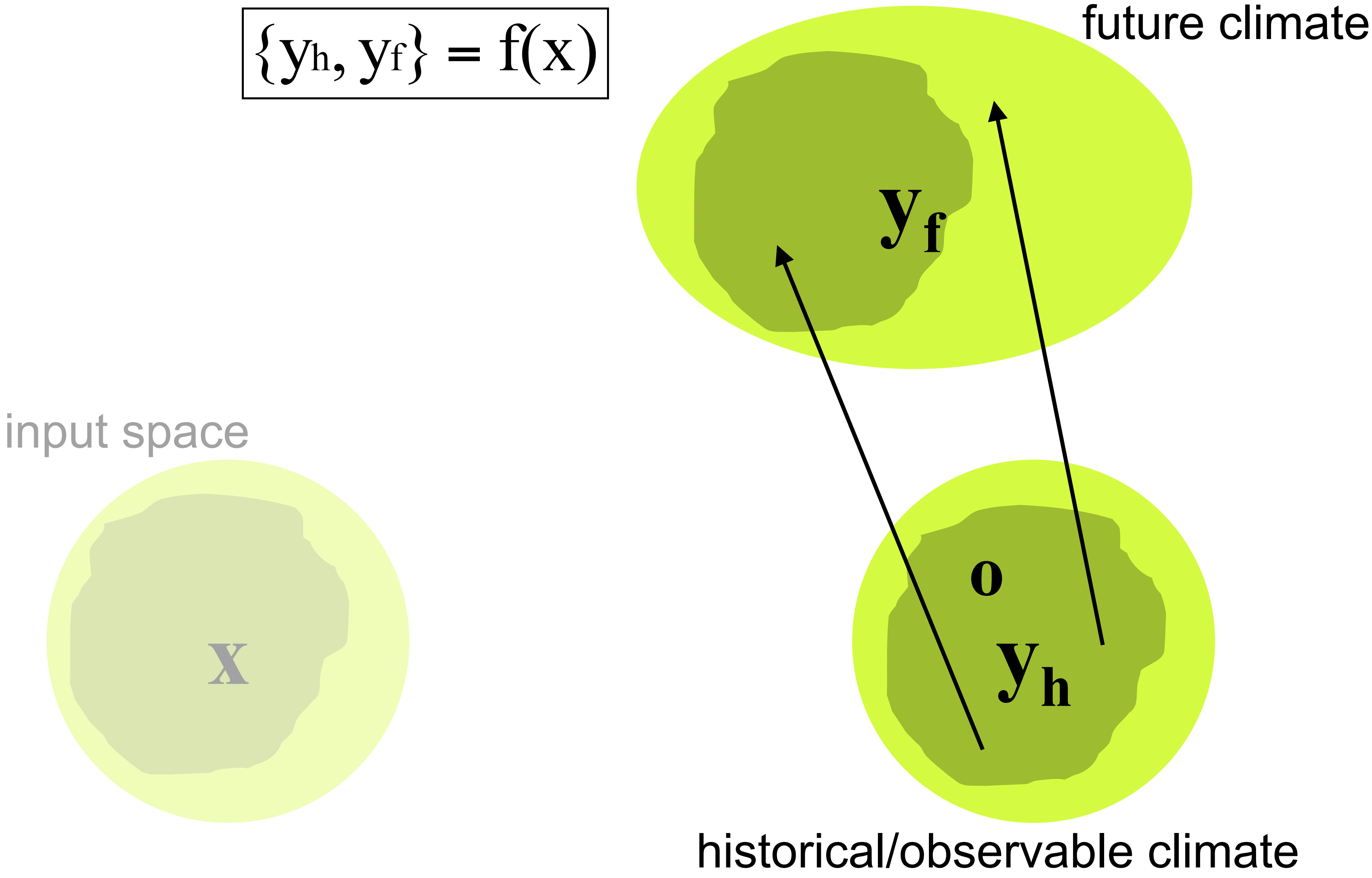
$$\frac{Dm_{cl}}{Dt} = S^{m_{cl}}, \tag{1.104}$$

$$\frac{Dm_{cf}}{Dt} = S^{m_{cf}}. \tag{1.105}$$

In a sense, (1.92)-(1.105) are the equations on which the Unified Model is based, since the transformations described in Section 2 are exact, and no terms are neglected.

Multi-Model Approach?

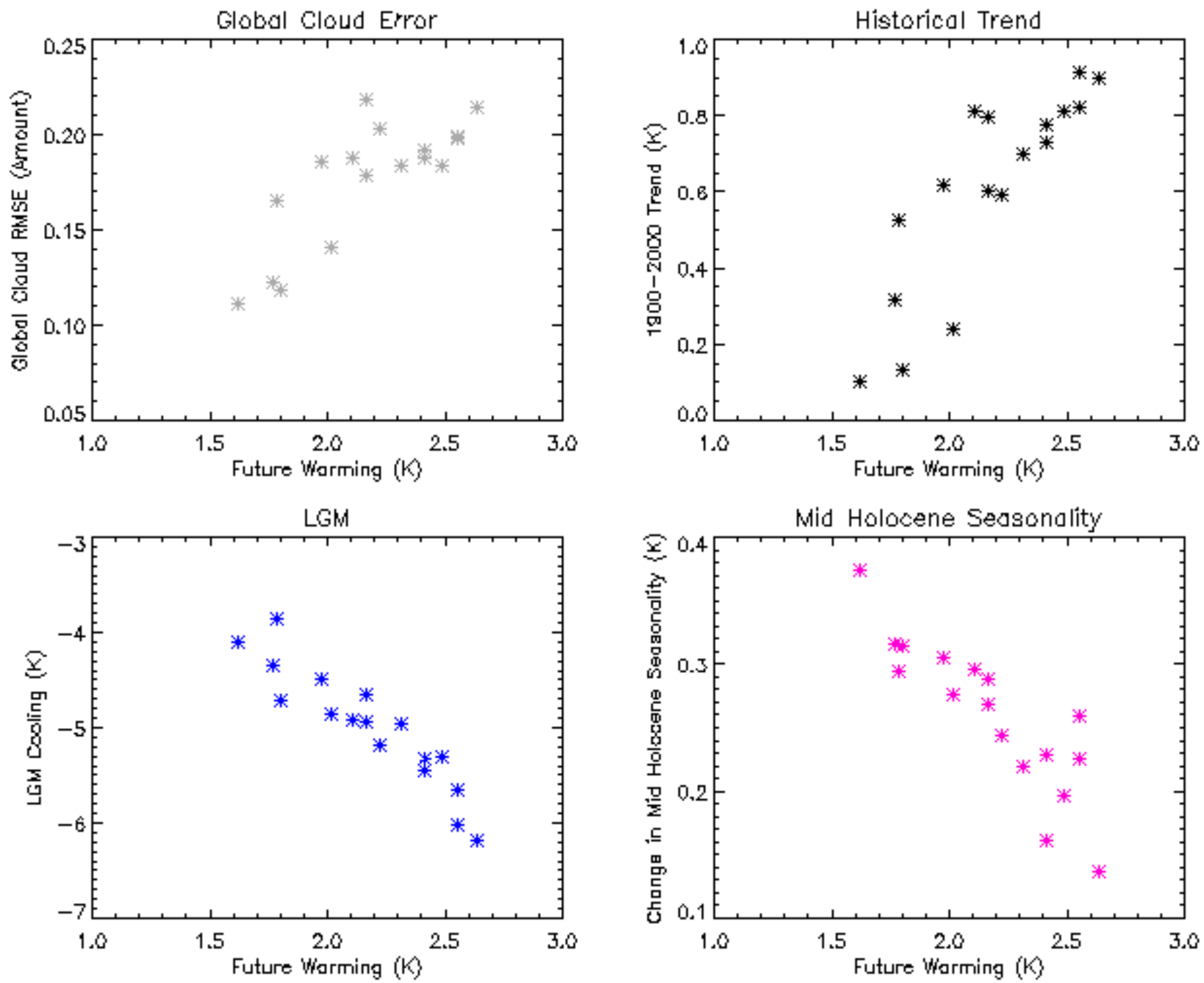
$$\boxed{\{y_h, y_f\} = f(x)}$$



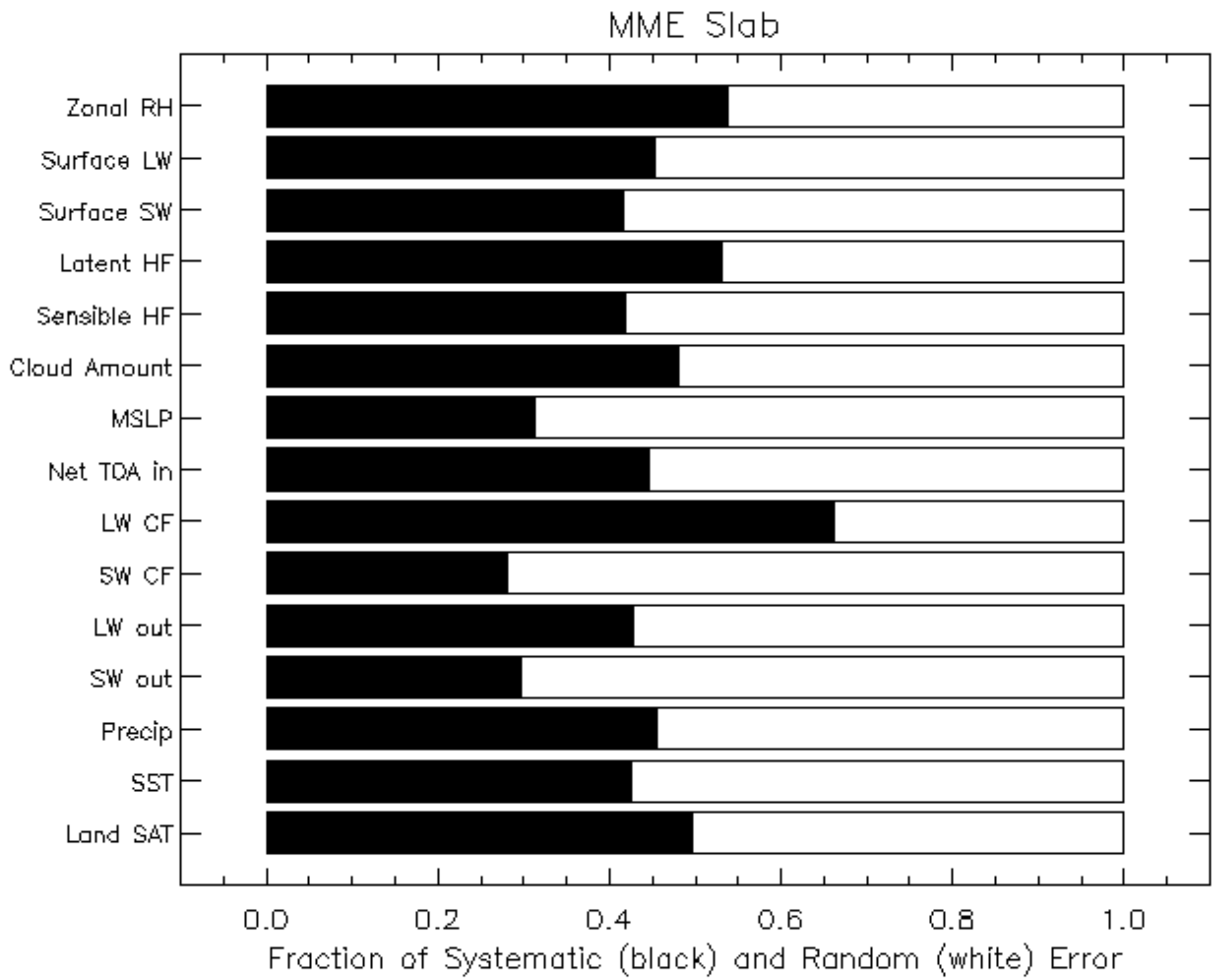
PalaeoQUMP: QUEST Project

- Tamsin Edwards, Sandy Harrison, Jonty Rougier, Michel Crucifix, Ben Booth, Philip Brohan, Ana María García Suárez, Mary Edwards, Michelle Felton, Heather Binney, …

- Aim: To use palaeoclimate data and simulations to constrain future projections



Systematic Errors in All Models



Collins et al. 2010

Bayesian Approach

$y = \mathrm{f}(\mathrm{x})$

- Vary uncertainty model input parameters x (*prior* distribution of y)
- Compare model output, m ('internal' model variables) with observations, o, to estimate the skill of each model version (*likelihood*)
- Form distribution of y weighted by likelihood (*posterior*)

Bayesian Notation: $p(y \mid \mathrm{o}) \propto p(y)\, p(\mathrm{o} \mid y)$

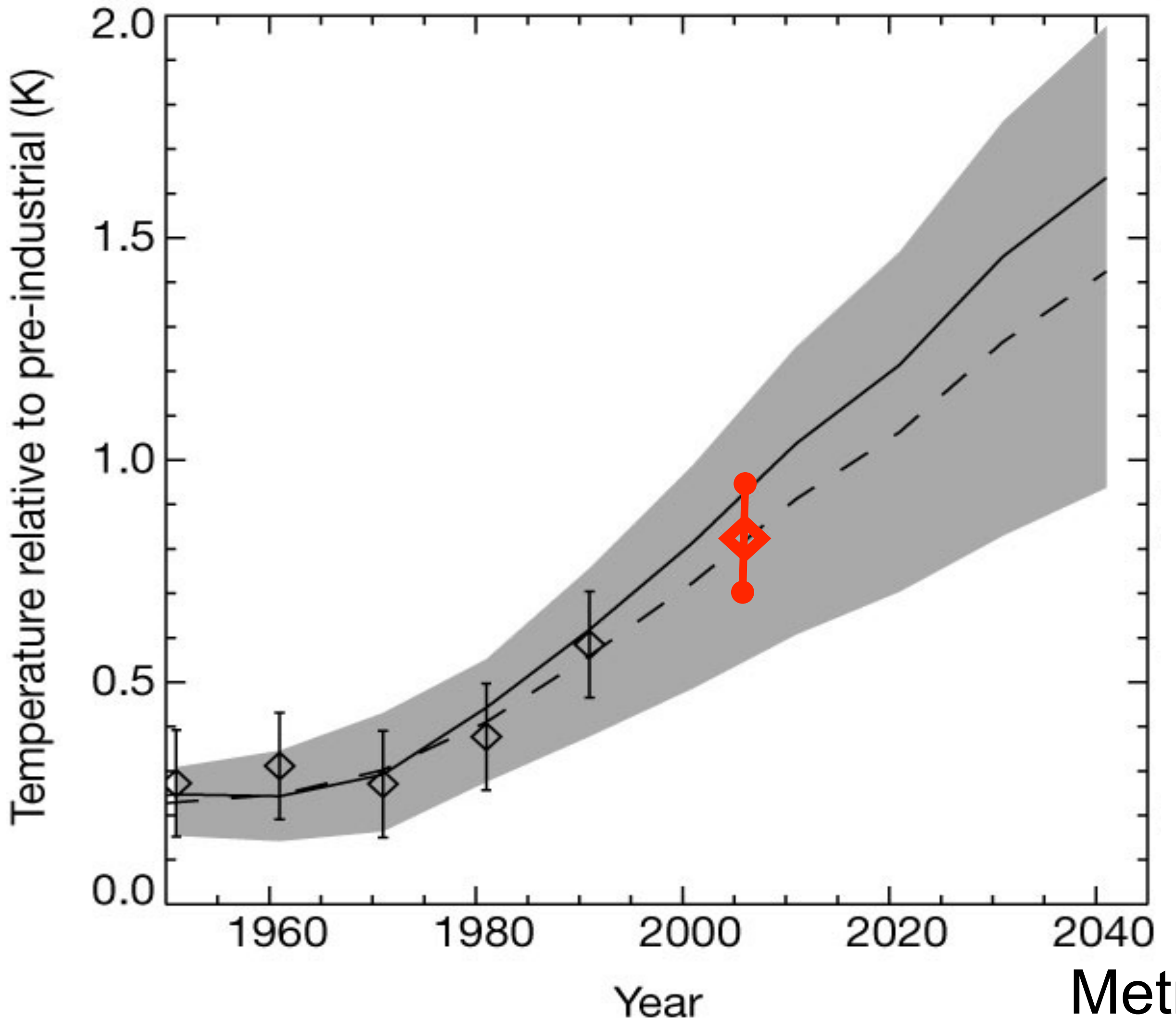
the *posterior* is proportional to the *prior* times the *likelihood*

Likelihood: $p(\mathrm{o} \mid y) = L_0(y)$

$\log L_0 = -\mathrm{g}(\mathrm{m} - \mathrm{o})$

$w = 1/\exp(\mathrm{g}(\mathrm{m} - \mathrm{o}))$

Global temperatures are evolving as predicted in response to human influence



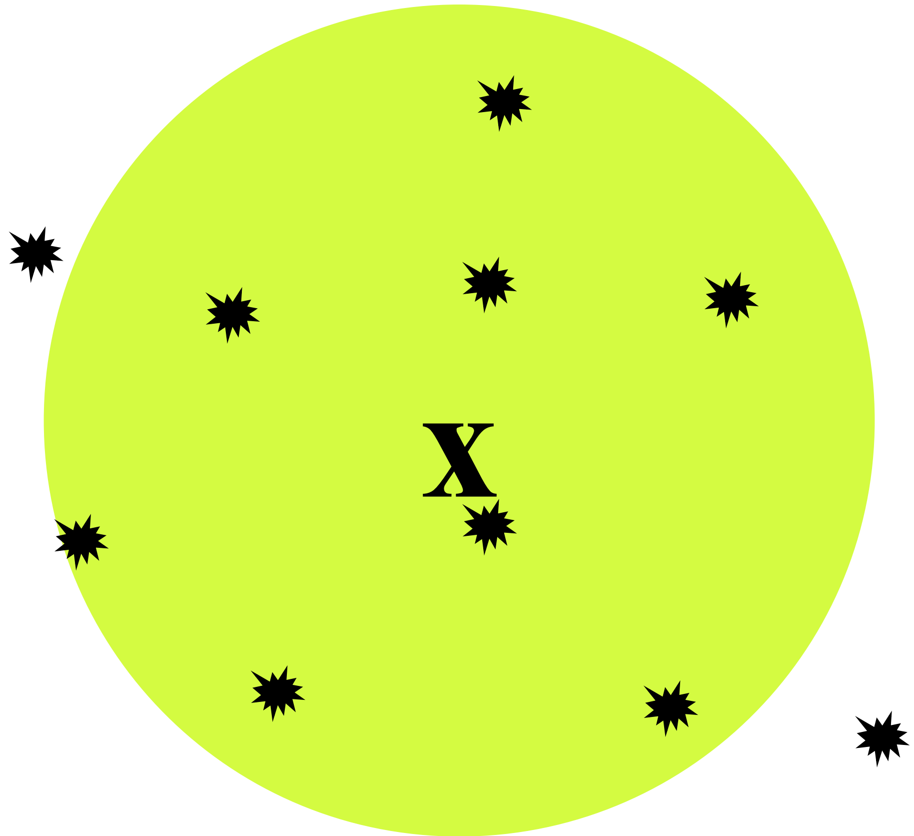
Global temperature response to greenhouse gases and aerosols
Solid: climate model simulation (HadCM2)
Dashed: recalibrated prediction using data to August 1996
(Allen, Stott, Mitchell, Schnur, Delworth, 2000)

Observed decadal mean temperature September 1999 to August 2009 inclusive

Metric=attributable warming

Emulator Schematic

- Use regression trained on ensemble runs to estimate past and future variables, $\{\mathbf{y}_h, \mathbf{y}_f\}$ at any point of parameter space, $\mathbf{x}$
- Use transformed variables and take into account some non-linear interaction terms
- Note – might need to run models at some quite “remote” regions of parameter space
- Keep account of emulator errors in the final PDFs



e.g. Rougier et al., 2009

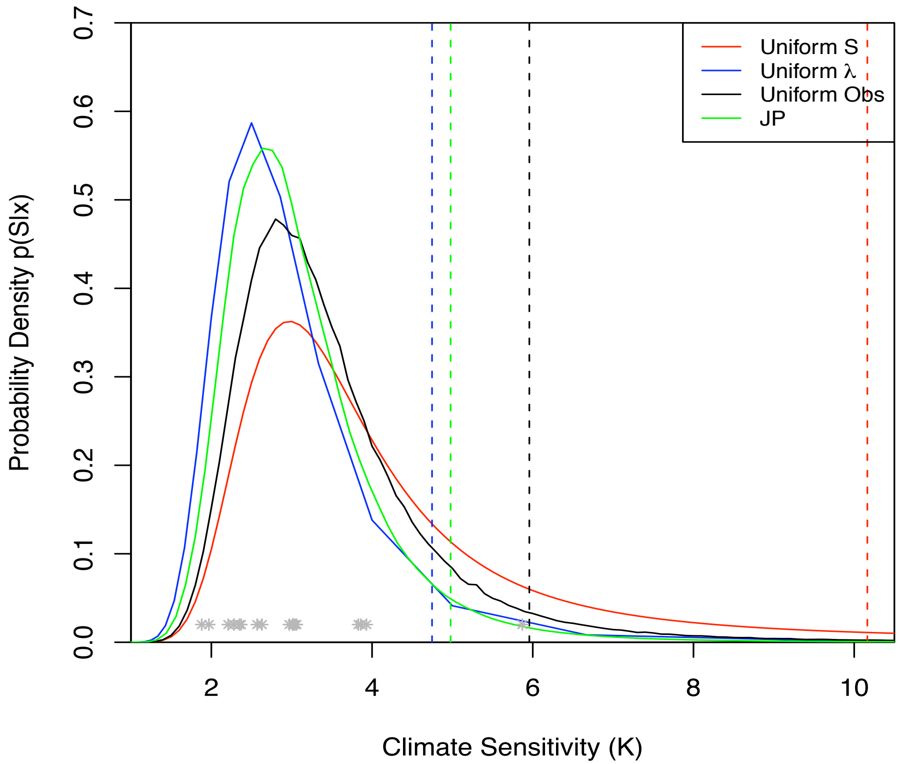
Estimating Discrepancy

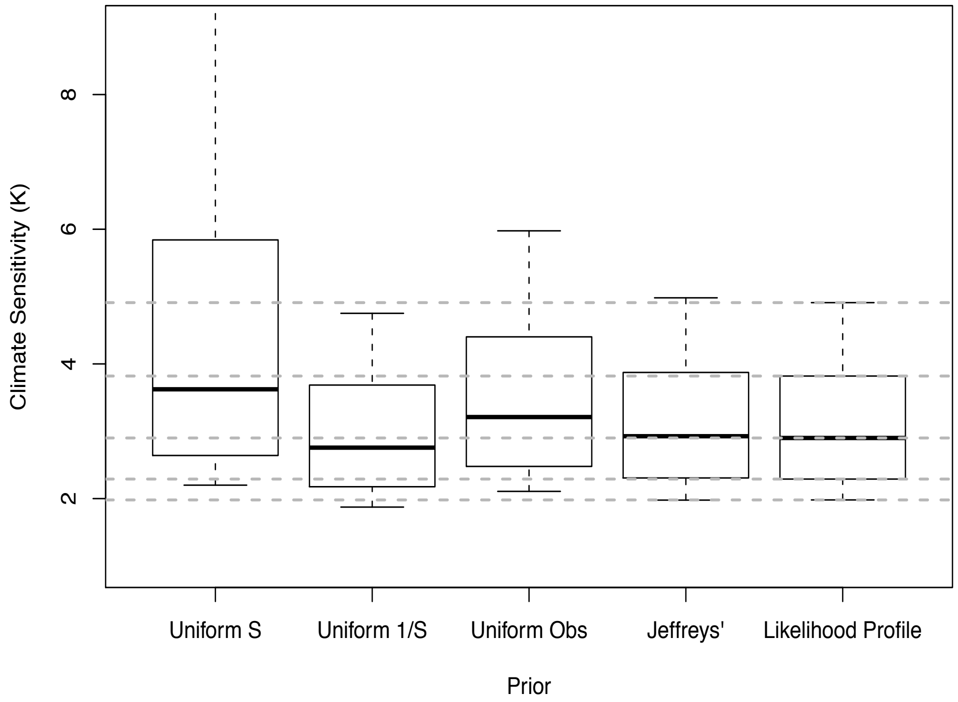
- Use the multi-model ensemble from IPCC AR4 (CMIP3) and CFMIP (models from different centres)
- For each multi-model ensemble member, find point in HadCM3 parameter space that is closest to that member
- There is a distance between climates of this multi-model ensemble member and this point in parameter space i.e. effect of processes not explored by perturbed physics ensemble
- Pool these distances over all multi-model ensemble members
- Uses model data from the past and the future

Objective Bayesian Approaches

- Climate model parameters are often "nuisance" and have no real world counterparts,
  - So how can we define a prior distribution over them?
  - Uniform priors are problematic (Annan & Hargreaves 2009), and arbitrary due to co-ordinate definition.
  - Expert priors also problematic in climate research (double counting).
- "Objective" Bayesian approaches use a rule/ algorithm to form prior aiming to,
  - Maximise information gain from the data.
  - Be Invariant to co-ordinate transformation.
  - Approximate frequentist "sampling" properties.
  - Account for geometry of model response.
- Already used (unknowingly?) in D&A based forecast studies (ASK).

EBM example (Rowlands 2010 in prep)

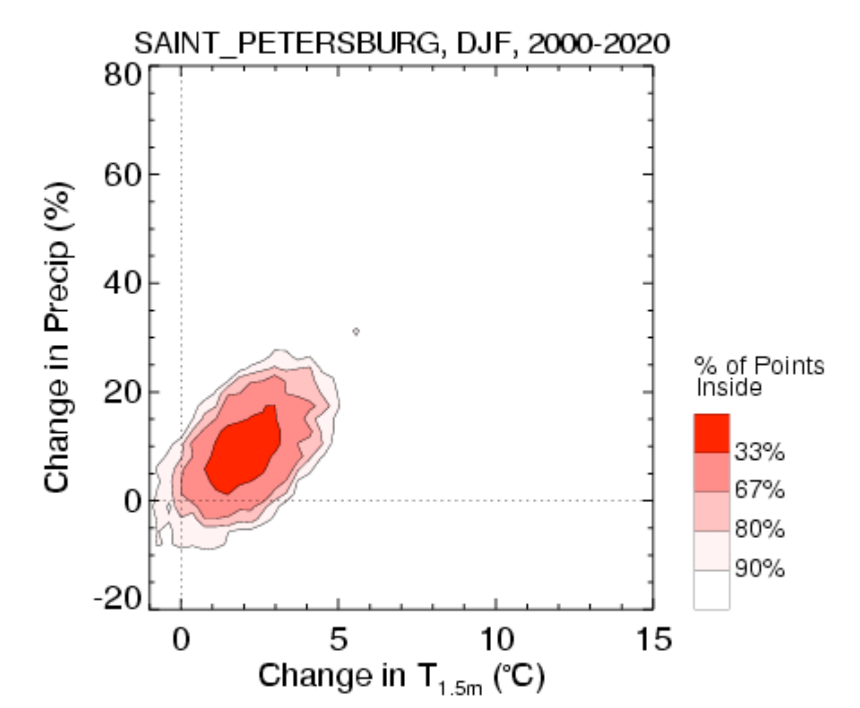


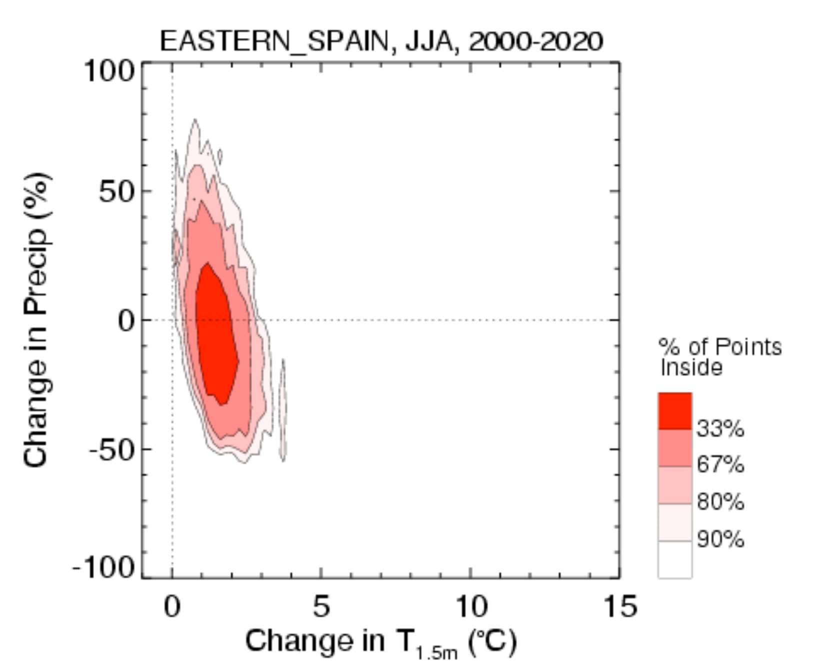


- Jeffreys' prior is the simplest approach (reference priors for the statistics aficionados).
- Gives 5-95% credible interval of 2.0-4.9 K for CS.
- Approximately matches likelihood profile.

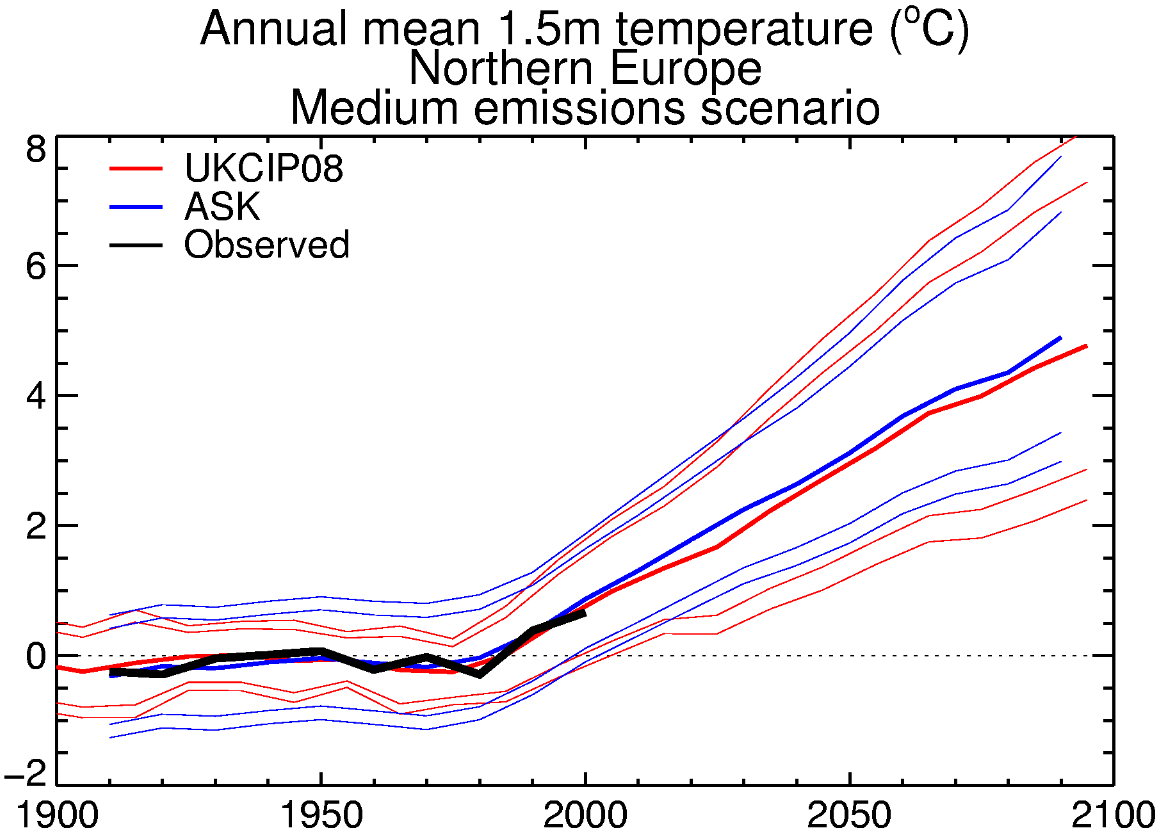
Probabilistic projections in response to A1B emissions

- Changes in temperature and precipitation for future 20 year periods, relative to 1961-90, at 300km scale.





Comparison with an Alternative Approach



Coloured lines show 2.5th, 10th, 50th (thick), 90th and 97.5th percentiles of projected past and future changes

Carbon cycle feedbacks omitted

Together with sensitivity tests, gives confidence in the projections

Peter Stott, UKCP Team