Organizers

Montserrat Fuentes (North Carolina State University)
Peter Guttorp (Department of Statistics, University of Washington)
Michael Stein (University of Chicago)

Objectives

The workshop is intended to be a forum for interaction between statisticians, stochasticists, climate modelers, ocean observers and data assimilators. The goal is to develop observation strategies and design computer experiments to better understand the model and data uncertainties that relate directly to oceans and ocean-related feedback mechanisms. The timing is good in that the studies that form the basis for the fifth assessment report of the IPCC are expected to be finished before the workshop.

Overview

A climate scientist turns to a policy maker and says:
"Our climate model is broken"
"Have you tried tuning it off and turning it back on again?"

The next day , the policy maker comes over and says
"Our government is broken"
"Have you tried tuning it off and turning it back on again?"

Climate science in support of policy making could benefit from something more.
Over the next hour, I am going to suggest that:

- we do not know the space and time scales on which today's climate simulations are informative for adaptation.
- significant improvement requires re-design, not mere reinterpretation.
- 2002 was the time to discuss this in public, but now is better than later.
- the ocean should justify more attention (both obs and cpu).

and

- there is a way forward, which is arguably safer, protects the credibility of science, and yields better policy and decision support. (?and science?)

The Forecast Problem: 1951

THE FORECAST PROBLEM

By H. C. WILLETT

Massachusetts Institute of Technology

INTRODUCTORY REMARKS

The Unsatisfactory Progress of Weather Forecasting as a Science. Probably there is no other field of applied science in which so much money has been spent to effect so little real progress as in weather forecasting.

COMPENDIUM
OF
METEOROLOGY

Prepared under the Direction of the
Committee on the Compendium of Meteorology
H. R. BYERS H. E. LANDSBERG H. WEXLER
B. HAURWITZ A. F. SPILHAUS H. C. WILLETT
H. G. HOUGHTON, Chairman

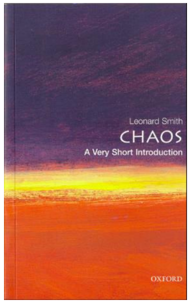
Edited by
THOMAS F. MALONE

AMERICAN METEOROLOGICAL SOCIETY
BOSTON, MASSACHUSETTS
1951

c. Mathematical techniques of extrapolation—based on various manipulations of the equations of motion and continuity. Accurate weather forecasting by mathematical computation is an ultimate objective for the attainment of which nearly every meteorologist hopes, but as a practical reality it appears today to be quite as distant as when Richardson [8] made his classical contribution to the problem in 1922. Richardson failed completely to derive, from the theoretical equations, satisfactory forecasts even of the short-range (6-hr) changes of the meteorological elements. This failure was doubtless caused in part by his efforts to deal with all of the variables at once, which complicated his calculations to a point where he was unable to identify the sources of his errors,

Science advances to provide actionable information.
It did not take weather forecasting another 30 years!
Failure to clearly distinguish where todays science is not actionable today harms both science and policy.

Types of Uncertainty
Kinds of Probability and
(Re)Designing Climate Simulations from Scratch



Leonard A. Smith
London School of Economics
&
Pembroke College, Oxford

Not Possible without A. Lopez, D.Stainforth, E. Suckling,
E. Thompson, E. Wheatcroft

Aims

Which kind of probabilities shall we consider this week?

Which types of uncertainties shall we consider this week?

Can we remain clear on what is it the probability of we are speaking, exactly?
(How do we get out of model-land to actionable (real world) information?)

Extracting quantitative insight from model large complex models involves discussion of:

Fidelity, Experimental Design, RDU, Prob(Big Surprize),
Distinguishing observables from the model name sakes,
Embracing model inadequacy,
A fundamental challenge to the interpretation of model-based probabilities

Bayesians, Doogians, and Physicists

Rational Decisions
I. J. Good
Journal of the Royal Statistical Society. Series B (Methodological)
Vol. 14, No. 1 (1952), pp. 107-114

No (few) physicists would argue that probabilities are not conditioned on the information available (something known/believed "now")

Consider Kelvin's Gambit:

It seems, therefore, on the whole most probable that the sun has not illuminated the earth for 100,000,000 years, and **almost certain** that he has not done so for 500,000,000 years. As for the future, we may say, with equal certainty, that inhabitants of the earth can not continue to enjoy the light and heat essential to their life for many million years longer **unless sources now unknown to us** are prepared in the great storehouse of creation.

On the Age of the Sun's Heat
Sir William Thomson (Lord Kelvin)
Macmillan's Magazine, vol. 5
(March 5, 1862), pp. 388-393.

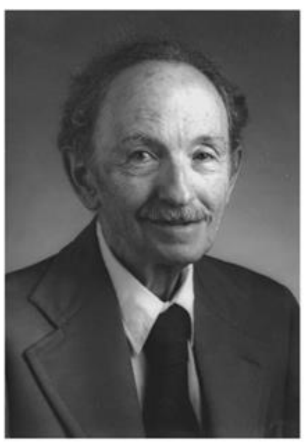


FIG. 4. *Jack Good, 1994.*

"A good Bayesian does better than an non-Bayesian, but a bad Bayesian gets clobbered."

Herman Rubin (1970) quoted by I.J. Good:

Types of Probability (Forecasts)

(o) Tautological Probability. A probability P(E|H) the value of which is specified in the definition of H. ("a fair coin", H is called "a simple statistical hypothesis")

(i) Physical Probability: P(x) "True probability" (Laplace's Demon/Inf Rat Org)

(ii) Psychological Probability: "Personal probability inferred from one's behaviour."

(iii) Subjective Probability: P(x|G) probability of x given our information G is true (Demon's Apprentice/?semi-finite Rational Org?)

(iv) Logical Probability: "Hypothetical subjective probability when you are perfectly rational and infinitely large . "Credibility" Russell (1948)
(Infinitely large Rational Org; Laplace Demon ?or Apprentice?)

(v) Dynamic Probability: $P_t(x| g_t<G)$ when an algorithm encapsulating *G* has not yet terminated (finite algorithm, merely still running).
Dynamic in the sense that this probability is expected to change *without any empirical information* (by reflection only).

(vi) (Im)Mature Probability: $P(x| g<G)$ when *G* is known (not) to be encapsulated in g.
Immature in that this probability is expected to change without addition reflection or additional empirical observation even after the algorithm finishes.

Rational Decisions *I. J. Good (1952) Journal of the Royal Statistical Society. Series B (Methodological) Vol. 14, No. 1 , pp. 107-114*

Types of Probability (Forecasts)

(o) Tautological Probability. A probability P(E|H) the value of which is specified in the definition of H. ("a fair coin", H is called "a simple statistical hypothesis")

(i) Physical Probability: P(x) "True probability" (Laplace's Demon/Inf Rat Org)

(ii) Psychological Probability: "Personal probability inferred from one's behaviour."

(iii) Subjective Probability: P(x|G) probability of x given our information G is true (Demon's Apprentice/?semi-finite Rational Org?)

(iv) Logical Probability: "Hypothetical subjective probability when you are perfectly rational and infinitely large . "Credibility" Russell (1948)
(Infinitely large Rational Org; Laplace Demon or Apprentice)

(v) Dynamic Probability: $P_t(x| g_t<G)$ when an algorithm encapsulating G has not yet terminated (finite algorithm, merely still running).
Dynamic in the sense that this probability is expected to change without any empirical information (by reflection only)

(vi) (Im)Mature Probability: $P(x| g<G)$ when G is known (not) to be encapsulated in g.
Immature in that this probability is expected to change without addition reflection or additional empirical observation even after the algorithm finishes.

.

Rational Decisions *I. J. Good (1952) Journal of the Royal Statistical Society. Series B (Methodological) Vol. 14, No. 1 , pp. 107-114*

Physical Probability and My Subjective Probability

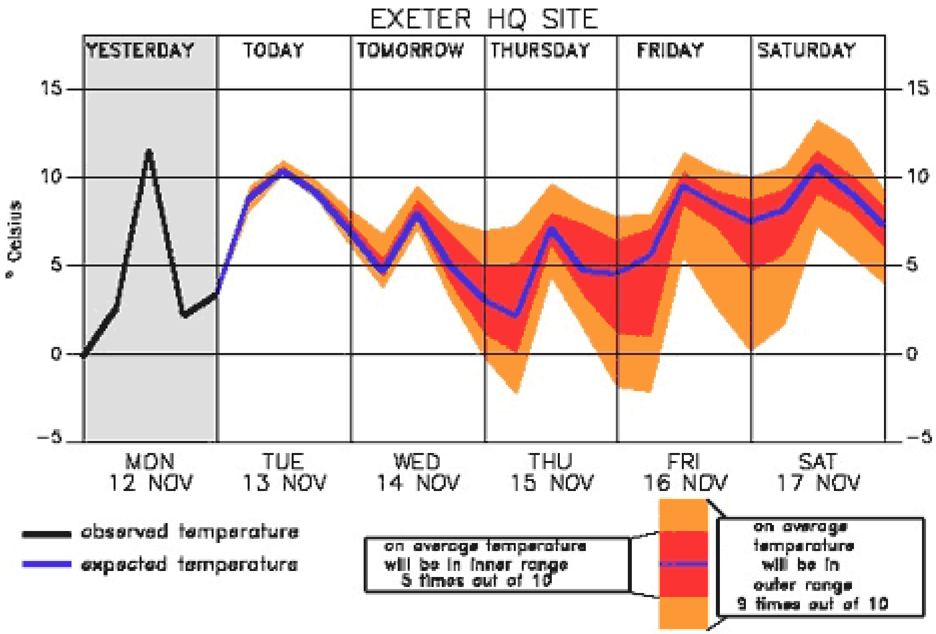
"Whether or not physical probability actually exists, it is often convenient to speak as if did." I.J. Good

"Physical probability automatically obeys axioms, subjective probability depends on axioms, and psychological probability neither obeys axioms nor depends very much on them." I.J. Good p74

My Subjective Probability is then my best attempt to estimate "the" Subjective Probability, just as Laplace saw astronomy related to his Demon.

The thinking of Doogian Bayesians is very close to that of Physicists.

Do weather forecasts provide actionable probabilities?



Percentages

A probability forecast can give a percentage of how likely a defined event is to occur, which can help users to assess the risks associated with particular weather events to which they are sensitive.

Ensembles are designed to estimate these probabilities by sampling the range of possible forecast outcomes. The probability of a particular event occurring is estimated by counting the proportion of ensemble members which forecast that event to occur. So if six out of the 24 members predict more than 5 mm of rain at a specified location in a defined period, we would estimate there to be a 1-in-4, or 25%, chance of the event happening.

http://www.metoffice.gov.uk/research/areas/data-assimilation-and-ensembles/ensemble-forecasting/decision-making

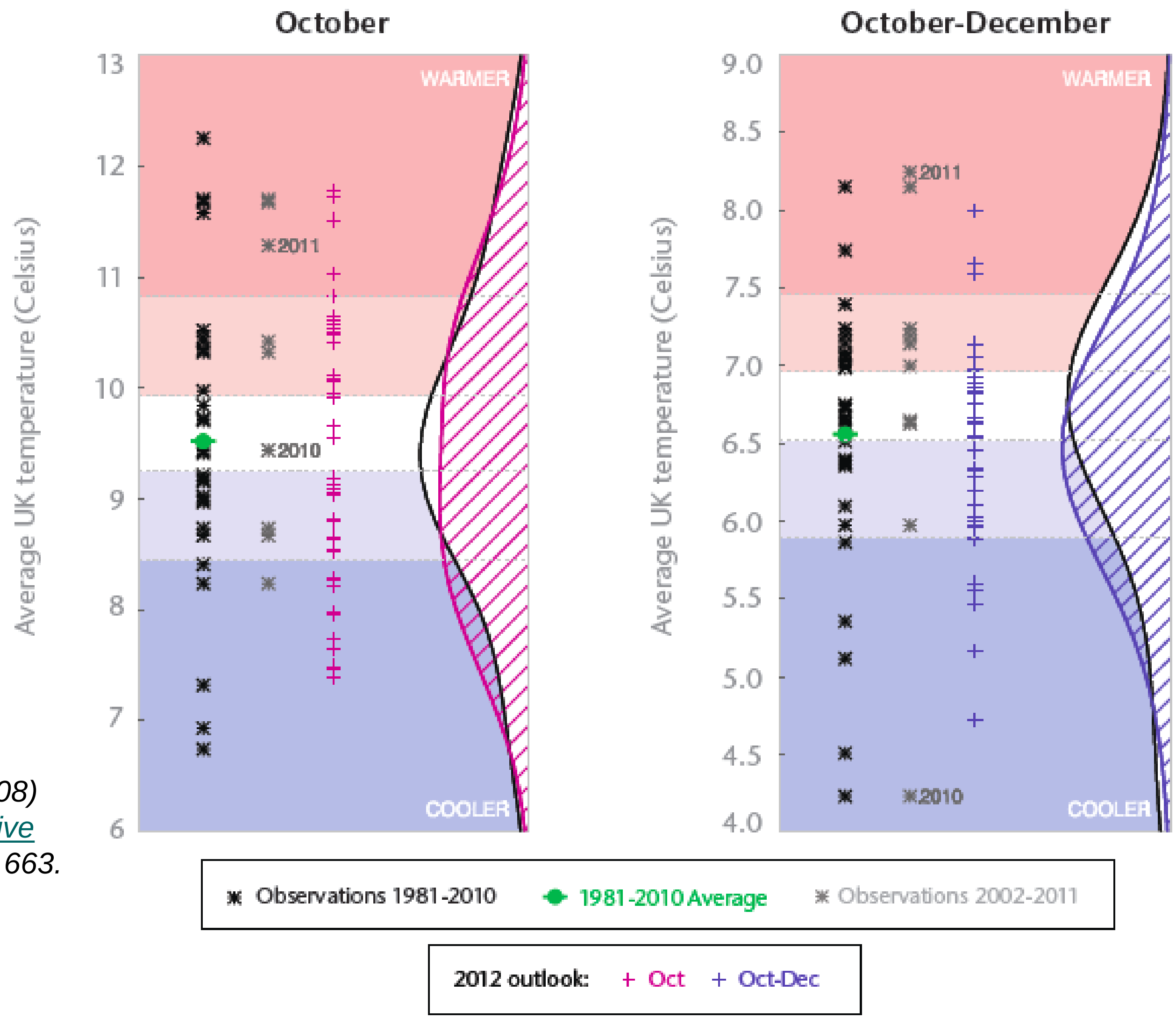
Bayesians can do better than this "naïve probability", but how much better?

Do seasonal forecasts provide actionable probabilities?

A very nice presentation of information.

Are these actionable probabilities?

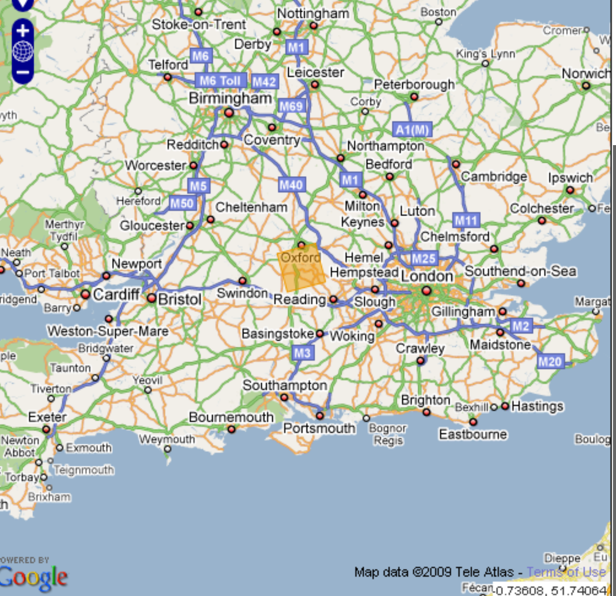
Perhaps, Yes!

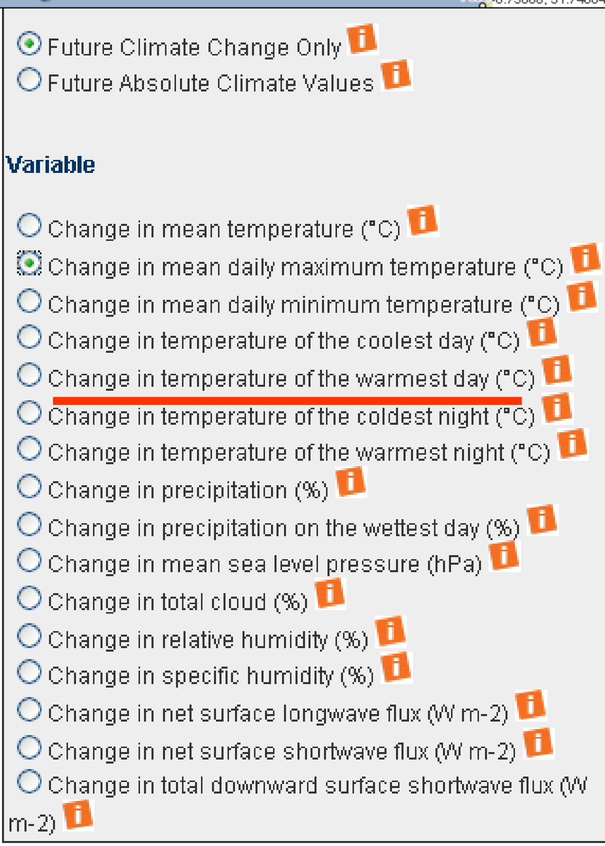


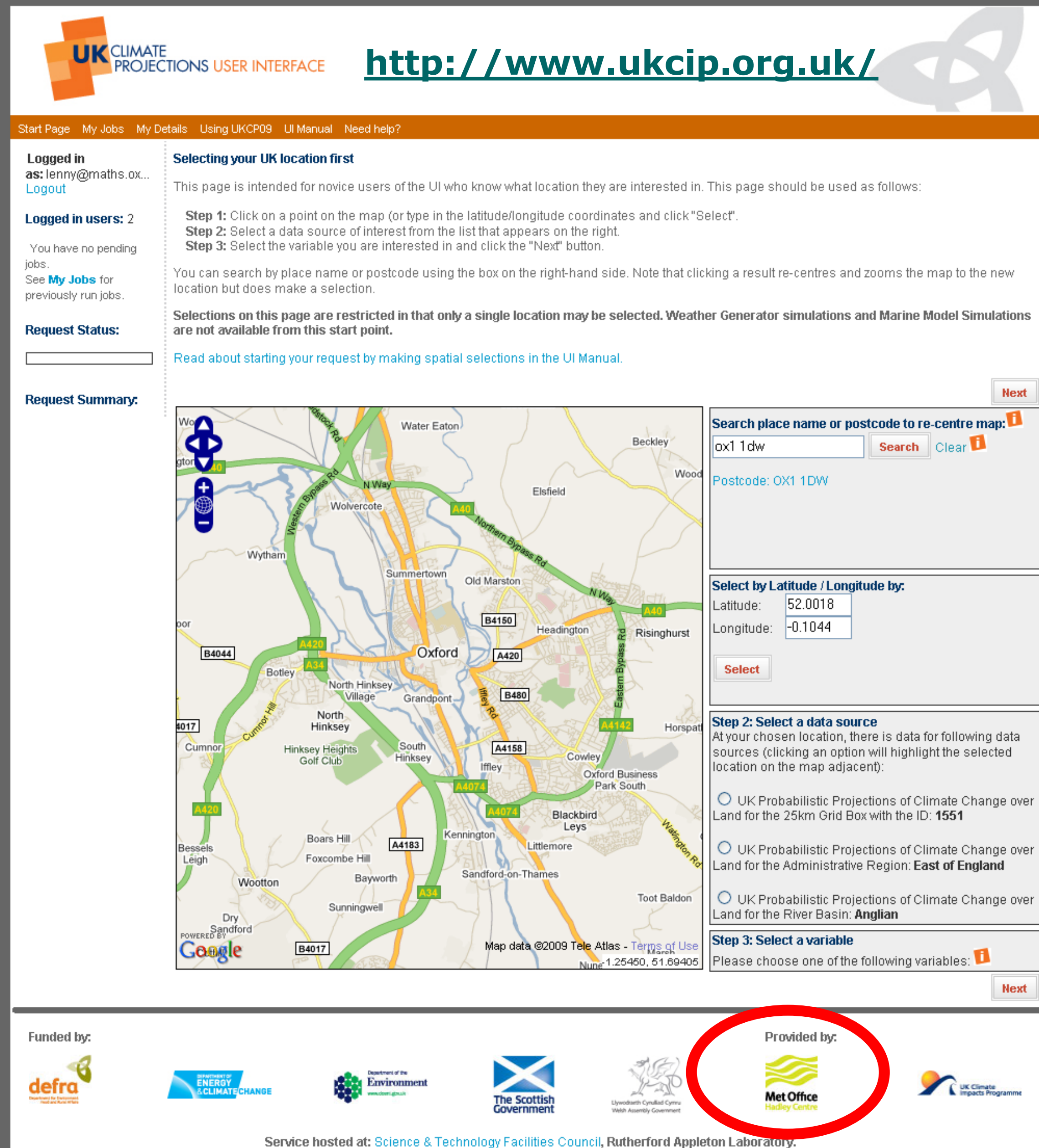
See: Bröcker, J. and Smith, L. A. (2008) [From Ensemble Forecasts to Predictive Distribution Functions](#) Tellus A 60(4): 663.

http://www.metoffice.gov.uk/media/pdf/n/3/A3-plots-temp-OND.pdf

Do UKCP09 PDFs for the stormiest summer day in 2080 Oxford provide actionable probabilities ?







Smith

Who could provide actionable probabilities?

Laplace's Demon (1814)

1) Perfect Equations of Motion (PMS)
2) Perfect noise-free observations
3) Unlimited computational power

Demon's Apprentice (2007)

1) Perfect Equations of Motion (PMS)
2) ~~Perfect noise-free observations~~
3) Unlimited computational power

Demon's Novice (2012)

1) ~~Perfect Equations of Motion (PMS)~~
2) ~~Perfect noise-free observations~~
3) Unlimited computational power

http://2darts.com/2dtuts/articles/50-terrifying-creatures/

Probability of what exactly?

Do UKCP09 provide actionable probabilities ?

Real World Event:

For example, if a projected temperature change of +4.5°C is associated with the 90% at a particular location in the 2080s for the UKCP09 medium emission scenario, this should be interpreted as it is projected that there is a 90% likelihood that temperatures at that location will be equal to or less than 4.5°C warmer than temperatures in the 1961–1990 baseline period. Conversely, there is a 10% likelihood that those temperatures will be greater than 4.5°C warmer than the baseline period.

Last updated: Sunday, 11 March 2012

http://ukclimateprojections.defra.gov.uk/23208 Oct 7,2013

Model Land:

- **Correct interpretation of UKCP09 probabilities:**

  An 80% probability level is indicating that 80% of the model runs fall at or below that value and 20% of the model runs are above that value.

Last updated: Tuesday, 03 April 2012

http://ukclimateprojections.defra.gov.uk/21680 Oct 7 2013

"…they are just tying themselves in knots…
The point has been accepted but has not been internalised."

Definitions

Predictability (working def$^{\underline{n}}$): The property of being able to make probabilistic forecasts (*in silico* or via reflection) with a rational expectation of having more skill than the most naïve rational forecast at hand (say, the climatological distribution).

Alt: Predictability: The possibility of making an "informative forecast" **P(x|g)**

g: Background Knowledge and information to hand, including $\boldsymbol{\mu}$**(x)** if it exists.

A naïve forecast in this case is $\mathbf{P_{clim}(x|g)}$**,** our best approximation of $\boldsymbol{\mu}$**(x)**

G : Ideal ("complete") background Knowledge (think of Laplace's Demon)

G: (True if incomplete) background Knowledge (the Demon's Apprentice)

Definitions

q(x|G**)**: An (ideal) subjective probability of x given (True if limited) information *G*
Good's Statistician's Stooge, Demon's Apprentice

p(x|g): An accessible ("my") subjective probability of x given g <= **G**.

Probability or Predictive Distribution: a positive definite function over observable values that integrates to one. Perhaps from a model.

(see Bröcker, J. and Smith, L. A. (2008)
From Ensemble Forecasts to Predictive Distribution Functions
Tellus A 60(4): 663)

Aim:
"Decision-relevant probability forecast": one that can be used rationally by decision makers as the probability of x (via the probability calculus, the tools of decision theory, …)

How would the novice know if he were doing badly?

An Illustrative example of the Information Deficit

$$\text{Empirical IGN} = \left\langle -\log_2 p(x_{obs}) \right\rangle_{obs} \qquad \text{Implied IGN} = \left\langle \text{Expected IGN} \right\rangle_{forecasts}$$

The Empirical Ignorance reflects the skill of the forecast in practice.

$$<\Sigma\, q(x) \log_2(p(x)/\mu(x))>$$

The Implied Ignorance is the skill the forecast claims to have, averaged of forecasts.

The Expected IGN is the expectation of IGN(p(x)) ***if*** *the outcome was drawn from p(x)*

$$<\Sigma\, p(x) \log_2(p(x)/\mu(x))>$$

If these first two values differ, then there is an "Information deficit" in the forecast system, which quantifies how overconfident the forecast is.

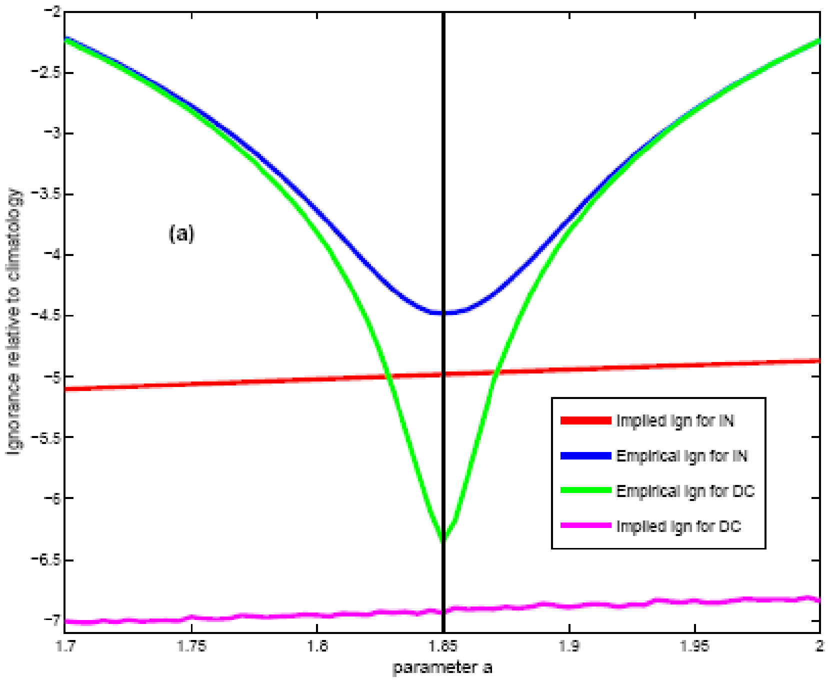
Information Deficit = *Empirical IGN – Implied IGN*

Unlike "Potential Predictability" the Information Deficit does not assume that the world becomes like the model: although incomplete, it can sometimes quantify overconfidence.

p(x) a forecast probability
q(x) ideal forecast | information
μ(x) "prior" or natural measure

Parameter Estimation: Correct Model Structure

Empirical IGN = $\langle -\log_2 p(x_{obs}) \rangle_{obs}$ Implied IGN = $\langle \text{Expected IGN} \rangle_{forecasts}$



Note that the Implied IGN

$$\langle \Sigma\, p(x) \log_2(p(x)/\mu(x)) \rangle$$

is less than the Empirical IGN

$$\langle \Sigma\, q(x) \log_2(p(x)/\mu(x)) \rangle$$

even at the correct value of a.

This Information Deficit(s) indicates that the (each) forecast scheme can still be improved.

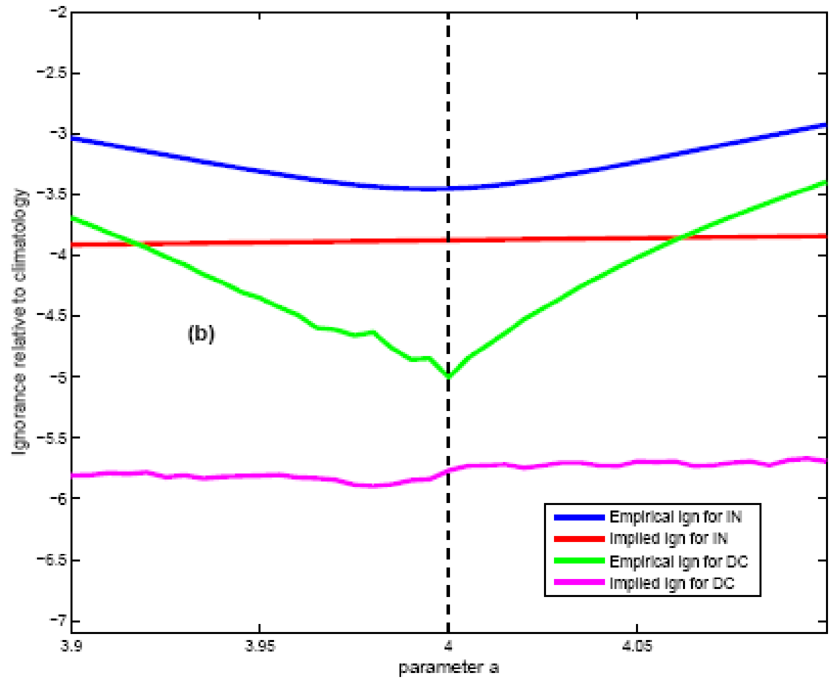
Perfect Model Structure

All Proper Scores agree
Data Assimilation Method Matters
Target uncertain (but exists)
Implied IGN reveals information deficit

H Du and L A Smith (2012) Parameter estimation using ignorance *Physical Review E 86, 016213*

But what is my model is structurally imperfect?
Parameter Estimation: Imperfect Model Structure

Empirical IGN = $\langle -\log_2 p(x_{obs}) \rangle_{obs}$ Implied IGN = $\langle \text{Expected IGN} \rangle_{forecasts}$



Structural Model Error:

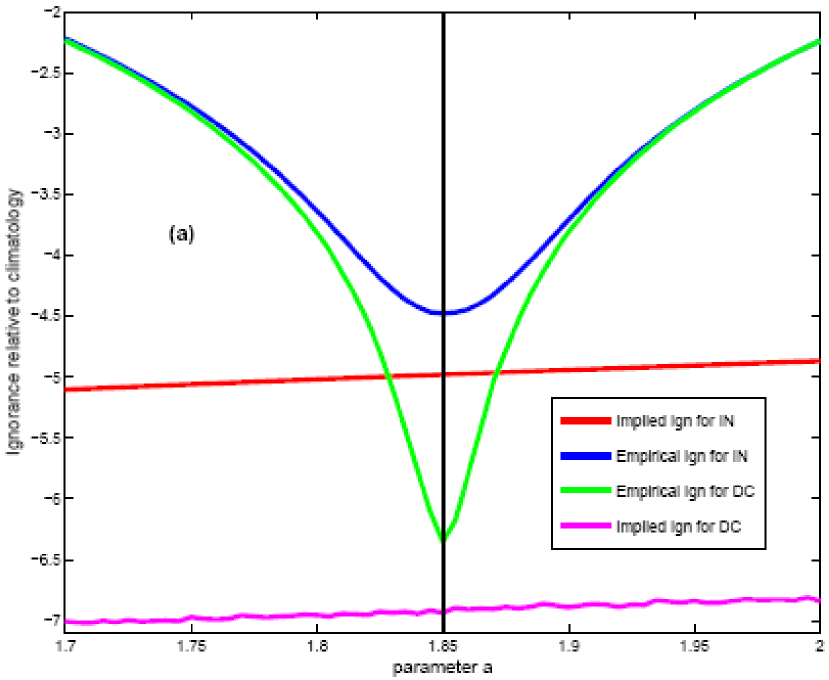
$$\text{Model Logistic Map: } l(x) = 4x(1-x)$$
$$\text{Quartic Map: } q(x) = \frac{16}{5}x(1-2x^2+x^3)$$
$$\text{System: } F(x) = (1-\epsilon)l(x) + \epsilon q(x) \text{ with } \epsilon = 0.1$$

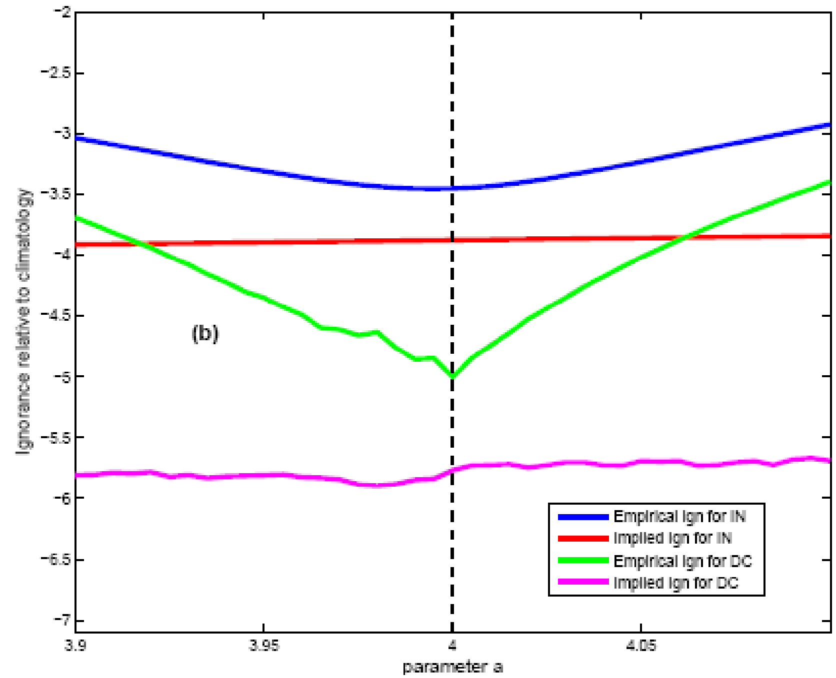
H Du and L A Smith (2012) Parameter estimation using ignorance *Physical Review E 86, 016213*

H Du and L A Smith (2012) Parameter estimation using ignorance *Physical Review E 86, 016213*

Parameter Estimation: IGN in the Logistic Map Model

Empirical IGN = $\langle -\log_2 p(x_{obs}) \rangle_{obs}$ Implied IGN = $\langle \text{Expected IGN} \rangle_{forecasts}$





Perfect Model Structure

All Proper Scores agree

DA Method Matters

Target uncertain (but exists)

Implied IGN reveals information deficit

Imperfect Model Structure

Score matters

DA Method matters

Lead-time matters

Target indeterminate (none exists)

Implied IGN reveals info deficit

Smith, L.A. and Stern, N. (2011) Uncertainty in science and its role in climate policy in Phil. Trans. R. Soc. A 369.

Kinds of Uncertainty

Good Probabilities here only

Policy makers deal regularly with these other uncertainties as well.

How well does climate modelling communicate these others?

— **Imprecision** (Knightian risk, conditional probability): related to outcomes which we do not know precisely, but for which we believe robust, decision-relevant probability statements can be provided. This is also called 'statistical uncertainty' [10–12].

— **Ambiguity** (Knightian uncertainty): related to outcomes (be they known, unknown or disputed), for which we are not in a position to make probability statements.[1] Elsewhere called 'recognized ignorance' [11,12] and 'scenario uncertainty' [10]. Ambiguity sometimes reflects uncertainty in an estimated probability, and is then referred to as 'second-order uncertainty'.

— **Intractability**: related to computations known to be relevant to an outcome, but lying beyond the current mathematical or computational capacity to formulate or to execute faithfully; also to situations where we are unable to formulate the relevant computations.

— **Indeterminacy**: related to quantities relevant to policy-making for which no precise value exists. This applies, for instance, with respect to a model parameter that does not correspond to an actual physical quantity. It can also arise from the honest diversity of views among people, regarding the desirability of obtaining or avoiding a given outcome. Noting indeterminacy reminds us of the difference between a situation where no fact of the matter exists from the case in which there is a fact of the matter but it is not known precisely.

[1] Some argue that a probability can always be assigned to any outcome. We wish to sidestep this argument, and restrict attention to decision-relevant probabilities in discussions of policy. Subjective probabilities may be the best ones available, and yet judged not good enough to quantitatively inform (as probabilities) the kinds of decisions climate policy considers. We return to this point below.

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

th

Insight without Quantitative Guidance

The robustness of the result suggests that even model-planets rather different from our best model-earth warm up about the same!

Because of the various simplifications of the model described above, it is not advisable to take too seriously the quantitative aspect of the results obtained in this study. Nevertheless, it is hoped that this study not only emphasizes some of the important mechanisms which control the response of the climate to the change of carbon dioxide,

The Effects of Doubling the CO_2 Concentration on the Climate of a General Circulation Model[1]

SYUKURO MANABE AND RICHARD T. WETHERALD

Geophysical Fluid Dynamics Laboratory/NOAA, Princeton University, Princeton, N.J. 08540

(Manuscript received 6 June 1974, in revised form 8 August 1974)

It is important to stress that our approach to the specification of discrepancy can only be expected to capture a subset of possible structural modelling errors and should be regarded as a lower bound. This is because models tend to share certain common systematic biases, which can be found in diverse elements of climate including multiannual means of basic quantities such as surface temperature,

Insight for mitigation is as good as it gets.
These quotes warn against using probabilities in adaptation.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles

J.M Murphy, B.B.B Booth, M Collins, G.R Harris, D.M.H Sexton and M.J Webb

Phil. Trans. R. Soc. A 2007 **365**, 1993-2028
doi: 10.1098/rsta.2007.2077

Structural uncertainty IS noted in the IPCC AR4:

A report of Working Group I of the Intergovernmental Panel on Climate Change

10

Global Climate Projections

The effects of uncertainty in the knowledge of Earth system processes can be partially quantified by constructing ensembles of models that sample different parametrizations of these processes. However, some processes may be missing from the set of available models, and alternative parametrizations of other processes may share common systematic biases. Such limitations imply that distributions of future climate responses from ensemble simulations are themselves subject to uncertainty (Smith, 2002), and would be wider were uncertainty due to structural model errors accounted for.

797

Admittedly, quantitative statement of the systematic errors are not easily found…

Limitations of (today's) science were clearer in the past.

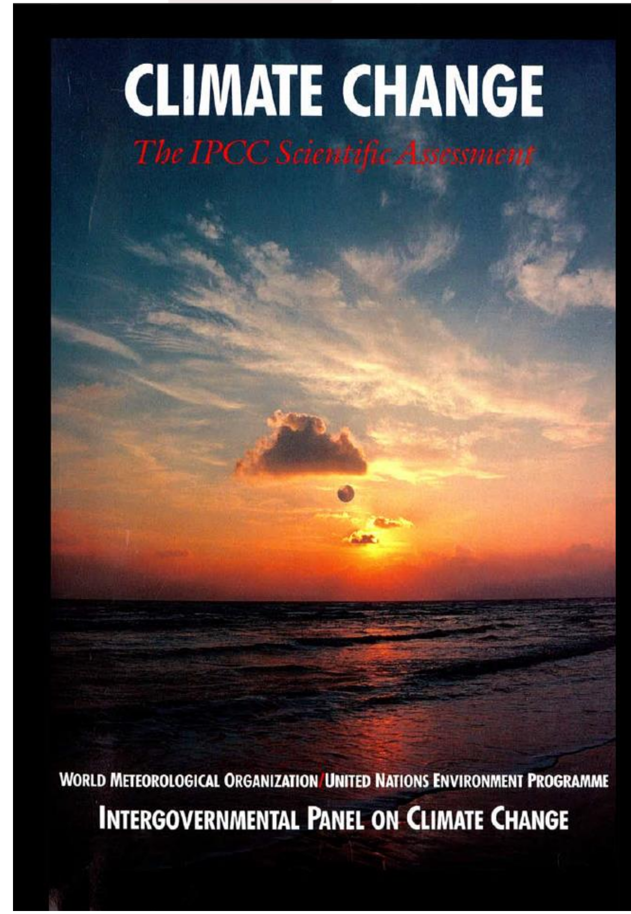
Carbon Dioxide and Climate: A Scientific Assessment

Report of an Ad Hoc Study Group on Carbon Dioxide and Climate
Woods Hole, Massachusetts
July 23-27, 1979
to the
Climate Research Board
Assembly of Mathematical and Physical Sciences
National Research Council

NATIONAL ACADEMY OF SCIENCES
Washington, D.C. 1979

The AR5 WG1 document will contain little new actionable information. It reassures us of things long known...

...but the pull for this information is increasing...



1990

In this century: An SPM resembles Highlights Approved by Policy Makers

Have we gone so far as seeking valueless numbers?

When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind…

Lord Kelvin (1883)
Lecture on "Electrical Units of Measurement" (3 May 1883),
Popular Lectures Vol. I, p. 73.

A next step will certainly be to extend Phillips' calculations to models containing more realistic energy sources and sinks and topographical inhomogeneities, in an attempt to explain some of the asymmetries in the mean circulation. It is clear by now that in the electronic computing machine we possess a most powerful weapon for attacking a great number of geophysical problems that have hitherto resisted solution because of their mathematical difficulties.

Charney (1956)

WASHINGTON CONFERENCE ON THEORETICAL GEOPHYSICS—1956 327

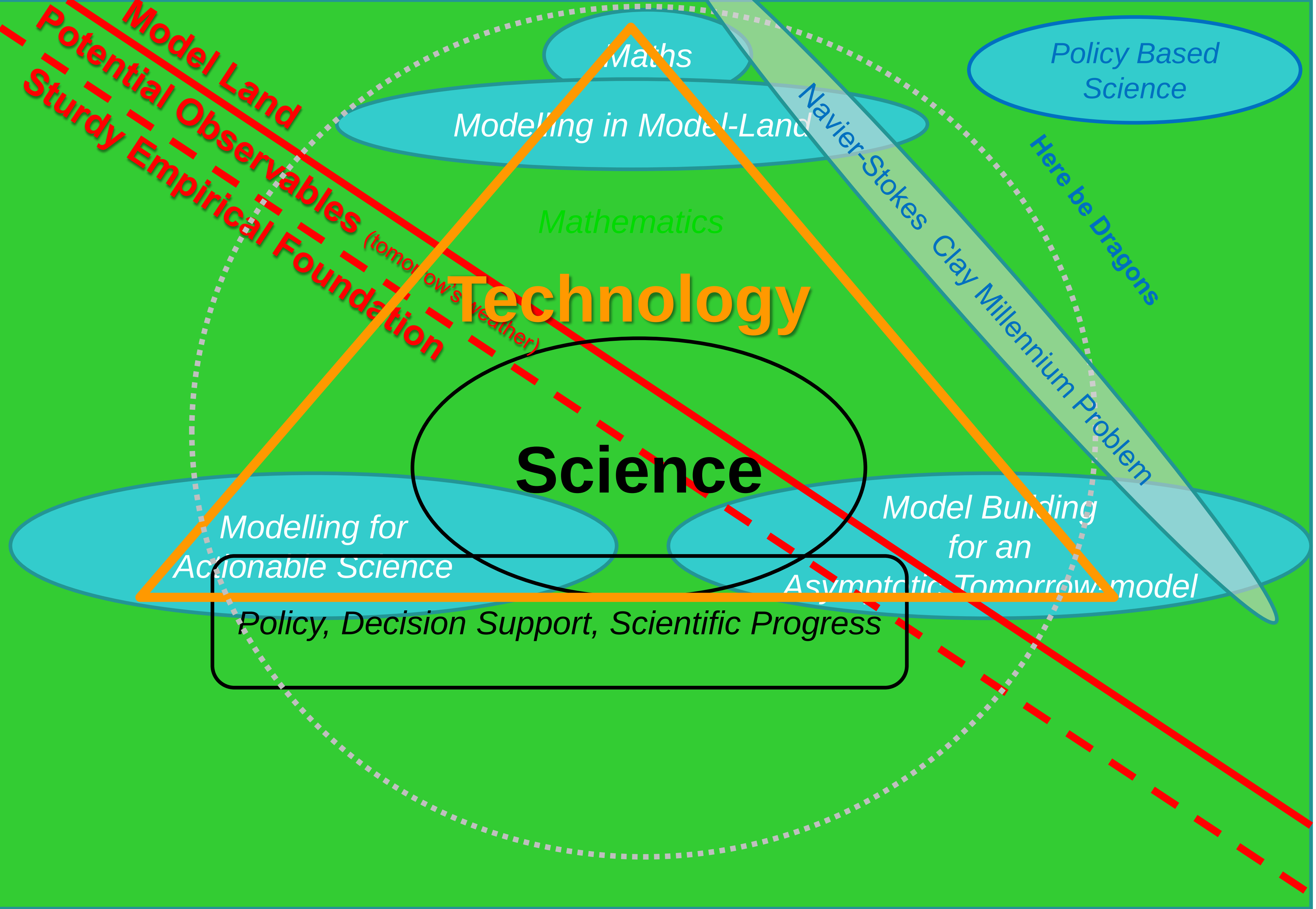
This line of argument may have been overplayed, at least in terms of forecasting (perhaps not in terms of understanding).

The provision of immature probabilities to decision makers risks the credibility of all science-based policy.

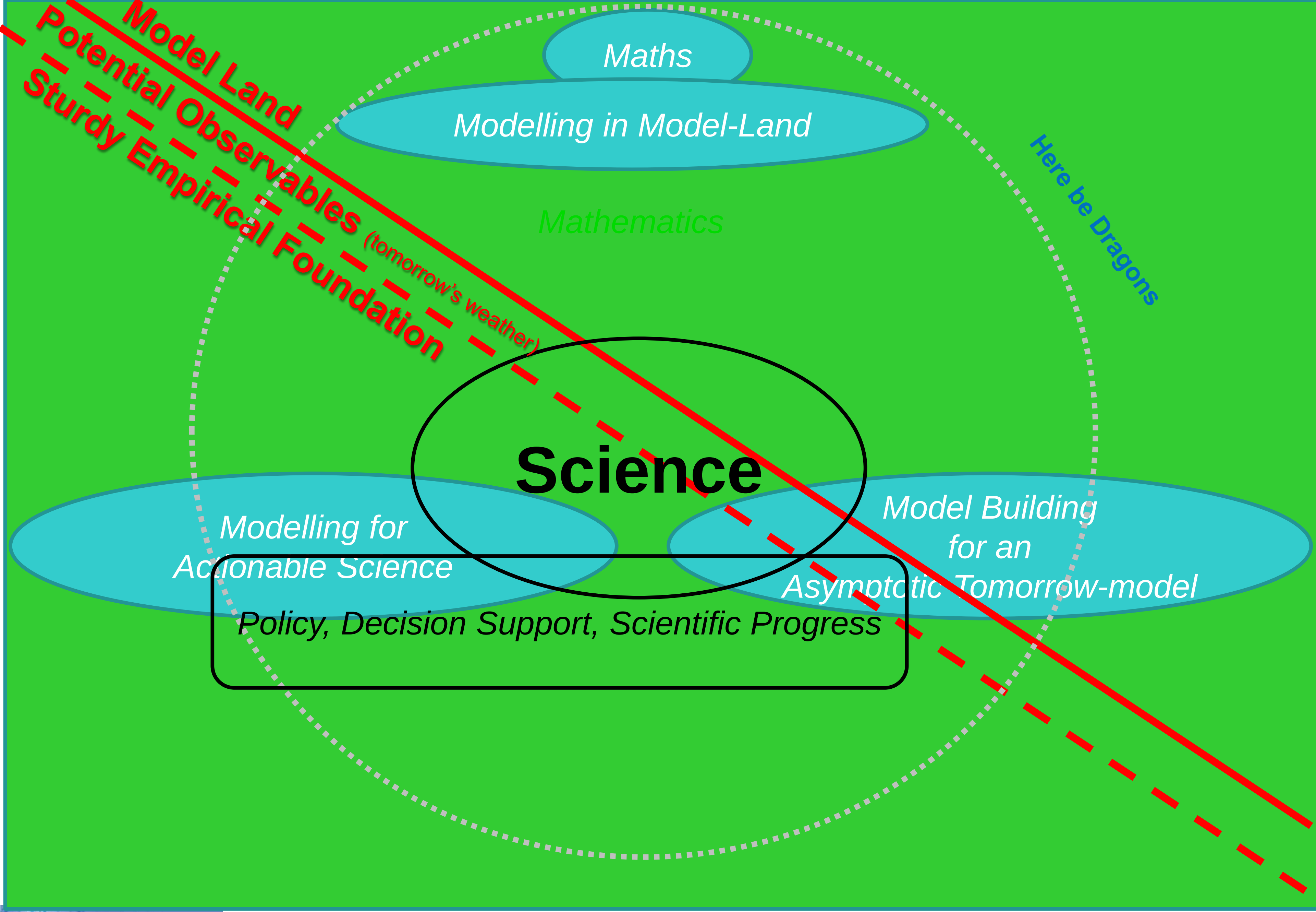
CENTRE FOR THE ANALYSIS OF TIME SERIES

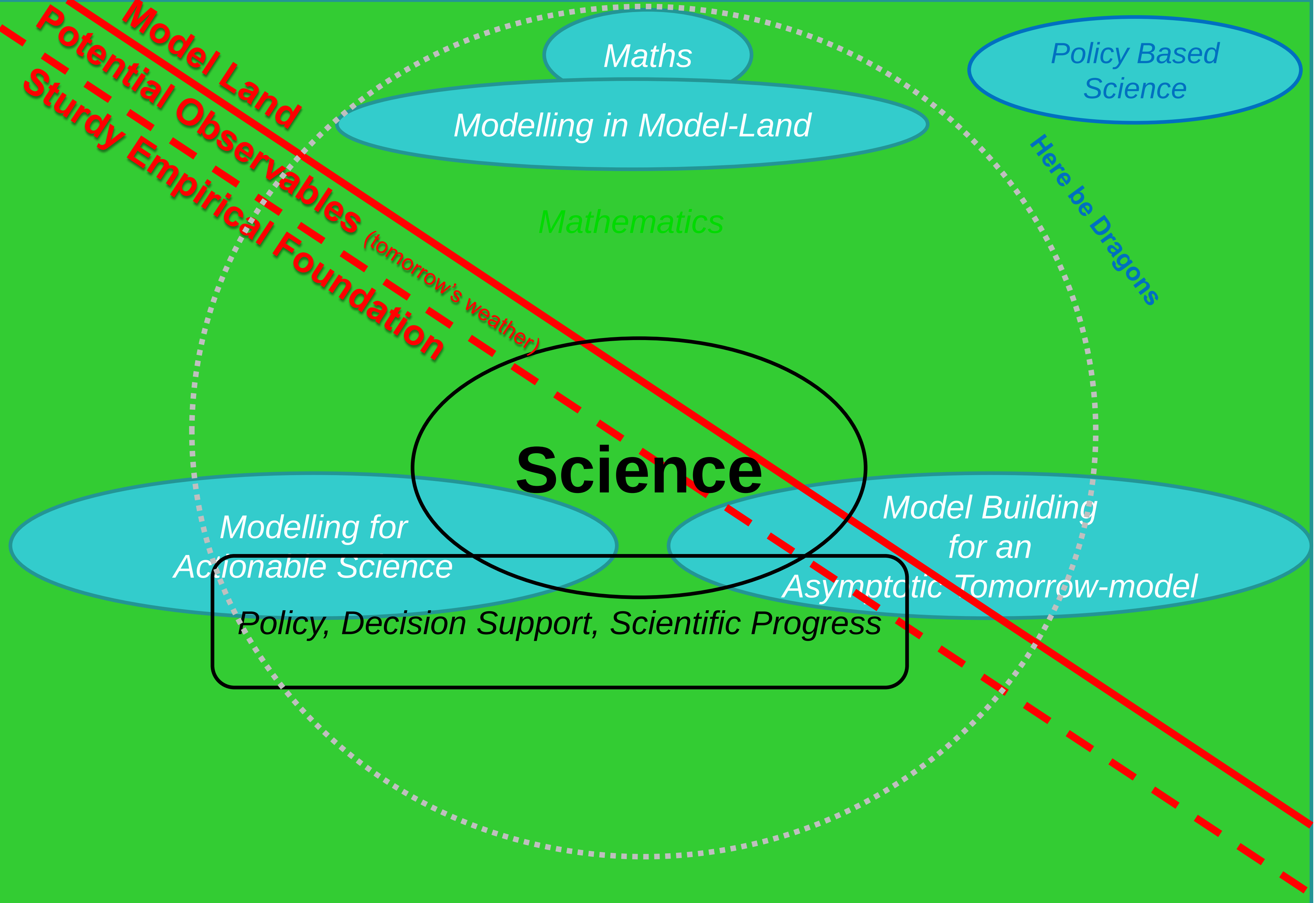
So how do we avoid this:

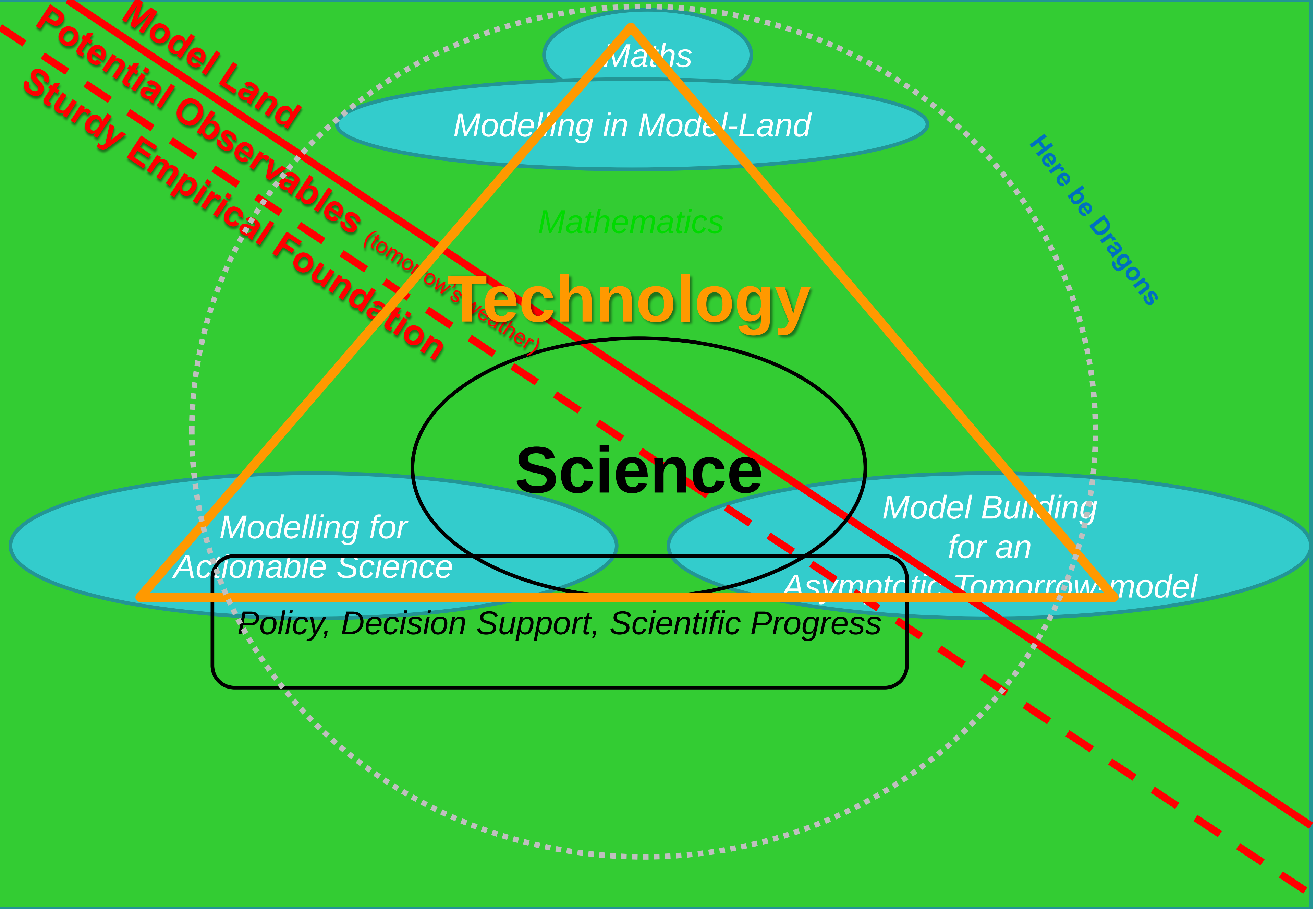
predictions are wrong
sorry for any inconvenience

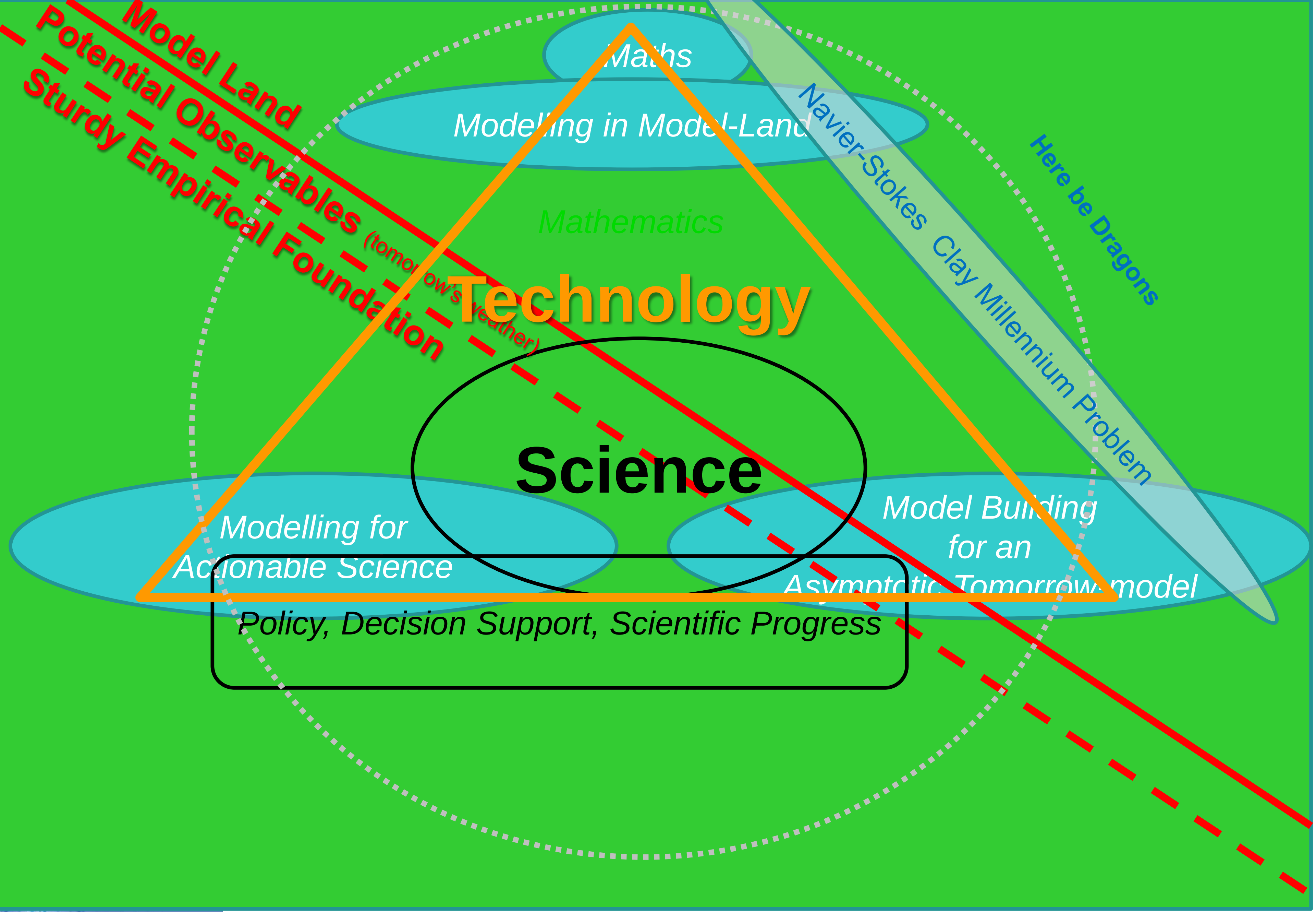


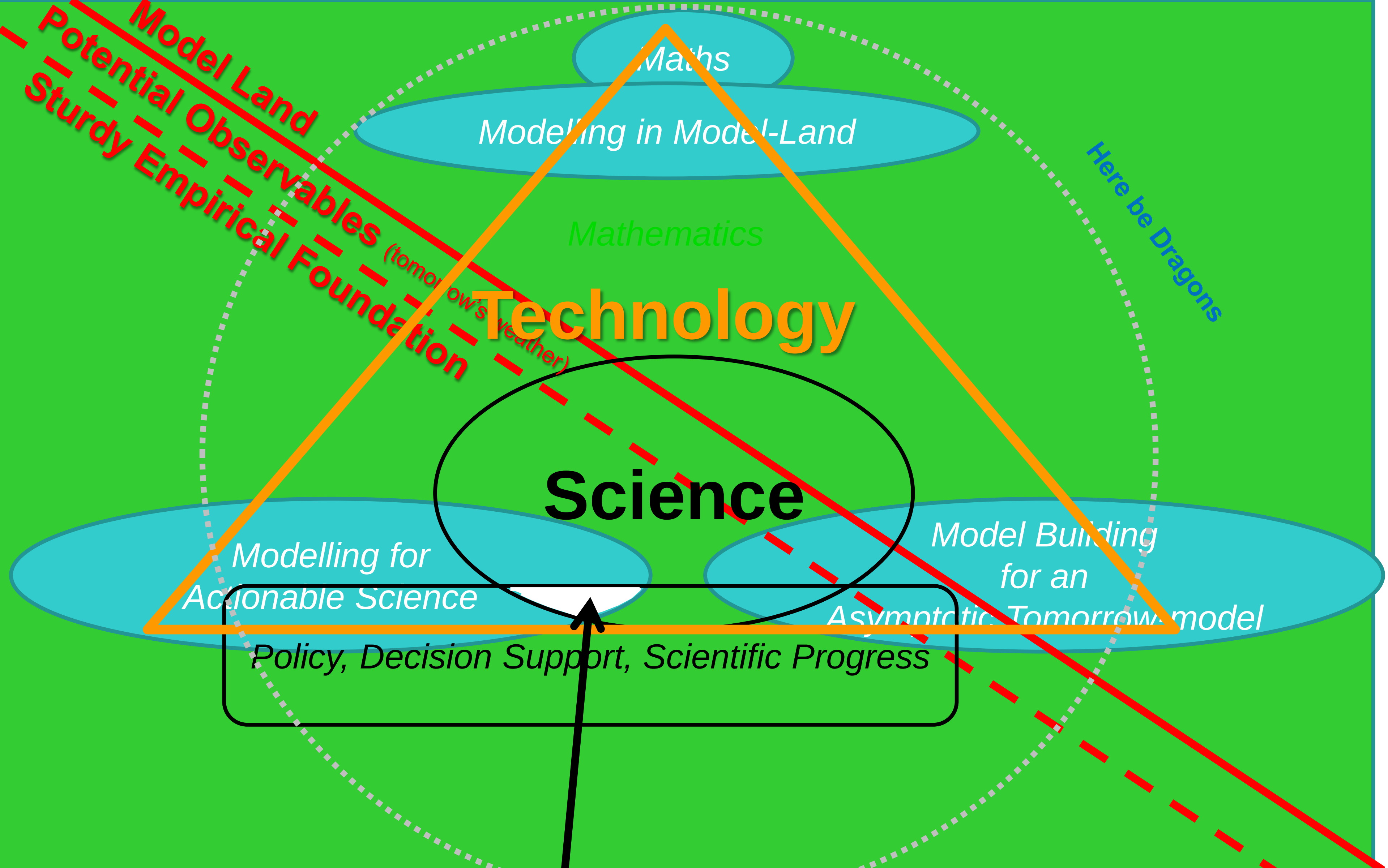
Model Land
Potential Observables (tomorrow's weather)
Sturdy Empirical Foundation
Maths
Modelling in Model-Land
Mathematics
Technology
Science
Navier-Stokes Clay Millennium Problem
Here be Dragons
Policy Based Science
Modelling for Actionable Science
Model Building for an Asymptotic Tomorrow model
Policy, Decision Support, Scientific Progress


Model Land
Potential Observables (tomorrow's weather)
Sturdy Empirical Foundation
Maths
Modelling in Model-Land
Mathematics
Here be Dragons
Science
Modelling for
Actionable Science
Model Building
for an
Asymptotic Tomorrow-model
Policy, Decision Support, Scientific Progress


Model Land
Potential Observables (tomorrow's weather)
Sturdy Empirical Foundation
Maths
Modelling in Model-Land
Policy Based Science
Here be Dragons
Mathematics
Science
Modelling for Actionable Science
Model Building for an Asymptotic Tomorrow-model
Policy, Decision Support, Scientific Progress


Model Land
Potential Observables (tomorrow's weather)
Sturdy Empirical Foundation
Maths
Modelling in Model-Land
Here be Dragons
Mathematics
Technology
Science
Modelling for Actionable Science
Model Building for an Asymptotic Tomorrow-model
Policy, Decision Support, Scientific Progress


Model Land
Potential Observables (tomorrow's weather)
Sturdy Empirical Foundation
Maths
Modelling in Model-Land
Mathematics
Technology
Science
Navier-Stokes Clay Millennium Problem
Here be Dragons
Modelling for Actionable Science
Model Building for an Asymptotic Tomorrow-model
Policy, Decision Support, Scientific Progress



Can we move climate modelling for decision support this intersection?

INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2006; 50:1159–1173
Published online 6 December 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fld.1136

Ocean Modelling

Forward-in-time upwind-weighted methods in ocean modelling

Matthew W. Hecht*,†

Los Alamos National Laboratory, Computers and Computational Sciences Division, MS B296, Los Alamos, NM 87545, U.S.A.

2. PRIMITIVE EQUATION OCEAN MODELS

The primitive equations are derived from the Navier Stokes equations, with hydrostatic and shallow approximations being made. The Boussinesq approximation is also usually, though not always, made. These issues are discussed thoroughly in Reference [8].

Consider the impact of these approximations in terms of systematic divergence of state space trajectories either **from the full equations** or from unrealistic **ocean drivers of non-ocean climatic processes**.

What are the time scales on which the model-ocean loses realism?

What are the time scales on which (perfectly modelled) non-ocean processes driven by this model-ocean lose fidelity and would invalidate model-climate for decision support?

Are these time scales (a) long relative to policy questions on the table? (b) swamped by other model-components behaving badly?

If not then the model ocean is a candidate for the RDU

Can we separate this line of reasoning from the study of the model itself?

Ocean Modelling

INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2006; **50**:1159–1173
Published online 6 December 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fld.1136

Forward-in-time upwind-weighted methods in ocean modelling

Matthew W. Hecht*,†

Los Alamos National Laboratory, Computers and Computational Sciences Division, MS B296, Los Alamos, NM 87545, U.S.A.

Bryan *et al.* [7] explained that the grid-Reynolds and grid-Peclet numbers must remain less than or equal to 2 in order to ensure that dissipation is sufficient to control this grid-point noise (this argument can also be found in the book of Griffies [8, Chapter 18]). In practice this constraint is most often violated, and with reasonable justification: One generally does not want to heavily smooth the entire model solution in order to control unphysical oscillations at a relatively few problem points.

Problem points, however, remain problematic.

It is in making decisions like this that the policy aims are central to the science.

Or more realistically: when the model output is deployed for policy relevance, it is important that someone recall the impacts of decisions "like this".

Intuitive examples: Planetary Dynamics

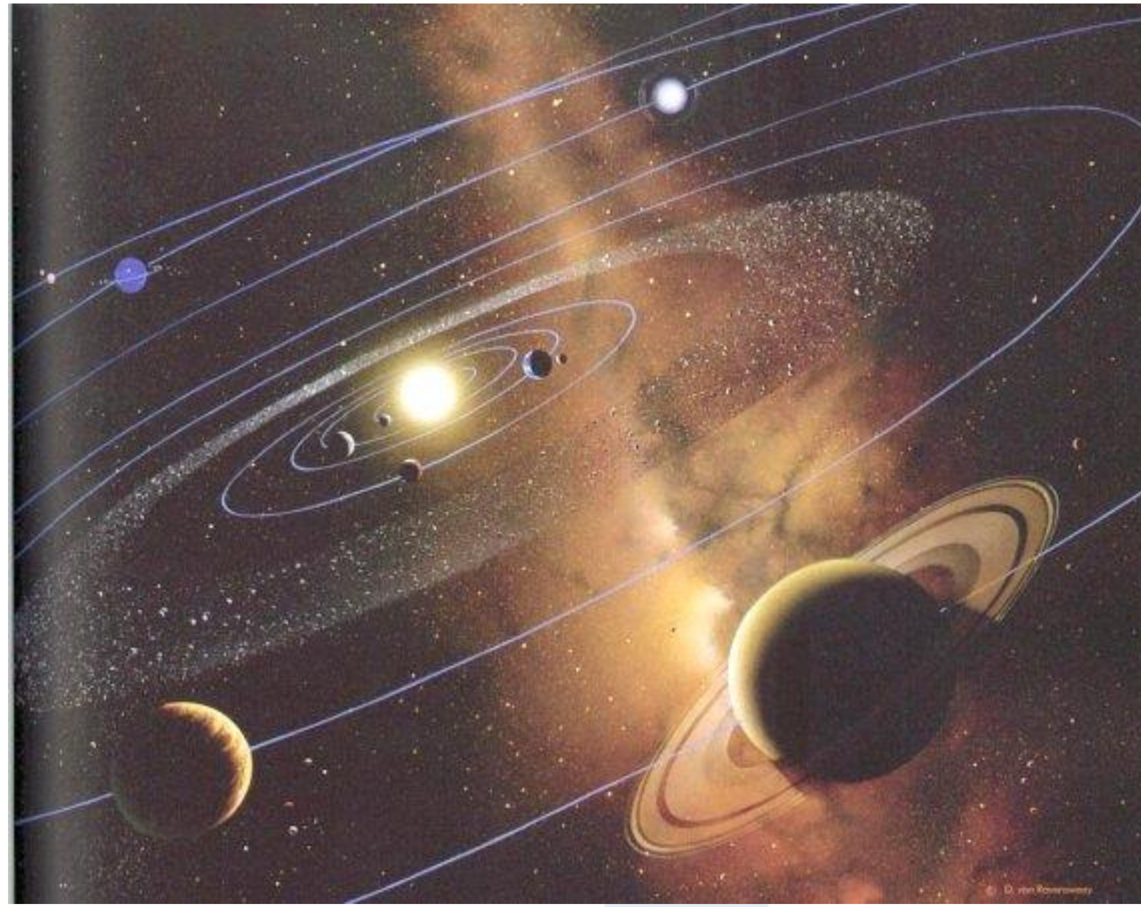
At best, our models hold only in certain circumstances. This is true even for our "Laws of Physics:" Newton's Laws are still celebrated for their successful prediction of the planet Neptune, although two historical facts (that one of the scientists who predicted Neptune also predicted the planet Vulcan, and that Vulcan was "observed" for many years) are less commonly found in physics texts. In the case of Vulcan, the then known Laws of Physics were applied outside their range of validity. By its very nature, this kind of failure is inconceivable before it is observed to have happened; because we cannot assign a meaningful probability to this occurrence, all results at the boundaries of our understanding must be treated as fundamentally uncertain.

Smith, L.A. (2002) What might we learn from climate forecasts? Proc. National Acad. Sci. USA 4 (99): 2487-2492.

Model Relevance and the Location of the Earth

If our computer model is based on Newton's Laws, and the Question is the location of the Earth in 10^8 years….

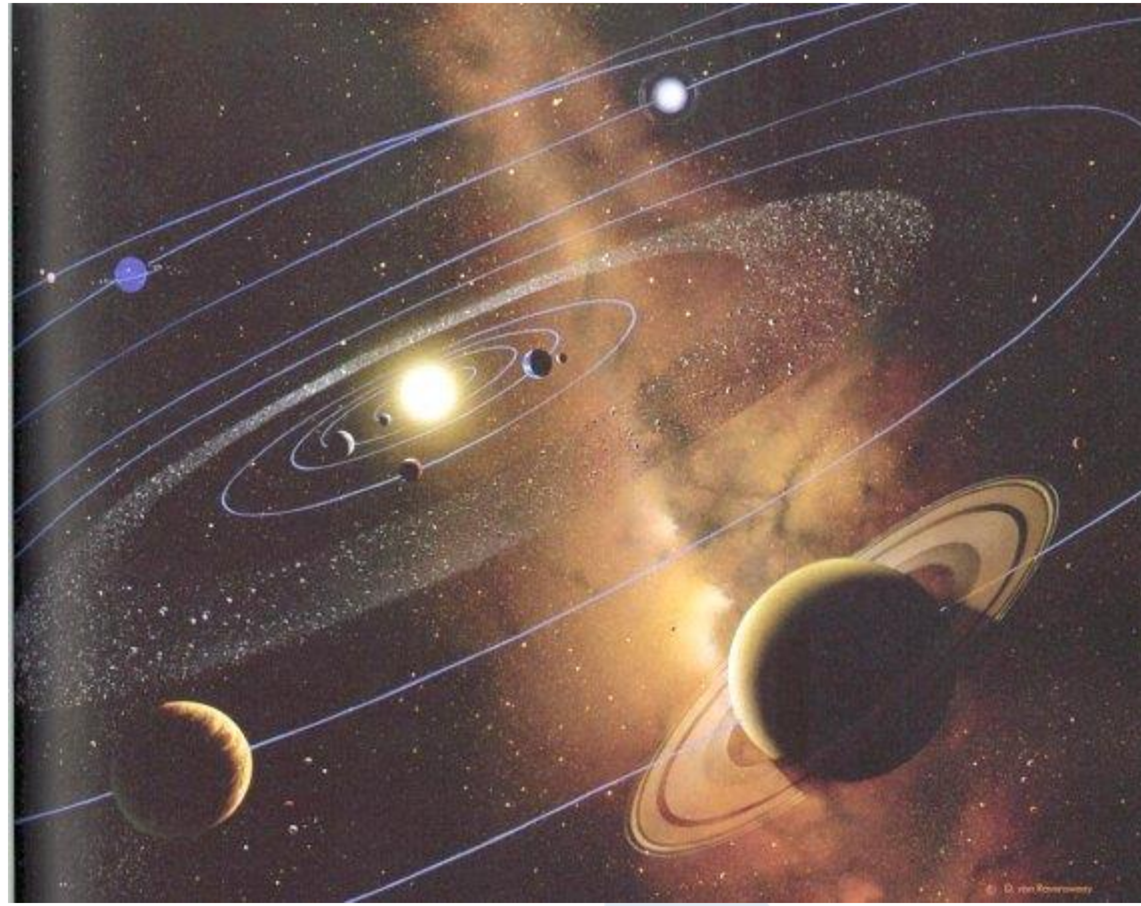
Monte Carlo sampling of from a prior distribution of initial condition and mass of each major planet will yield a final time distribution one thousand years hence, from which we can form a sensible PDF.



Bayesian Discrepancy terms, based on different integration schemes, compilers and computer hardware **may yield valuable information**.

Model Relevance and the Location of the Earth

If our computer model is based on Newton's Laws, and the Question is the location of the Earth in 10^6 years....

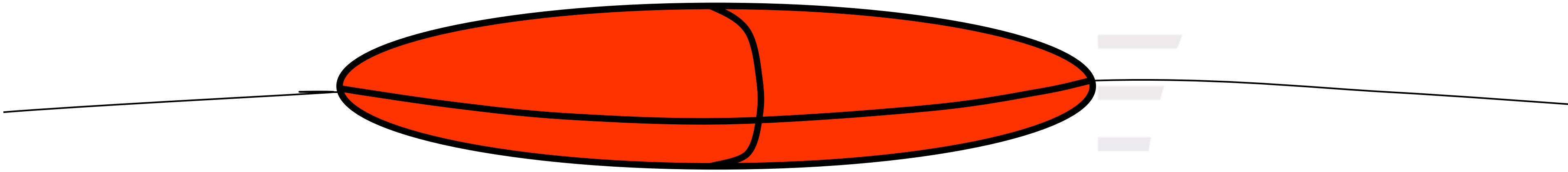


Monte Carlo sampling of from a prior distribution of initial condition and mass of each major planet will yield a final time distribution one thousand years hence, from which we can form a sensible PDF.

Bayesian Discrepancy terms, based on different integration schemes, compilers and computer hardware **may yield valuable information**.

The Probability of being outside the 99% level?

Our models allow us to draw a balloon in space that captures 99% of the model-earth trajectories.



And our/my subjective probability that our Earth will fall outside that corresponding real-world balloon might rationally be ~1%

The orange balloon corresponds to a probability contour for capturing the Earth.

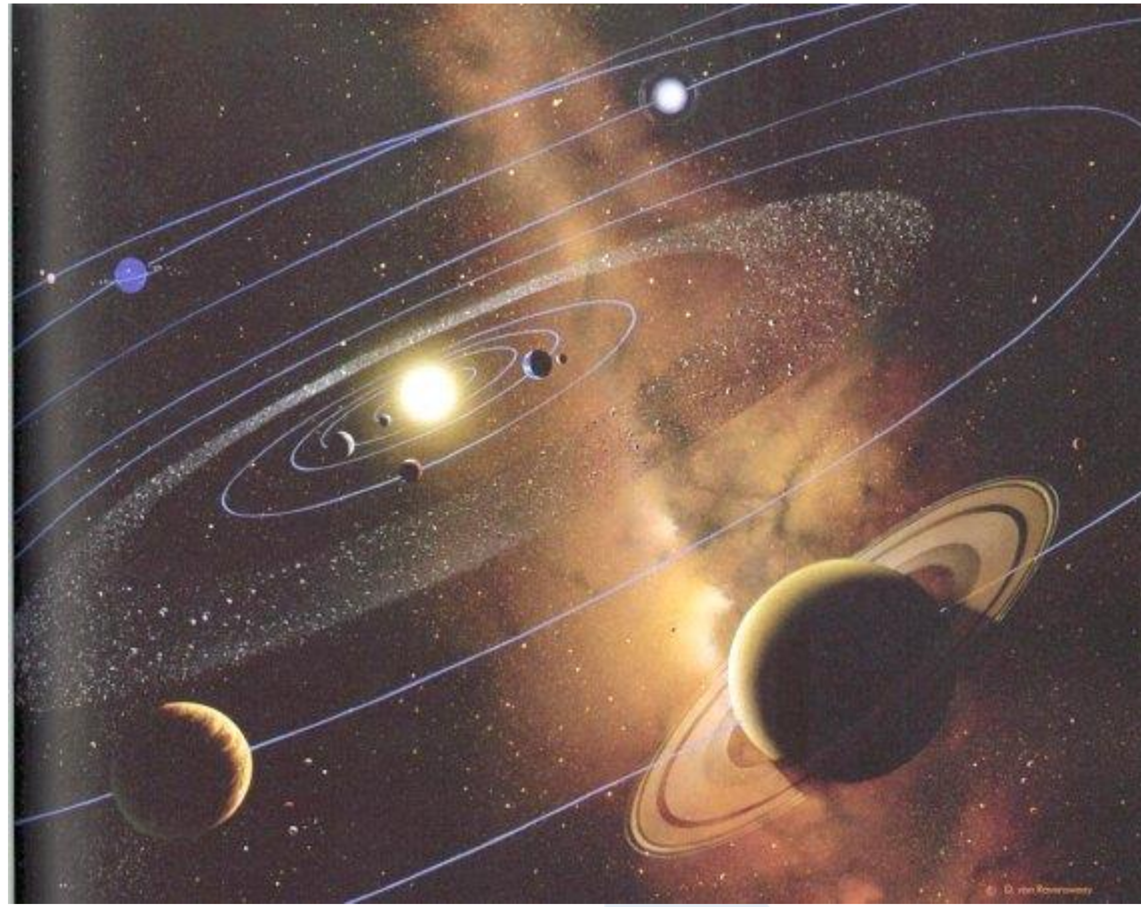
Model Relevance and the Location of Mercury

If our computer model is based on Newton's Laws, and the Question is the location of **Mercury**….

Then this approach is absurd.

Newton's Laws are long known to fail for Mercury.

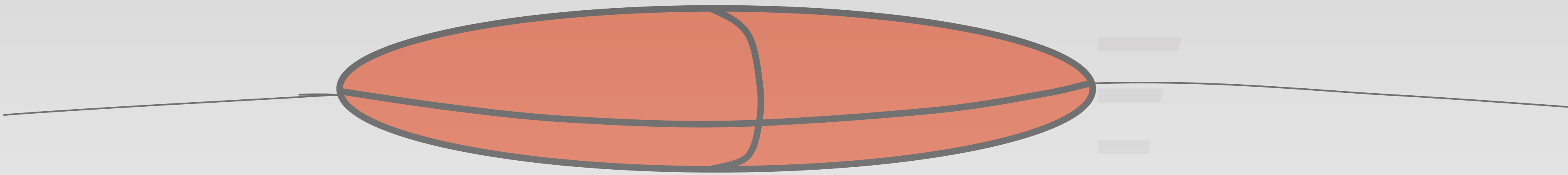
And science could warn us the Newtonian-PDF was misleading long before General Relativity provided an empirically adequate model!



Discrepancy based on different integration schemes, etc, will lead to over-confidence, belief in spurious accuracy and bad decision making!

The Probability of being above the 99% level?

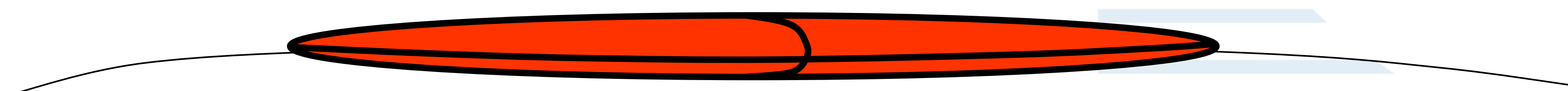
Our model allows us to draw a balloon in space that captures 99% of the model-earth trajectories.



And our subjective probability that the real earth will fall outside that corresponding real-world balloon might be rationally be ~1%

And for Mercury?
Our subjective probability that the real Mercury will fall outside the real-world balloon corresponding to the volume that captures 99% of our model-Mercury trajectories… might be **say, ~90%!**

For Mercury, relevant science is more informative than the Newtonian model.



Sit and think, don't just simulate and count!

The AR4 acknowledges this shortcoming explicitly

Uncertainty and the IPCC Sixty-Forty Rule

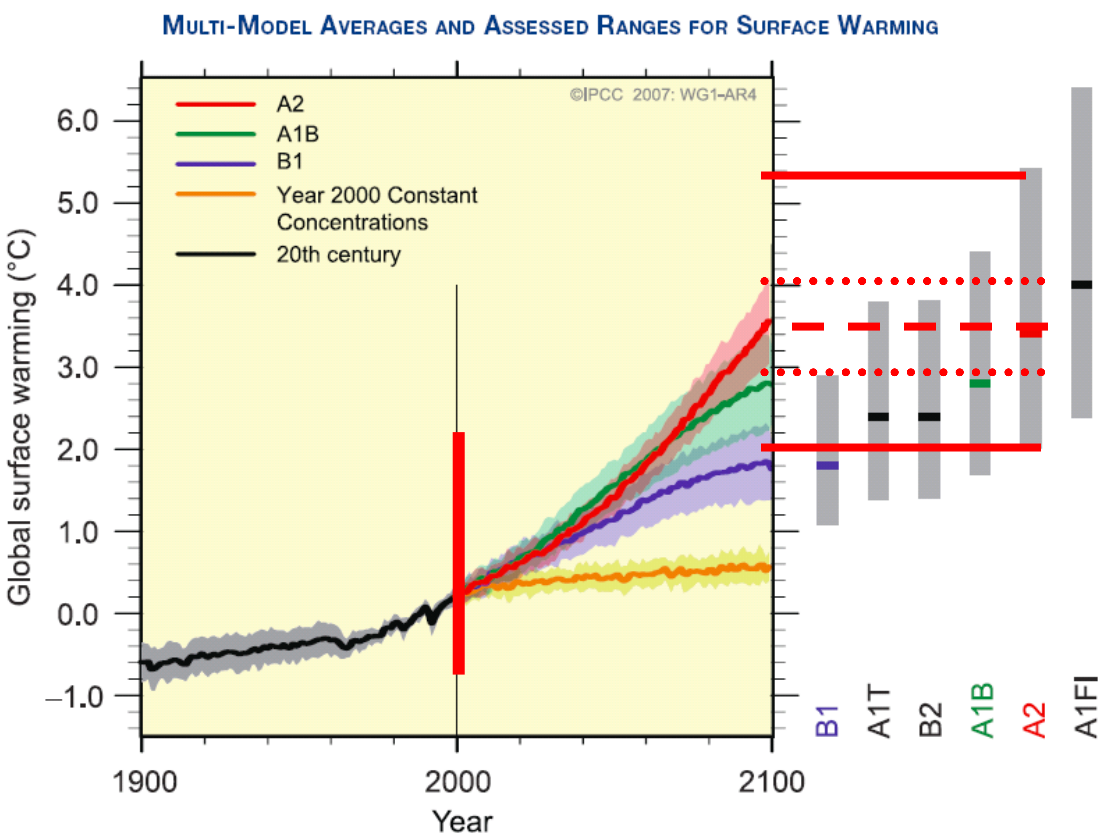


Figure SPM.5. *Solid lines are multi-model global averages of surface warming (relative to 1980–1999) for the scenarios A2, A1B and B1, shown as continuations of the 20th century simulations. Shading denotes the ±1 standard deviation range of individual model annual averages. The orange line is for the experiment where concentrations were held constant at year 2000 values. The grey bars at right indicate the best estimate (solid line within each bar) and the **likely** range assessed for the six SRES marker scenarios. The assessment of the best estimate and **likely** ranges in the grey bars includes the AOGCMs in the left part of the figure, as well as results from a hierarchy of independent models and observational constraints. {Figures 10.4 and 10.29}*

The conditional forecasts (projections) are the grey bars (right); they differ from the ensemble distributions left and centre.

On what space and time scales can decision makers (and economic modellers) have rational confidence in model based probabilities?

The IPCC does not interpret the diversity of ensembles as directly reflecting the probability distribution of future Global Mean Temperature.

An intuitive example: The Last Solar Eclipse

So for Newton's Laws, the story goes something like this:

At best, our models hold only in certain circumstances. This is true even for our "Laws of Physics:" Newton's Laws are still celebrated for their successful prediction of the planet Neptune, although two historical facts (that one of the scientists who predicted Neptune also predicted the planet Vulcan, and that Vulcan was "observed" for many years) are less commonly found in physics texts. In the case of Vulcan, the then known Laws of Physics were applied outside their range of validity. By its very nature, this kind of failure is inconceivable before it is observed to have happened; because we cannot assign a meaningful probability to this occurrence, all results at the boundaries of our understanding must be treated as fundamentally uncertain.

Our model does not have sufficient fidelity for Mercury on time scales of either target question.
This lack of fidelity will surely impact the locations of the other planets,
Estimate of this impact can be obtained via back of the envelope calculation.

For the particular question of the last solar eclipse this appears (say) not to be the Relevant Dominant Uncertainty:

- *Mercury's violation of Newton's Laws does not limit us here.*
- *Focus Resources elsewhere.*
- *Quantify the (shorter) time-to-(in)fidelity of our models of tidal breaking.*

Do we even need Mercury in the model for the last solar eclipse?!?

For the particular question of the future location of Mercury, this is arguably the RDU:

- *Is there a multi-model emulation fix :* ***No***
- *Is there an empirical fix: ?unlikely on these time scales?*
- *Obtain (or wait for) resources to code General Relativity.*

There is no quantitative decision-relevant info, other than the fact itself.

In the second case, the rational policy approach is risk management, not cost benefit analysis

Structural Model Error: Internal Consistency

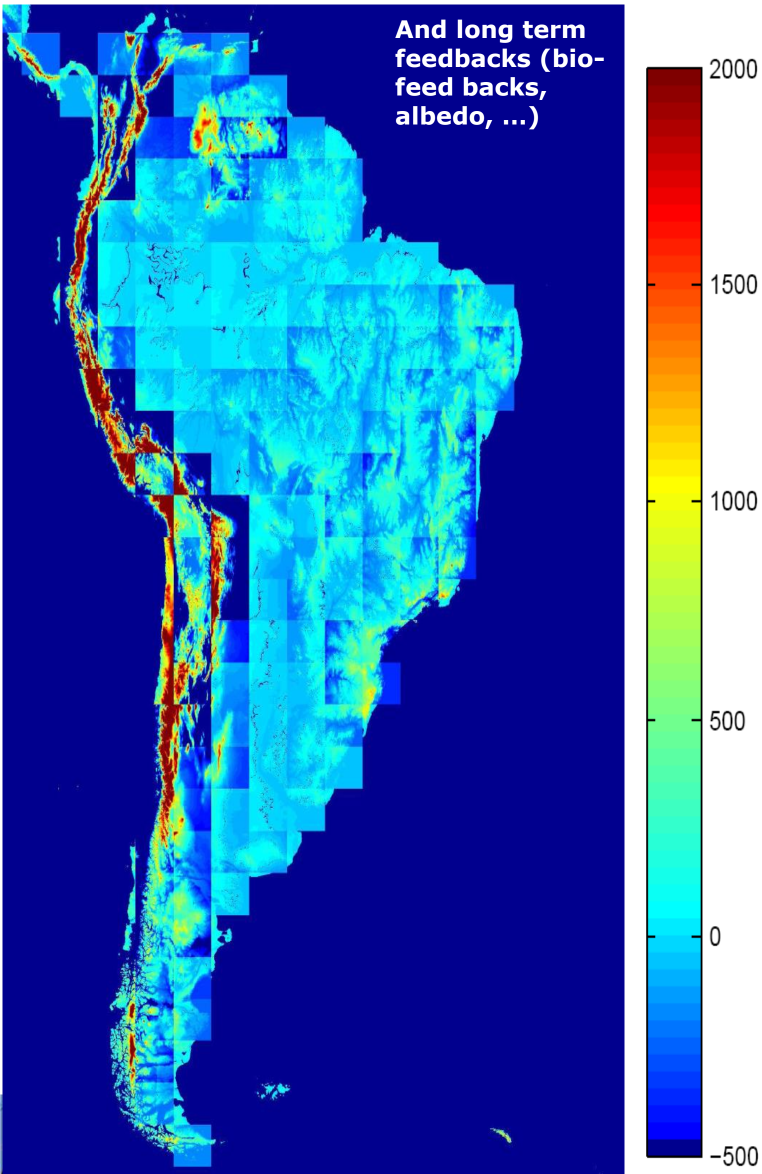
Smith, L.A. (2002) [What might we learn from climate forecasts?](#) Proc. National Acad. Sci. USA 4 (99): 2487-2492.

we can never make objective probability statements on the basis of our climate simulations. What we can do is establish their internal consistency: we can determine for which phenomena and on which time scales our models might reflect reality.

. Real statisticians will immediately object, of course, that capturing the phenomena "in-sample" does not guarantee our ability to capture it "out-of-sample," that is, in the future. This is true, but we are seeking only a necessary condition: if our models cannot capture the phenomena of interest over the data period from which the model was constructed, say 1950–2000, then those interested only in economic impacts should not even look at the statistics of those phenomena in 2000–2050.

Shortcomings of State of the Art Models

Missing 2km tall walls of rock!



Observed Height – HadCM3 Height

At what lead times do inadequacies in downstream flow (or precipitation) result in feedbacks with beyond local impacts? alter extremes? &c?

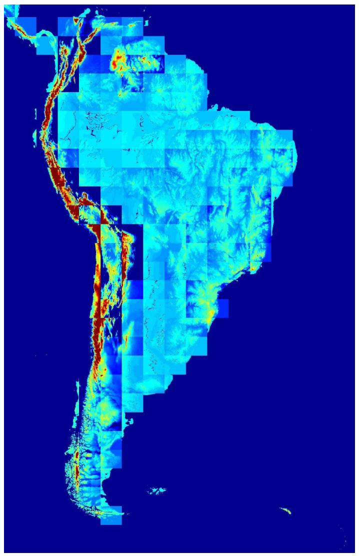
At what lead times is it no longer reasonable to interpret the diversity of climate models as reflecting the uncertainty in the future climate?

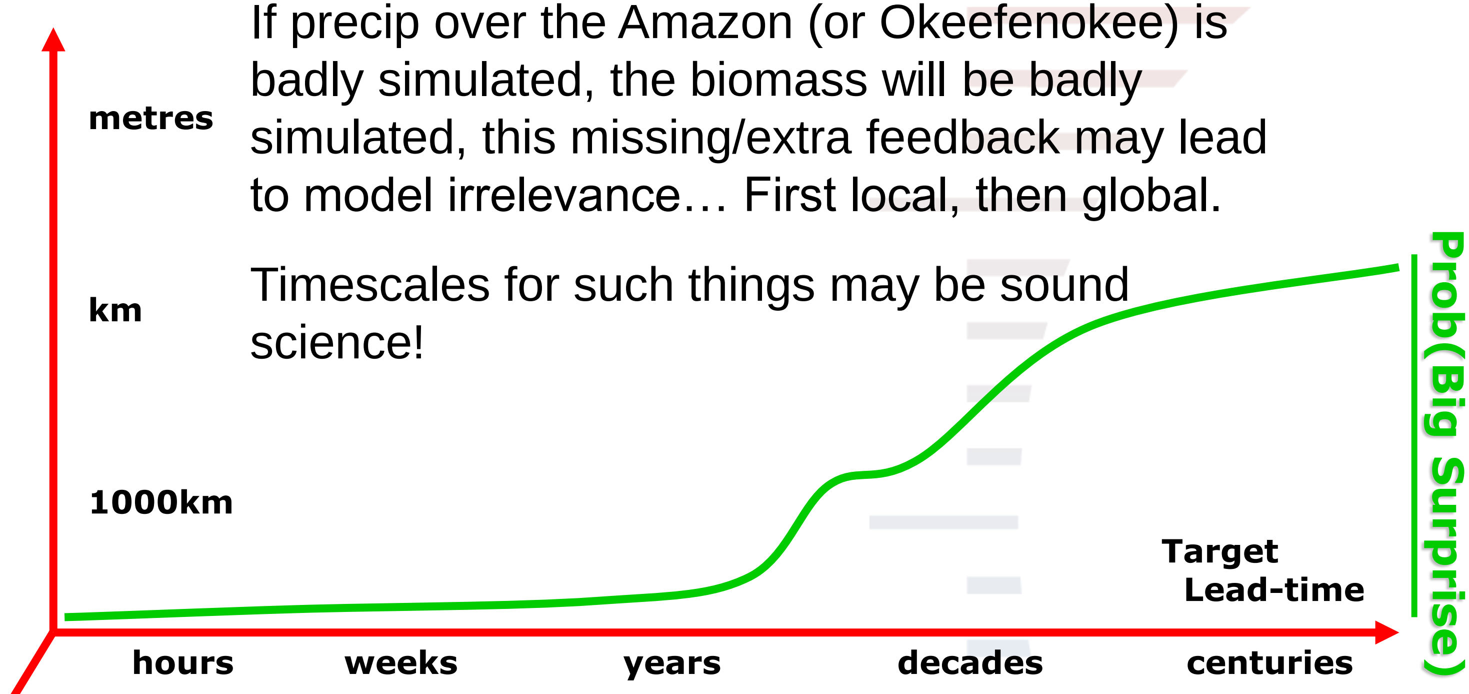
We can quantify these limits as it continues to develop methodology to make the dynamics of state-of-the-art models available to decision makers and economists.

Or we can redeploy resources used beyond these limits to better inform policy.

A decision makers needs one and only one thing: A deadline.

Model-based probability forecasts are incomplete without a quantitative measure of the likelihood of model irrelevance.

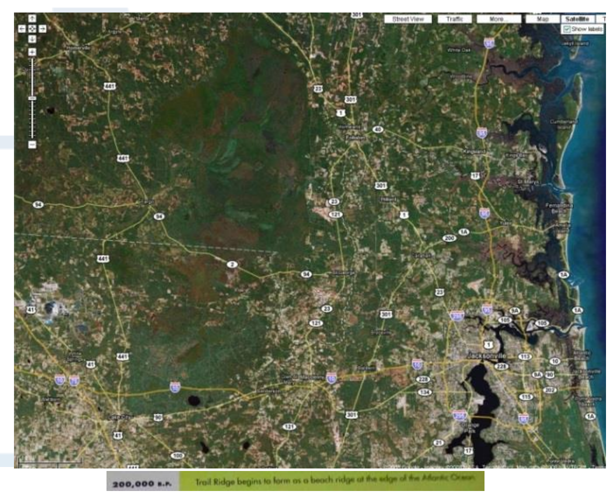




If precip over the Amazon (or Okeefenokee) is badly simulated, the biomass will be badly simulated, this missing/extra feedback may lead to model irrelevance… First local, then global.

Timescales for such things may be sound science!

"No presentation of model-based probabilities is complete without an expression of model irrelevance."



Relevant Dominant Uncertainty (RDU):

The RDU is the source of uncertainty or error that is the primary limit on the fidelity of our probability forecasts. It could be

- *knowledge of natural measure (more obs)*
- *knowledge of simulation PDF (from the ensemble)*
- *knowledge of most relevant current model states (DA)*
- *model fidelity w.r.t reality for a given forecast target (MI)*

Communicating Dynamic or Immature Probability forecasts will carry less risk to the credibility science-based decision support when the RDU is discussed and your Probability of a Big Surprise is provided, explicitly.

How Would We (Re)design for Policy?

Start with a decision/policy target question:

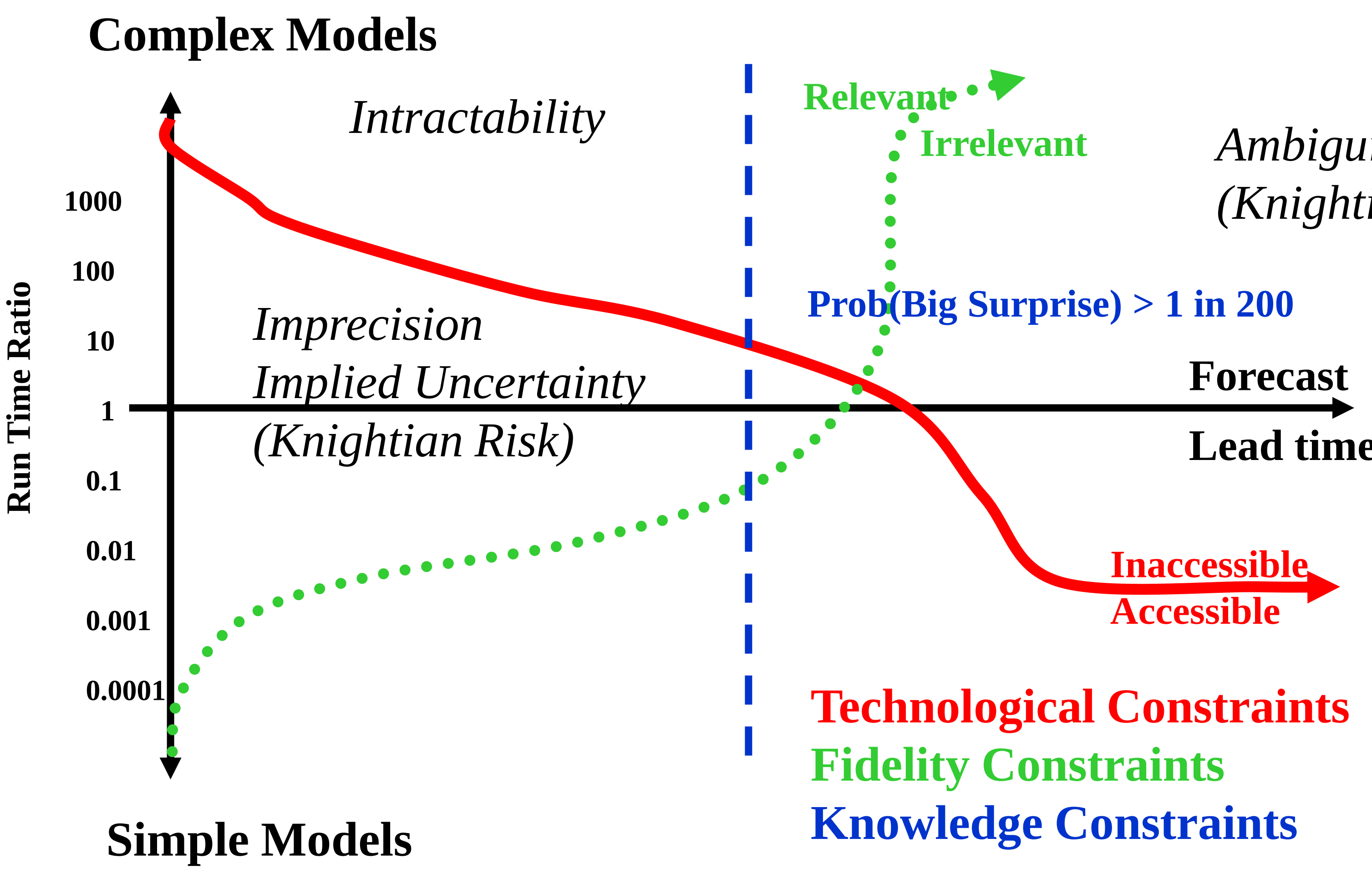
What is the target question?
What are the relevant phenomena?
On what time scales do we have high fidelity simulation?
Might these reach the time scales of interest? **If yes: Simulate!**
What is the Relevant Dominant Uncertainty (RDU)?
What kind of uncertainty limits the fidelity of simulation?
What is the Probability of a Big Surprise?
Is the available simulation sufficiently informative to justify the
the resources required to make it?
If yes: Simulate!

Clear Disclosure: Are Actionable Distributions In-Hand (or likely to be available soon)

Weather models look more like simplified climate models.

The Constraints on Simulation Modelling for Prediction

What are the challenges we face with interpreting model simulations in different regions of this schematic?



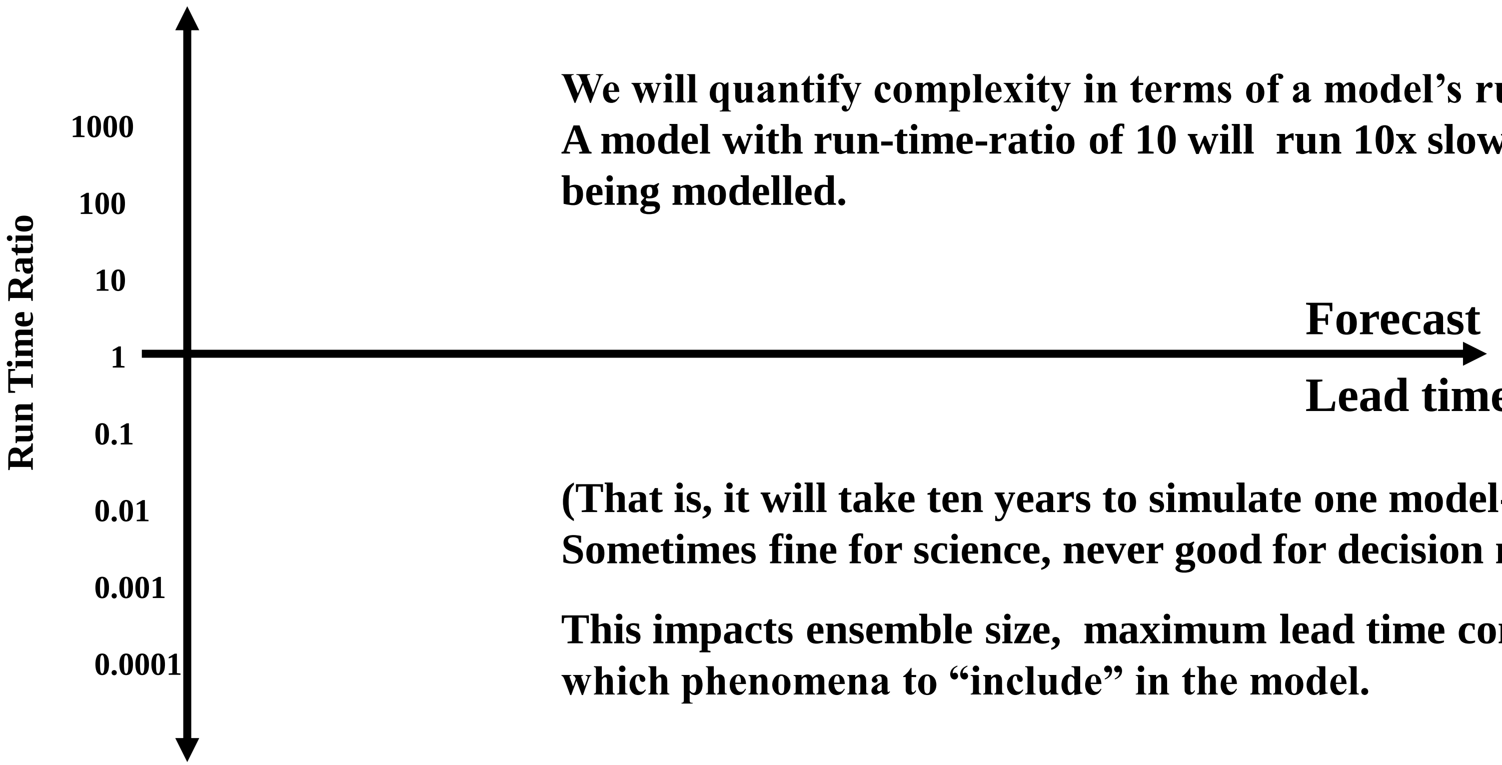
LA Smith & N Stern (2011) Uncertainty in science and its role in climate policy Phil. Trans. R. Soc. A (2011), 369, 1-24.

Leonard Smith

What are you constrained by?

For decision support, the model has to run faster than real time.
The larger the lead time, the fewer ensemble members you can run to examine sensitivity.

Complex Models



We will quantify complexity in terms of a model's run-time-ratio. A model with run-time-ratio of 10 will run 10x slower than the system being modelled.

(That is, it will take ten years to simulate one model-year. Sometimes fine for science, never good for decision makers.)

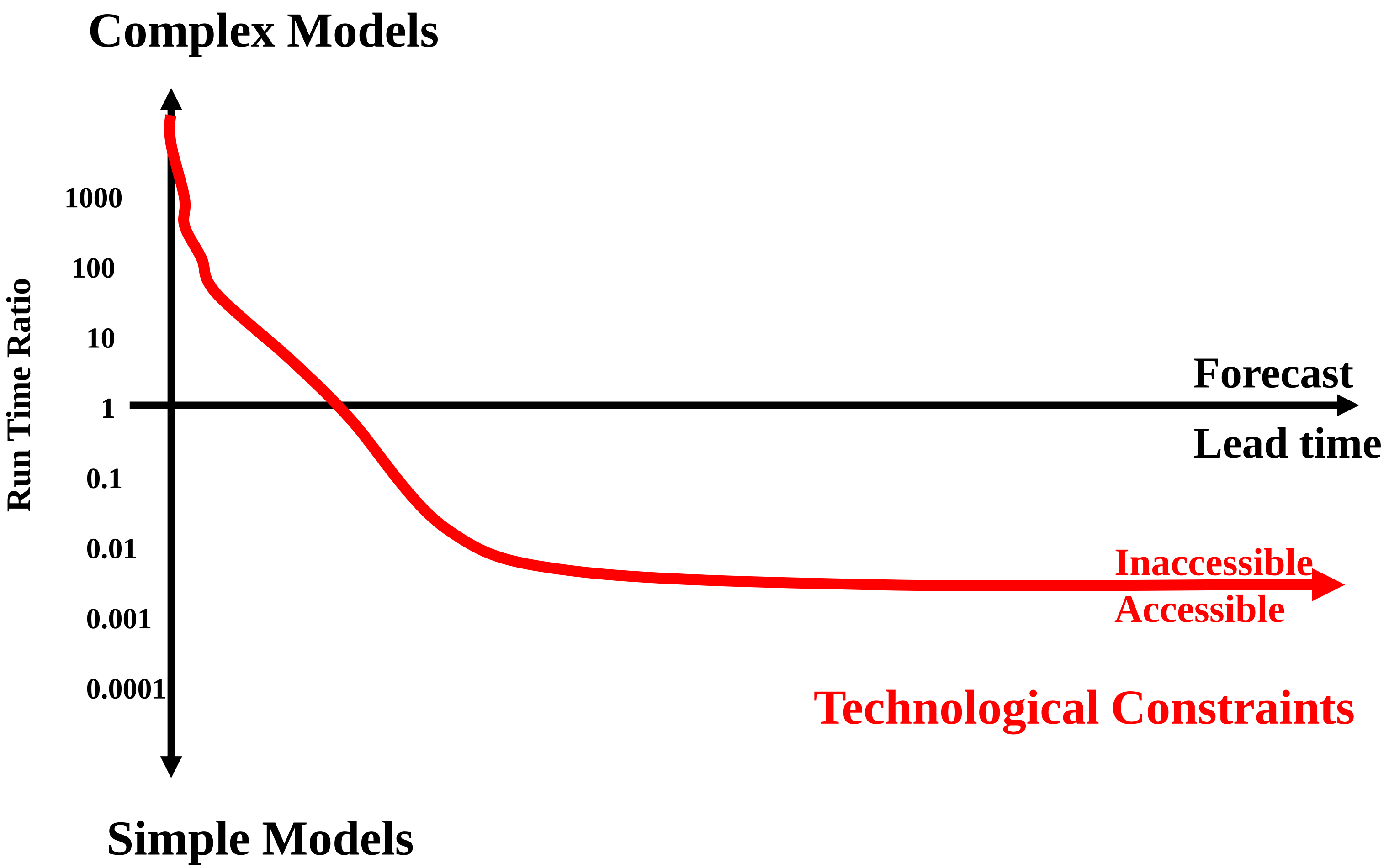
This impacts ensemble size, maximum lead time considered, and which phenomena to "include" in the model.

Simple Models

What are you constrained by?

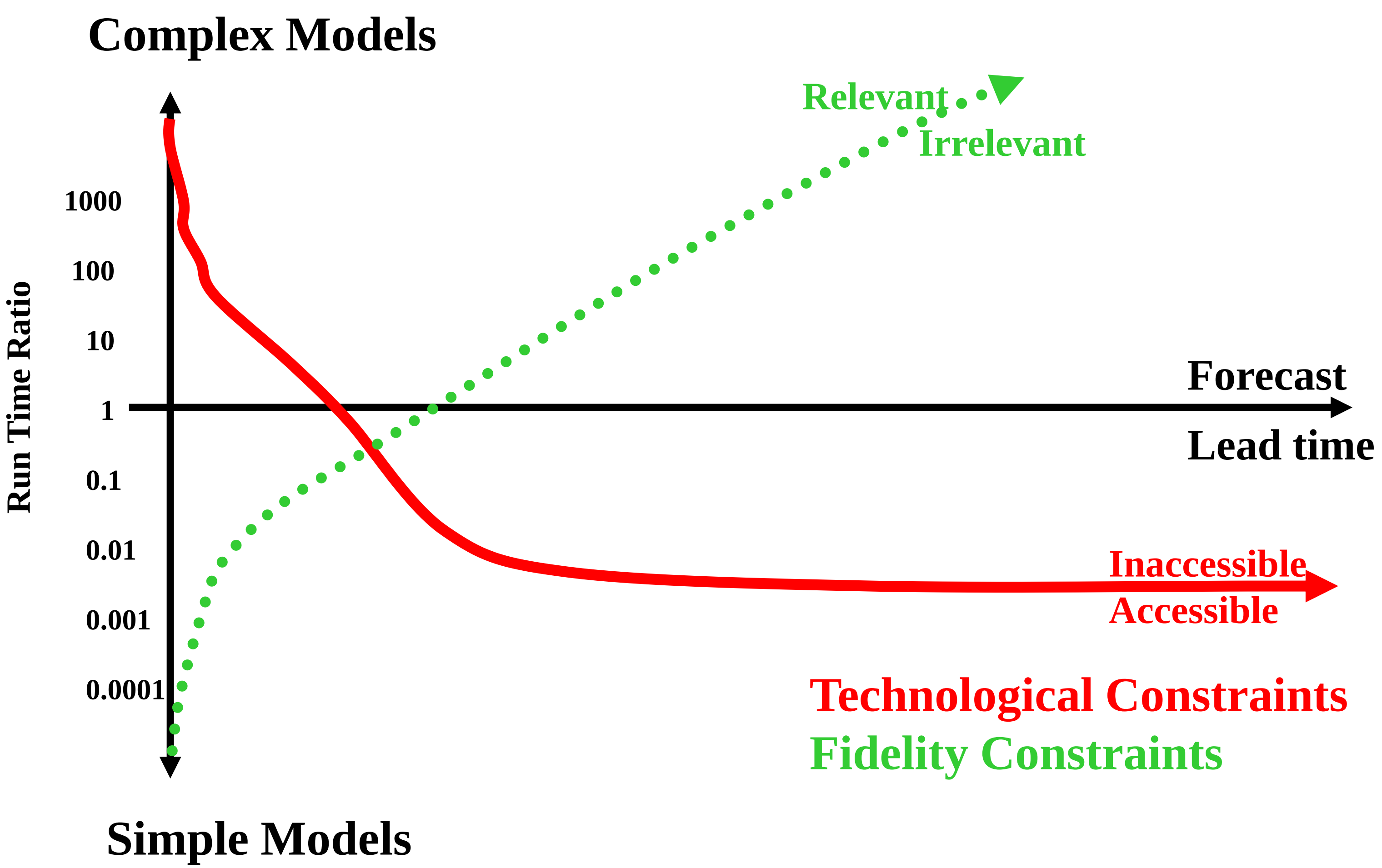
this leads to immature probabilities.

Complex models may not fit in current hardware, even if you know what you would build.
And the more complex your model, the fewer "simulation hours" you will have.



What are you constrained by?

Requirements for model fidelity sets a lower bound on the complexity with lead time.
Almost always, the model is required to grow more complex at larger lead times.

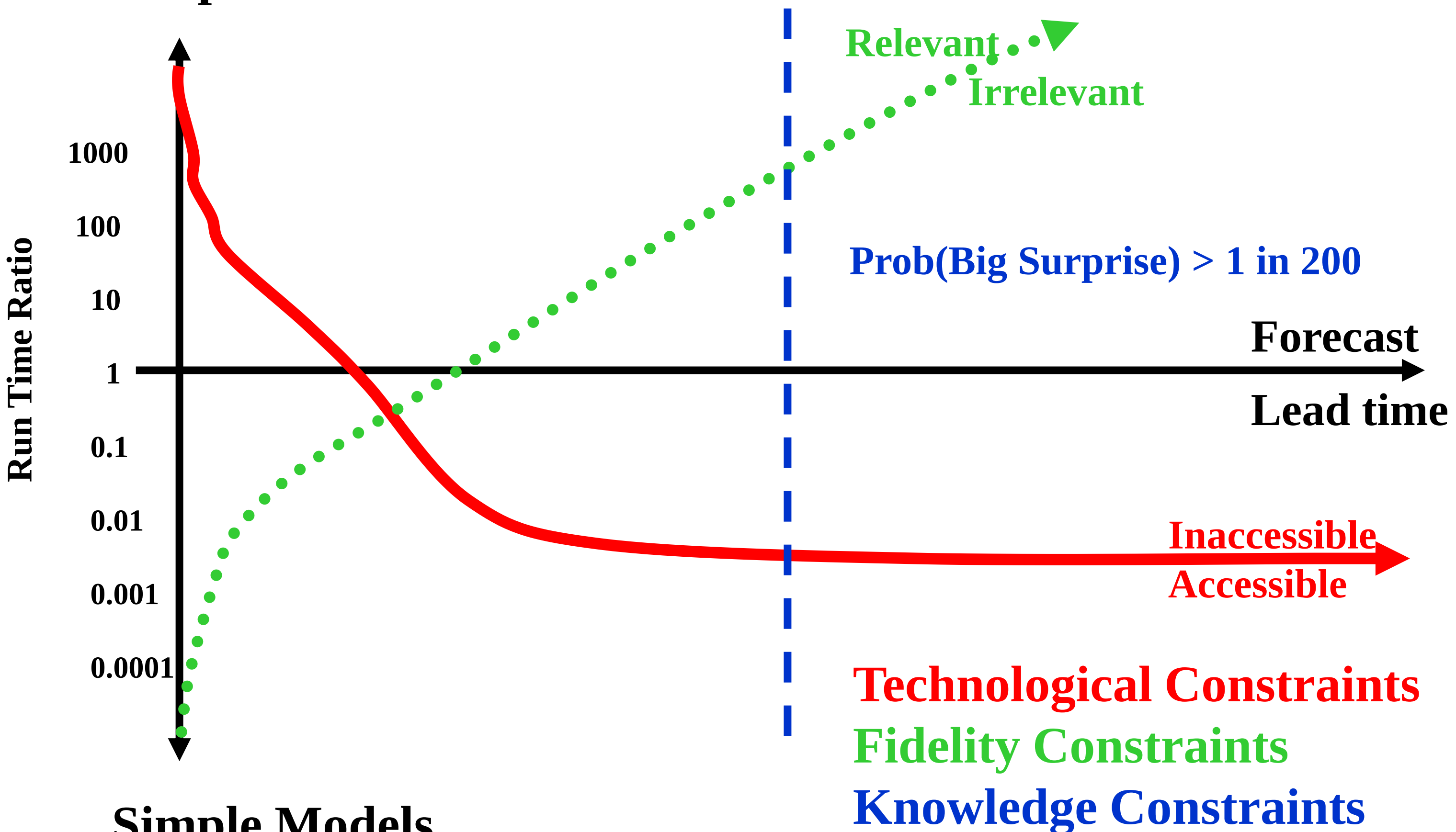


Leonard Smith

What are you constrained by?

Limits of current scientific/economic/mathematical knowledge mean the model may *be expected to* prove inadequate. We will tolerate this as long as the Prob(Big Surprise) < 0.05 (Basel III/Solvency II)

Complex Models

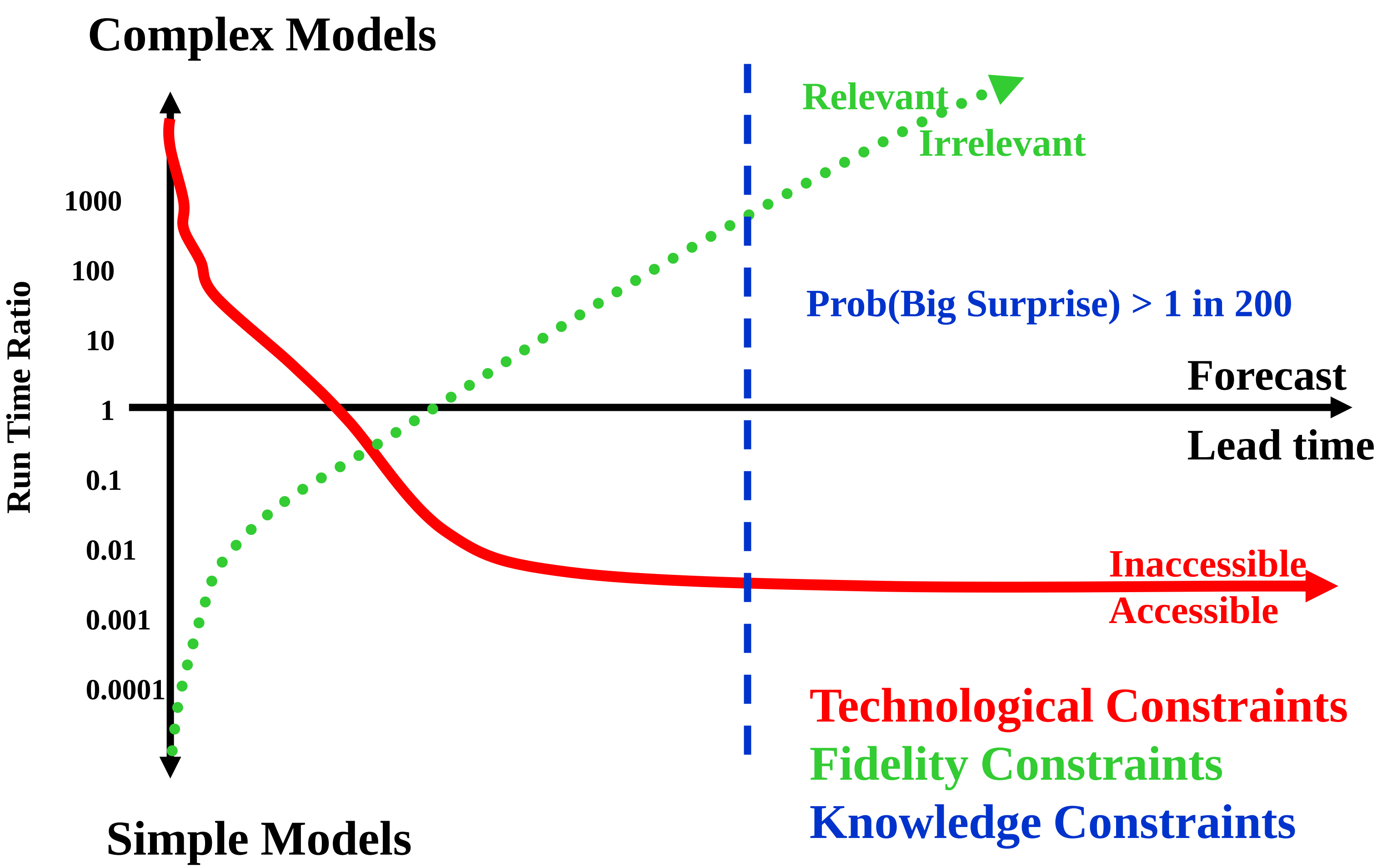


Simple Models

Leonard Smith

What are you constrained by?

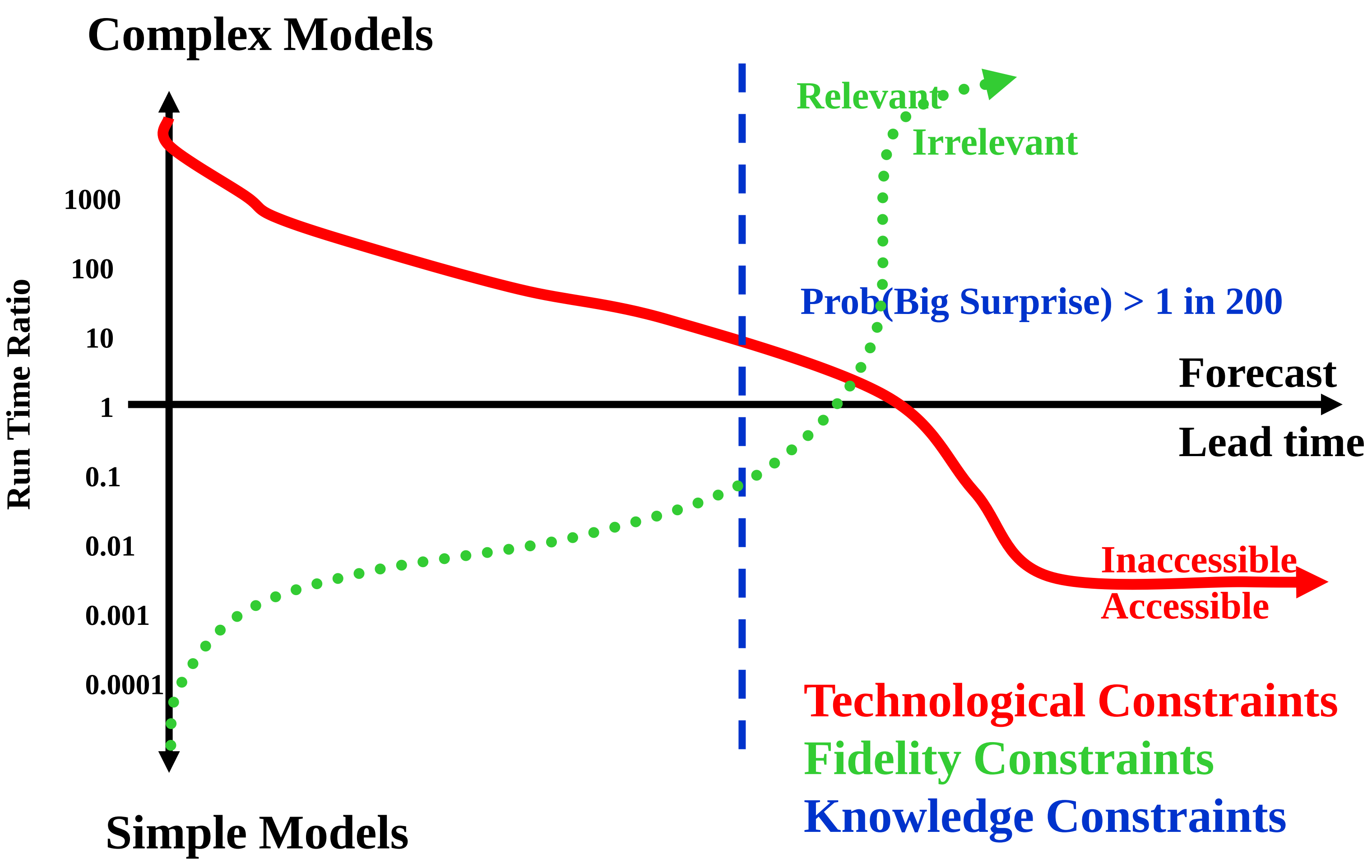
The decision you take will depend on how these three curves lie.



Leonard Smith

What are you constrained by?

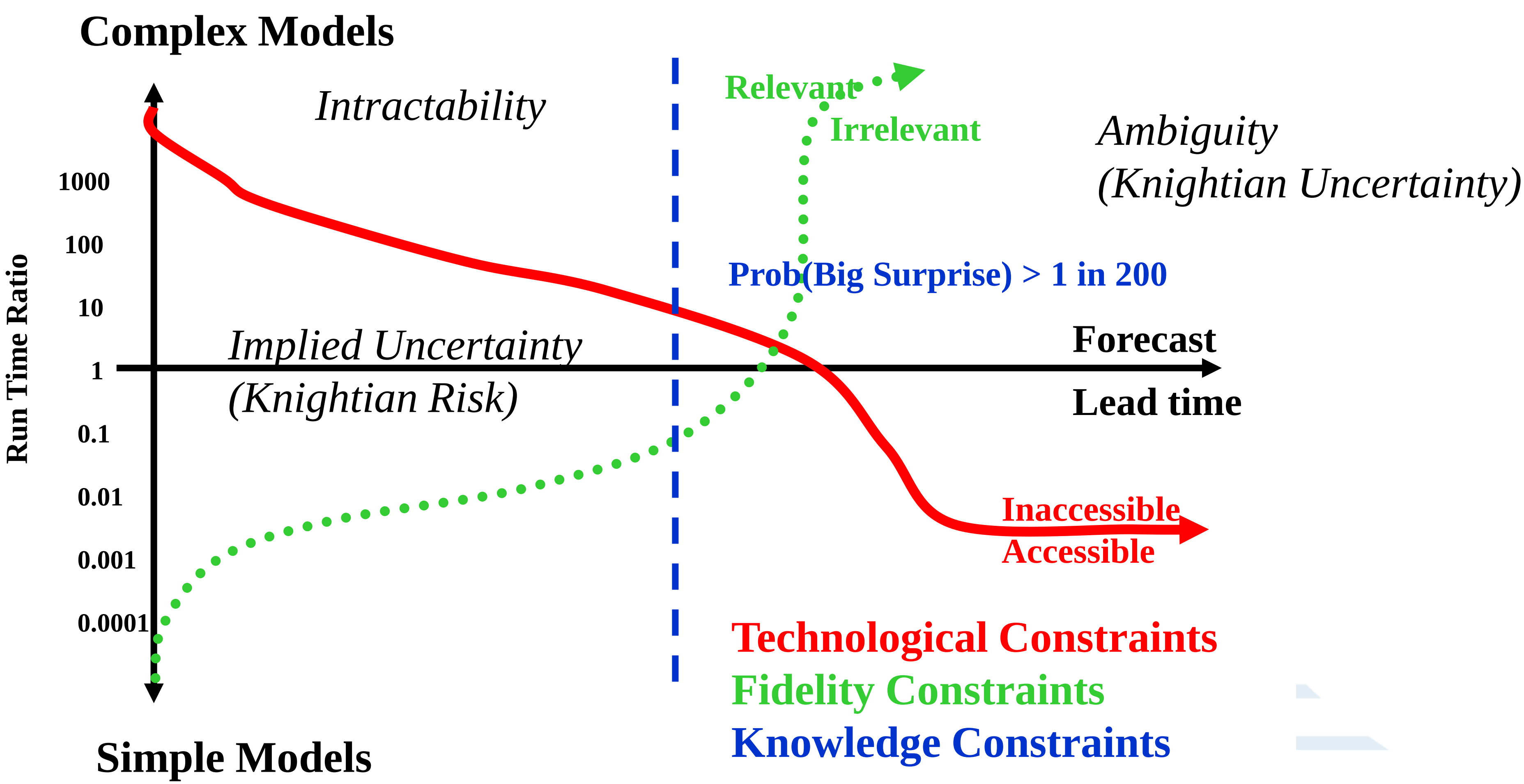
The decision you take will depend on how these three curves lie.



Leonard Smith

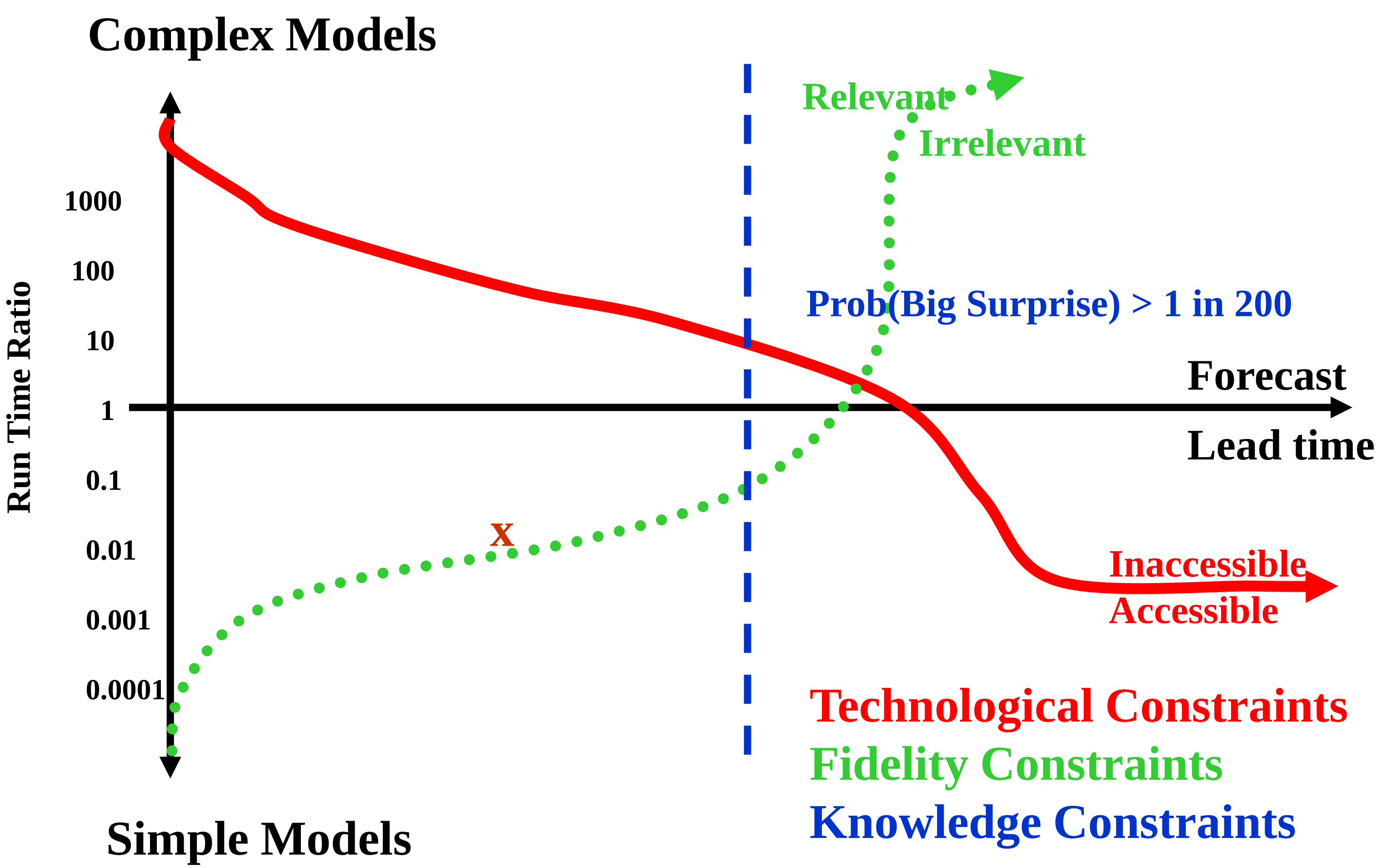
What are you constrained by?

What are the challenges we face with interpreting model simulations in different regions of this schematic?



What are you constrained by?

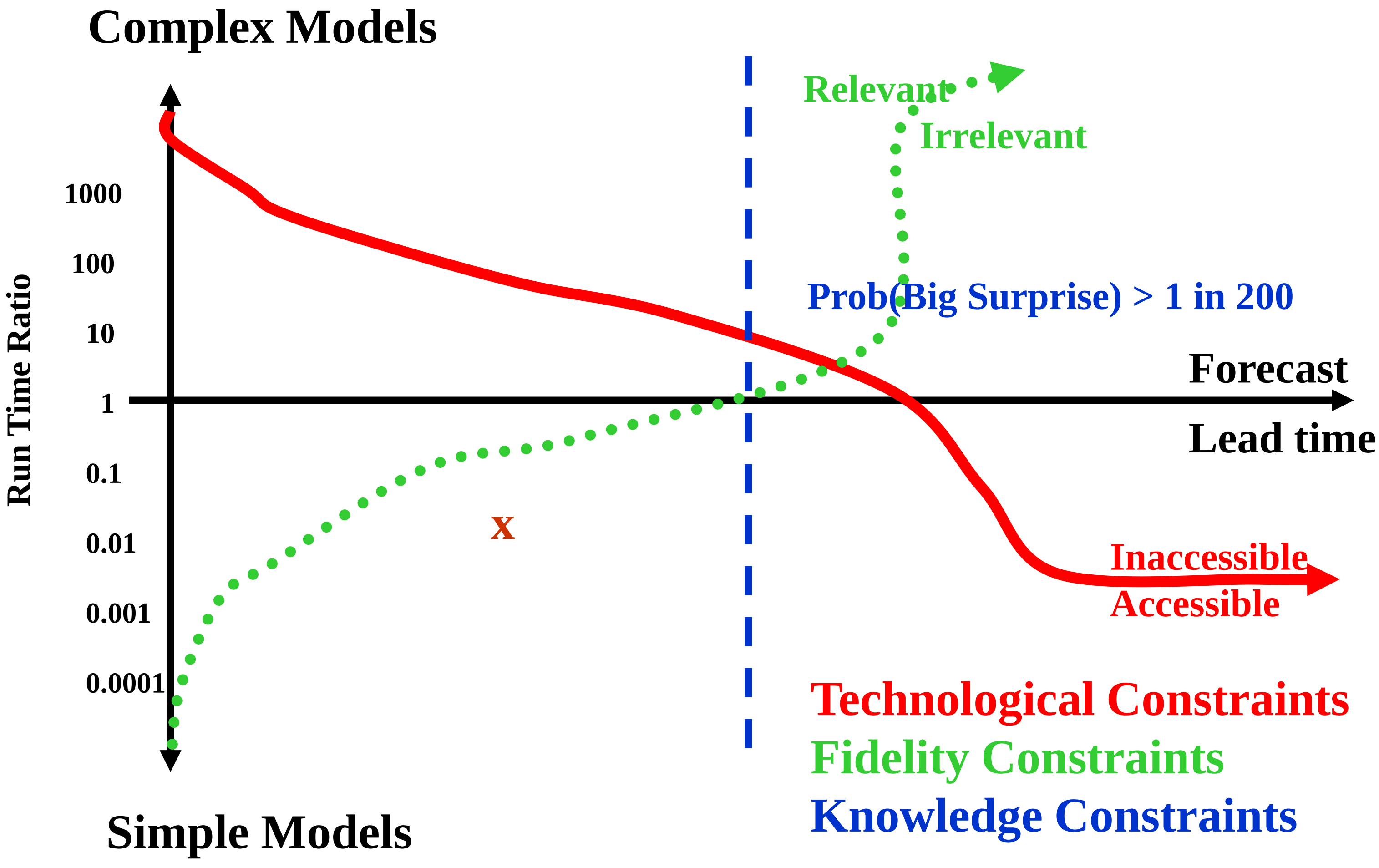
We need to be above the green line, below the red, and to the left of the blue.
So we could make **a relevant 100 day simulation** and have it a tomorrow.



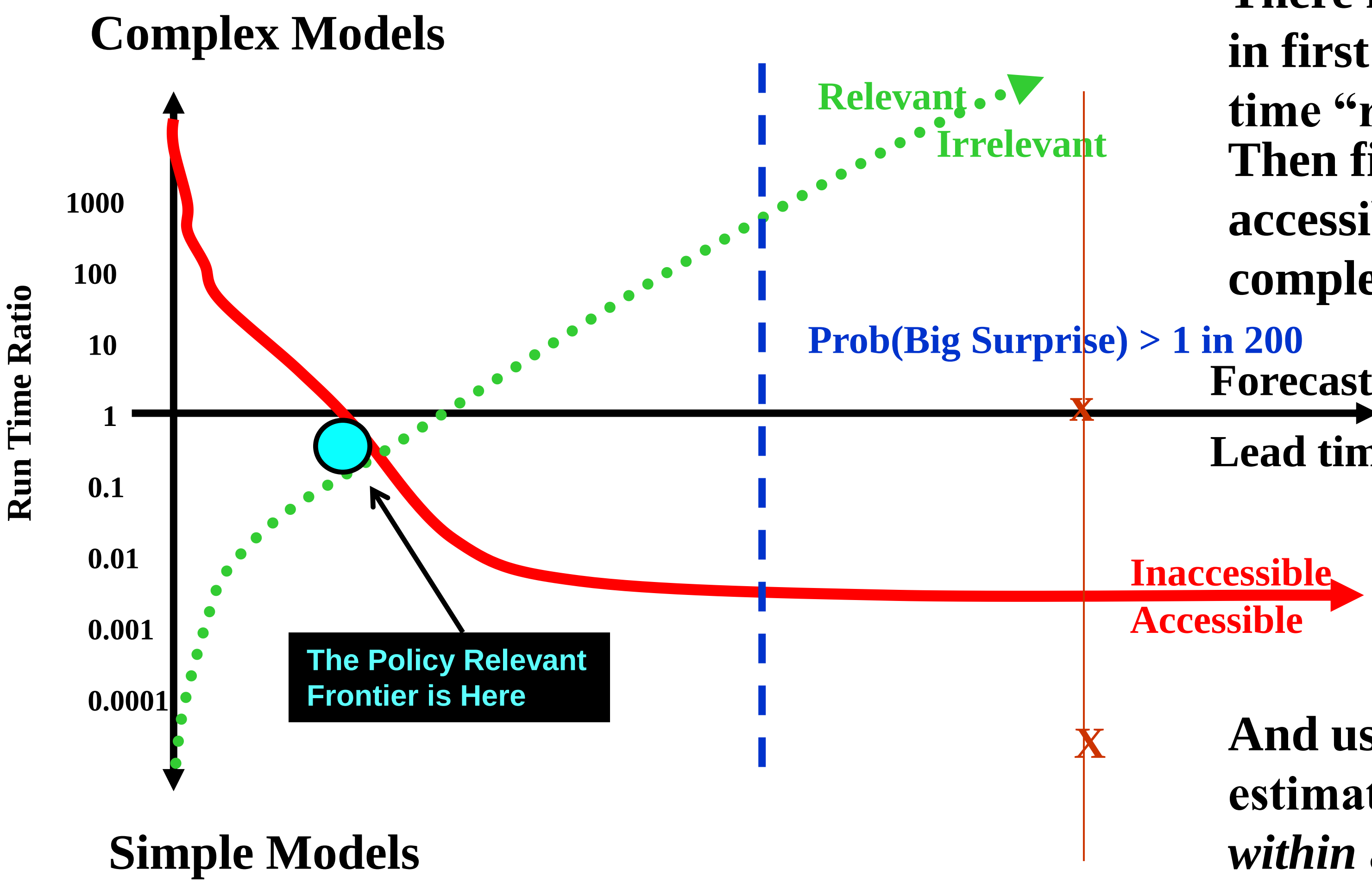
Leonard Smith

What are you constrained by?

But in this case, this "100 day" model is out of our reach.
Of course we can build it anyway, call it "best available" knowing it is both best and irrelevant; and pass it on (saying clearly that Prob(B.S.)~1)



Decision Support Model (Designed to deliver)



There is some danger in first picking the lead time "required."
Then finding an accessible level of complexity

And using ensembles to estimate "uncertainty" *within* an irrelevant model.

Technological Constraints
Fidelity Constraints
Knowledge Constraints

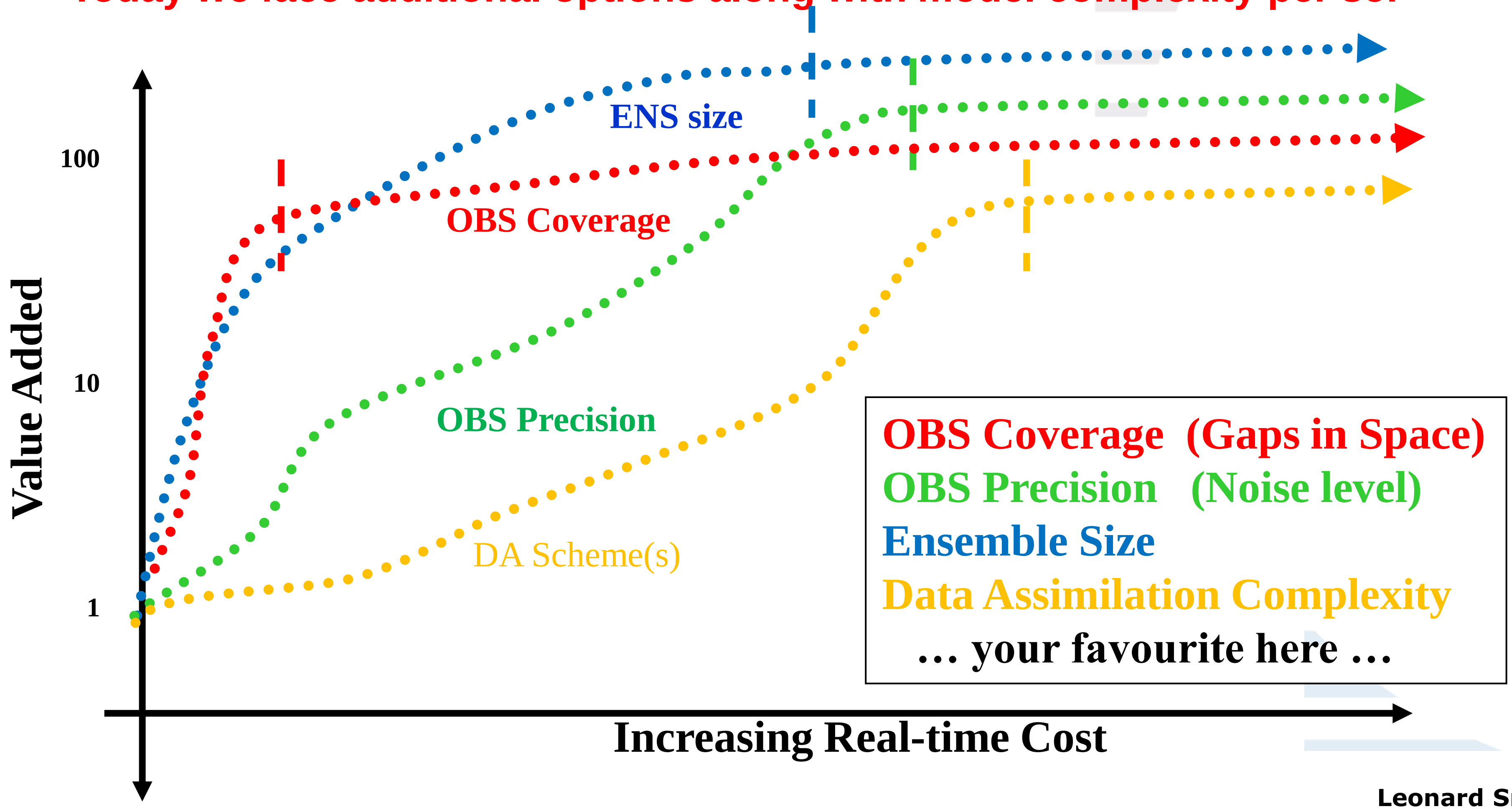
A Resource Allocation Methodology for Forecasting

Thompson (1957) investigated the improvement of US weather forecasting as a resource allocation problem.

How should a given investment be spread between:

(a) better obs, (b) better theory, (c) faster computers?

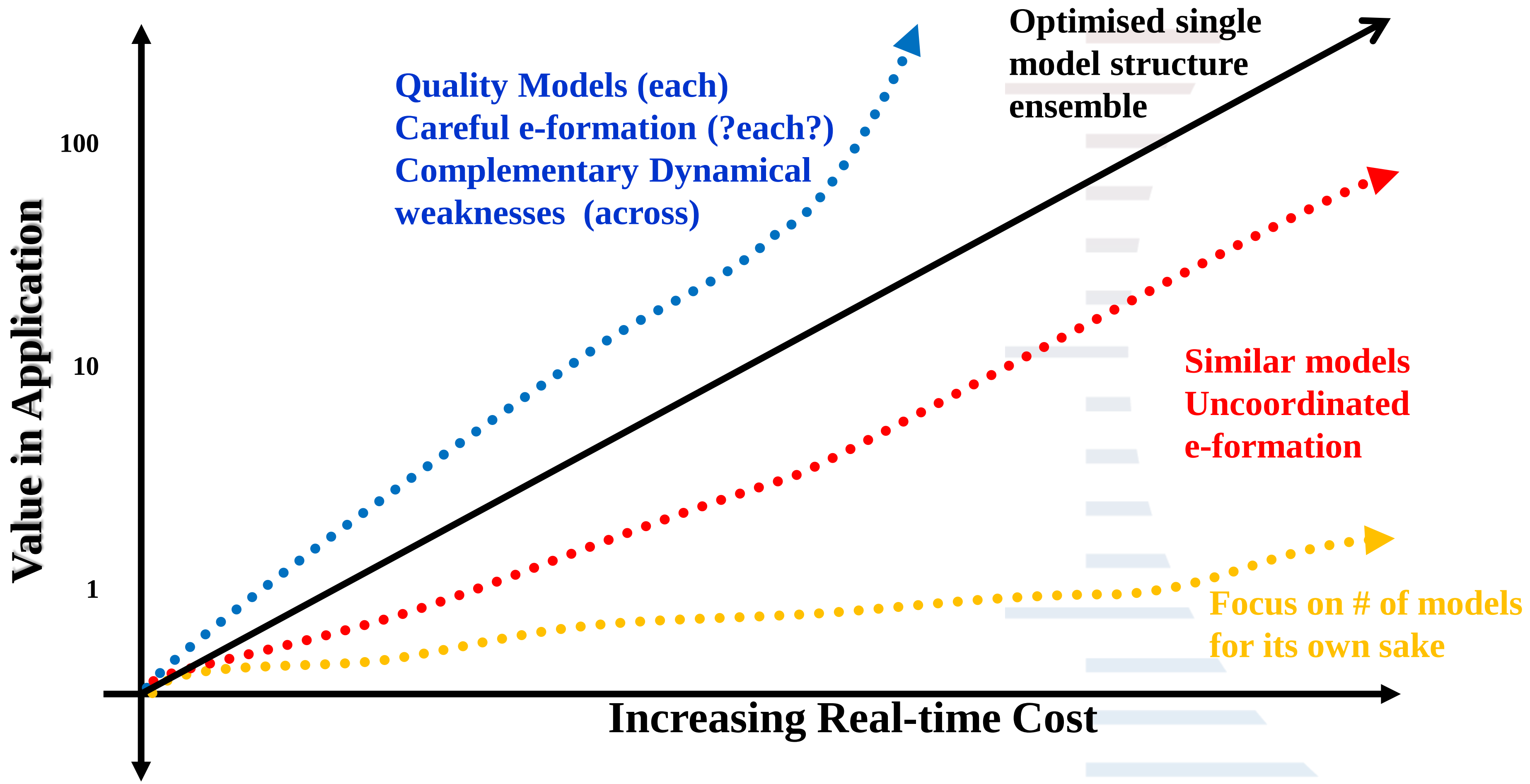
Today we face additional options along with model complexity per se:



Leonard Smith

What about "the" case for Multi-models?

The answer depends on the strategy



How might structural independence be enhanced? (in space stations?)

Could there ever be a general result?

Climate in Practice: In-sample examples.

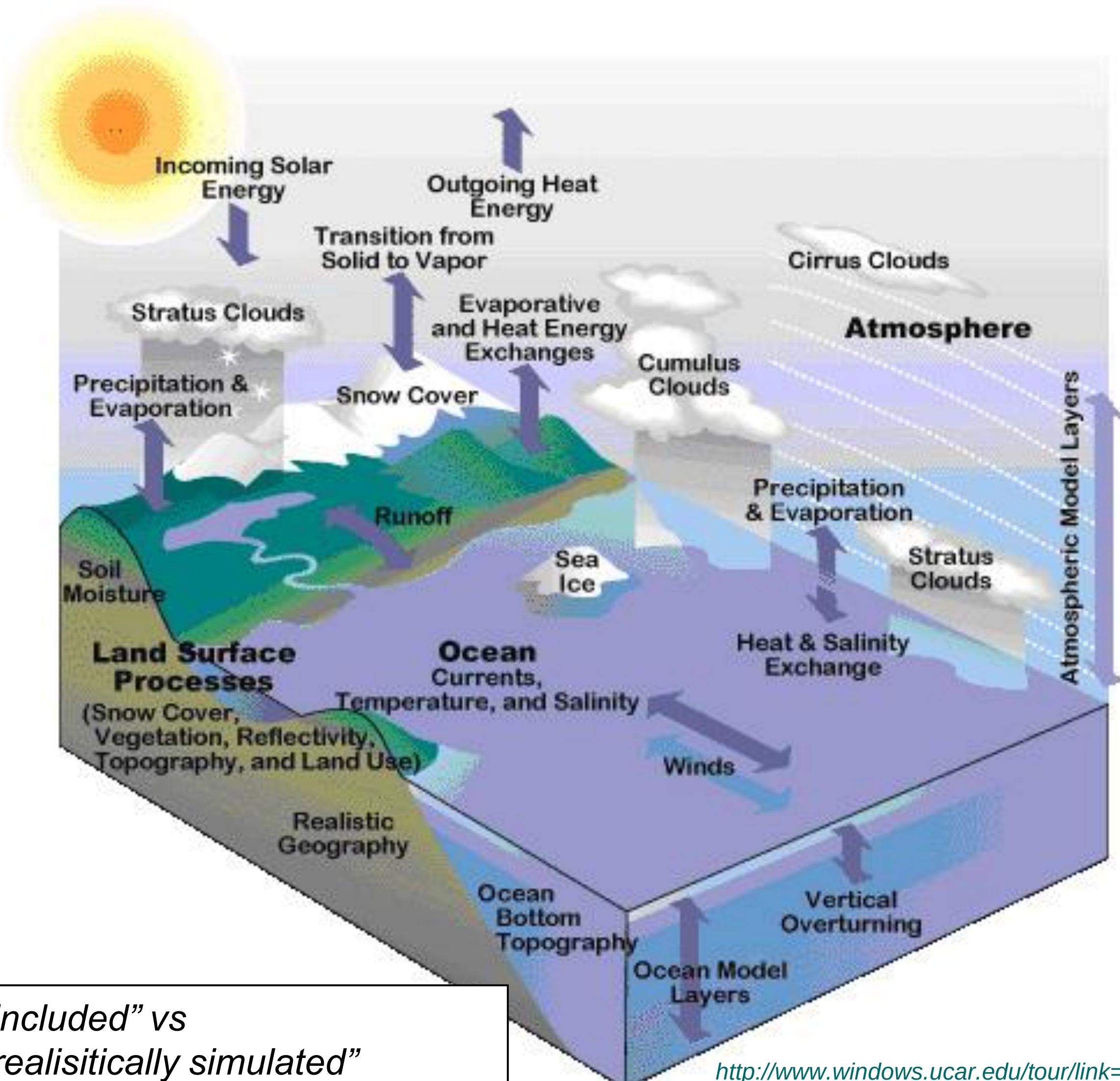
Climate models are based on well-established physical principles and have been demonstrated to reproduce observed features of recent climate (see Chapters 8 and 9) and past climate changes (see Chapter 6). There is considerable confidence that Atmosphere-Ocean General Circulation Models (AOGCMs) provide credible quantitative estimates of future climate change, particularly at continental and larger scales. Confidence in these estimates is higher for some climate variables (e.g., temperature) than for others (e.g., precipitation). This summary highlights areas of progress since the TAR:

Page 591

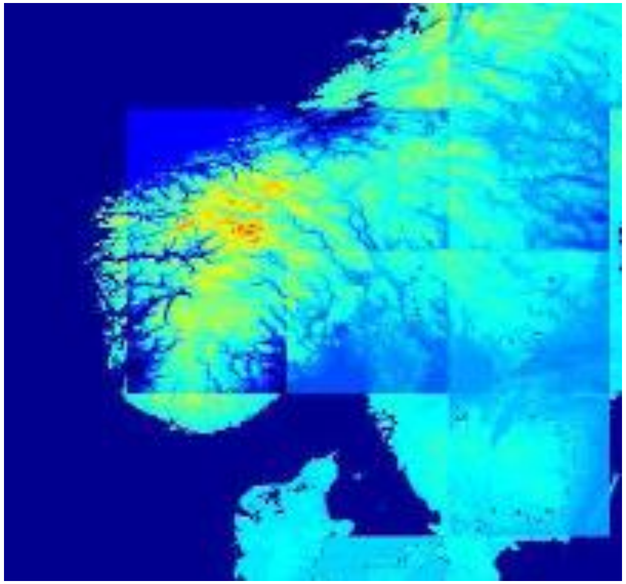
A report of Working Group I of the Intergovernmental Panel on Climate Change

A dangerously schematic schematic

Climate Model Schematic



Climate Model (the squares)



HadCM3 missing elevation

2min x 2min obs – HadCM3

“included” vs “realisitically simulated”

http://www.windows.ucar.edu/tour/link=/earth/climate/cli_models3.html

Towards Comprehensive Earth System Models
Past present and future

1975	1985	1992	1997	Present	
Atmosphere	Atmosphere	Atmosphere	Atmosphere	Atmosphere	Atmosphere
	Land surface	Land surface	Land surface	Land surface	Land surface
		Ocean & sea-ice	Ocean & sea-ice	Ocean & sea-ice	Ocean & sea-ice
			Sulphate aerosol	Sulphate aerosol	Sulphate aerosol
				Non-sulphate aerosol	Non-sulphate aerosol
				Carbon cycle	Carbon cycle
					Atmospheric chemistry

Stamp Collecting?
Must get one first…
then find a better one.

In 1975 each was known to be potentially important.

Off-line model development

Ocean & sea-ice model

Sulphur cycle model

Non-sulphate aerosols

Land carbon cycle model

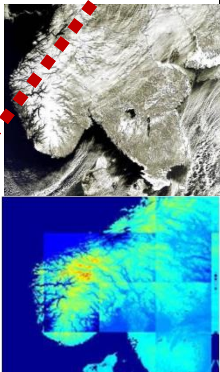
Ocean carbon cycle model

Carbon cycle model

Atmospheric chemistry

Atmospheric chemistry

Strengthening colours denote improvements in models



What could be released if we have a maximum 40 year lead time due to model fidelity? 20 year?

NCAR

Models on Stage: Model Depth and Equidismality

Model depth (relevance of being "based on the laws of physics" to skill) can be explored through comparison with Simple Surrogate models.

Smith, L.A. (1992) Identification and prediction of low dimensional dynamics Physica D 58 (1-4): 50-76.

That is **not** to say a model more in accord with the laws of physics will not ultimately provide the ideal forecast, merely that the models **in hand today** can lay no a priori claim on doing so!

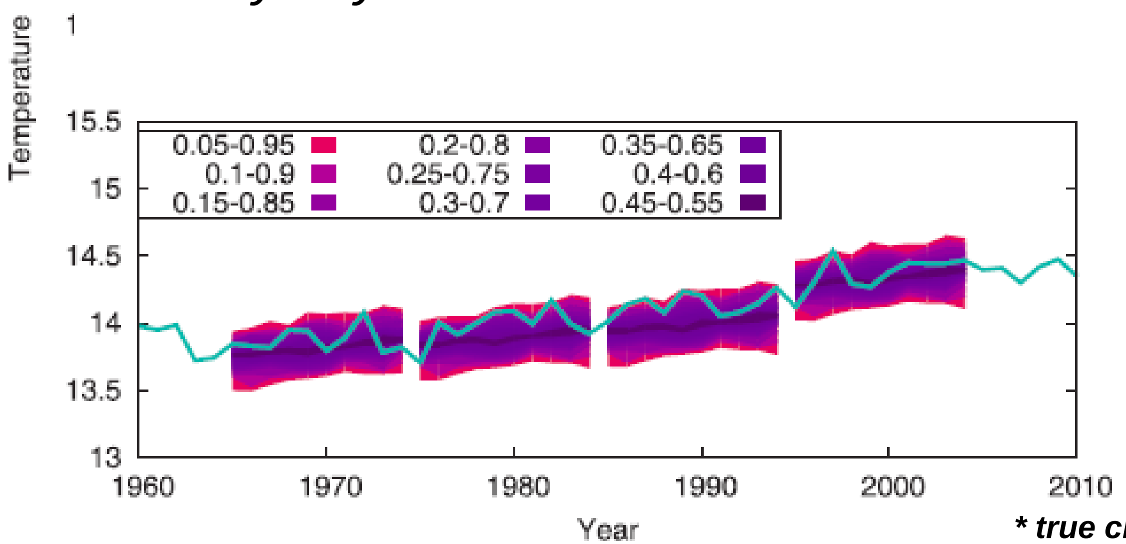
Surrogate models aide in distinguishing equifinality from equidismality.

Models for science based policy should always be benchmarked against simple model never merely against each other!

Climate Forecasting by Hand

- *Take the last hundred twenty nine years of global mean temperature, T_i.*
- *Find the 128 one-year changes by subtracting T_(i-1) from T_i*
- *Add each of these 128 numbers to this years global mean temperature to get a 128 member "ensemble forecast" for next year.*
- *Draw a distribution* over these 128 points and call it a probability forecast for next year.*
- *Repeat for 2 year (direct) forecasts, then 3 years, 4, 5, … 10.*

You can do this in a day. By hand:



**** true cross-validation throughout***

FIG. 5. As in Fig. 4, but for every fifth launch from the DC model.

JOBNAME: JCLI 00#0 2013 PAGE: 1 SESS: 8 OUTPUT: Wed Jul 31 14:17:29 2013 Total No. of Pages: 14
/ams/jcli/0/jcliD1200485

An Evaluation of Decadal Probability Forecasts from State-of-the-Art Climate Models*

EMMA B. SUCKLING AND LEONARD A. SMITH

Centre for the Analysis of Time Series, London School of Economics, London, United Kingdom

(Manuscript received 26 July 2012, in final form 12 May 2013)

ABSTRACT

AU1

While state-of-the-art models of Earth's climate system have improved tremendously over the last 20 yr, nontrivial structural flaws still hinder their ability to forecast the decadal dynamics of the Earth system realistically. Contrasting the skill of these models not only with each other but also with empirical models can reveal the space and time scales on which simulation models exploit their physical basis effectively and quantify their ability to add information to operational forecasts. The skill of decadal probabilistic hindcasts for annual global-mean and regional-mean temperatures from the EU ENSEMBLES project is contrasted with several empirical models. Both the ENSEMBLES models and a "dynamic climatology" empirical model show probabilistic skill above that of a static climatology for global-mean temperature. The dynamic climatology model, however, often outperforms the ENSEMBLES models. The fact that empirical models display skill similar to

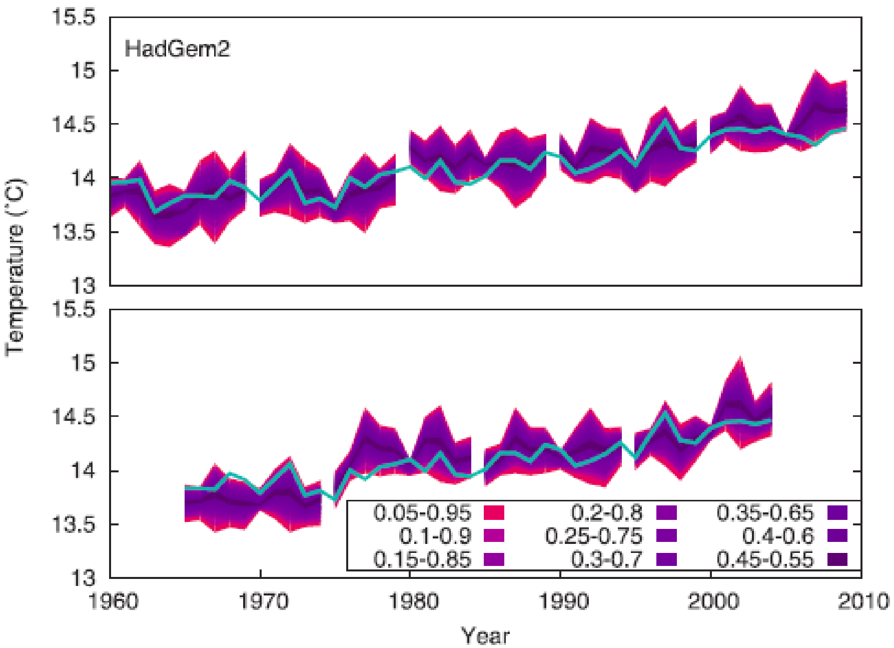


FIG. 4. Forecast distributions for HadGEM2 (UKMO) for the 5th–95th percentile. The HadCRUT3 observed temperatures are shown in blue. The forecasts are 10 yr long and launched every 5 yr, and so the fan charts would overlap; to avoid this they are presented in two panels: forecasts launched in 10-yr intervals from (top) 1960 and (bottom) 1965.

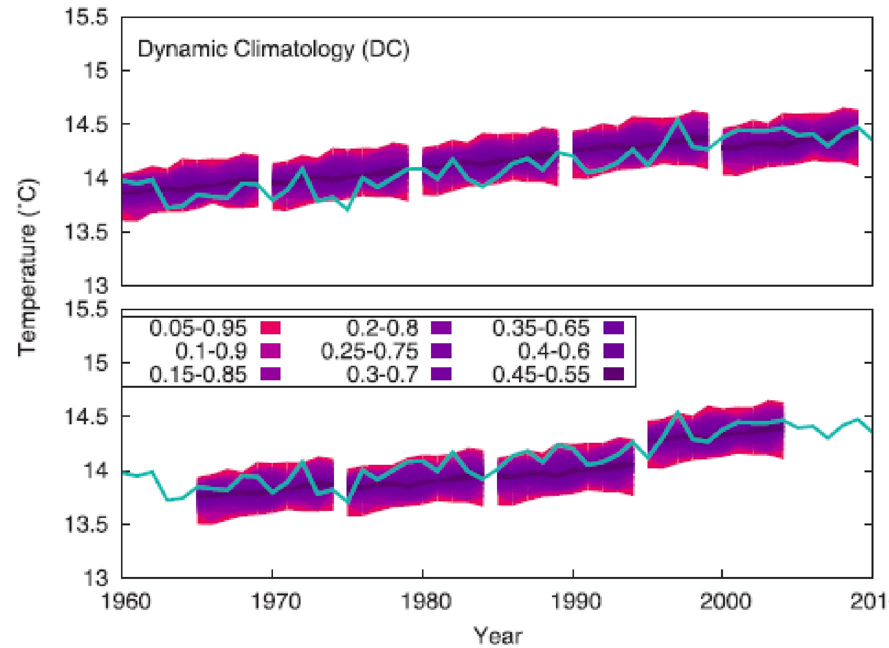


FIG. 5. As in Fig. 4, but for every fifth launch from the DC model.

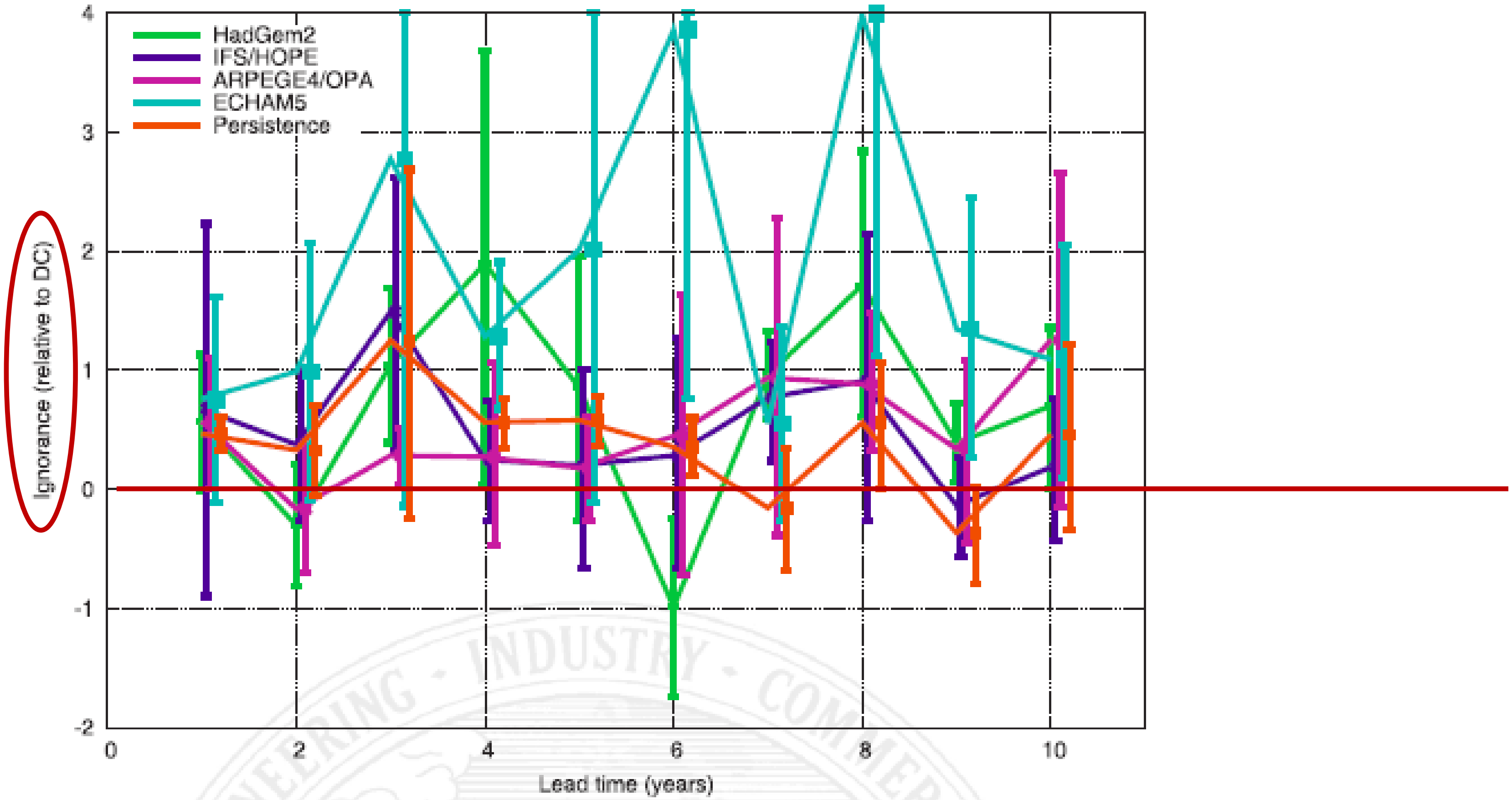


FIG. 9. Ignorance of the ENSEMBLES models relative to DC as a function of lead time. The bootstrap resampling intervals are illustrated at the 10th–90th percent level. Note that the simulation models tend to have positive scores (less skill) than the DC model at every lead time.

Is this just one over-tuned empirical model?

On decadal time scales, rather simple empirical models produce probability forecasts which systematically outperform state-of-the-art GCMs.

Of course GCMs provide insight into **mechanisms and phenomena** otherwise inaccessible.

Overselling the fidelity of "physics-based" models puts the credibility of all science-based policy making at risk. Needlessly.

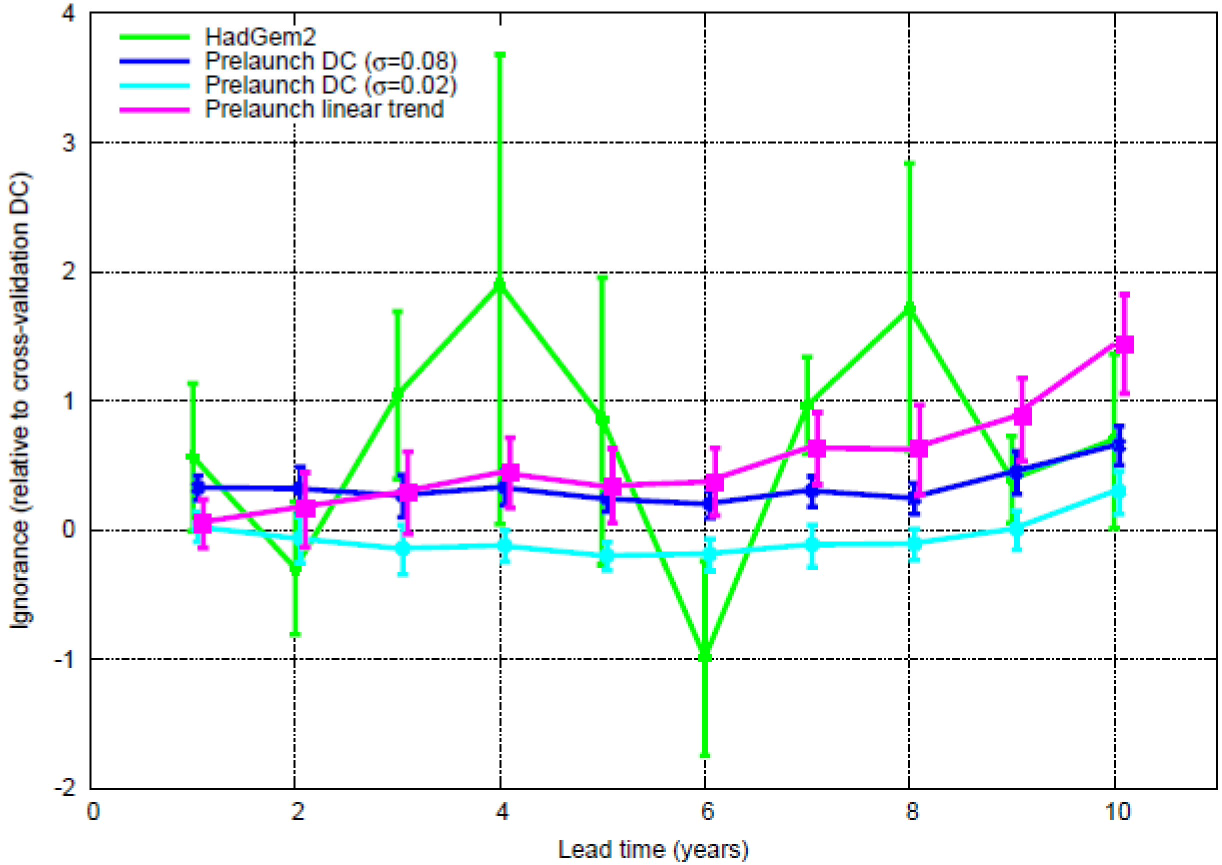


FIG. 10. Ignorance of the Prelaunch DC and Prelaunch Trend models relative to the standard DC model as a function of lead time. The HadGem2 model from ENSEMBLES is also shown. It is shown that the Prelaunch DC model is not significantly less skillful than the standard DC model and is robust to variations in parameter tuning. The Prelaunch linear trend model is, however, generally shown to be less skillful than the standard DC model. The bootstrap resampling intervals are illustrated at the 10-90th percent level.

Suckling and Smith (2013) J Climate, in press

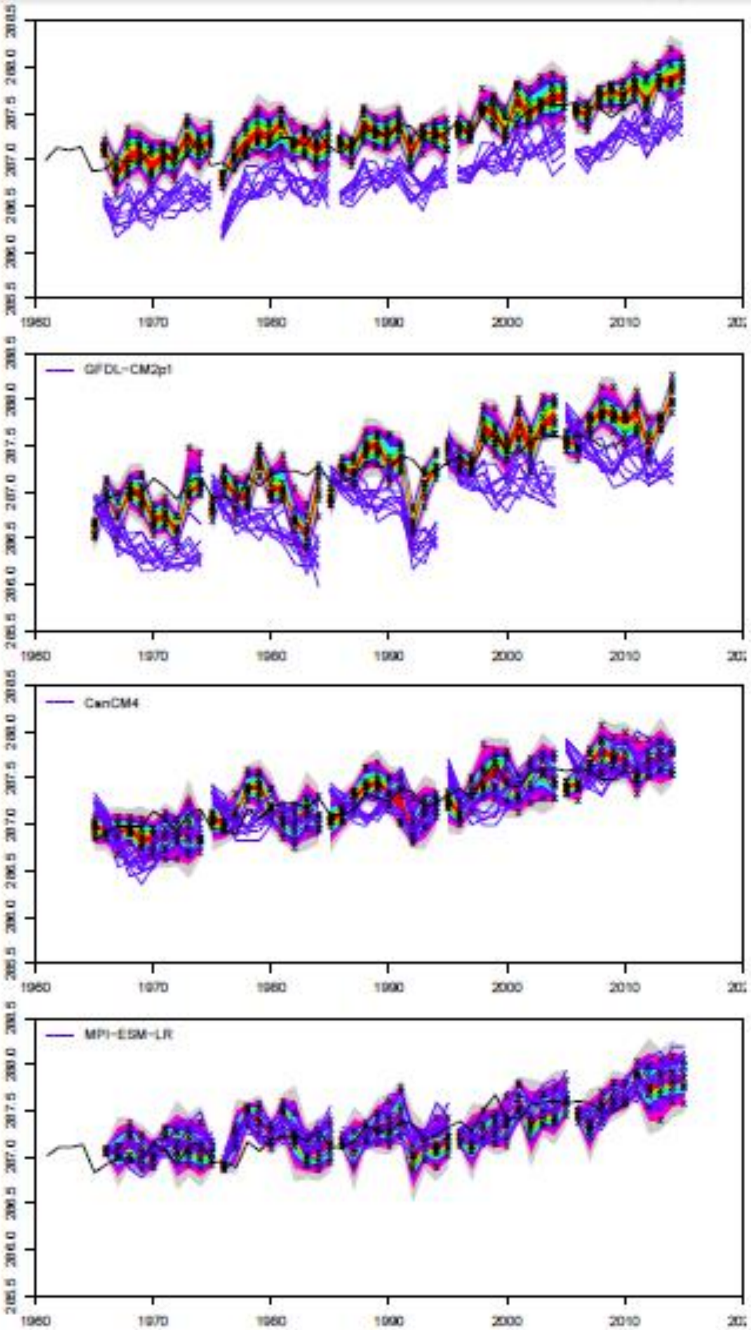


Figure 2: Fan charts showing model forecasts using fitted kernel width and offsets. Original ensembles also shown before offset (blue lines) and after offset (black crosses).

CMIP5 All-Stars

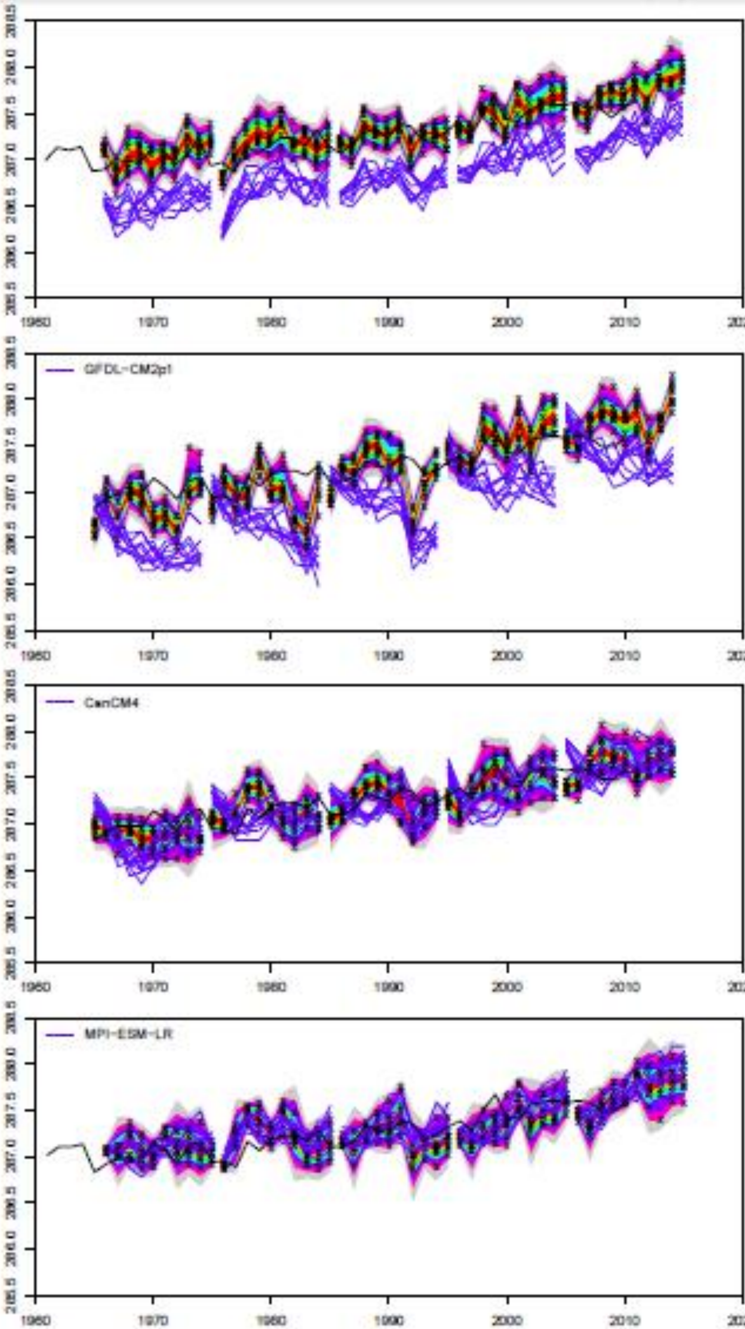


Figure 2: Fan charts showing model forecasts using fitted kernel width and offsets. Original ensembles also shown before offset (blue lines) and after offset (black crosses).

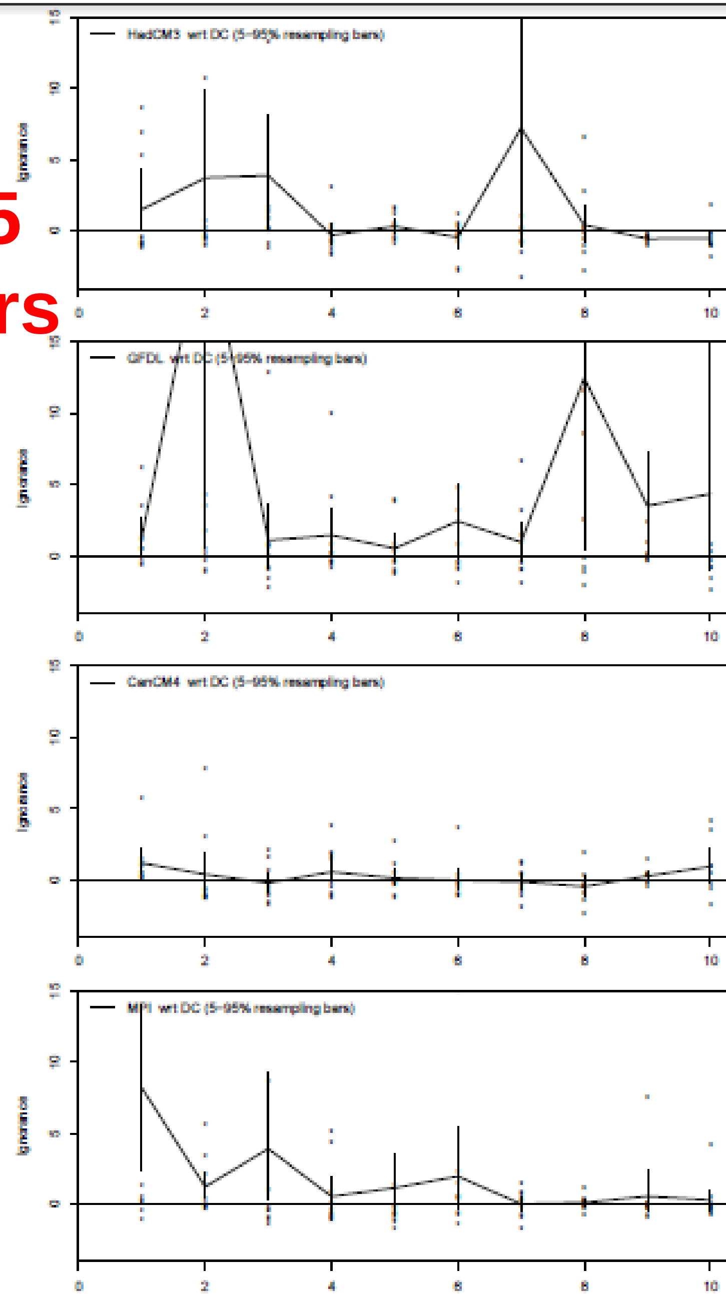


Figure 8: Relative ignorance of model wrt DC, as a function of lead time (years), showing individual data points and 5-95% bootstrap ranges of the mean.

Take home message

On the ten year forecast time scale, empirical models more than hold their own in head-to-head forecast evaluation with simulation models.

On longer time scales, physics-based simulation models will be required to obtain hi-fidelity forecasts.

But that fact is not evidence that today's physics-based simulation models can provide hi-fidelity forecasts on either decadal or century time scales.

Users might embrace this for better climate services.

But why weren't these simple straw men models in the AR3?

(None of this challenges the basic physics of global warming, of course.)

This looks a lot like:

From an Oxford Bus Shelter:

X30 N30 predictions are wrong
sorry for any inconvenience

The traditional approach

Smith, L.A. (2002) What might we learn from climate forecasts? Proc. National Acad. Sci. USA 4 (99): 2487-2492.

The traditional approach to climate modeling is to build the most complicated model that will fit inside the largest computer available, run it once, and see what happens. This approach yiclds a singlc "bcst-gucss" forccast. Yct cvcn in high school physics, we learn that an answer without "error bars" is no answer at all.

So what do we find in the AR5?

The Danger of Rank Order Beauty Contests

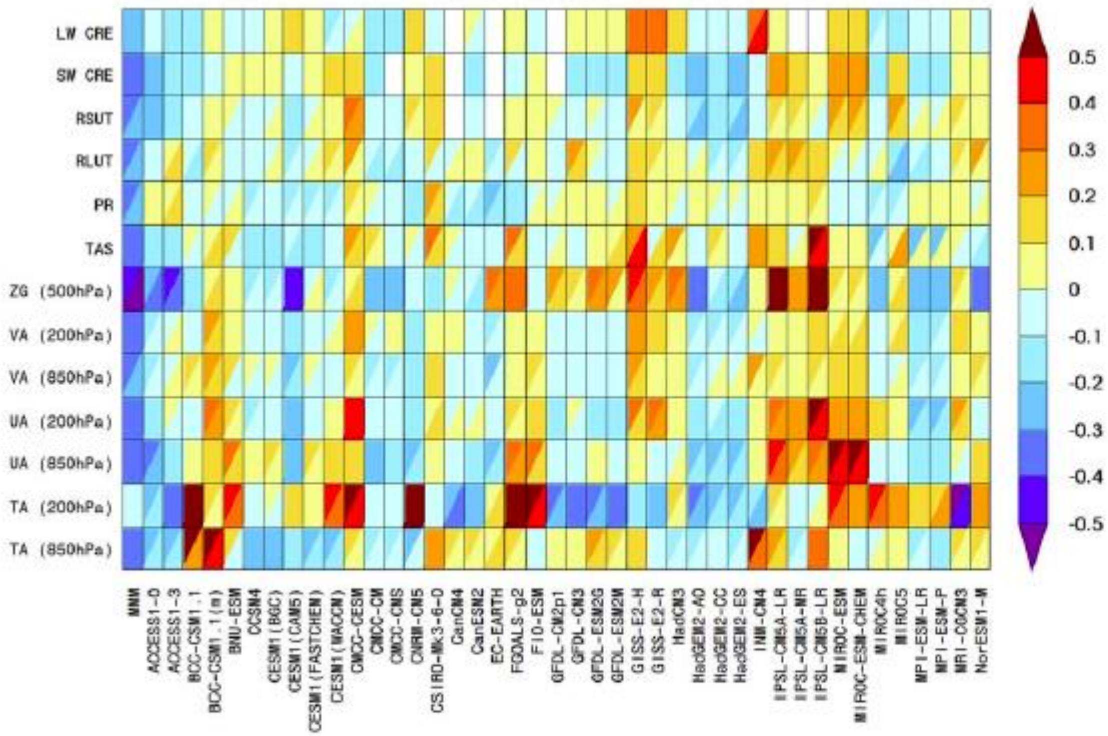
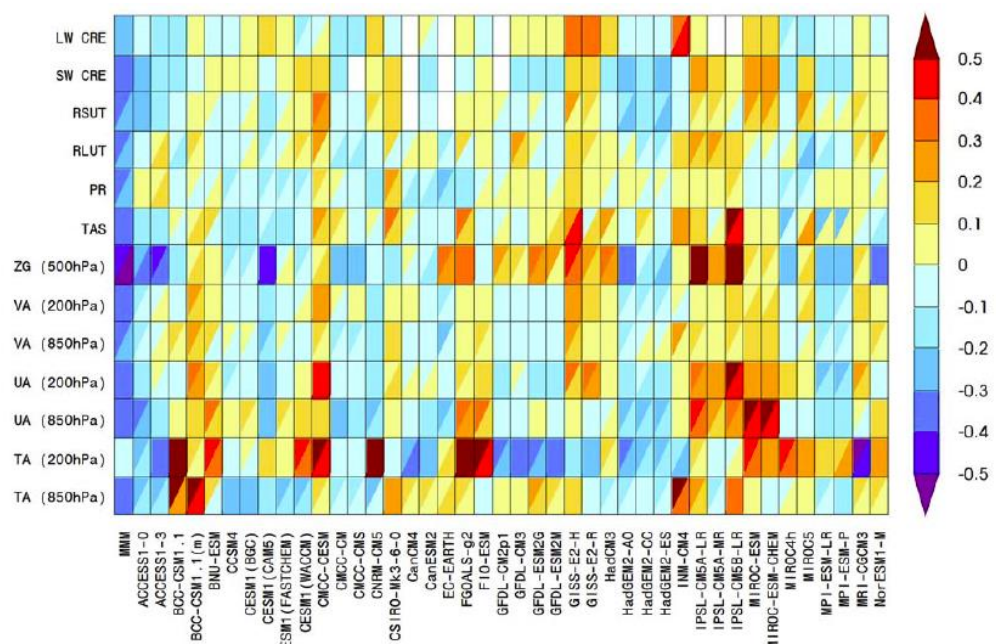


Figure 9.7: Relative error measures of CMIP5 model performance, based on the global seasonal-cycle climatology (1980–2005) computed from the historical experiments. Rows and columns represent individual variables and models, respectively. The error measure is a space–time root-mean-square error (RMSE), which, treating each variable separately, is portrayed as a relative error by normalizing the result by the median error of all model results (P. Gleckler, Taylor, & Doutriaux, 2008). For example, a value of 0.20 indicates that a model's RMSE is 20% larger than the median CMIP5 error for that variable, whereas a value of –0.20 means the error is 20% smaller than the median error. No colour (white) indicates that model results are currently unavailable. A diagonal split of a grid square shows

The Danger of Rank Order Beauty Contests

Figure 9.7: Relative error measures of CMIP5 model performance, based on the global seasonal-cycle climatology (1980–2005) computed from the historical experiments. Rows and columns represent individual variables and models, respectively. The error measure is a space–time root-mean-square error (RMSE), which, treating each variable separately, is portrayed as a relative error by normalizing the result by the median error of all model results (P. Gleckler, Taylor, & Doutriaux, 2008). For example, a value of 0.20 indicates that a model's RMSE is 20% larger than the median CMIP5 error for that variable, whereas a value of –0.20 means the error is 20% smaller than the median error. No colour (white) indicates that model results are currently unavailable. A diagonal split of a grid square shows the relative error with respect to both the default reference data set (upper left triangle) and the alternate (lower right triangle). The relative errors are calculated independently for the default and alternate data sets. All reference data used in the diagram are summarized in Table 9.3.

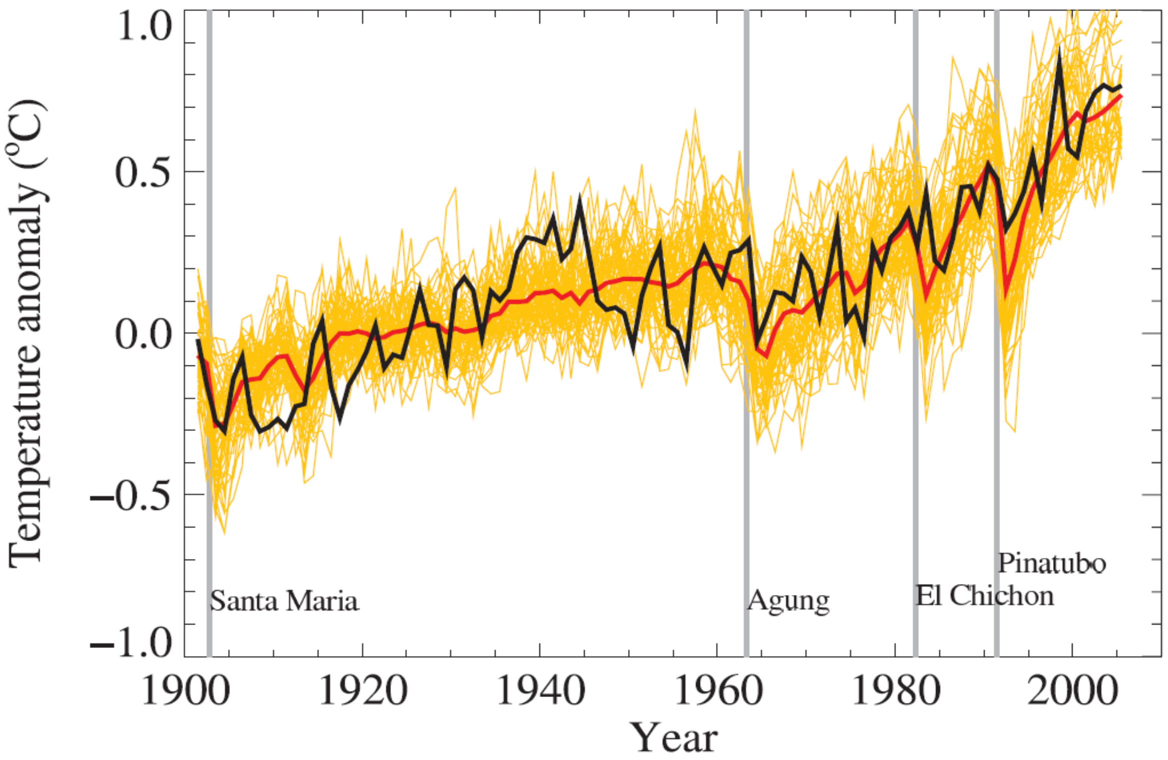
What does this plot on its own actually tell me?



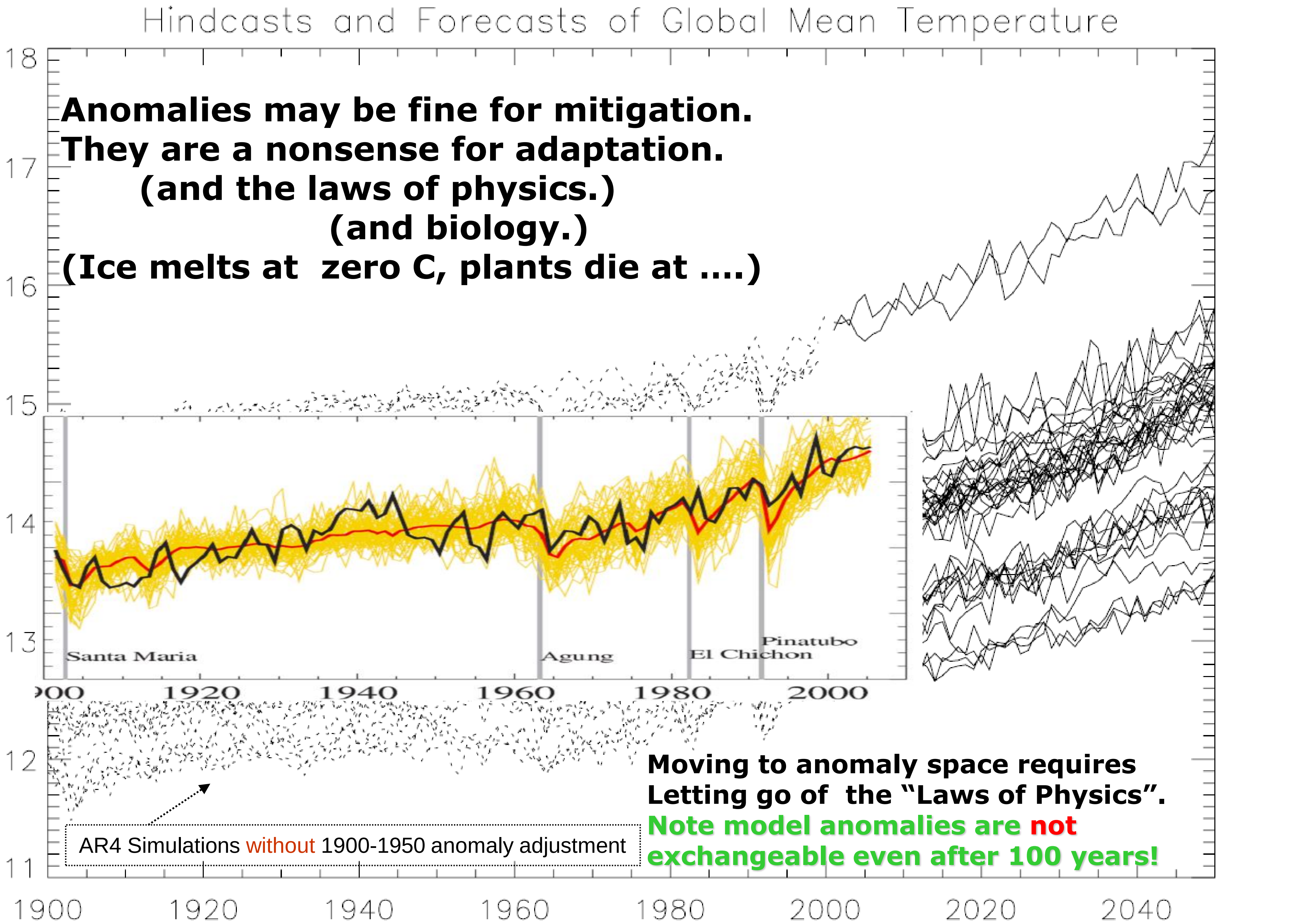
Climate in Practice: In-sample examples.

This graph tends to leave the impression models do rather well.

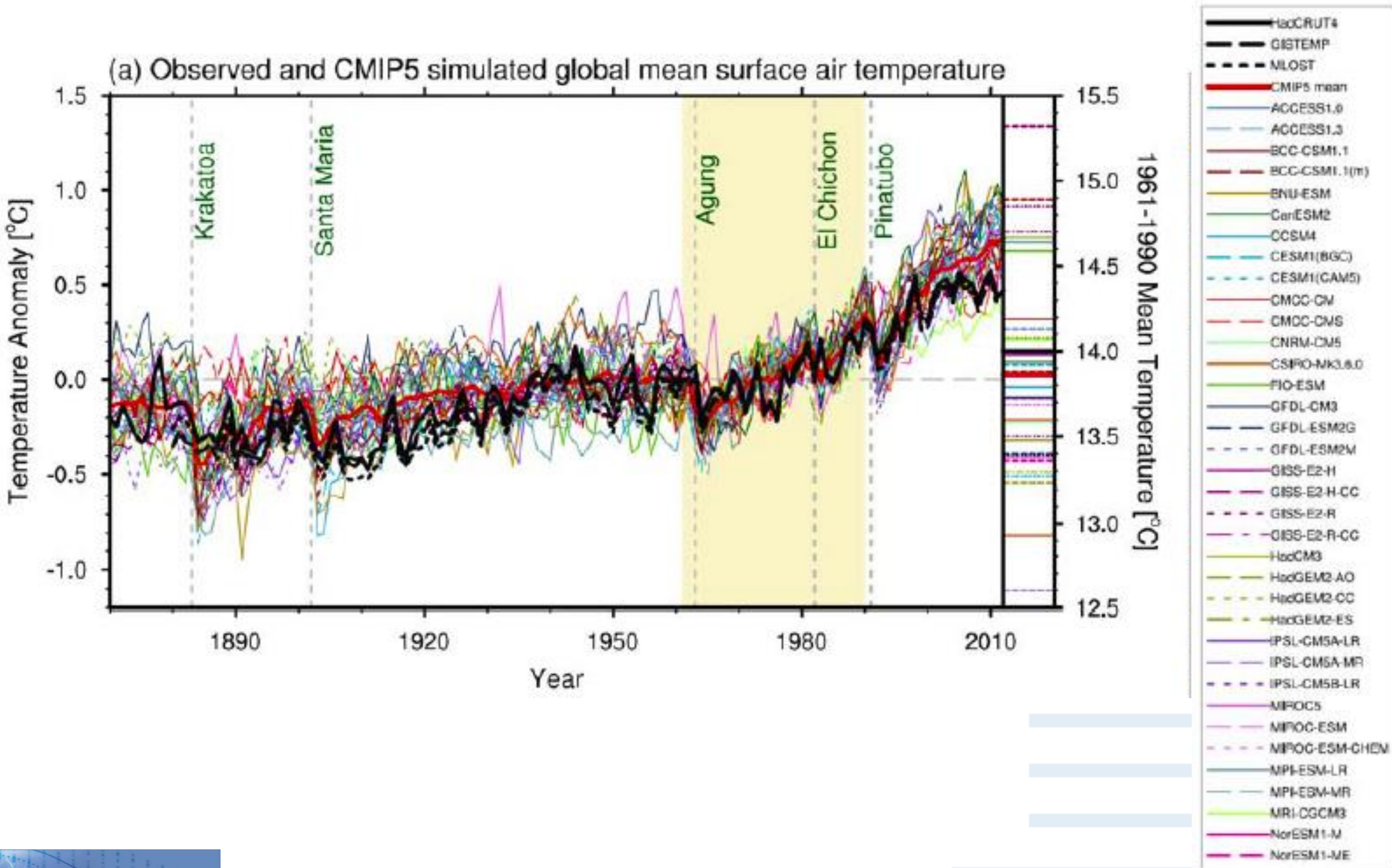
FAQ 8.1, Figure 1. *Global mean near-surface temperatures over the 20th century from observations (black) and as obtained from 58 simulations produced by 14 different climate models driven by both natural and human-caused factors that influence climate (yellow). The mean of all these runs is also shown (thick red line). Temperature anomalies are shown relative to the 1901 to 1950 mean. Vertical grey lines indicate the timing of major volcanic eruptions. (Figure adapted from Chapter 9, Figure 9.5. Refer to corresponding caption for further details.)*



While systematic errors are larger than the observed effect



Communication II Anomalies



Anomalies

Final Draft (7 June 2013) Chapter 9 IPCC WGI Fifth Assessment Report

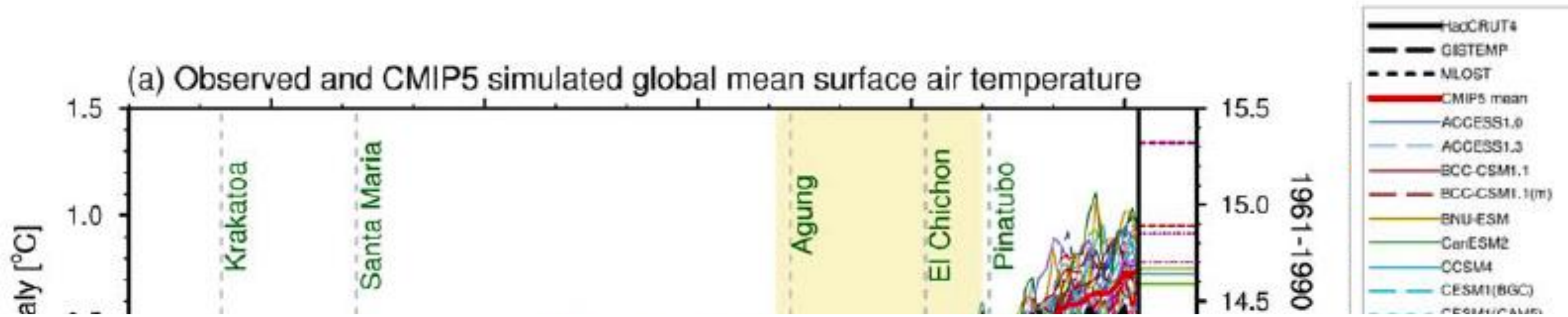
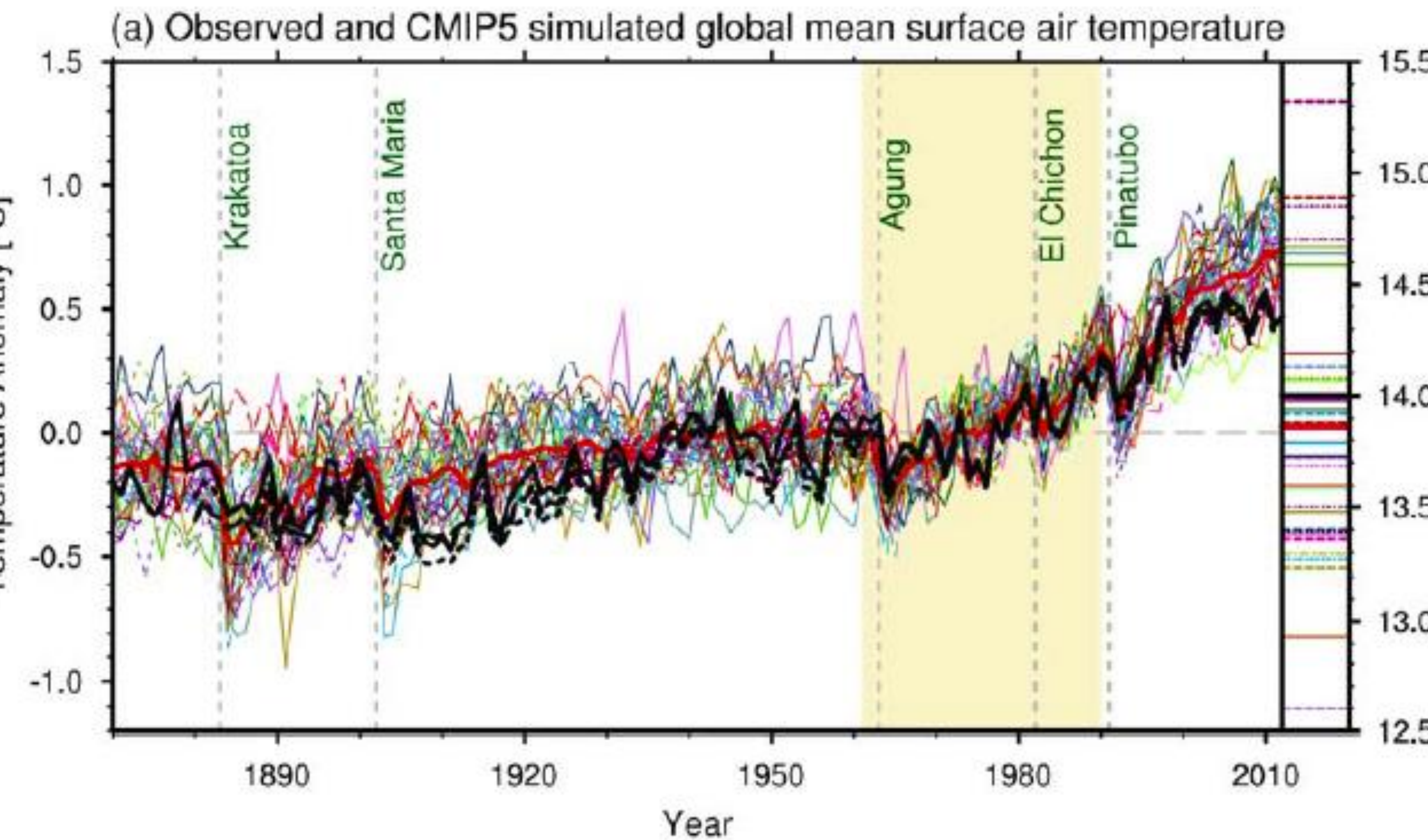
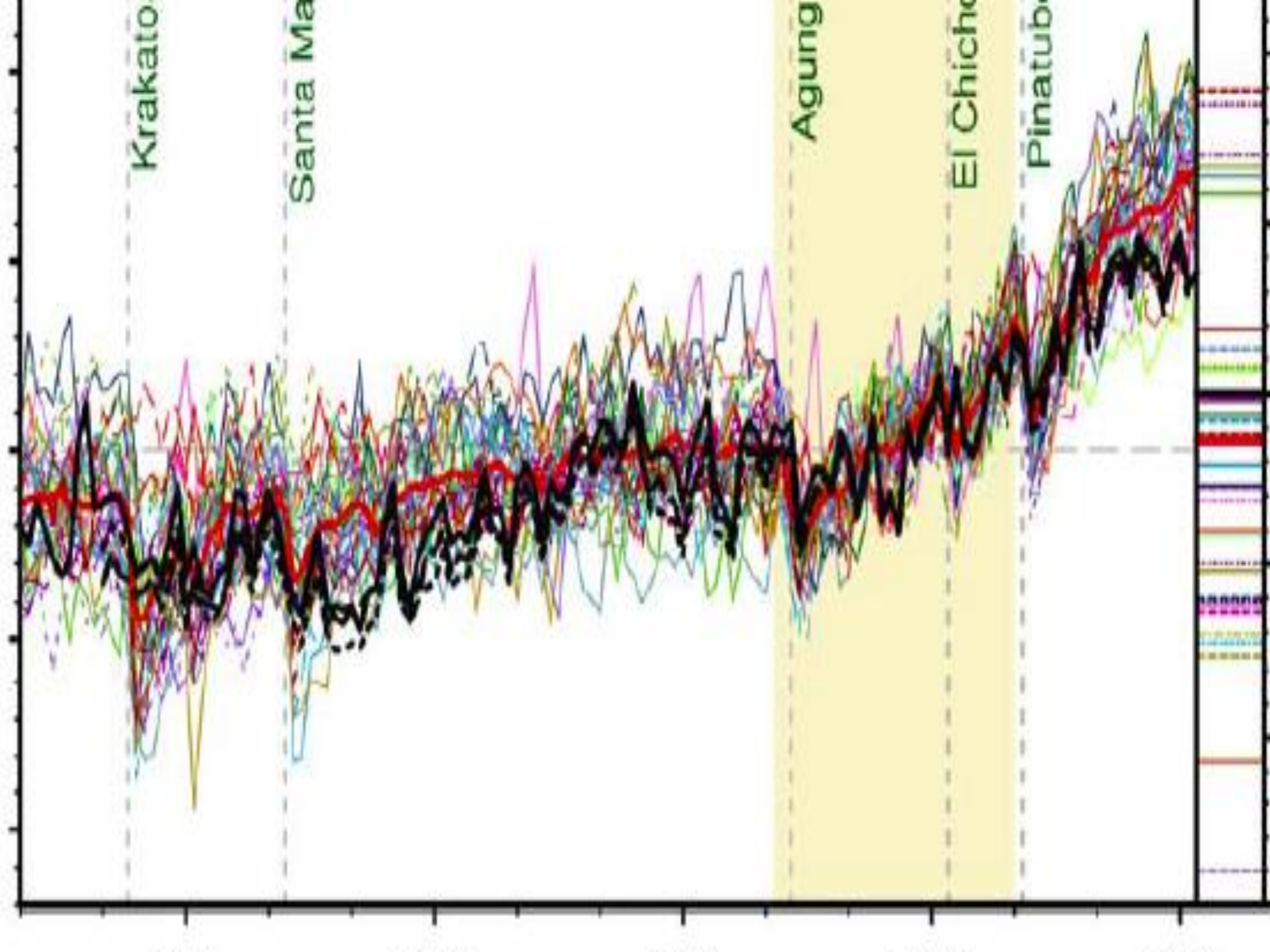


Figure 9.8: Observed and simulated time series of the anomalies in annual- and global-mean surface temperature. All anomalies are differences from the 1961–1990 time-mean of each individual time series. The reference period 1961–1990 is indicated by yellow shading; vertical dashed grey lines represent times of major volcanic eruptions. (a) Single simulations for CMIP5 models (thin lines); multi-model mean (thick red line); different observations (thick black lines). Observational data (see Chapter 2) are HadCRUT4 (Morice, Kennedy, Rayner, & Jones, 2012), GISTEMP (Hansen, Ruedy, Sato, & Lo, 2010), and MLOST (Vose et al., 2012) and are merged surface temperature (2 m height over land and surface temperature over the ocean). All model results have been sub-sampled using the HadCRUT4 observational data mask (see Chapter 10). Following the CMIP5 protocol (Taylor et al., 2012), all simulations use specified historical forcings up to and including 2005 and use RCP4.5 after 2005 (see Figure 10.1 and note different reference period used there; results will differ slightly when using alternative RCP scenarios for the post-2005 period). (a) Inset: the global-mean surface temperature for the reference period 1961–1990, for each individual model (colours), the CMIP5 multi-model mean (thick red), and the observations (thick black, P. D. Jones, New, Parker, Martin, and Rigor (1999)). Bottom: single simulations from available EMIC simulations (thin lines), from Eby et al. (2013). Observational data are the same as in (a). All EMIC simulations ended in 2005 and use the CMIP5 historical forcing scenario. (b) Inset: Same as in (a) but for the EMICs.

NorESM1-ME


(a) Observed and CMIP5 simulated global mean surface air temperature
1.5
1.0
0.5
0.0
-0.5
-1.0
15.5
15.0
14.5
14.0
13.5
13.0
12.5
Krakatoa
Santa Maria
Agung
El Chichon
Pinatubo
1890
1920
1950
1980
2010
Year



Krakato
Santa Ma
Agung
El Chich
Pinatub

(Re)Designing Climate Modelling for Policy

The basic idea is simply to consider the limited fidelity of the model before submitting the climate model runs; then redeploy resources due to shorter runs.
And explicitly discussing model inadequacy, big surprises, RDUs, and other central questions with policy makers at the beginning of the simulation process, not the end.
The limitations can be quantified through shadowing experiments, but at the moment reflection on the processes involved suggestion the duration exceeds rational expectations of fidelity (NARCAP divergence).

Pseudo-orbit Data Assimilation can provide information on the structure of model error. (and might prove rather useful in ocean models and reanalysis independently.

It opens up the options of doing time slice and temperature slice experiments with operational weather model resolutions.

In short, following Charney's triad, with fewer obs but aided by physical insight

Establishing what we can in fact do with fidelity, rather than using any means required to get "an" answer, and then (falsely) calling it "best available".

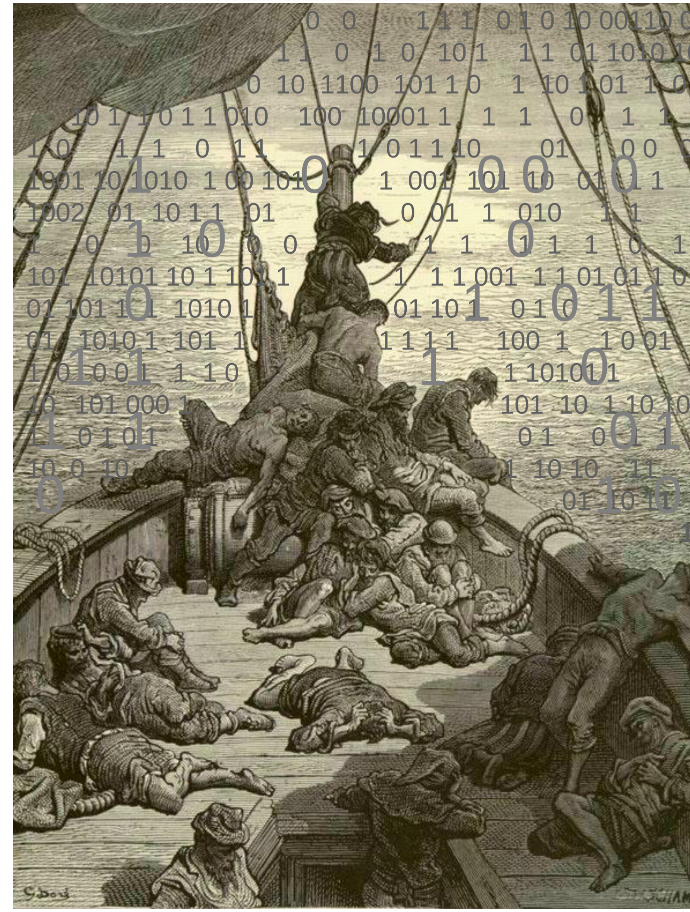
Data Data Everywhere, and Not a Bit to Bank On

One is sometimes surrounded by model-based probabilities from models known unlikely to be adequate to inform the questions we must answer.

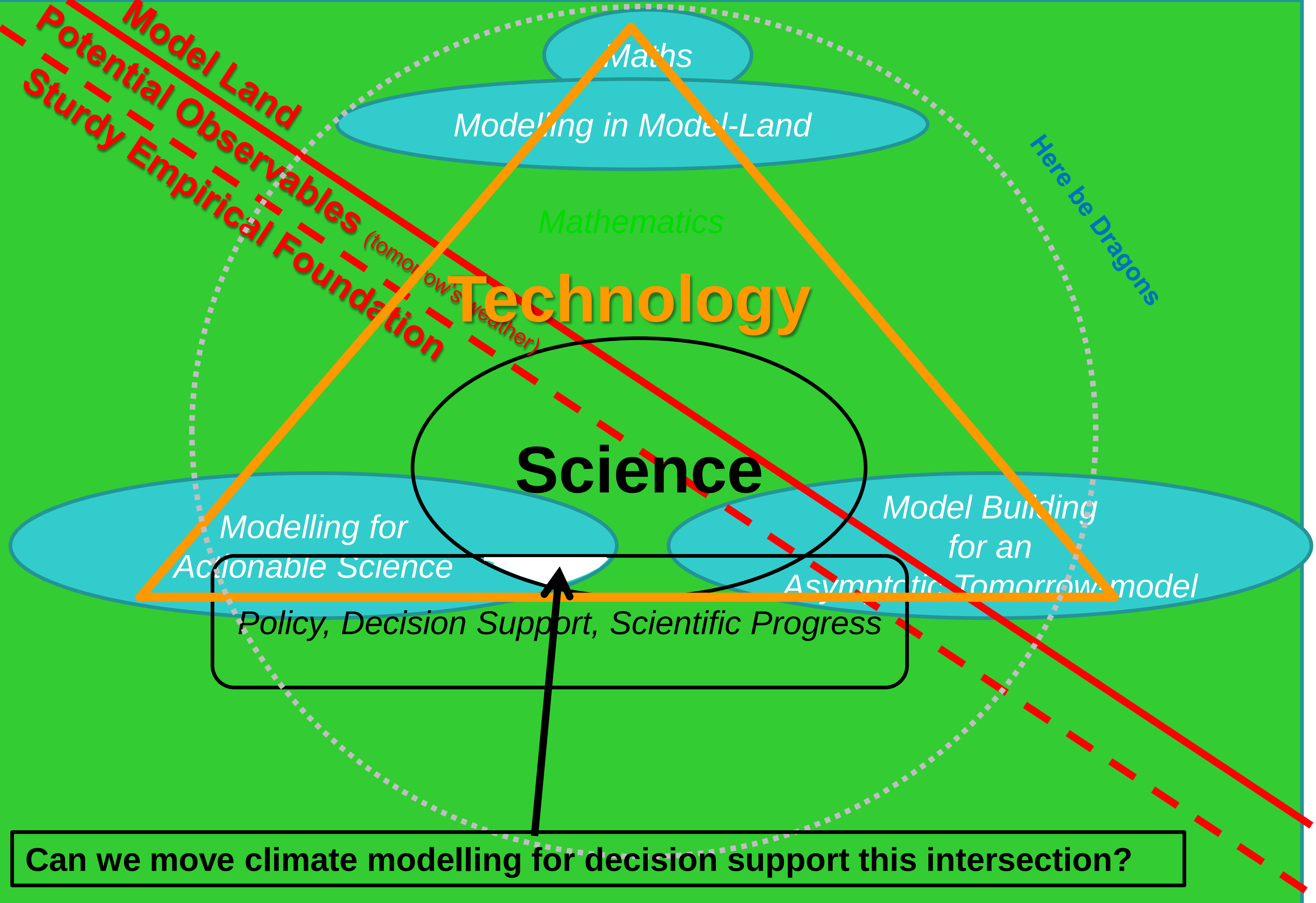
What is the rational path forward when the "best available" model is believed not to be adequate for purpose?

Estimate the probability that your model probability is misleading?
That is, **state the P(Big Surprise)**

Or refocus computer experiments in the limited arena where they truly are policy relevant, and state the P(Big Surprise)



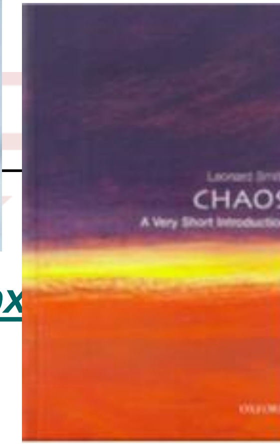
Inspired by Andy Morse (thx andy)
http://4umi.com/coleridge/rime/1


Model Land
Potential Observables (tomorrow's weather)
Sturdy Empirical Foundation
Maths
Modelling in Model-Land
Mathematics
Technology
Here be Dragons
Science
Modelling for
Actionable Science
Model Building
for an
Asymptotic Tomorrow-model
Policy, Decision Support, Scientific Progress
Can we move climate modelling for decision support this intersection?

Thank you

From an Oxford Bus Shelter:

predictions are wrong
sorry for any inconvenience



Publications www2.lse.ac.uk/CATS/Publications/Leonard Smith Publications.aspx

L A Smith and D A Stainforth (2012) Clarify the limits of climate models in Nature Vol. 489

Du, H. and Smith, L. A. (2012) Parameter estimation using ignorance **Physical Review E 86, 016213**
Bröcker, J. and Smith, L. A. (2008) From Ensemble Forecasts to Predictive Distribution Functions **Tellus A 60(4): 663**

D J Rowlands, et al (2012) Broad range of 2050 warming from an observationally constrained large climate model ensemble. Nature Geoscience
K Bevan, W Buytaert & L A Smith (2012) On virtual observatories and modelled realities Hydrol. Process., 26: 1905–1908

Smith, LA and Stern, N (2011) Uncertainty in science and its role in climate policy Phil. Trans. R. Soc. A (2011), 369, 1-24

R Hagedorn and LA Smith (2009) Communicating the value of probabilistic forecasts with weather roulette. MeteoRl App 16 (2): 143
K Judd, CA Reynolds, LA Smith & TE Rosmond (2008) The Geometry of Model Error. J of Atmos Sci 65 (6), 1749-1772. Abstract
DA Stainforth, MR Allen, ER Tredger & LA Smith (2007) Confidence, uncertainty and decision-support relevance in climate predictions, Phil. Trans. R. Soc. A, 365, 2145-2161. Abstract

LA Smith (2006) Predictability past predictability present. Chapter 9 of Predictability of Weather and Climate (eds T. Palmer and R Hagedorn). Cambridge, UK. Cambridge University Press.

LA Smith, (2002) What Might We Learn from Climate Forecasts? Proc. National Acad. Sci. USA 4 (99): 2487-2492. Abstract

D Orrell, LA Smith, T Palmer & J Barkmeijer (2001) Model Error in Weather Forecasting, Nonlinear Processes in Geophysics 8: 357
LA Smith (2000) 'Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems' in Nonlinear Dynamics and Statistics, ed. Alistair I Mees, Boston: Birkhauser, 31-64. Abstract
LA Smith, C Ziehmann & K Fraedrich (1999) Uncertainty Dynamics and Predictability in Chaotic Systems, Quart. J. Royal Meteorological Soc. 125: 2855-2886.

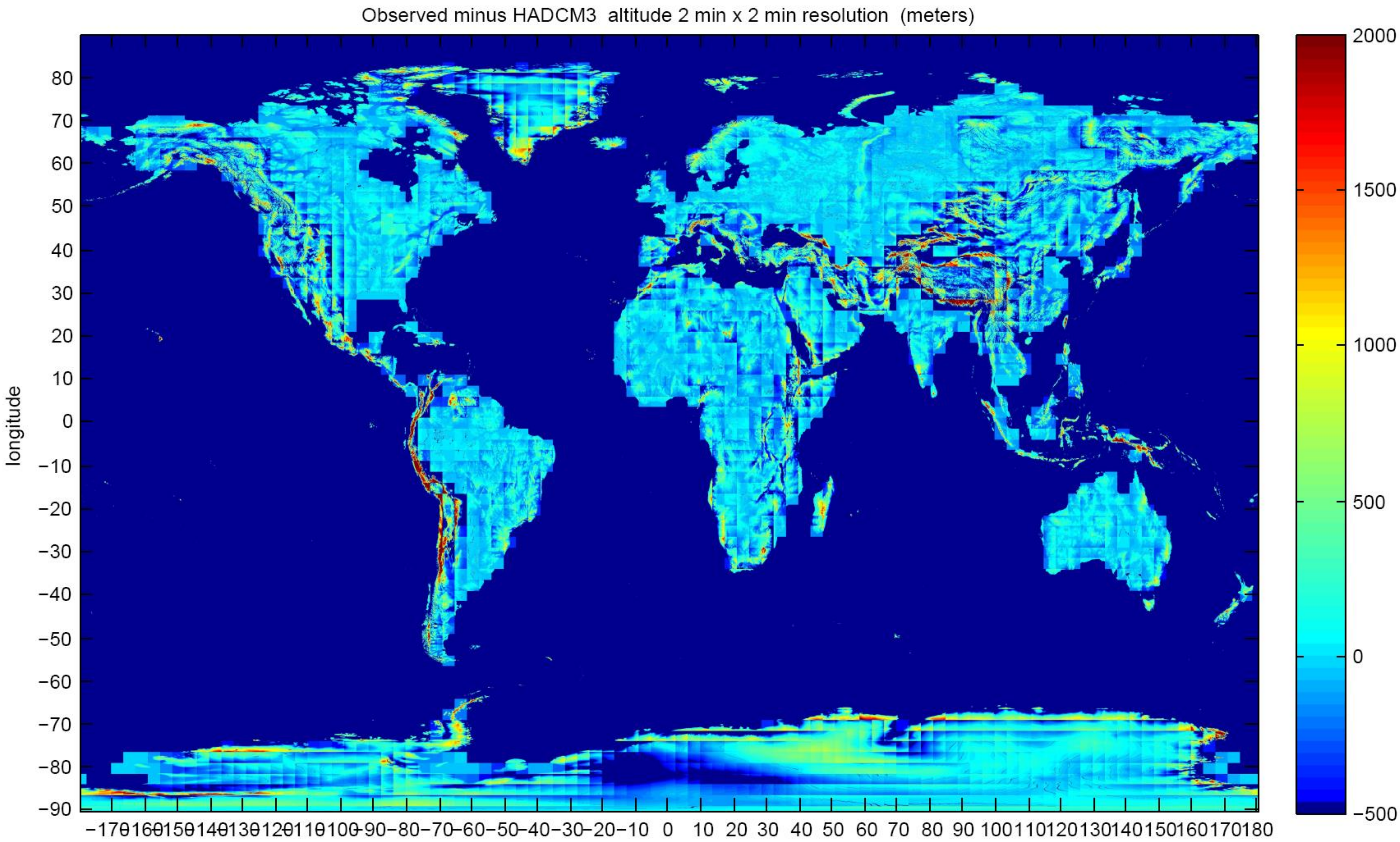
END

Objection has been taken to such forecasts, because they cannot be always exactly correct,—for all places in one district. It is, however, considered by most persons that general, comprehensive expressions, in aid of local observers, who can form independent judgments from the tables and *their own instruments*, respecting their immediate vicinity, *though not so well for distant places*, may be very useful, as well as interesting: while to an unprovided or otherwise uninformed person, an idea of the kind of weather thought *probable* cannot be otherwise than acceptable, provided that he is in no way *bound* to act in accordance with any such views, against his own judgment.

Like the storm signals, such notices should be merely *cautionary*—to denote anticipated disturbance *somewhere* over these islands,—without being in the least degree compulsory, or interfering arbitrarily with the movements of vessels or individuals.

Certain it is, that although our conclusions may be incorrect—our judgment erroneous—the laws of nature, and the signs afforded to man, are invariably true. Accurate interpretation is the real deficiency.

Fitzroy, 1862


Observed minus HADCM3 altitude 2 min x 2 min resolution (meters)
longitude
80
70
60
50
40
30
20
10
0
−10
−20
−30
−40
−50
−60
−70
−80
−90
−170−160−150−140−130−120−110−100−90−80−70−60−50−40−30−20−10 0 10 20 30 40 50 60 70 80 90 100110120130140150160170180
2000
1500
1000
500
0
−500

What is a "Big Surprise"?

Big Surprises arise when something our simulation models cannot mimic turns out to have important implications for us.

In some cases where we are "leaking probability" when a fraction of our model runs explore conditions which we know they cannot simulate realistically. (Science can warn of "known unknowns" even when the magnitude remains unknown)

Big Surprises invalidate (not update) model-based probability forecasts, the g in P(x|𝔾) .   "Bayes" is irrelevant outside of probability theory.

How might we better communicate inadequacy as well as imprecision?

Condition explicitly on the euro not collapsing [Bank of England].

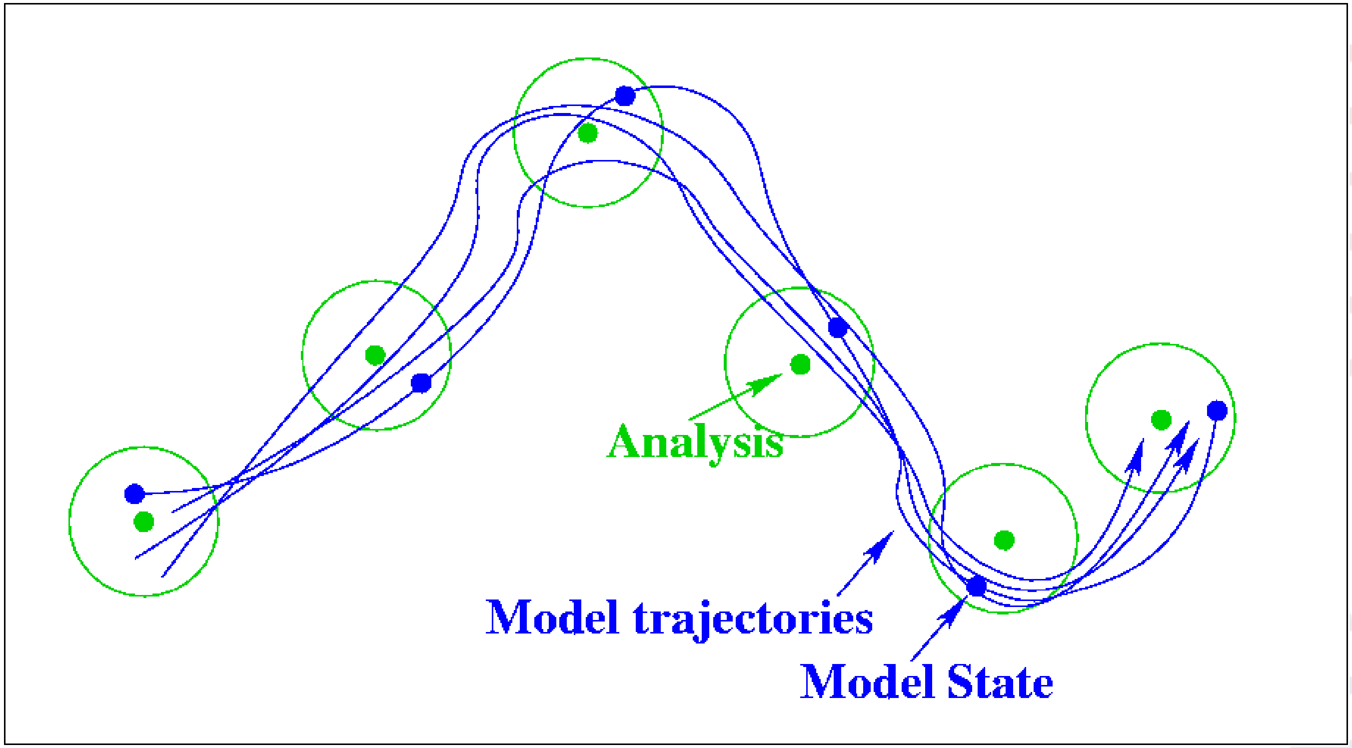
Financial and energy market assumptions

Provide subjective estimates of the probability that the model is misinformative in the future [for example, the P(BS)].

Refuse to issue a quantitative forecast, probability or otherwise [UKML].

Measures of Predictability

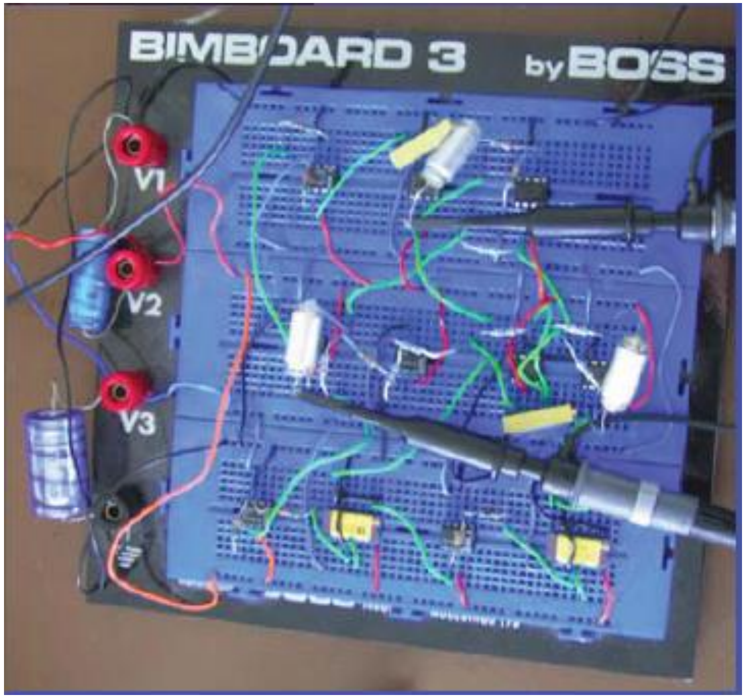
Lyapunov Time, Doubling Times, Shadowing Times, ~~Decay of Predictability~~

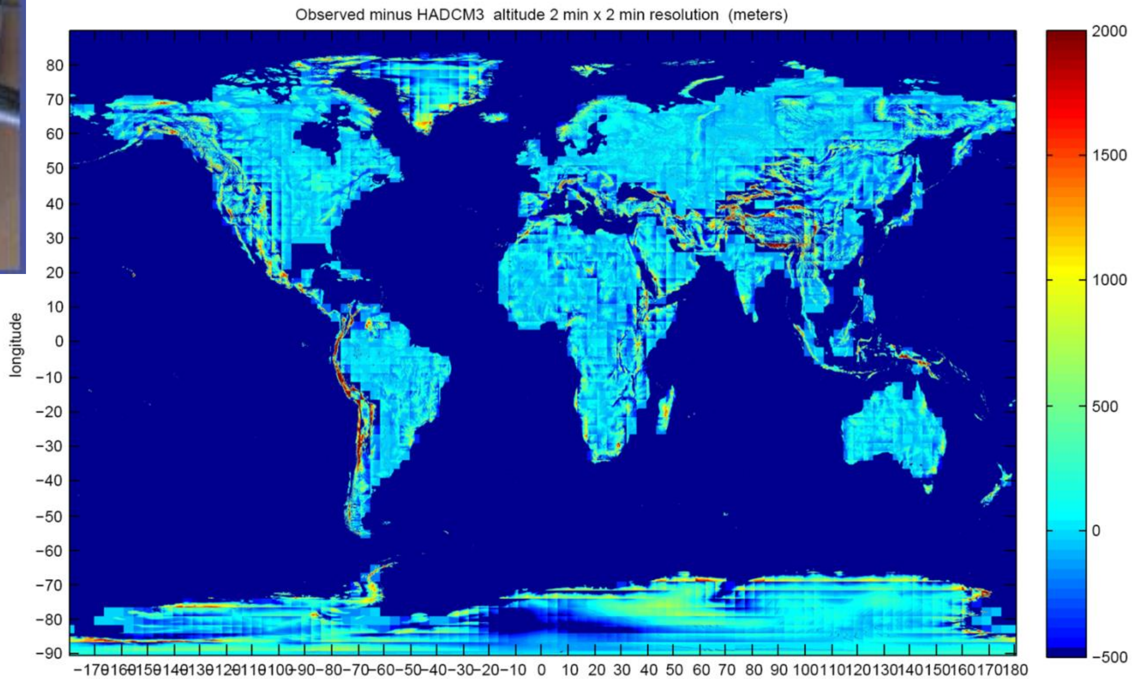


The distribution of i-shadowing times provides an excellent upper bound on predictability. But they are expensive, perhaps undefined in a forward forecast context, and if the model is perfect they are all infinite!

Smith, L. A. (2000) 'Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems' in Nonlinear Dynamics and Statistics, ed. Alistair I Mees, Boston: Birkhauser, 31-64
Smith, L. A. (1996) Accountability and Error in Ensemble Forecasting. In 1995 ECMWF Seminar on Predictability. Vol 1, pg 351-368. ECMWF, Reading

Is chaos the dominant uncertainty in practice?





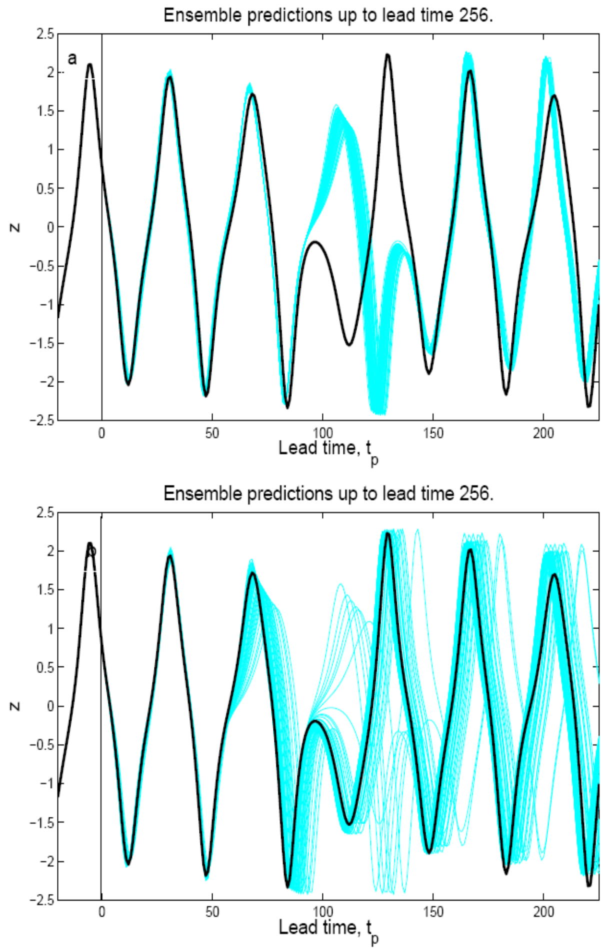
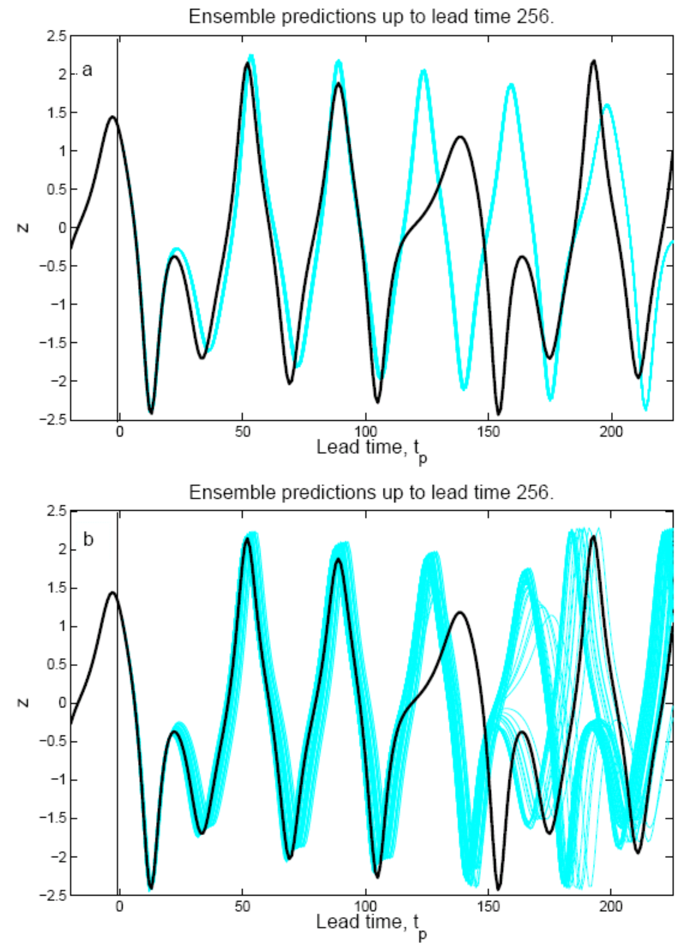


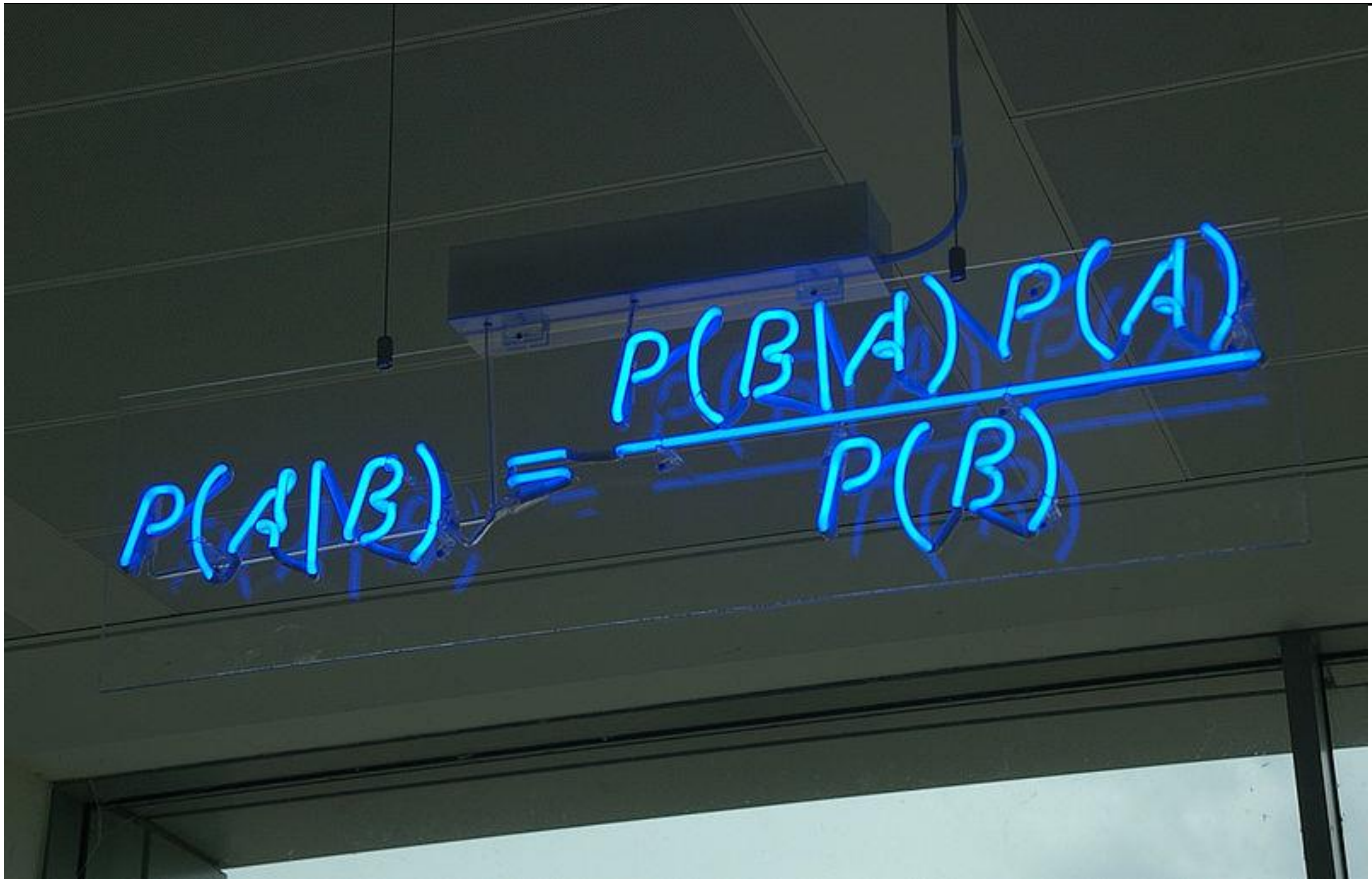
Figure 7: Ensemble predictions using (a) model 1 and (b) model 2. The



Ensemble predictions using (a) model 1 and (b) model 2. The

From "distance" to climatology to Forecast evaluation:
The IGN relative to climatology only reflects information content when the distribution is a "good forecast".

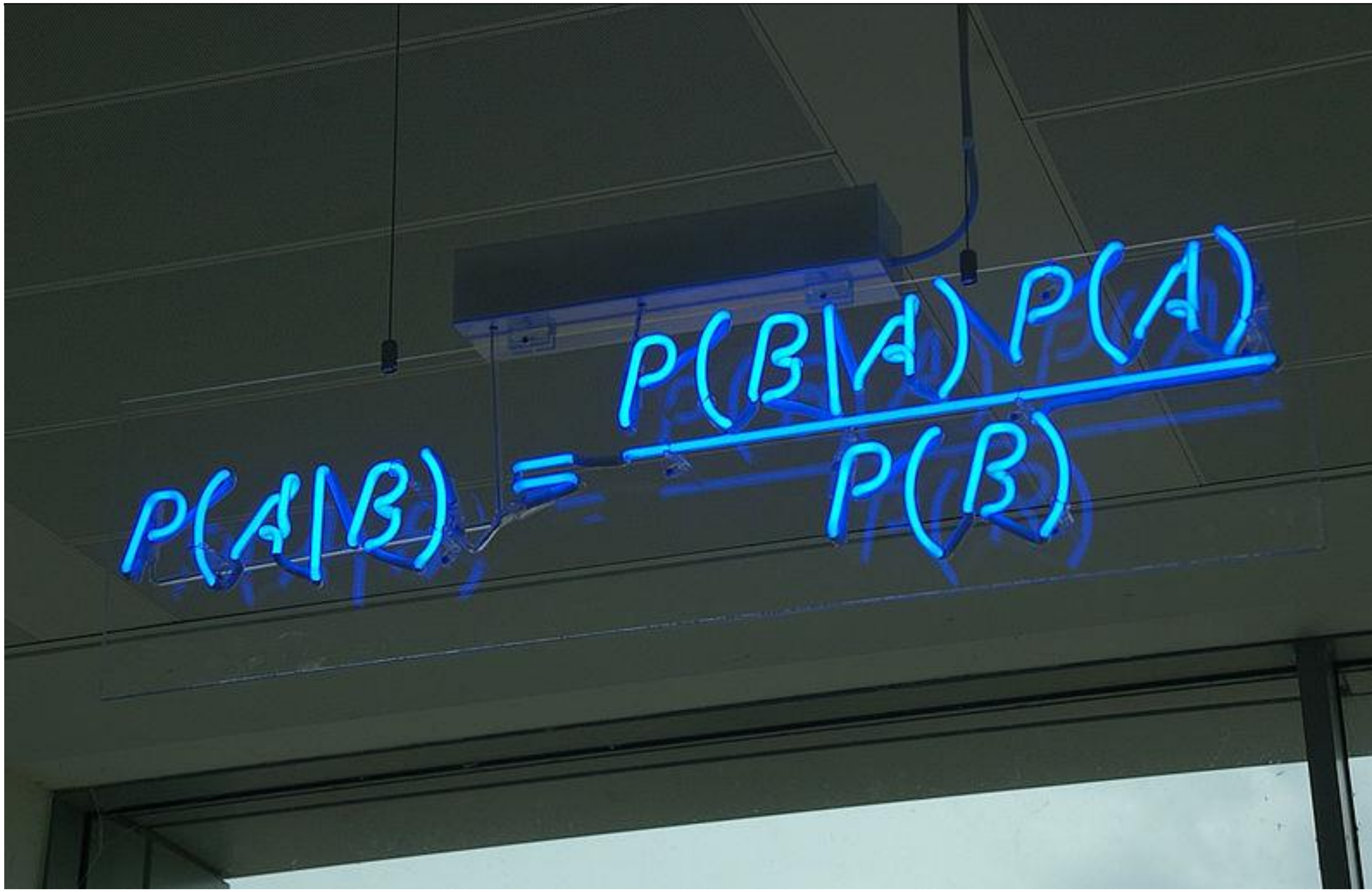
Target uncertain (but exists) : Be a Bayesian
Target indeterminate (none exists): Bayes (& the probability calculus) irrelevant.



$$P(\alpha \mid Data, \boldsymbol{I}) = P(Data \mid \alpha, \boldsymbol{I})\, P(\alpha \mid \boldsymbol{I}) \;/\; P(Data \mid \boldsymbol{I}) \propto P(Data \mid \alpha, I)\, P(\alpha \mid I)$$

Target uncertain (but exists) : Be a Bayesian

Target indeterminate (none exists): Bayes (& the probability calculus) irrelevant.



$$P(\alpha \mid Data, \mathbf{I}) = P(Data \mid \alpha, \mathbf{I})\, P(\alpha \mid \mathbf{I}) \;/\; P(Data \mid \mathbf{I}) \;\cancel{\propto}\; P(Data \mid \alpha, I)\, P(\alpha \mid I)$$

$= 0$

Free exchange

Hot air

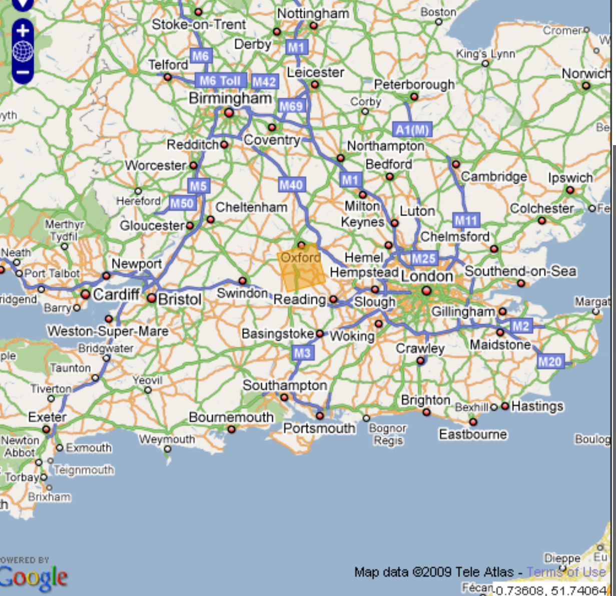
Are models that show the economic effects of climate change useless?

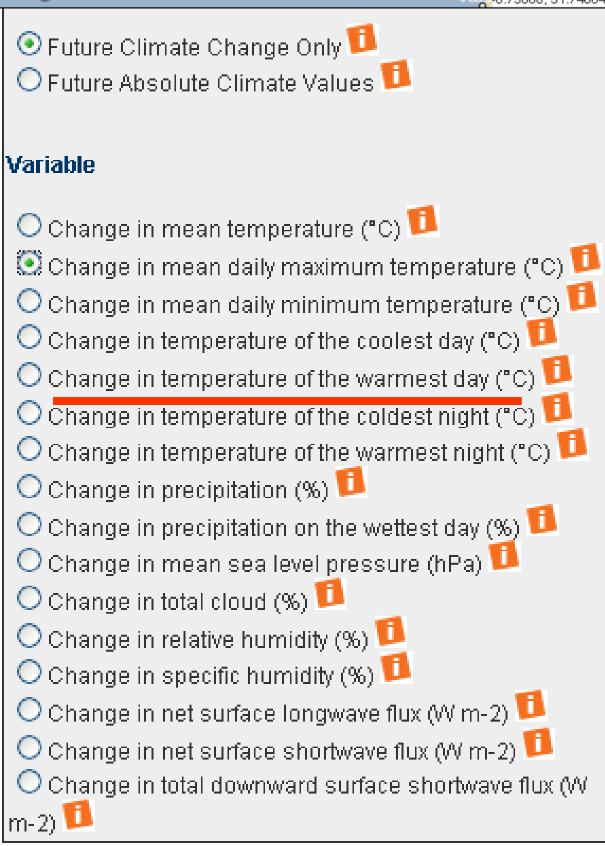
Oct 5th 2013 | From the print edition

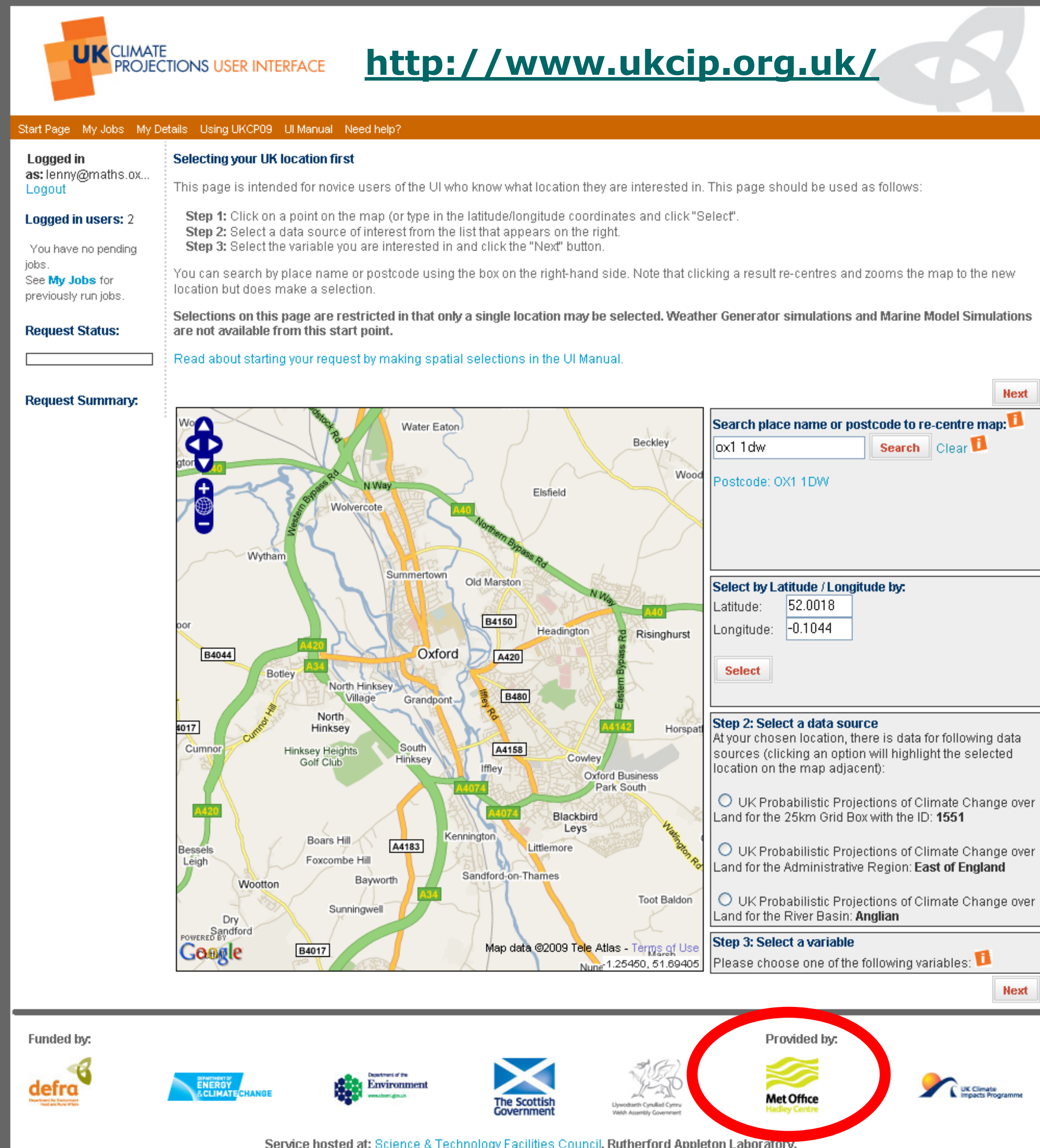
Like 87 Tweet 97

MODELS simplify. They are supposed to. It is a feature, not a bug. Their formulae may be complex but models deliberately omit some things in order to focus on others. They also

Is it plausible to provide a PDF of hottest or stormiest summer day in 2080's Oxford???







Smith

Publications **[www2.lse.ac.uk/CATS/Publications/Leonard Smith Publications.aspx]**

L A Smith and D A Stainforth (2012) Clarify the limits of climate models in Nature Vol. 489

Du, H. and Smith, L. A. (2012) *Parameter estimation using ignorance Physical Review E 86, 016213*

D J Rowlands, et al (2012) Broad range of 2050 warming from an observationally constrained large climate model ensemble. Nature Geoscience

K Bevan, W Buytaert & L A Smith (2012) On virtual observatories and modelled realities Hydrol. Process., 26: 1905–1908

Broker J and LA Smith (2008) *From Ensemble Forecasts to Predictive Distribution Functions Tellus A 60(4): 663*

Smith, LA and Stern, N (2011) Uncertainty in science and its role in climate policy Phil. Trans. R. Soc. A (2011), 369, 1-24

R Hagedorn and LA Smith (2009) Communicating the value of probabilistic forecasts with weather roulette. MeteoRl App 16 (2): 143

K Judd, CA Reynolds, LA Smith & TE Rosmond (2008) The Geometry of Model Error. J of Atmos Sci 65 (6), 1749-1772. Abstract

DA Stainforth, MR Allen, ER Tredger & LA Smith (2007) Confidence, uncertainty and decision-support relevance in climate predictions, Phil. Trans. R. Soc. A, 365, 2145-2161. Abstract

LA Smith (2006) Predictability past predictability present. Chapter 9 of Predictability of Weather and Climate (eds T. Palmer and R Hagedorn). Cambridge, UK. Cambridge University Press.

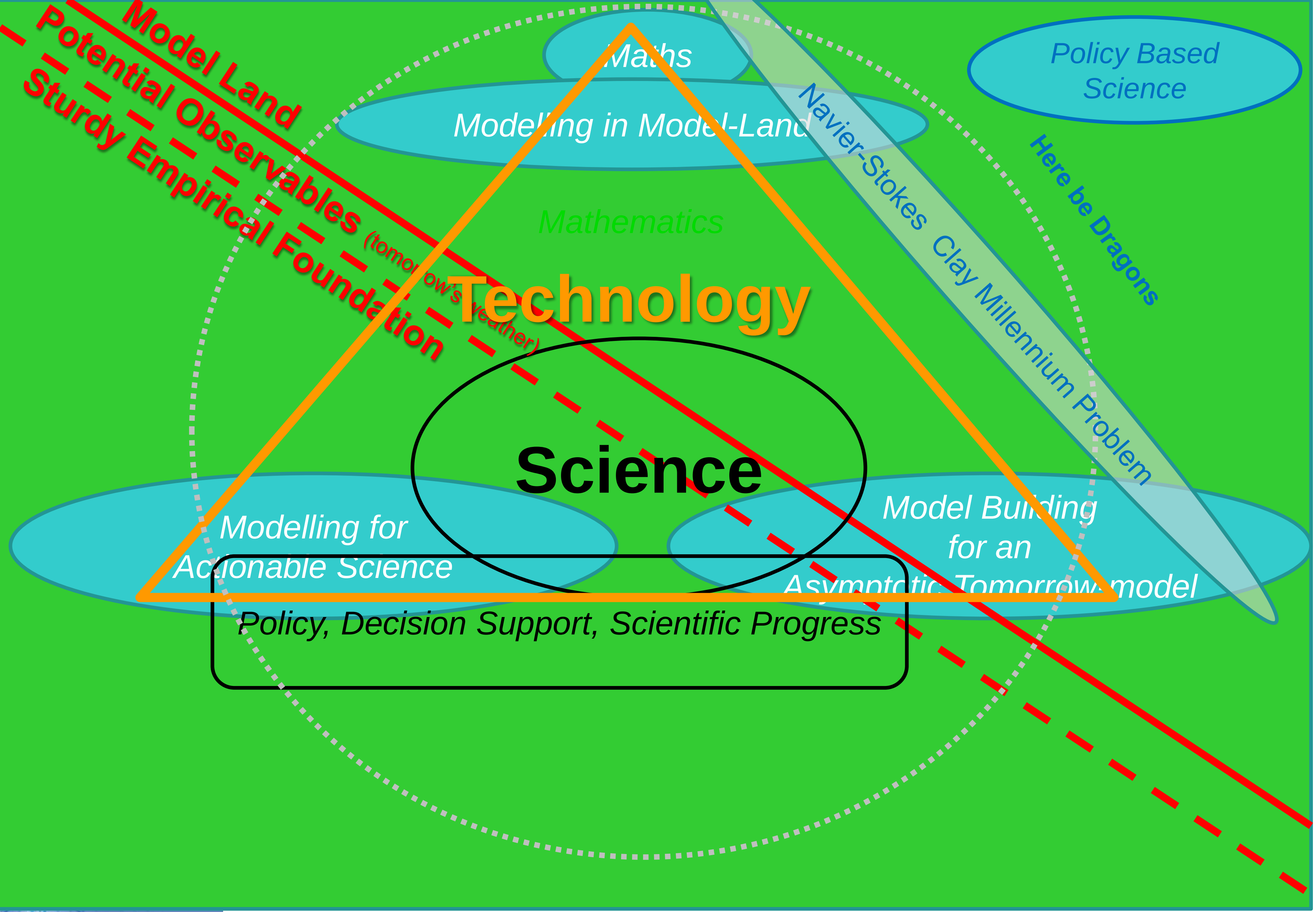
LA Smith (2006) Predictability past predictability present. Chapter 9 of Predictability of Weather and Climate (eds T. Palmer and R Hagedorn). Cambridge, UK. Cambridge University Press.

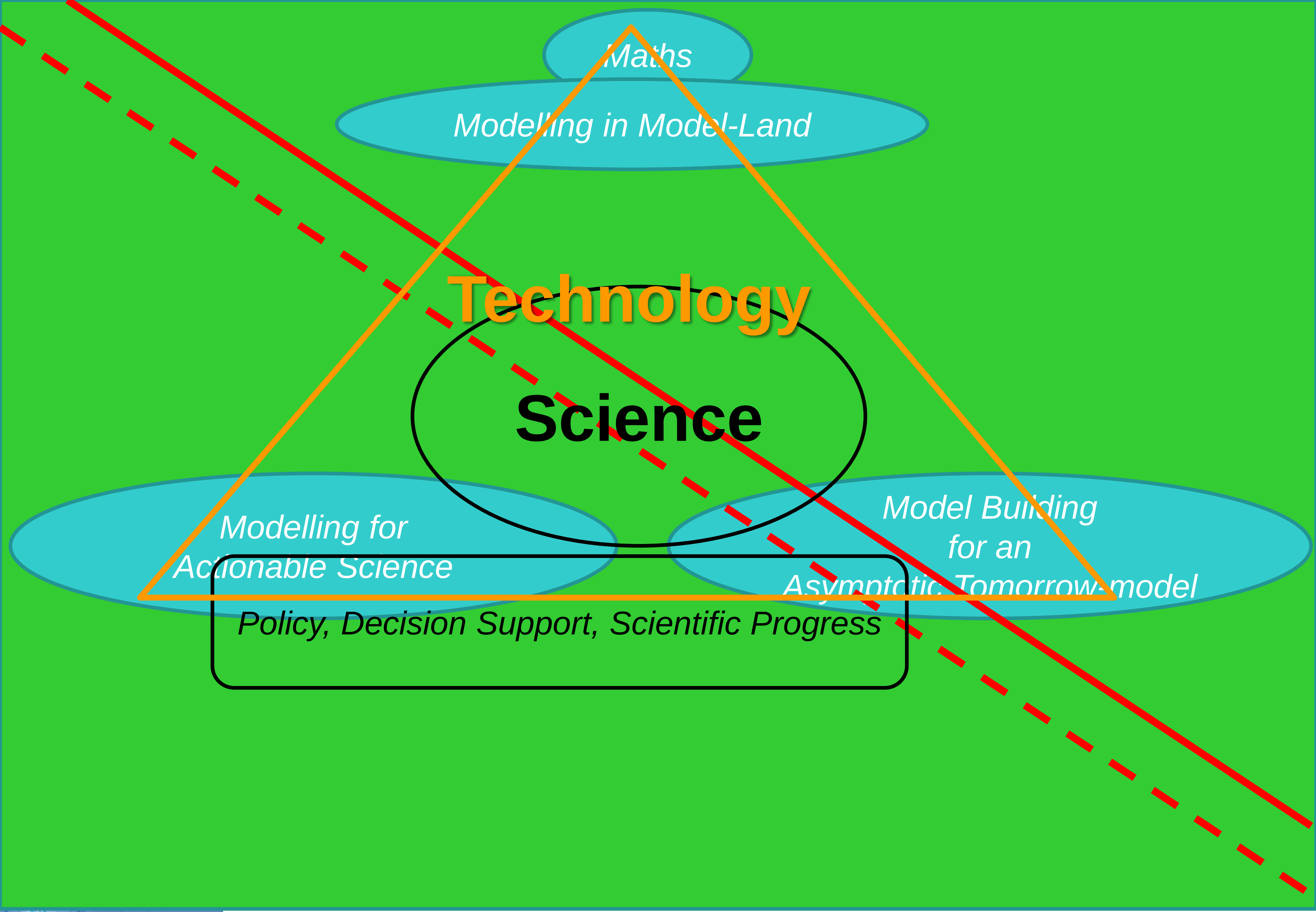
LA Smith, (2002) What Might We Learn from Climate Forecasts? Proc. National Acad. Sci. USA 4 (99): 2487-2492. Abstract

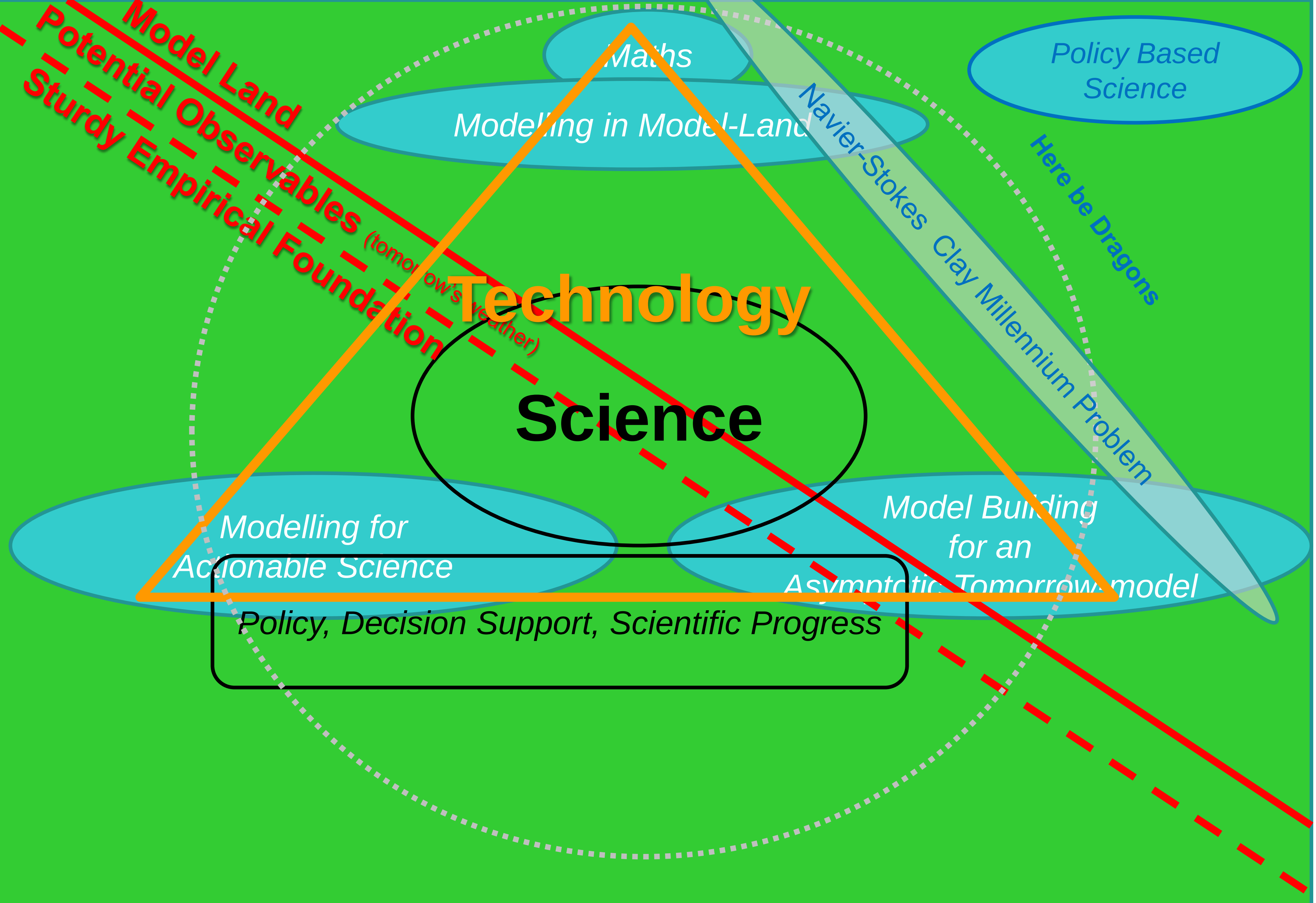
D Orrell, LA Smith, T Palmer & J Barkmeijer (2001) Model Error in Weather Forecasting, Nonlinear Processes in Geophysics 8: 357

LA Smith (2000) 'Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems' in Nonlinear Dynamics and Statistics, ed. Alistair I Mees, Boston: Birkhauser, 31-64. Abstract

LA Smith, C Ziehmann & K Fraedrich (1999) Uncertainty Dynamics and Predictability in Chaotic Systems, Quart. J. Royal Meteorological Soc. 125: 2855-2886.

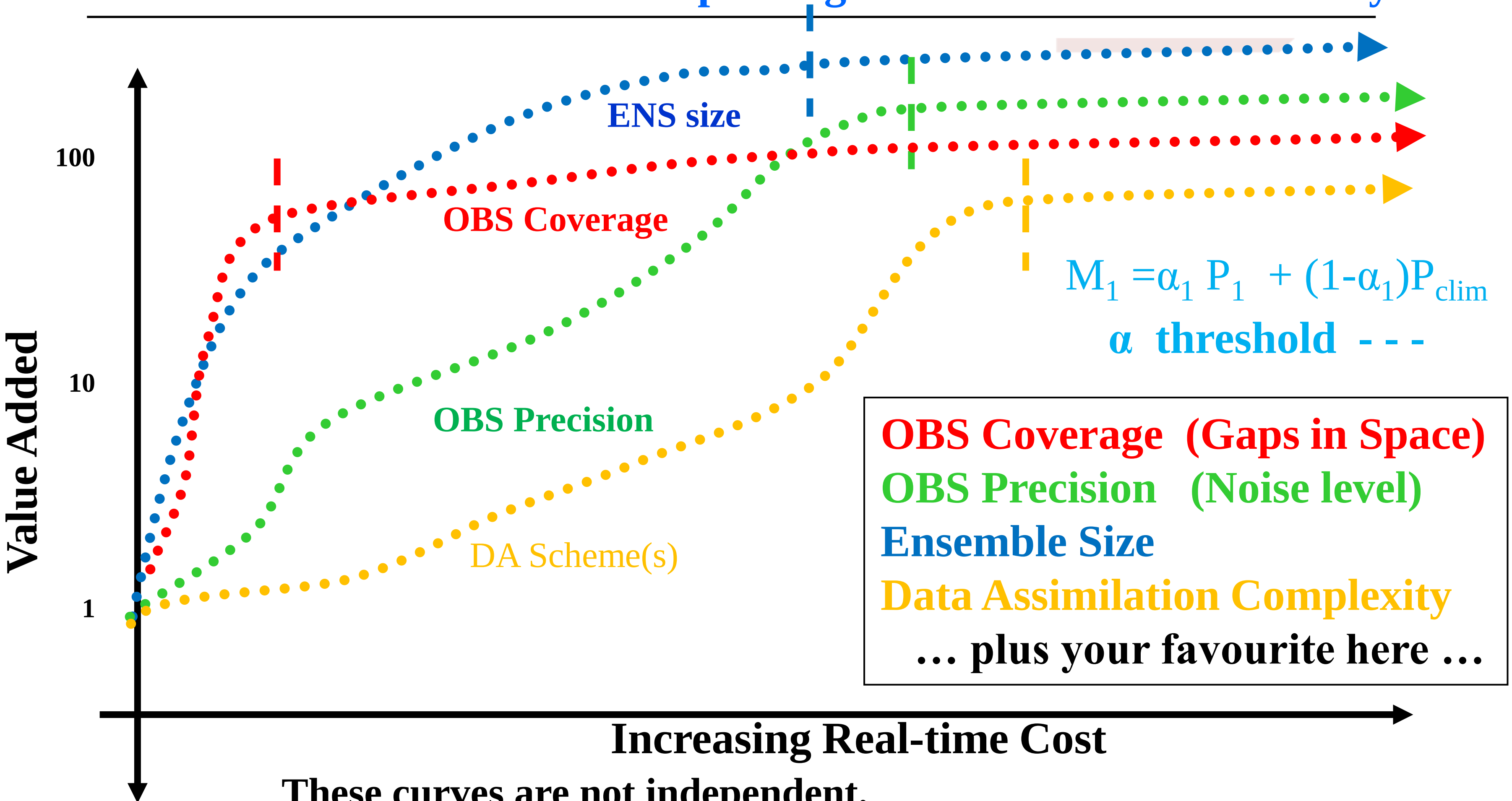

Model Land
Potential Observables (tomorrow's weather)
Sturdy Empirical Foundation
Maths
Modelling in Model-Land
Policy Based Science
Navier-Stokes Clay Millennium Problem
Here be Dragons
Mathematics
Technology
Science
Modelling for Actionable Science
Model Building for an Asymptotic Tomorrow model
Policy, Decision Support, Scientific Progress


Maths
Modelling in Model-Land
Technology
Science
Modelling for
Actionable Science
Model Building
for an
Asymptotic Tomorrow-model
Policy, Decision Support, Scientific Progress


Model Land
Potential Observables (tomorrow's weather)
Sturdy Empirical Foundation
Maths
Modelling in Model-Land
Navier-Stokes Clay Millennium Problem
Policy Based Science
Here be Dragons
Technology
Science
Modelling for Actionable Science
Model Building for an Asymptotic Tomorrow model
Policy, Decision Support, Scientific Progress

Improving Predictability

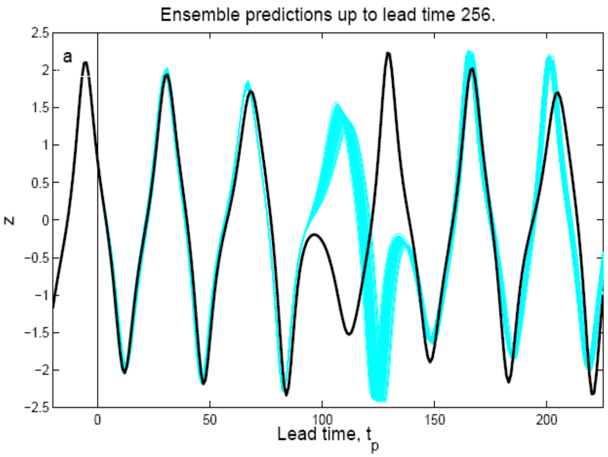
Schematic view of value added for improving initial condition uncertainty.

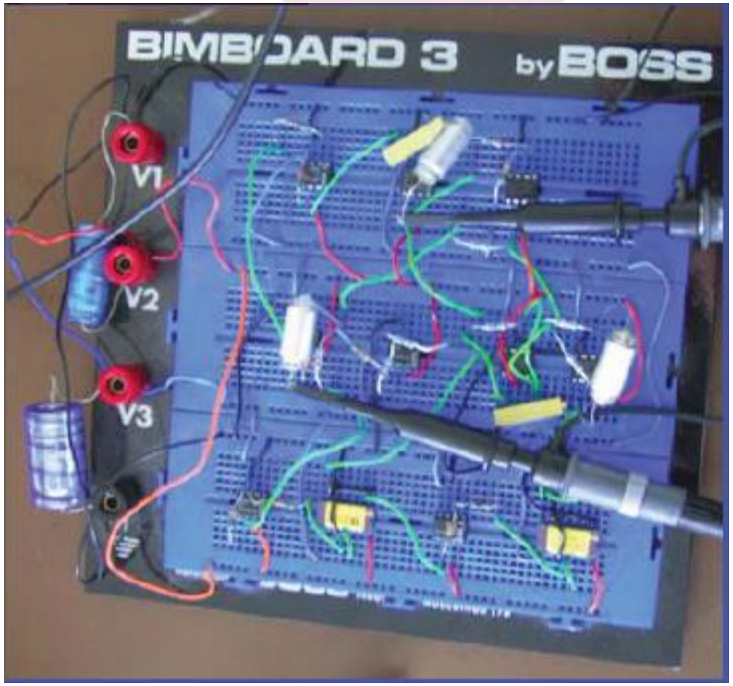


These curves are not independent.
The curves vary with the target.
Development costs start from different legacy baselines
Historically these "optimised" separately (?draw on separate budgets?)
How to measure "value added" in this context?

Examples of Probability Forecasts

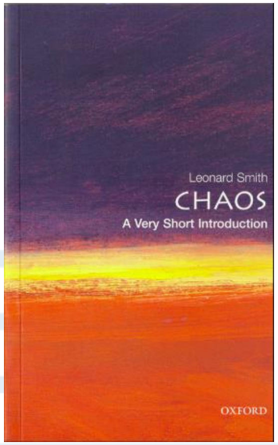
Perfect Model Scenario   *Weather-like Tasks*   ~~Climate-like~~ Tasks





In these forecasts of a simple "chaotic" circuit, the limitations on predictability come from model inadequacy (structural model error) not from chaos.

This model could not shadow past the point where the ensemble departs from the future trajectory.



How did we get to post code probabilities in 2080s?

(It would be interesting to trace how the idea that climate models could provided quantitative insight came about.)

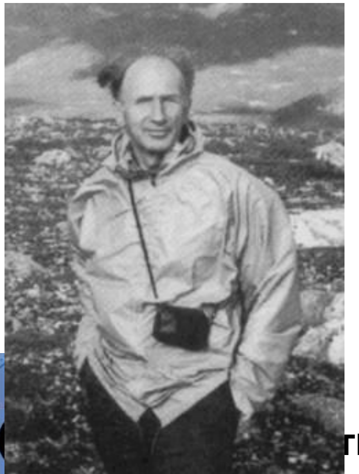
Because of the various simplifications of the model described above, it is not advisable to take too seriously the quantitative aspect of the results obtained in this study. Nevertheless, it is hoped that this study not only emphasizes some of the important mechanisms which control the response of the climate to the change of carbon dioxide.

The Effects of Doubling the CO_2 Concentration on the Climate of a General Circulation Model[1]

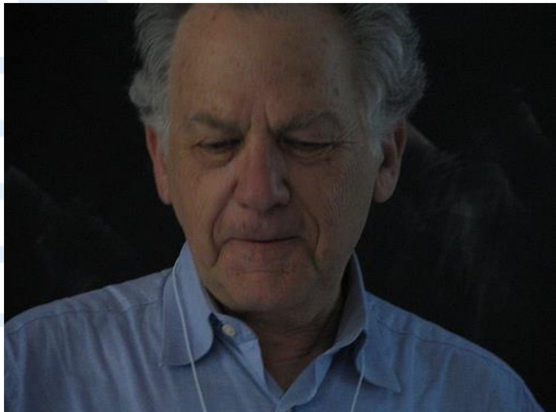
SYUKURO MANABE AND RICHARD T. WETHERALD

Geophysical Fluid Dynamics Laboratory/NOAA, Princeton University, Princeton, N.J. 08540

(Manuscript received 6 June 1974, in revised form 8 August 1974)



Mechanisms == Insight



Distinguishing Weather-like and Climate-like tasks

Weather-like forecasting tasks:
- model lifetime is long in comparison to the typical forecast lead-time
- large archive of truly out-of-sample forecast-outcome pairs
- arguably extrapolation in time but interpolation in state space

Here the same model is deployed many times in similar circumstances and one can learn from past mistakes.

Climate-like forecasting tasks:
- lead-times of interest are far longer than the lifetime of model
- forecast-outcome archive is very small, arguably empty
- lead-times of interest are long compared to the career of a researcher.

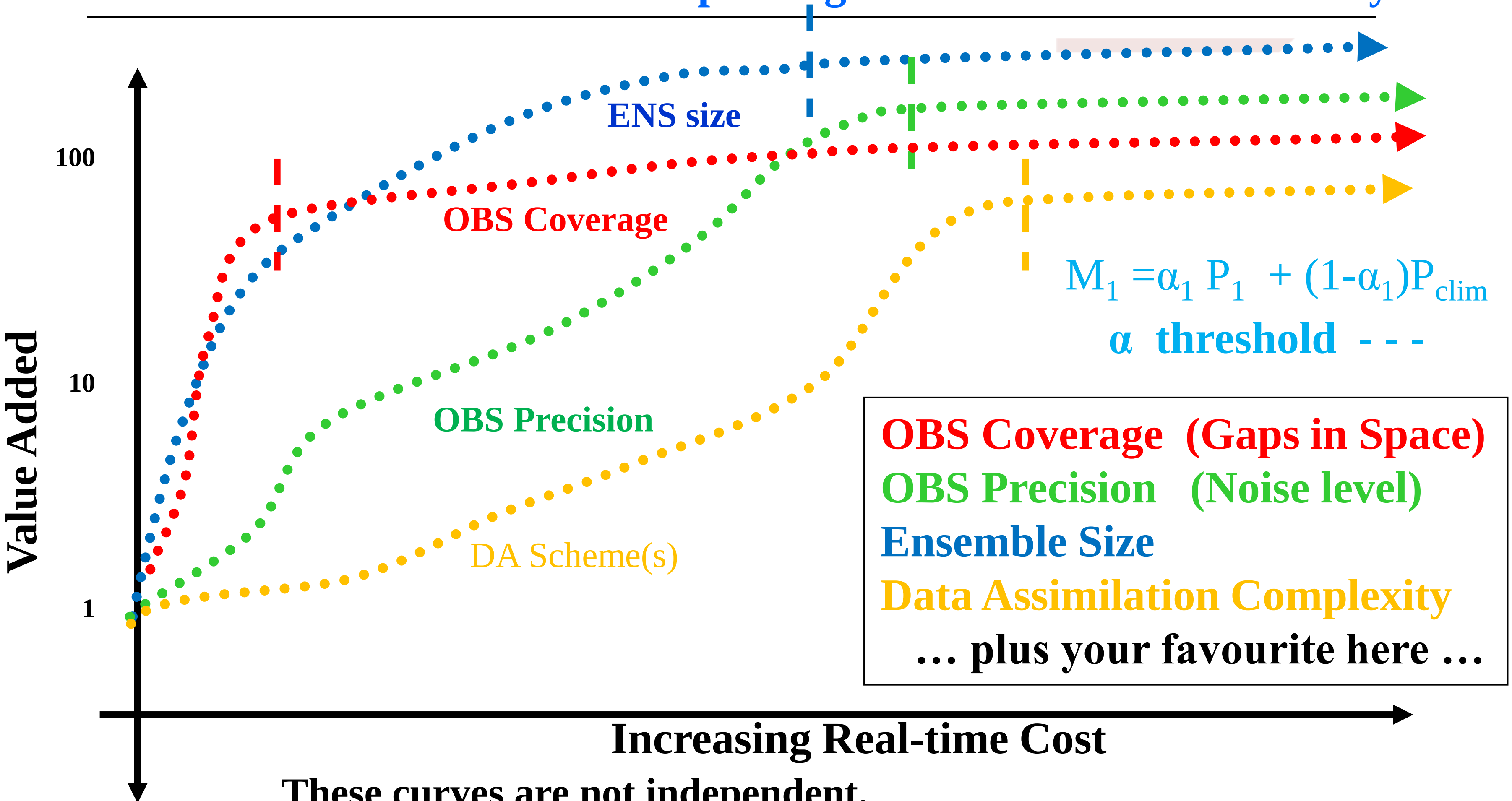
By the nature of the problem there are <u>no true out-of-sample observations</u>.

Best practice principles of forecasting differ in these two settings.

There is also the **Perfect Model Scenario**, where known mathematical systems (known if forgotten) play the role of the model and of the target system.

Improving Predictability

Schematic view of value added for improving initial condition uncertainty.



These curves are not independent.
The curves vary with the target.
Development costs start from different legacy baselines
Historically these "optimised" separately (?draw on separate budgets?)
How to measure "value added" in this context?

Decision Support Given only Immature Probabilities

Provide Stories in addition to (?rather than?) Numbers

Distinguish Immature Probability from Physical Probability

Balance “Based on the Laws of Physics” with P(BS)

Clearly distinguish statements regarding the probability of the next model run from those on the probability of the true target.

Demonstrate relative skill of empirical surrogate models
(Smith, 1992, 1997; Suckling and Smith 2012)

Identify and discuss the RDU and the timescales for tackling it.

Structural Model Error

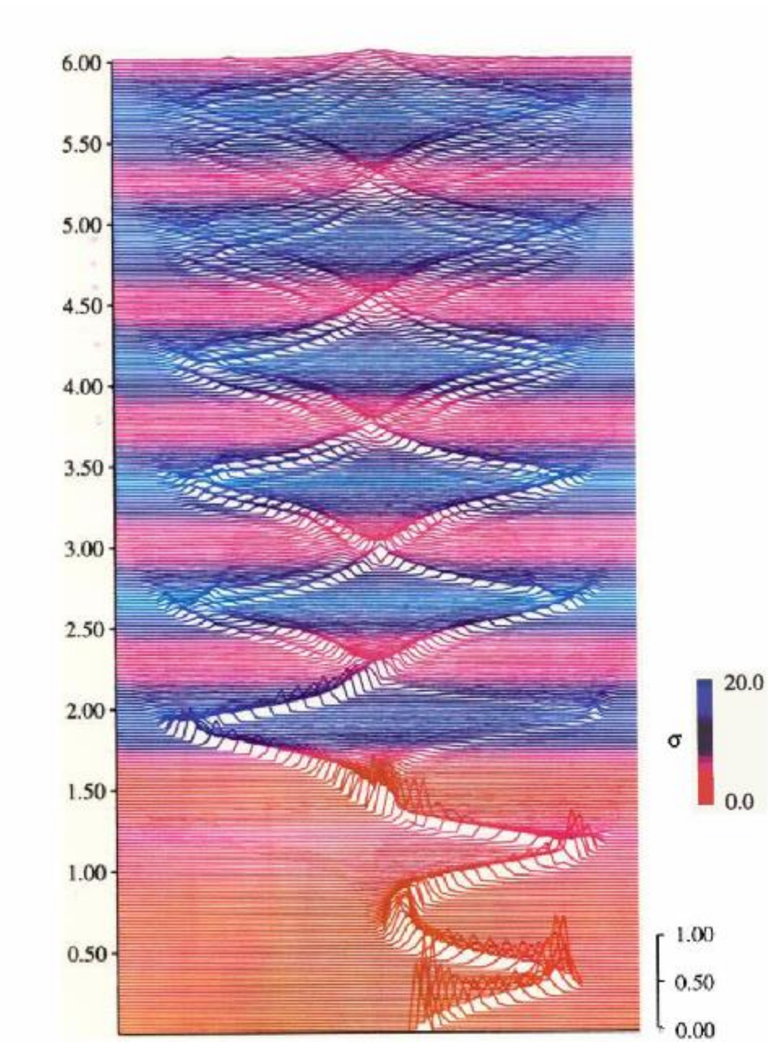
Smith, L.A. (2002) What might we learn from climate forecasts? Proc. National Acad. Sci. USA 4 (99): 2487-2492.

Most climate models are large dynamical systems involving a million (or more) variables on big computers. Given that they are nonlinear and not perfect, what can we expect to learn from them about the earth's climate? How can we determine which aspects of their output might be useful and which are noise? And how should we distribute resources between making them "better," estimating variables of true social and economic interest, and quantifying how good they are at the moment? Just as "chaos" prevents accurate weather forecasts, so model error precludes accurate forecasts of the distributions that define climate, yielding uncertainty of the second kind. Can we estimate the uncertainty in our uncertainty estimates? These questions are discussed. Ultimately, all uncertainty is quantified within a given modeling paradigm; our forecasts need never reflect the uncertainty in a physical system.

"Laws, where they do apply, hold only *ceteris paribus*."
Nancy Cartwright (1)

Laplace's Demon, its Apprentice and Good's ∞ Rational Org

Perfect Model Scenario *Weather-like Tasks* *Climate-like Tasks*



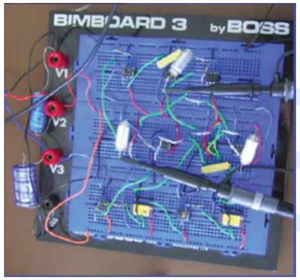
Smith (2002) Chaos and Predictability in *Encyc Atmos Sci*

Traditional risk management under uncertainty is straightforward in a chaotic system (given a perfect model and large observational archive).

It is only a matter of investment, to extract parameters, probability forecasts that are actionable, &c

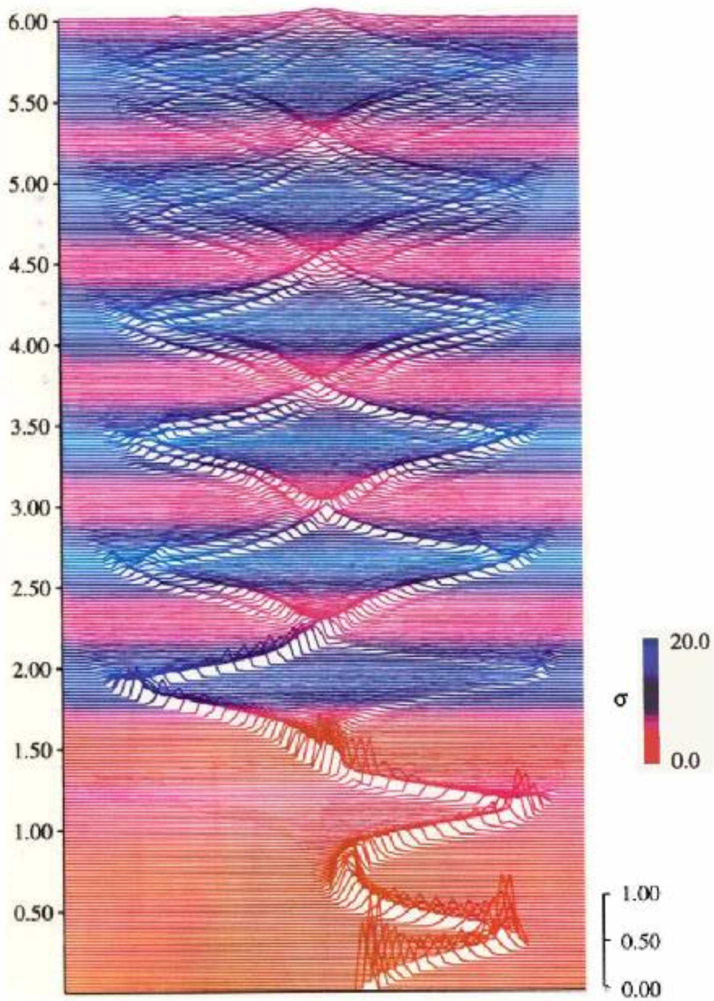
Judd, K. and Smith, L. A. (2001) Indistinguishable States I: The Perfect Model Scenario, Physica D 151: 125-141.

These results do not generalise to even the most simple real-world systems best modelled by chaotic models.



Measures of Predictability

Lyapunov Time, Doubling Times, Shadowing Times, Decay of Predictability



In this talk, consider the time required before an initial ensemble provides negligible information given the natural measure.

$$- \Sigma\ p(x) \log_2(p(x)/\mu(x))$$

If either the ensemble or the model is imperfect, the relevant time is related to

$$- \Sigma\ q(x) \log_2(p(x)/\mu(x))$$

And perhaps the most interesting quantity is

$$- \Sigma\ q(x) \log_2(p(x)/q(x))$$

where

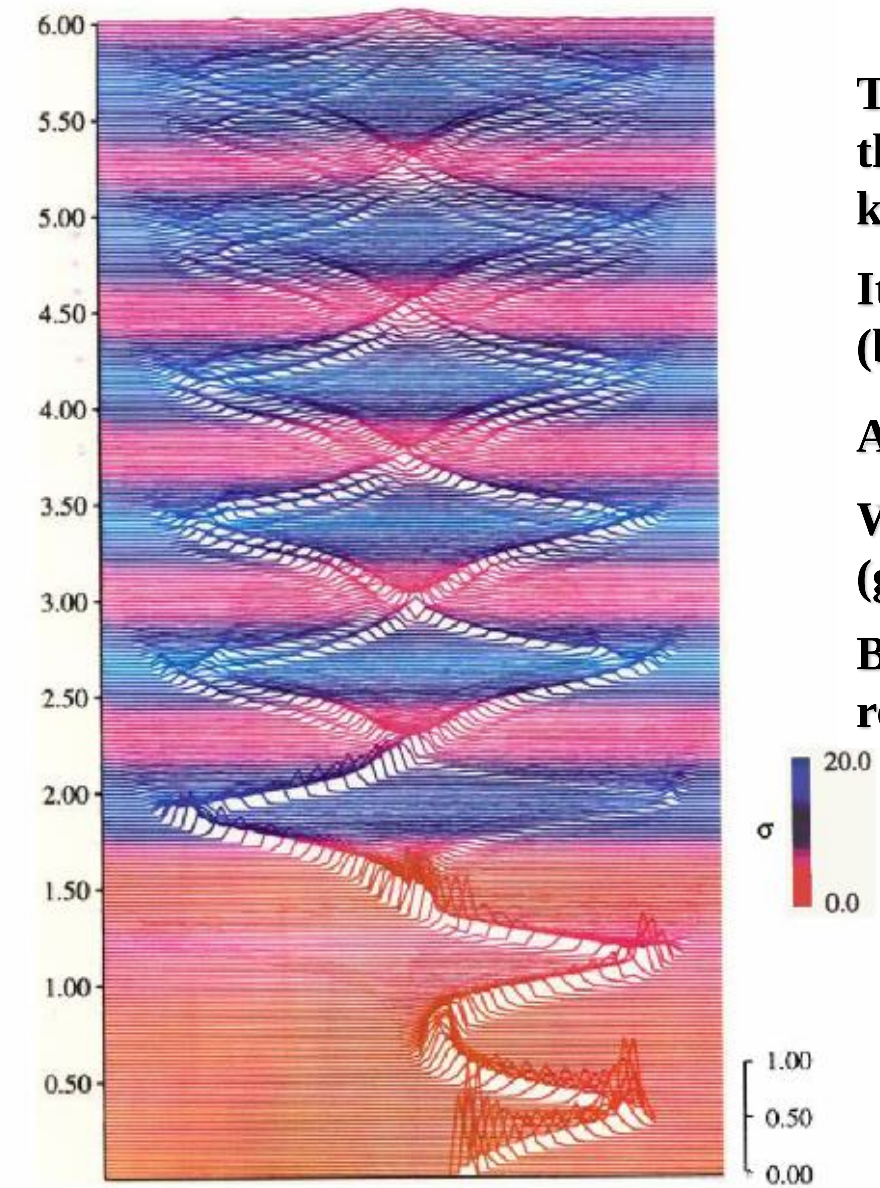
p(x) a forecast probability
q(x) ideal forecast | information
μ(x) “prior” or natural measure

Smith, L. A. (1996) Accountability and Error in Ensemble Forecasting. In 1995 ECMWF Seminar on Predictability. Vol 1, pg 351-368. ECMWF, Reading

we will return to this point.

Smith (2002) Chaos and Predictability in Encyc Atmos Sci

Probability Forecasts: Chaos



Smith (2002) Chaos and Predictability in *Encyc Atmos Sci*

The evolution of this probability distribution for the chaotic Lorenz 1963 system tells us all we can know of the future, given what we know now.

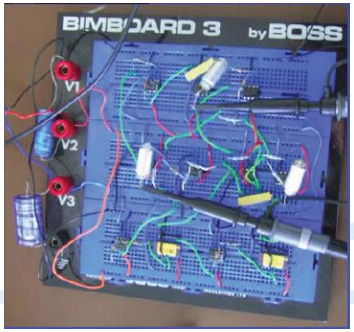
It allows prudent quantitative risk management (by brain-dead risk managers)

And sensible resource allocation.

We can manage uncertainty for chaotic systems (given a perfect model).

But how well do we manage uncertainty in the real world? For GDP? Weather? Climate?

Do we have a single example of a nontrivial system where anyone has succeeded (and willing to offer odds given their model-based PDFs?)



Ocean Modelling

Structural biases affecting the AMOC?

Danny Williamson, Adam Blaker

March 25, 2013

When quantifying the uncertainty in model based projections of the AMOC or any climate process it is important to identify and quantify the uncertainty introduced by structural deficiencies in the climate model. At a

Ocean Modelling

INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2006; **50**:1159–1173
Published online 6 December 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fld.1136

Forward-in-time upwind-weighted methods in ocean modelling

Matthew W. Hecht*,†

Los Alamos National Laboratory, Computers and Computational Sciences Division, MS B296, Los Alamos, NM 87545, U.S.A.

The motivation given by Gerdes *et al.* for using FCT was the concern over unphysical extrema in the tracer fields, identified in the Gulf of Guinea. Similar concerns with spurious tracer extrema have since motivated its use in many modelling studies, and FCT remains a supported option in newer versions of the model code [6]. Concerns with unphysical extrema, oscillatory behaviour, temporal accuracy and data structure have motivated the development and subsequent use of other consistently forward-in-time upwind-weighted methods for tracer transport. A number of problems in ocean modelling have been addressed with

Parameters Estimation via Forecasting P(x)

How might we use probability forecasting to estimate parameters?

a) Form a series of ICE ensembles for a given parameter value
b) Compute a series of probability forecasts
c) Select a proper score: -log(p(x)/μ(x)) (IJ Good, 1952)
d) Compute the score as a function of parameter value.

H Du and L A Smith (2012) Parameter estimation through ignorance *Physical Review E 86, 016213*

Physical Probability and My Subjective Probability

Whether or not physical probability actually exists, it is often convenient to speak as if did. *I.J. Good*

Logical Probability/Credibility: "Hypothetical subjective probability when you are perfectly rational and infinitely large. (?requires a perfect model?)

"Logical probabilities are liable to be unknown in practice"

I.J. Good p74

"Whether or not physical probability is regarded as distinct from …., it is often convenient to talk as if it were distinct."

I.J. Good p66

Physical probability automatically obeys axioms, subjective probability depends on axioms, and psychological probability neither obeys axioms nor depends very much on them." *I.J. Good p74*

My Subjective Probability is then my best attempt to estimate "the" Subjective Probability, just as Laplace saw astronomy related to his Demon.

The thinking of Doogian Bayesians is very close to that of Physicists.

Types of probability (after I.J. Good)

Physical Probability: this is the actual probability of the outcome.

P(BS) zero

Subjective Probability : a(n IJ) Good Bayesian Probability
One Laplace's Demon's Apprentice or a Rational "Org" would strive for.
An accountable ensemble and an actionable probability. **P(BS) small**

Dynamic (Evolving) Probability: This is a probability that is expect to change without any additional empirical information, as when a chess playing computer is stopped early, or only half of your ensemble has run.

Mature Probability: A mature probability encapsulates all the information implied by your knowledge, more compute power is not expected to make an unexpected different. **P(BS) small**

If your model is computationally constrained and you would expect a significant change in the PDF given a different model on a bigger machine, then your probability is immature. **P(BS) required!**

Targets of Probabilistic Forecasting:

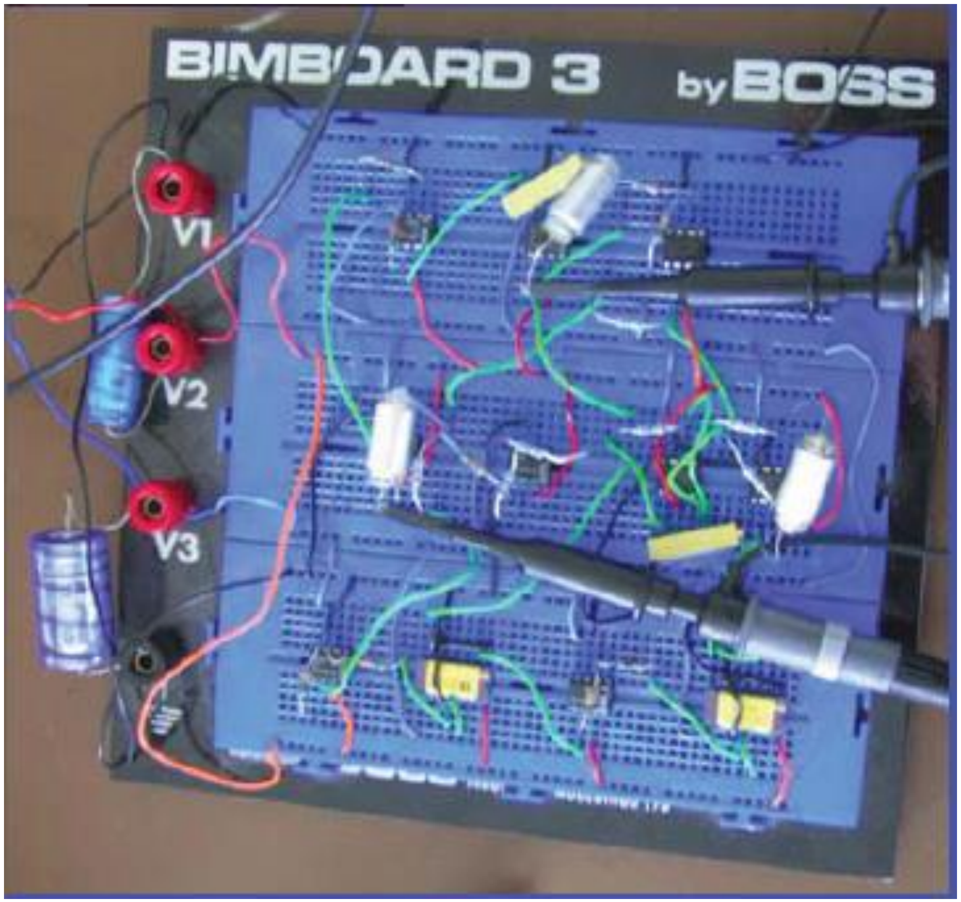
- the past : LHR temp on 14 July 1779
- the next model run: Climate sensitivity of the next cpdn run to come home
- the future: LHR temp on 14 July 2013
- a Real*8 number: Best available parameter value for a given forecast target

Our failure to clearly distinguish the probability distribution of temperature-in-the-model, vs temperature as measured on a thermometer causes serious confusion amongst decision makers.

Is uncertainty in a past value of the same type as uncertainty in a future value? Certainly the approaches to reducing it can be vastly different, as can the elements contributing it.

The Circuit and Ensemble Size

Is chaos the dominant uncertainty in practice?



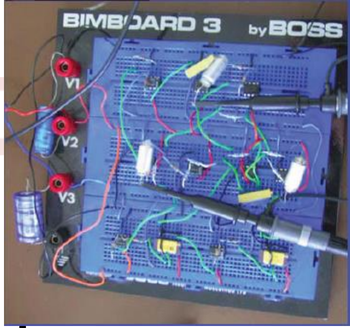
There is a long standing claim in meteorology that going to ensembles larger than ~16 adds nothing tangible to the accuracy for the forecast.

Consider a house that offers odds based on a 16-member forecast, and a player who Kelly bets based on a larger ensemble…

T

te.

Take home questions (in the paper)



How might we better communicate model diversity given the possibility that we cannot get probabilities useful as such!

Do we have a single example of a nontrivial system where anyone has succeeded (and willing to bet on their model-based probabilities?)

At what lead times do inadequacies drive (or fail to drive) feedbacks yielding local impacts? extremes? global impacts?
How far to one go with a simulation model (when to stop: in time? space?)

How can we best deal with models behaving badly?

What prevents the provision of Prob(Big Surprise) with lead time?

How can we improve the communication of insights from simulations without falling afoul of forecasting good practice?

How to distinguish the value of improvement from the utility of prediction?

Might the provision of probability be maladaptive?

How might we better communicate the inadequacy as well as imprecision

Is the value of qualitative insight at risk of being discarded in favour of quantitative mis-information?

Forecast Breakdowns, Estrangement & Shadowing

circuit

i) model quality
ii) model depth

Predictability of the Zeroth kind: Shadowing (after the fact)
On time scales the model cannot phi shadow, no operational justification to run it in terms of decision-relevant probabilistic forecasts.
-> experimental design

ii) model depth (relevance of being "based on the laws of physics" to skill)
Comparison with Simple Surrogate models. (emma)

That is not to say a model more in accord with the laws of physics will not ultimately provide the ideal forecast, merely that the models in hand can lay no claim to doing so!

Aims and Discussion given Mature (Subjective) Probability

Deduce IGN: Additional Information (Obs, DA, N_ens, parameters, Multi-model)
Report Info Deficit,
Report FDE
Identify State Dependent Model skill?
Report Shadowing time Distribution
Determine/Research RDU

?Ikeda Figure?

Communicating the Relevant Dominate Uncertainty

No scientist is admired for failing in the attempt to solve problems that lie beyond his competence."
P.D. Medawar

Good science can significantly improve the science in a model without decreasing Prob(BS)

Following Medawar's advice, scientists typically avoid the intractable parts of a problem, even when uncertainties there dominate the overall uncertainty of the simulation.

Clarifying the uncertainty most relevant to the decision maker, in terms of dominating the uncertainty in the outcome whether, modelled or not, would aid the use of projections in decision support.

Alternatives better than the probability of a big surprise would be welcome.

Proper Scores for formation and evaluation

A score S(p(x), X) is proper if, for any two probability densities p(x) and q(x):

$$\int S(p(x), z)q(z)\mathrm{d}z \geq \int S(q(x), z)q(z)\mathrm{d}z. \tag{8}$$

In words: the minimum of the left hand side over all possible choices of $p(x)$ is obtained if $p(x) = q(x)$ for all x. A score is strictly proper if this happens *only* if $p(x) = q(x)$ for all x.

So the expected score will be a minimum when the verification is drawn from the forecast distribution being evaluated.

This does not imply there is a "true" density function, nor that the forecaster is human (and so might "hedge" her forecast).

(How might a parameter estimation algorithm "hedge"?)

J Bröcker, LA Smith (2007) Scoring Probabilistic Forecasts: On the Importance of Being Proper
Weather and Forecasting **22 (2), 382-388**

Proper Scores for formation and evaluation

$$\text{IGN} = -\log(p(X))$$ Good(1952)

$$S(p(x), X) = \int p(z)^2\, dz - 2\, p(X)$$ Proper linear score

Ignorance and the proper linear score are proper scores, but the proper linear score is not local.

Within PMS, proper scores agree that the true parameters are "best."

If the model structure is incorrect, proper scores need not agree on the optimal parameter value.

J Bröcker, LA Smith (2007) Scoring Probabilistic Forecasts: On the Importance of Being Proper
Weather and Forecasting **22 (2), 382-388**

Local Scores and Distant Scores

$$IGN = -\log(p(X))$$ Good(1952)

Local Scores: Local scores depend only on the value of p(z) at z = X.

IGN is the only proper local score for continuous variables.

$$S(p(x), X) = \int p(z)^2\,dz - 2\,p(X)$$

Distant Scores: The proper linear score is distant in that the score depends on the structure of p(z) far from the outcome x.

All proper polynomial scores are distant: the score includes a term that rewards the forecaster for the shape of the distribution independently of p(x).

J Bröcker, LA Smith (2007) Scoring Probabilistic Forecasts: On the Importance of Being Proper *Weather and Forecasting* 22 (2), 382-388

Evaluating Probability Forecasts

In practise, the vast majority of simulation based probability forecasts target a(n effectively) continuous variable (the temperature at LHR, wind speed at ORD, wave height at Dover, GDP of Great Britain …)

Approaches to forecast evaluation (also called "verification") include:

a) First reducing a PDF to a scalar value, and evaluating that number.
b) Reducing the target variable to a binary variable via a threshold
c) Casting to target into terciles, quartiles, … (hereafter few-tiles).
d) Evaluation of the continuous target or fine-grained proxy.

In forecast evaluation, it is critical to separate the role of the context (scalar, threshold, few-tile, continuous) from the role of the score. (Each should be justified prior to the analysis.)

Weather in Practice: Demonstrating true skill out-of-sample

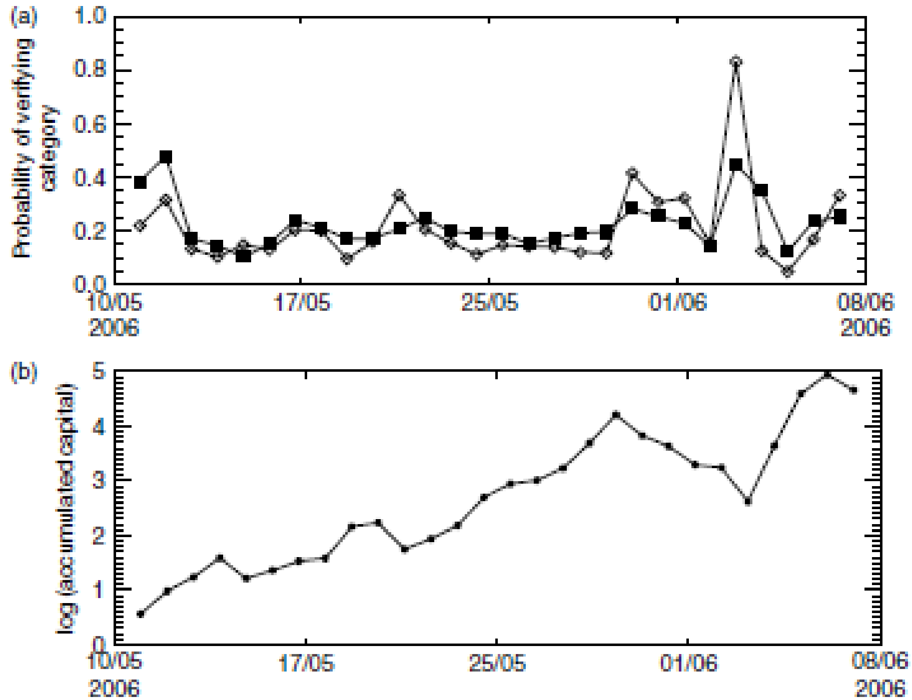


Figure 5. Illustration of the results of Weather Roulette for predicting with a 10-day lead time in which quintile-category the 2 m temperatures at London-Heathrow will fall during the period 11 May–7 June 2006. (a) Probabilities of the verifying categories as predicted by the dressed HRES forecast (open diamonds) and the dressed EPS forecasts (filled squares); (b) logarithm of the accumulated capital when playing the dressed EPS against the HRES forecast under the 'fully proper' variant.

Skill, in bits or interest rates, of ensemble forecasting LHR temperatures.
Similar results are found in the 3 years after this paper was published.

R Hagedorn and LA Smith (2009) Communicating the value of probabilistic forecasts with weather roulette. *Meteorological Applications* 16 (2): 143-155.

Take-home Points and Questions

Measures of predictability need to be local, if predictability varies with the state.

It is useful to distinguish mathematical, weather-like and climate-like tasks.

Information content (relative entropy) of the forecast is a good measure of predictability in weather-like tasks.

The information deficit is a useful statistic for understanding the fidelity of probability forecasts.

How would you approach resource allocation to improve operational forecasting?

Real-world probability forecasts consistently fail consistency tests.

Relevant Dominant Uncertainty (RDU):

In Reality (FDE): RDU is chaos
RDU is Model error
RDU is initial distribution

Weather : Blend
Climate: think!

Downside: go bust/over confident over-committed

Resource Allocation

Use the Information Deficit to Improve Ikeda Forecasts

Lastly, we can improve the forecast in a variety of different ways, and use the information deficit to judge which buys the best forecast improvement:

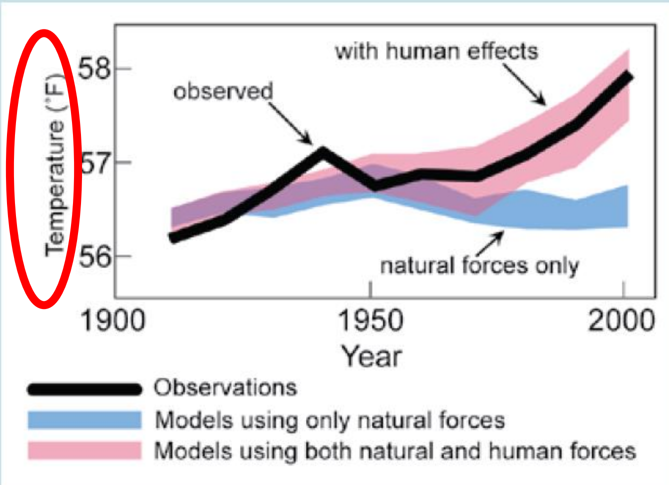
a) Reducing/Increasing the observational noise
b) Increasing the ensemble size
c) Improving the ensemble interpretation scheme
d) Other…

] Smith, L. A. (2002) What Might We Learn from Climate Forecasts? Proc. National Acad. Sci. USA 4 (99): 2487-2492. Abstract
[45] Orrell, D., Smith, L. A., Palmer, T. and Barkmeijer, J. (2001) Model Error in Weather Forecasting, Nonlinear Processes in Geophysics 8: 357-371. Abstract
[44] Hansen, J. A. and Smith, L. A. (2001) Probabilistic Noise Reduction. Tellus 53 A (5): 585-598. Abstract
[43] Gilmour, I., Smith, L. A. and Buizza, R. (2001) Linear Regime Duration: Is 24 Hours a Long Time in Synoptic Weather Forecasting? J. Atmos. Sci. 58 (22): 3525-3539. Abstract
[42] Judd, K. and Smith, L. A. (2001) Indistinguishable States I: The Perfect Model Scenario, Physica D 151: 125-141. Abstract
[41] McSharry, P. E., Ellepola, J. H., von Hardenberg, J., Smith, L. A., Kenning, D. B. R. and Judd, K. (2002) Spatio-temporal Analysis of Nucleate Pool Boiling: Identification of Nucleation Sites using Non-orthogonal Empirical Functions (NEFs). Int. J. Heat & Mass Transfer 45 (2): 237-253. Abstract
[40] Smith, L. A. (2000) 'Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems' (PDF) in Nonlinear Dynamics and Statistics, ed. Alistair I Mees, Boston: Birkhauser, 31-64. Abstract
[39] Hansen, J. A. and Smith, L. A. (2000) The role of Operational Constraints in Selecting Supplementary Observations, J. Atmos. Sci., 57 (17): 2859-2871. Abstract
[38] Smith, L. A. (2000) Limits to Predictability in 2000 and 2100, in Proceedings of IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, ed. S. Haykin (IEEE, Piscataway), pg 129-134. (Figure 1). Abstract
[37] Ziehmann, C., Smith, L. A. and Kurths, J. (2000), Localized Lyapunov Exponents and the Prediction of Predictability, Phys. Lett. A, 271 (4): 237-251. Abstract
[36] McSharry, P. and Smith, L. A. (1999) Better nonlinear models from noisy data: Attractors with maximum likelihood, Phys. Rev. Lett 83 (21): 4285-4288. Abstract
[35] Smith, L. A., Ziehmann, C. and Fraedrich, K. (1999) Uncertainty Dynamics and Predictability in Chaotic Systems, Quart. J. Royal Meteorological Soc. 125: 2855-2886. Abstract
[34] Ziehmann, C., Smith, L. A. and Kurths, J. (1999) The Bootstrap and Lyapunov Exponents in Deterministic Chaos, Physica D, 126 (1-2): 49-59. Abstract
[33] McSharry, P. and Smith, L. A. (1998) Just Do It. Reductionism, Modelling and Black-box Forecasting In International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling: Theory and Applications with Time-Series Prediction Competition, Suykens, J. A. K. and Vandewalle, J. (eds.) pp. 106-111, Leuven, K. U. Belgium, Kluwer Academic Publishers. Abstract
[32] Smith, L. A. and Gilmour, I. (1998) Accountability and internal consistency in ensemble formation. In the Proceedings of the ECMWF Workshop on Predictability, 1997. ECMWF, Reading, UK.
[31] Smith, L. A. (1997) The Maintenance of Uncertainty. In Proc International School of Physics "Enrico Fermi", Course CXXXIII, pg 177-246, Societ'a Italiana di Fisica, Bologna, Italy.

United States Global Change Research Program

IPCC INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE WMO UNEP

Separating Human and Natural Influences on Climate



As the blue band indicates, without human influences, global average temperature would actually have cooled slightly over recent decades. With human influences, it has risen strongly (black line), consistent with expectations from climate models (pink band).

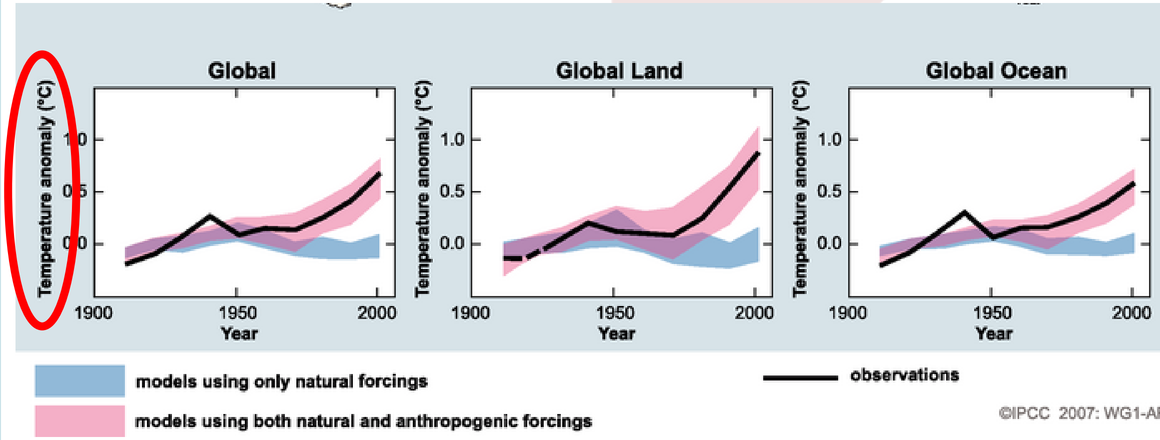


Figure SPM.4. Comparison of observed continental- and global-scale changes in surface temperature with results simulated by climate models using natural and anthropogenic forcings. Decadal averages of observations are shown for the period 1906 to 2005 (black line) plotted against the centre of the decade and relative to the corresponding average for 1901–1950. Lines are dashed where spatial coverage is less than 50%. Blue shaded bands show the 5–95% range for 19 simulations from five climate models using only the natural forcings due to solar activity and volcanoes. Red shaded bands show the 5–95% range for 58 simulations from 14 climate models using both natural and anthropogenic forcings. {FAQ 9.2 Figure 1}

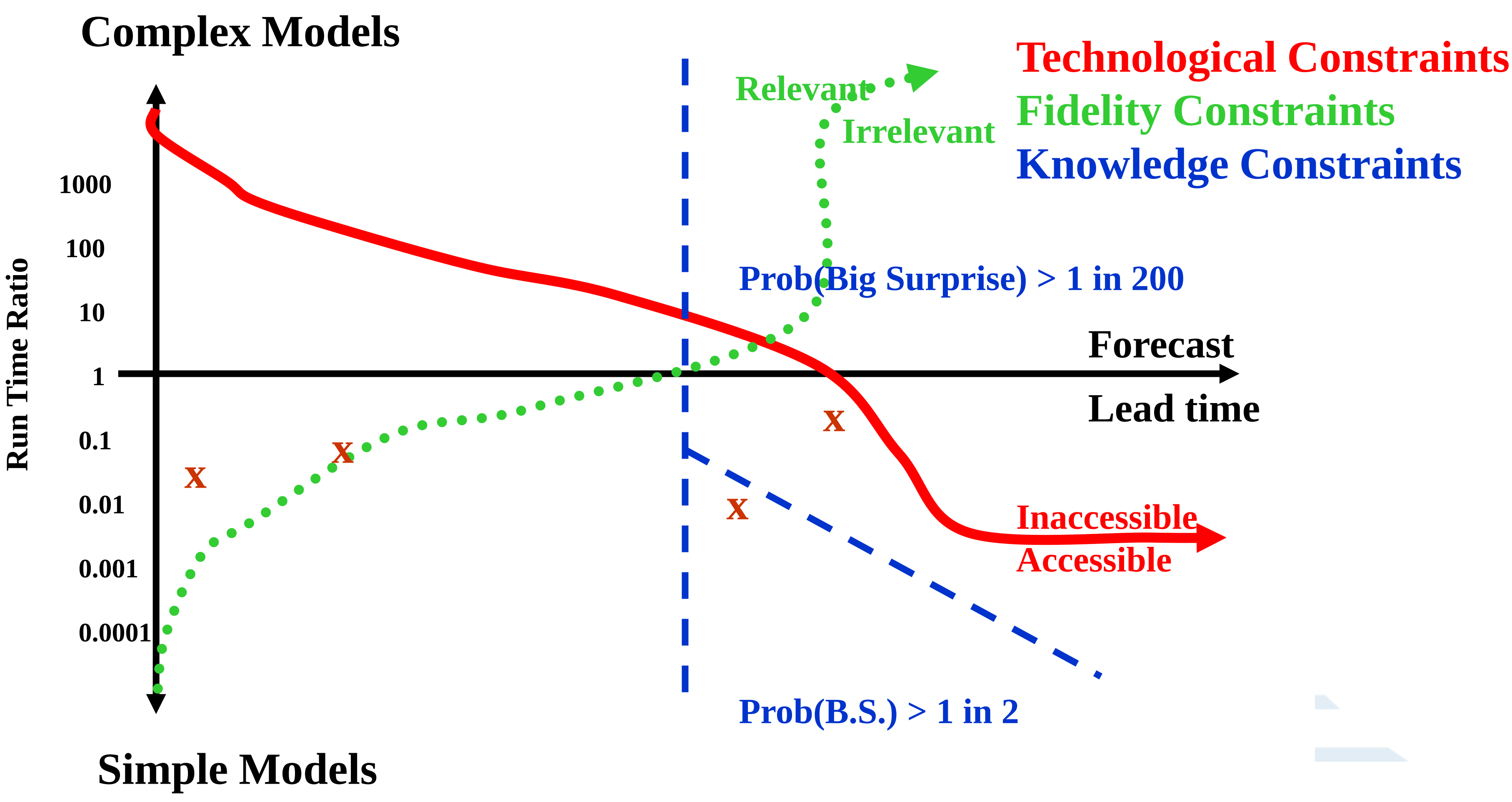
http://www.ipcc.ch/publications_and_data/ar4/wg1/en/figure-spm-4.html

http://www.globalchange.gov/images/cir/pdf/20page-highlights-brochure.pdf

Statistical post-processing: These are anomalies, not temperatures. Parameterization of cloud formation is a bit of a distraction when we are missing two kilometre tall walls of rock…

Where have we designed operational models?

My subjective view of operational weather (< 10 days), seasonal (< 18 months), and hires Climate (< 80 years) models each fall.



Leonard Smith

Types of Probability (Forecasts)

(o) Tautological Probability. A probability P(E|H) the value of which is specified in the definition of H. ("a fair coin", H is called "a simple statistical hypothesis")

(i) Physical Probability: P(x) "True probability" (Laplace's Demon/Inf Rat Org)

(ii) Psychological Probability: "Personal probability inferred from one's behaviour."

(iii) Subjective Probability: P(x|G) probability of x given our information G is true (Demon's Apprentice/?semi-finite Rational Org?)

(iv) Logical Probability: "Hypothetical subjective probability when you are perfectly rational and infinitely large . "Credibility" Russell (1948) (Infinitely large Rational Org; Laplace Demon ?or Apprentice?)

(v) Dynamic Probability: $P_t(x| g_t<G)$ when an algorithm encapsulating G has not yet terminated (finite algorithm, merely still running).
Dynamic in the sense that this probability is expected to change without any empirical information (by reflection only)

(vi) (Im)Mature Probability: $P(x| g<G)$ when G is known (not) to be encapsulated in g.
Immature in that this probability is expected to change without addition reflection or additional empirical observation even after the algorithm finishes.

Rational Decisions *I. J. Good (1952) Journal of the Royal Statistical Society. Series B (Methodological) Vol. 14, No. 1 , pp. 107-114*