Addressing Climate-like Questions Scientifically Requires doing Science in the Dark

Geoscientists are not alone

Climate-like Questions are also found in:

Nuclear Stewardship, Novel Engineering Design National Intelligence, Reactor Safety, Americas Cup Design, Vehicle Crash-worthiness, Waste Disposal, (Re)Insurance

Scientists are forced to "violate" traditional best practice guidance if such violations are imposed by the nature of the question being addressed.

One cannot wait 50 years for out-of-sample observations.

It is a brute fact that a climate model's lifetime is less than it's forecast lead-time!

The physics underlying CO_2 induced warming remains as solid as science gets.

Other groups working in the dark sometimes embrace model inadequacy more than we do, and speak much more tentatively.

LSE CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

www.lsecats.ac.uk

Clarity before Consensus

Good Karma (Climate) Science Communication

Can Climate Scientists Play Nice(r) in Public?

"In model-land, no one can hear you scream"

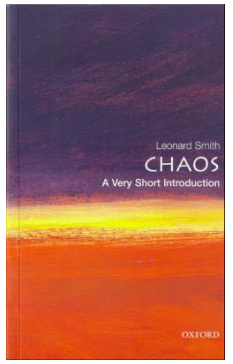
What would Judy Curry Say?

Do we agree on more than we agree on?

Thinking it might be 5 need not mean you're stupid

Leonard A. Smith

London School of Economics &
Pembroke College, Oxford



Possible only due to work with CATS
Erica Thompson, H Du. Ana Lopez
Dave Stainforth, and Ed Wheatcroft …

with special thanks to **Jim Berger, Dave Higdon**
Gavin S, Judy C, Liz M, Ray P and Reto K

ENSEMBLES

EQUIP

Wouldn't it be nice if…

…we could move from this

(who do you want to be in this picture?)

towards this

even if it is less fun to watch?

http://fc07.deviantart.net/fs71/i/2012/175/4/2/aliens_wars_by_byzko_wader-d54q3rp.jpg
http://www.wallpapervortex.com/wallpaper-15262_aliens_xenomorph.html#.VEtqJBbYcuc
http://theinspirationroom.com/daily/print/2009/2/aliens_vs_predator_chess.jpg

I want to discuss how we communicate, internally & externally.

While I liked Gavin's AGU lecture very much, I wish not to advocate.

Science to Motivate | Science to Inform

Clear Facts (Anomalies); Lowering the Bar (Science in the DARK)

"I've got, that sinking feeling" (think: Righteous Brothers)
(or the theme from Alien)

I want to discuss how we communicate, internally & externally.

While I liked Gavin's AGU lecture very much, I wish not to advocate.

Science to Motivate | Science to Inform

Clear Facts (Anomalies); Lowering the Bar (Science in the DARK)

"I've got, that sinking feeling" (think: Righteous Brothers)
(or the theme from Alien)

IP_{CC} and the Shirleys
Surely is doesn't matter if…
Perhaps not, but clearer Experimental Design is basic good practice.

"My Model is not Purple"
And Gavin is an honourable man, we needn't argue over maths questions that are not truly relevant!

I want to discuss how we communicate, internally & externally.

While I liked Gavin's AGU lecture very much, I wish not to advocate.

Science to Motivate | Science to Inform

Clear Facts (Anomalies); Lowering the Bar (Science in the DARK)

"I've got, that sinking feeling" (think: Righteous Brothers)
(or the theme from Alien)

IP_{CC} and the Shirleys
Surely is doesn't matter if…
Perhaps not, but clearer Experimental Design is basic good practice.

"My Model is not Purple"
And Gavin is an honourable man, ?we need to argue closely over maths questions that are truly relevant?

Come on, tell us what you really think:
Clarity over/before consensus.
?Agreement? if we avoid talking past each other/point scoring.

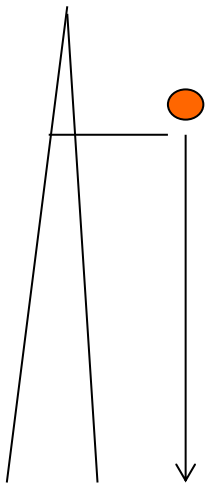
Raising our game (AAAS/UKCP09/Climate Day on Capital Hill)

Some facing climate-like science challenges seek out model inadequacy

In climate-like tasks the burden of proof lies with the analyst. The importance of credibility and trust is closely guarded by others with climate-like tasks; some search out ways to explore model inadequacy.

How accurately do you think we could predict the drop time?

Perhaps we could learn by doing simple experiments, like dropping a ball from a tower?



What is the real drop-time?

(Many thanks to Dave Higdon.)

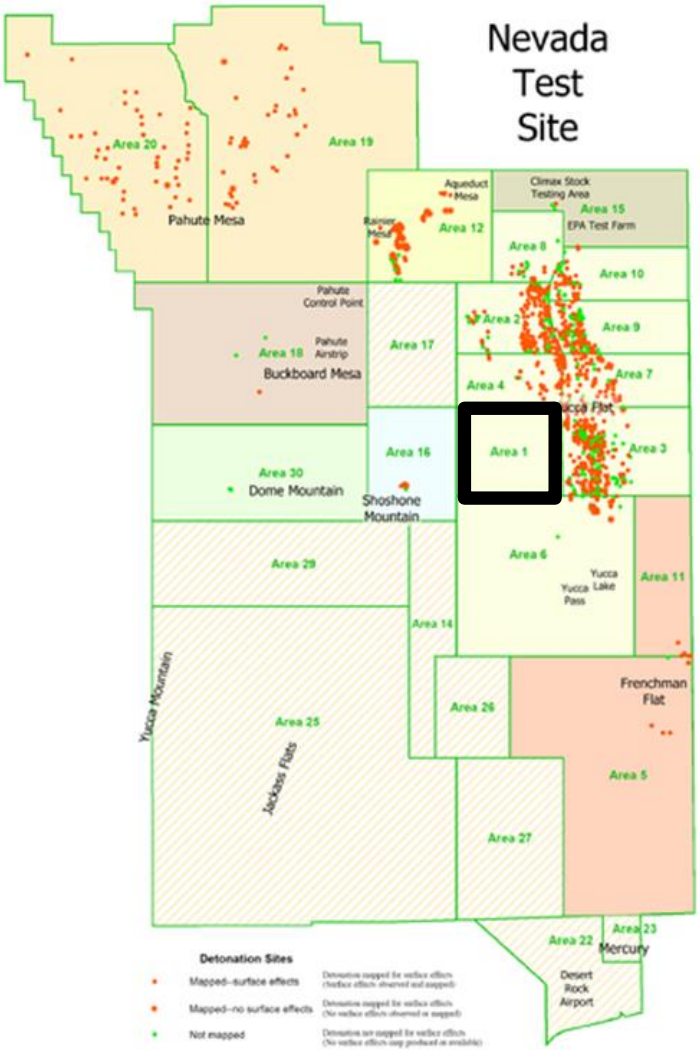
Chasing Model Inadequacy

1000 ft Ua1 shaft

Ball(s)

2 bowling balls
3 basketballs
2 golf balls
3 wiffle balls
… (no duck)

detectors



http://www2.nstec.com/Documents/Fact%20Sheets/U1a%20Facility.pdf

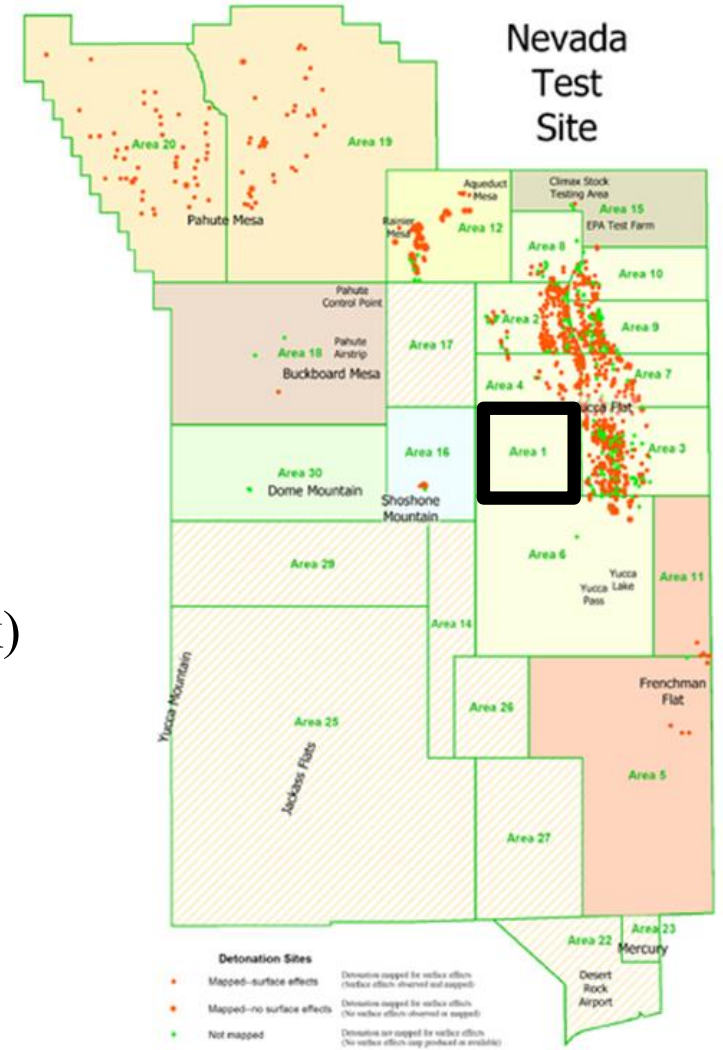
Chasing Model Inadequacy

1000 ft Ua1 shaft

Ball(s)

2 bowling balls
3 Basketballs
2 golf balls
3 Wiffle balls
… (no rubber duck)

…ors



http://www2.nstec.com/Documents/Fact%20Sheets/U1a%20Facility.pdf

How close was our median time for basketballs?

A. < 0.01 sec

B. <0.1 sec

C. < 1 sec

D. < 1 min

E. > 1 min

Basket ball.
Initial velocity zero.
1000 ft "tower".
Laser sheet timing.

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Q3.1

How close was our median time for basketballs?

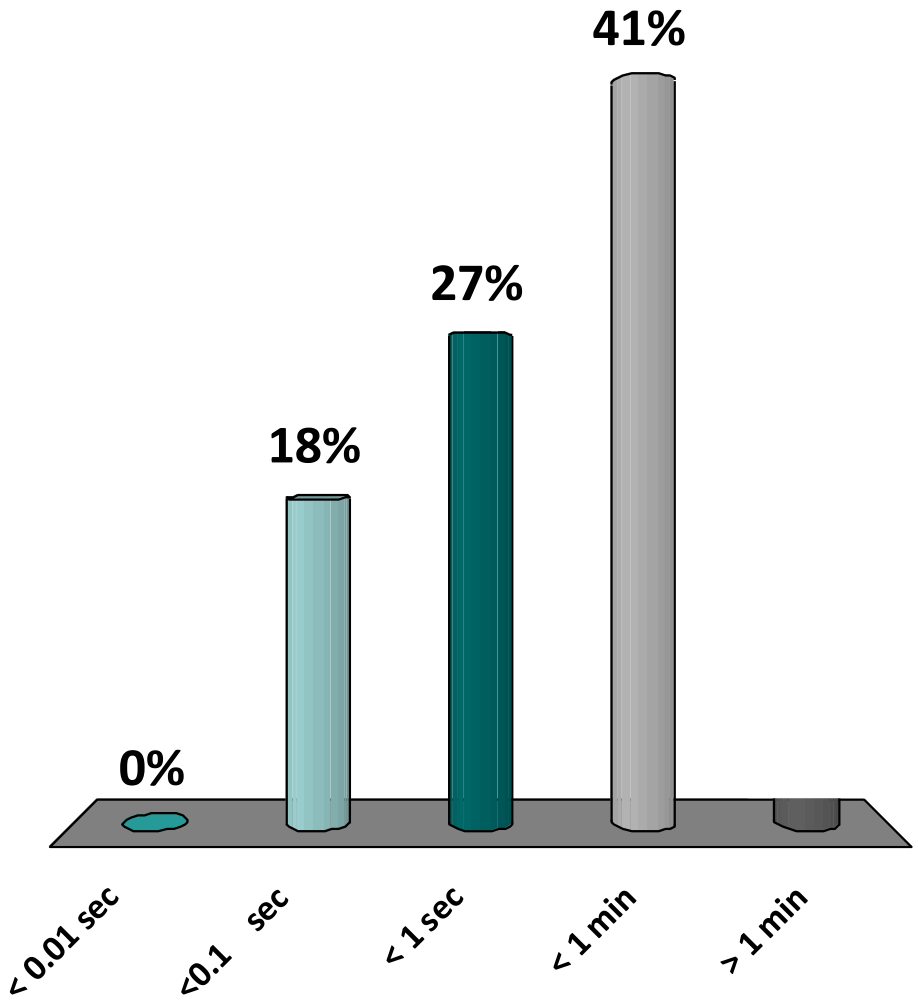
A. < 0.01 sec

B. <0.1 sec

C. < 1 sec

D. < 1 min

E. > 1 min



Basket ball.
Initial velocity zero.
1000 ft "tower".
Laser sheet timing.

Q3.1

How close was our median time for basketballs?

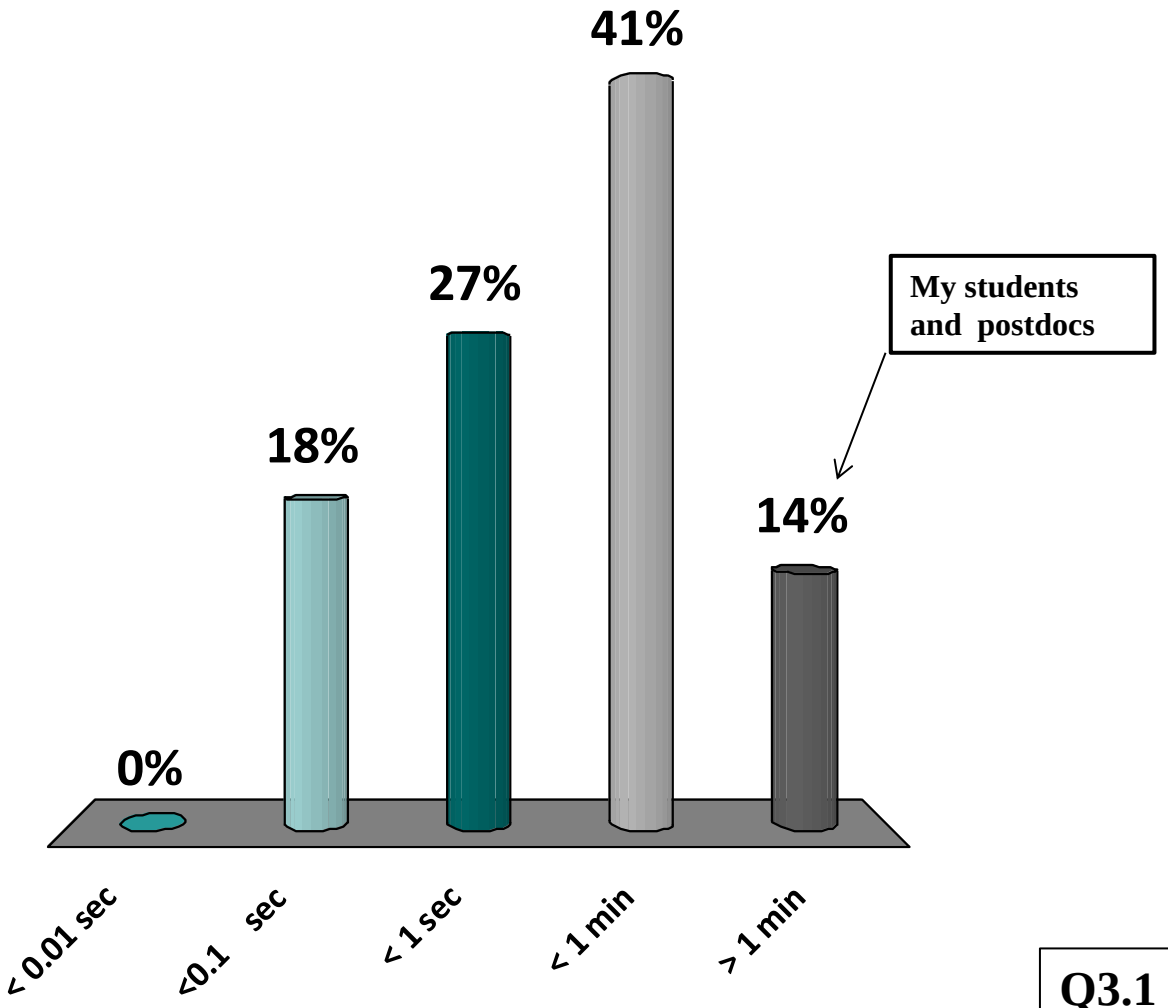
A. < 0.01 sec

B. <0.1 sec

C. < 1 sec

D. < 1 min

E. > 1 min



Basket ball.
Initial velocity zero.
1000 ft "tower".
Laser sheet timing.

Q3.1

So: How accurate were our drop time estimates?

I only have preliminary results, but I think they make my point rather better than any data I could have made up:

"The bowling ball was completely destroyed."

"One of the basket balls failed to make it to the bottom."

A "Big Surprise" is when something your model doesn't reflect is important: We thought surface roughness was the main Unknown.

Sometimes, scientists can estimate Prob(BS)
(but not within the models, of course)

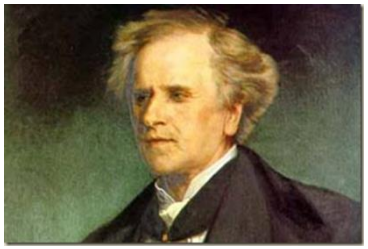
CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

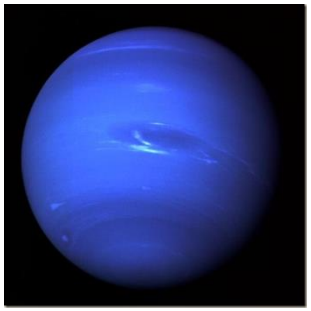
Science education and Not-so-famous Failures

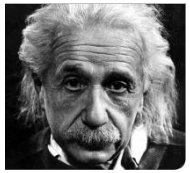
Kelvin's Gambit: "As for the future, we may say, with equal certainty, that inhabitants of the earth can not continue to enjoy the light and heat essential to their life for many million years longer **unless sources now unknown to us** are prepared in the great storehouse of creation."

William Thomson (1862)
Macmillan's Magazine **5** 388

This clarification (conditioning on **I** explicitly) is a source of strength, not of weakness.
We can ask what the Prob(Big Surprise) is thought to be.
And whether or not a probability forecast is thought to be **mature**.







There is no planet Vulcan!
Newton's Laws fail near the sun.

Why do we so rarely communicate the failures of science in early science education?

http://upload.wikimedia.org/wikipedia/commons/3/3d/El_Chich%C3%B3n.jpg
http://simpleclimate.files.wordpress.com/2010/01/pinatubo91_eruption_plume_06-12-91.jpg

Lawrence's "Intellectual Phase Locking"

Birge (1957) wrote that E. O. Lawrence stated that "in any highly precise experimental arrangement there are initially many instrumental difficulties that lead to numerical results far from the accepted value of the quantity being measured. It is, in fact, just such wide divergences that are the best indication of instrumental errors of one kind or another. Accordingly, the investigator searches for the source or sources of such errors, and continues to search until he gets a result close to the accepted value. *Then he stops*" (p. 51).

Psychology of Science: Implicit and Explicit Processes
Robert W. Proctor, E.J. Capaldi

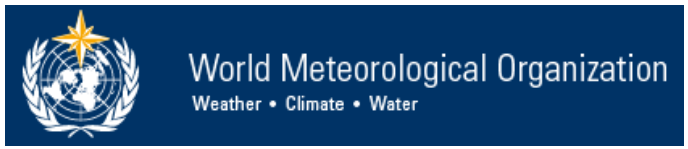
If this is a documented problem for estimating the speed of light, how could it not be for the "climate sensitivity" of a model?

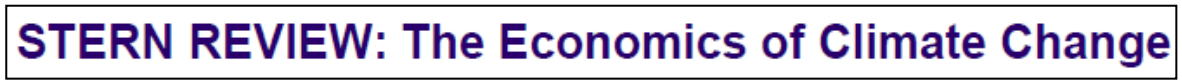
There is no shame in considering this!

Selective Publication Challenges even Out-of-Sample Forecasting

These Challenges range wider than climate

This is a subset of the people I talk to.

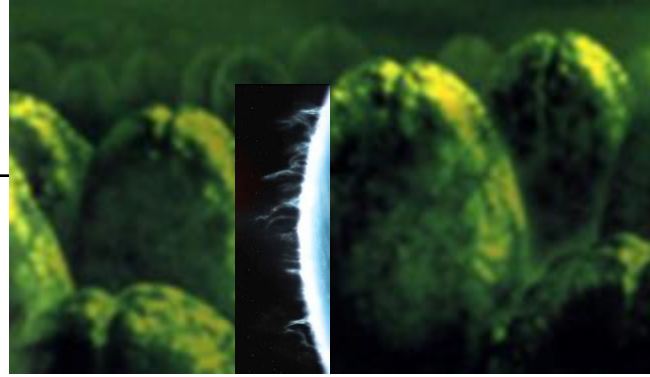




Many embrace discussing Prob(Big Surprise).
In weather-like tasks they run into it in practice.

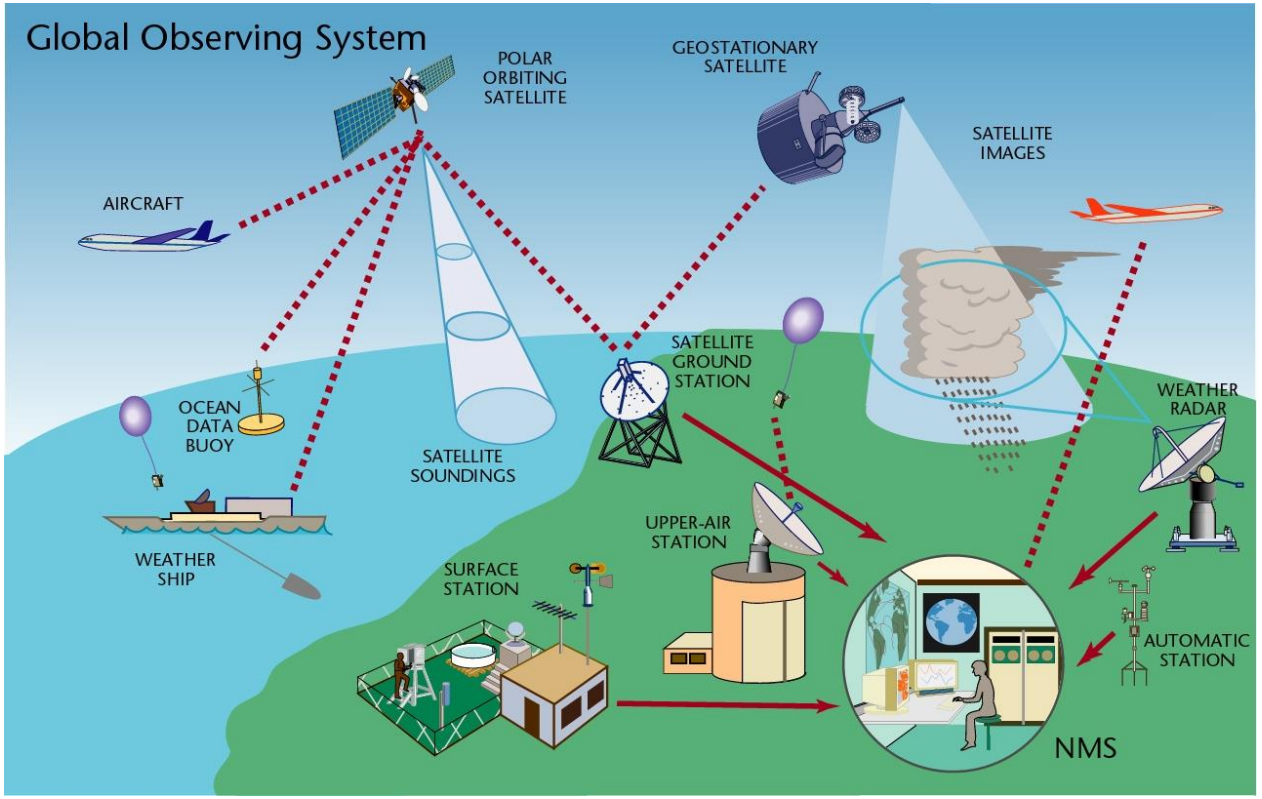
I've got, that sinking feeling…

(think of the Righteous Bothers)



When I see easily misinterpreted schematics which misrepresent (over-sell) our models.

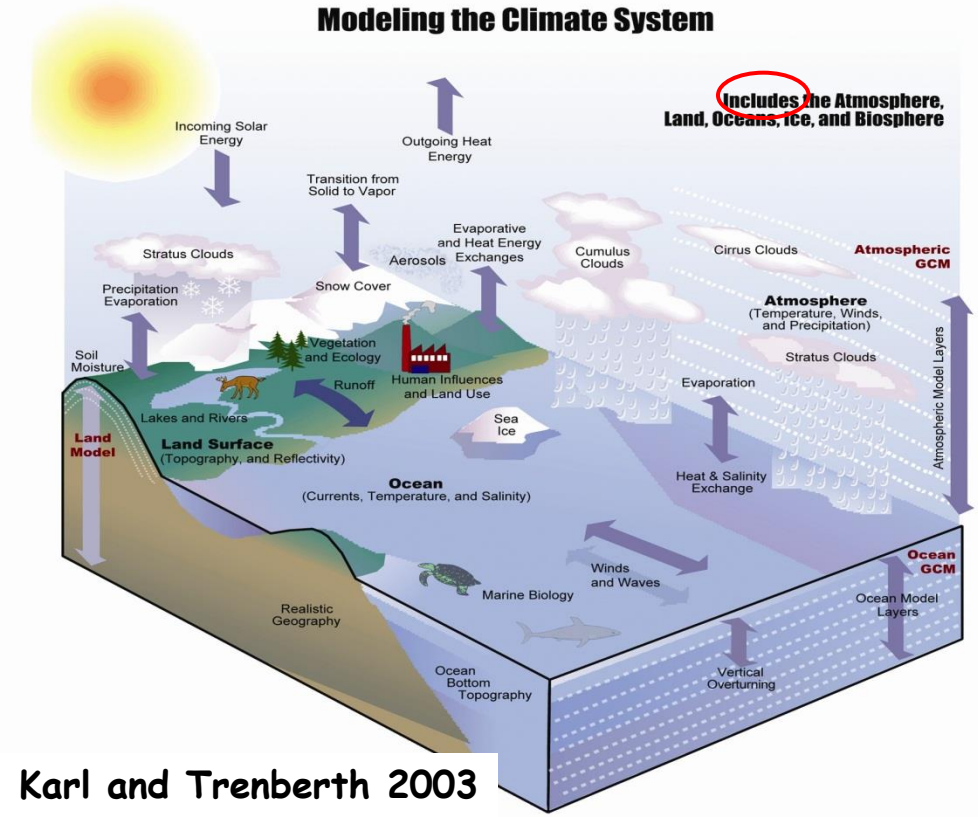
World Weather "Observing System" Schematic



http://blog.metservice.com/wp-content/uploads/2010/03/GOS.jpg

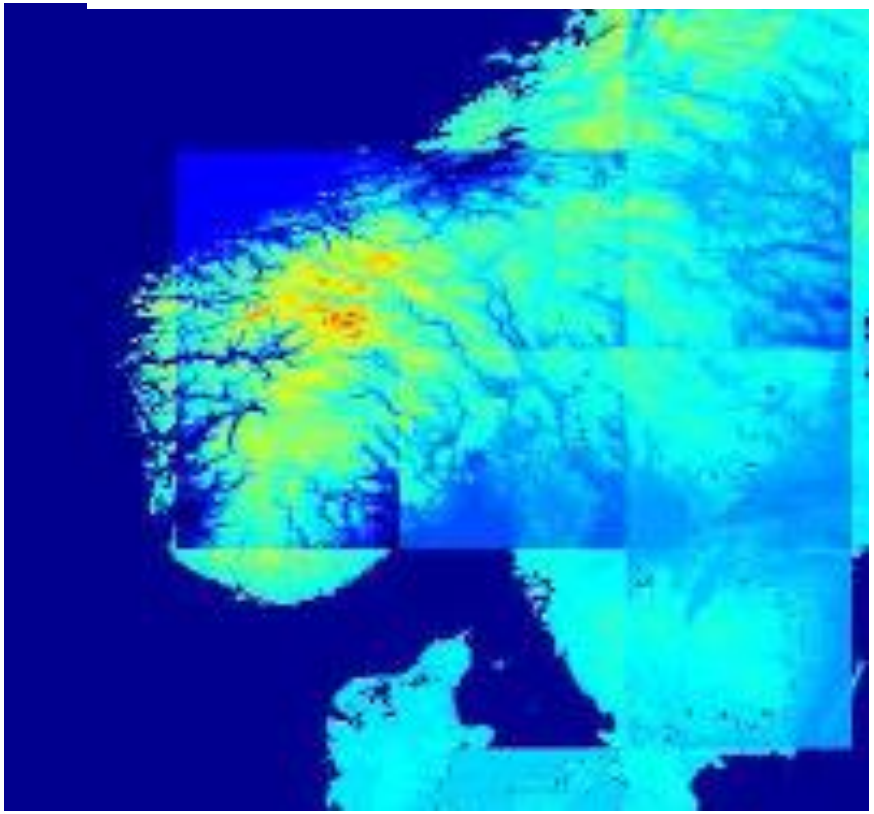
Limits to Transparency: Dangerously schematic schematics

Climate Model Schematic



Karl and Trenberth 2003

Climate Model Points(the squares)
(**What you see is NOT in the model**)



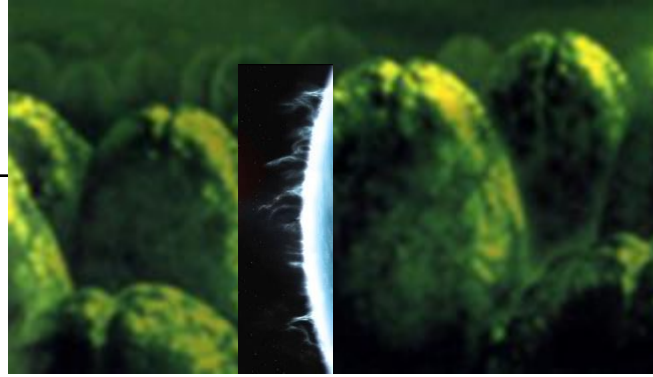
A very schematic schematic reflecting phenomena the model "includes".

"included" vs "realistically simulated"

The detail you see above is what is ***missing*** in HadCM3: the large squares reflect model grid resolution, the detail reflects the difference between the observed surface height and the model surface height, "constant" "within" a grid point.

Insurance Company with a snowfall question…

I've got, that sinking feeling…



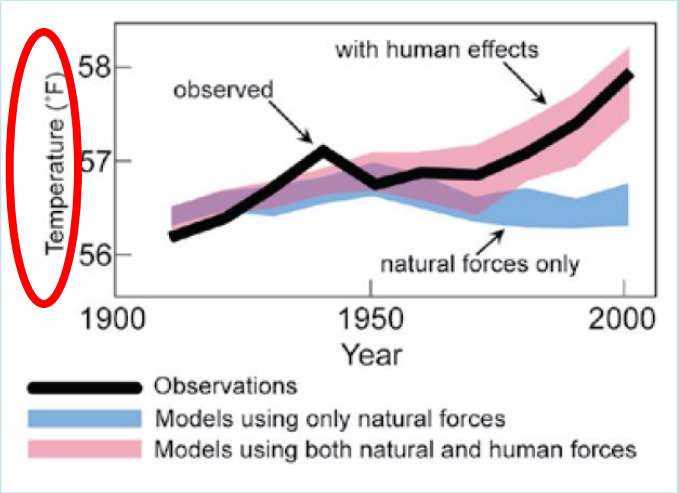
When anomalies are not identified as such; or
the magnitude of systematic errors hidden by taking anomalies is:

1) Hidden
2) Implied to fall within in the observational noise
3) Suggested unimportant in modelling for adaptation

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).

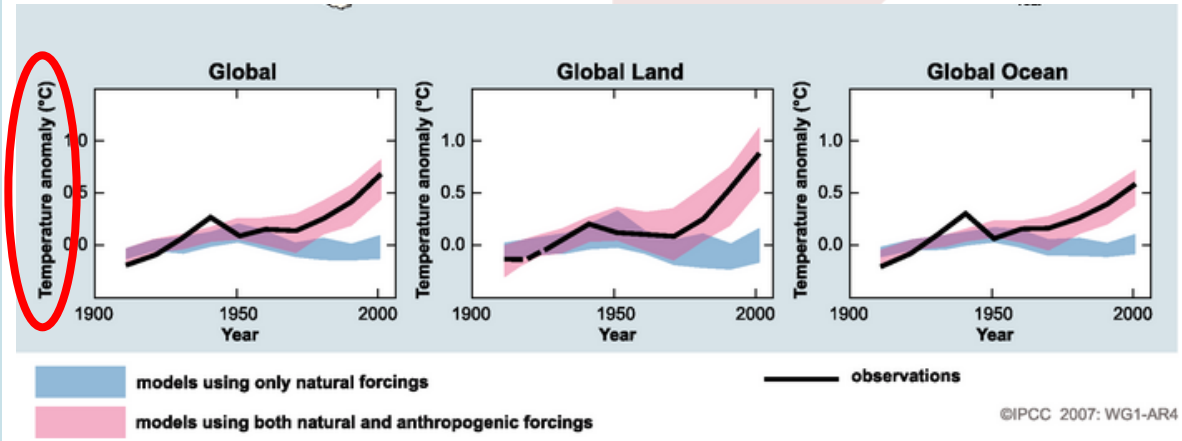


Figure SPM.4. Comparison of observed continental- and global-scale changes in surface temperature with results simulated by climate models using natural and anthropogenic forcings. Decadal averages of observations are shown for the period 1906 to 2005 (black line) plotted against the centre of the decade 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

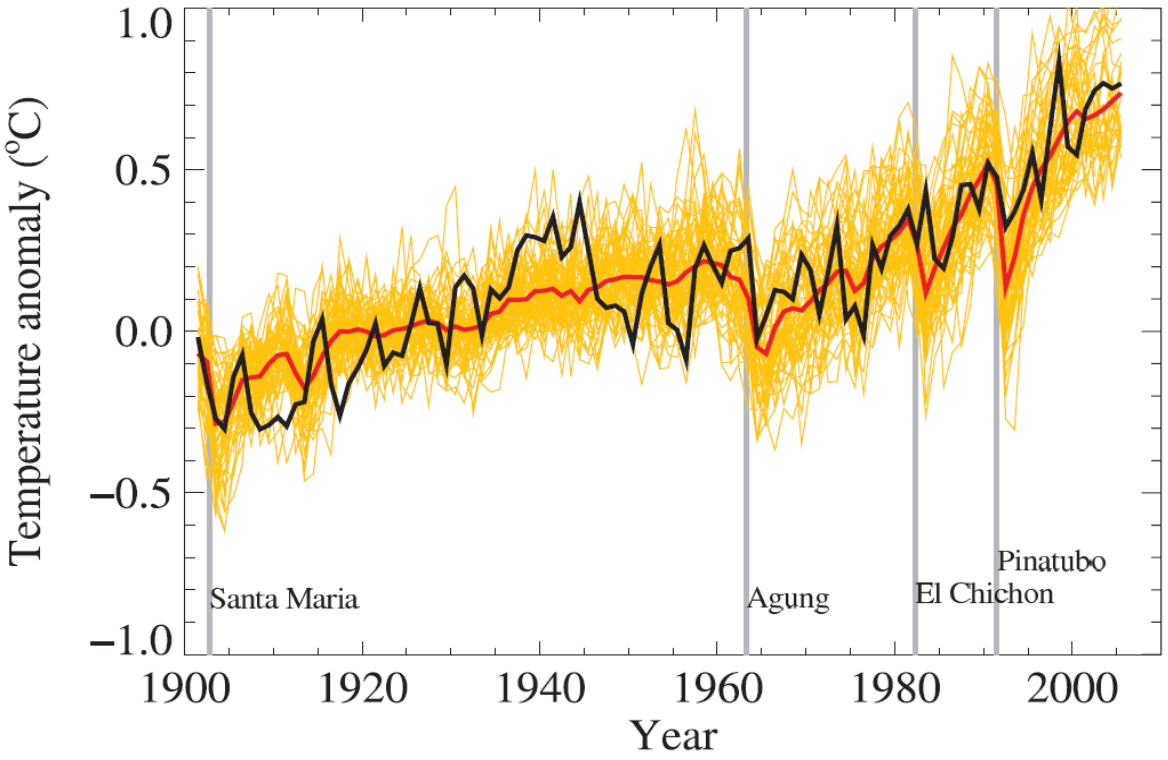
http://www.globalchange.gov/images/cir/pdf/20page-highlights-brochure.pdf

Statistical post-processing: These are anomalies, not temperatures. How can we ignore an range of anomaly corrections wider than what would cause "dangerous climate change"?

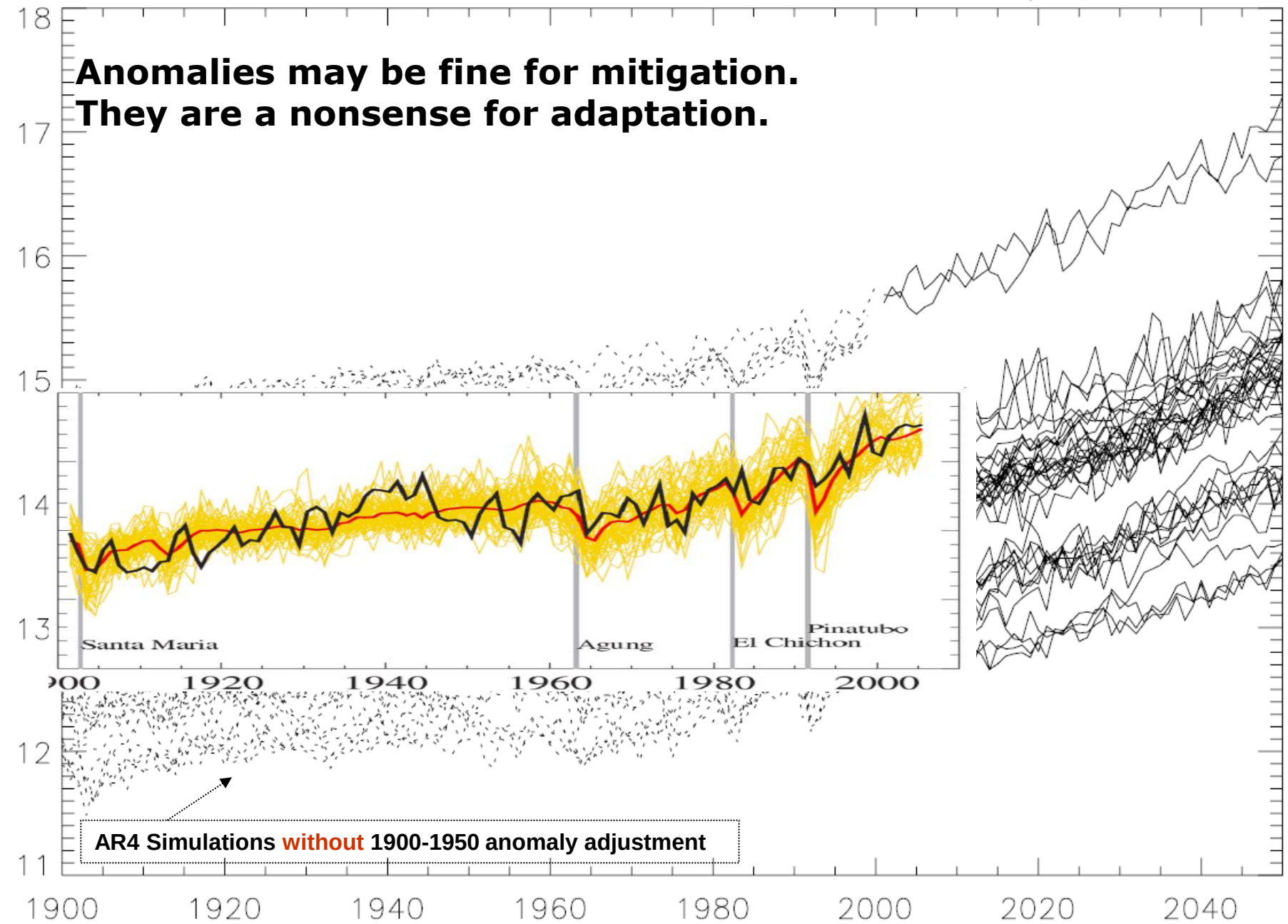
Limited Transparency: Systematic Errors

Models agree that a wide range of sorta-Earthlike planets warm about the same amount under the observed forcing. This is evidence for mitigation policy strategy…

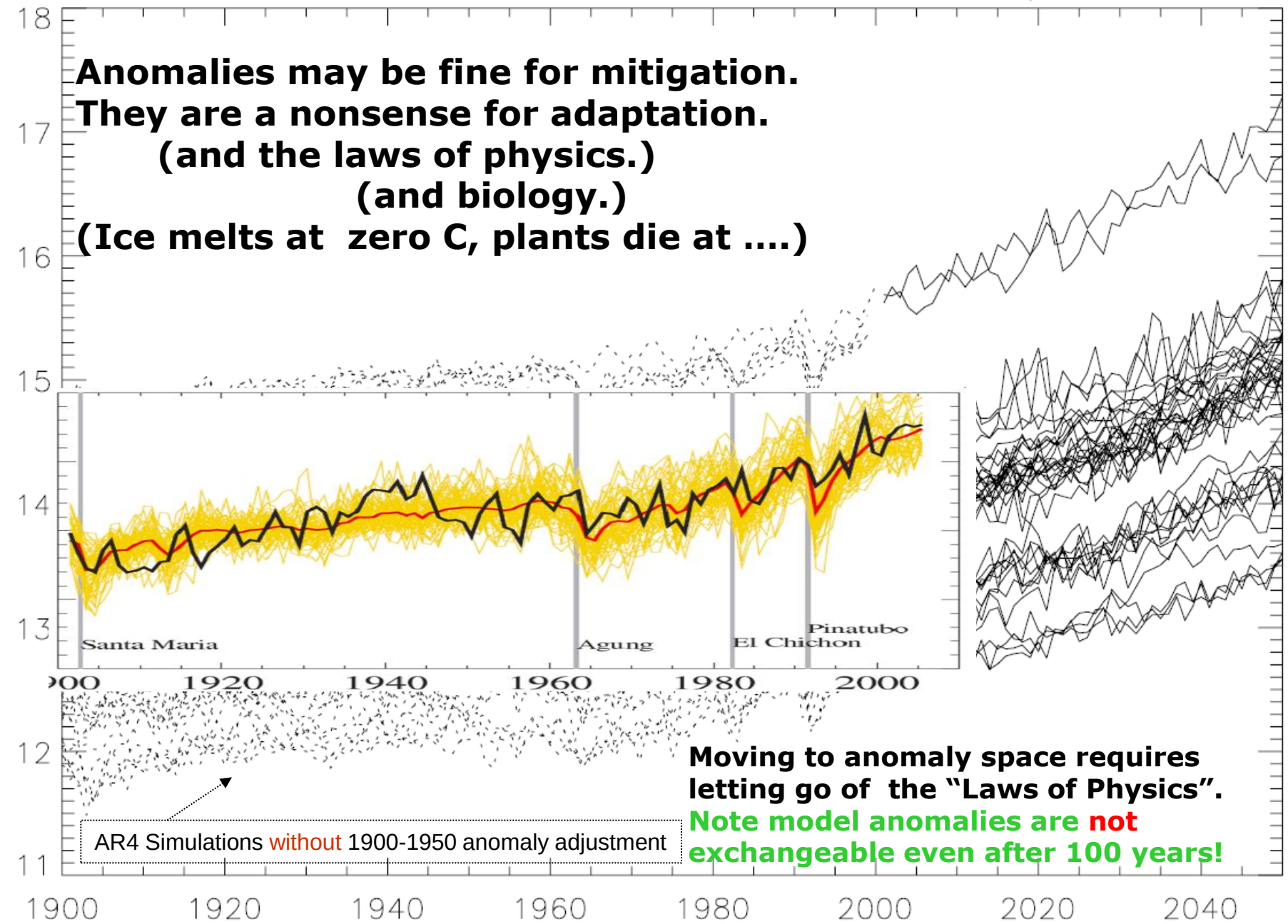
FAQ 8.1, Figure 1. *Global mean near-surface temperatures over the 20th century from observations (black) and as obtained from 58 simulations produced by 14 different climate models driven by both natural and human-caused factors that influence climate (yellow). The mean of all these runs is also shown (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.)*



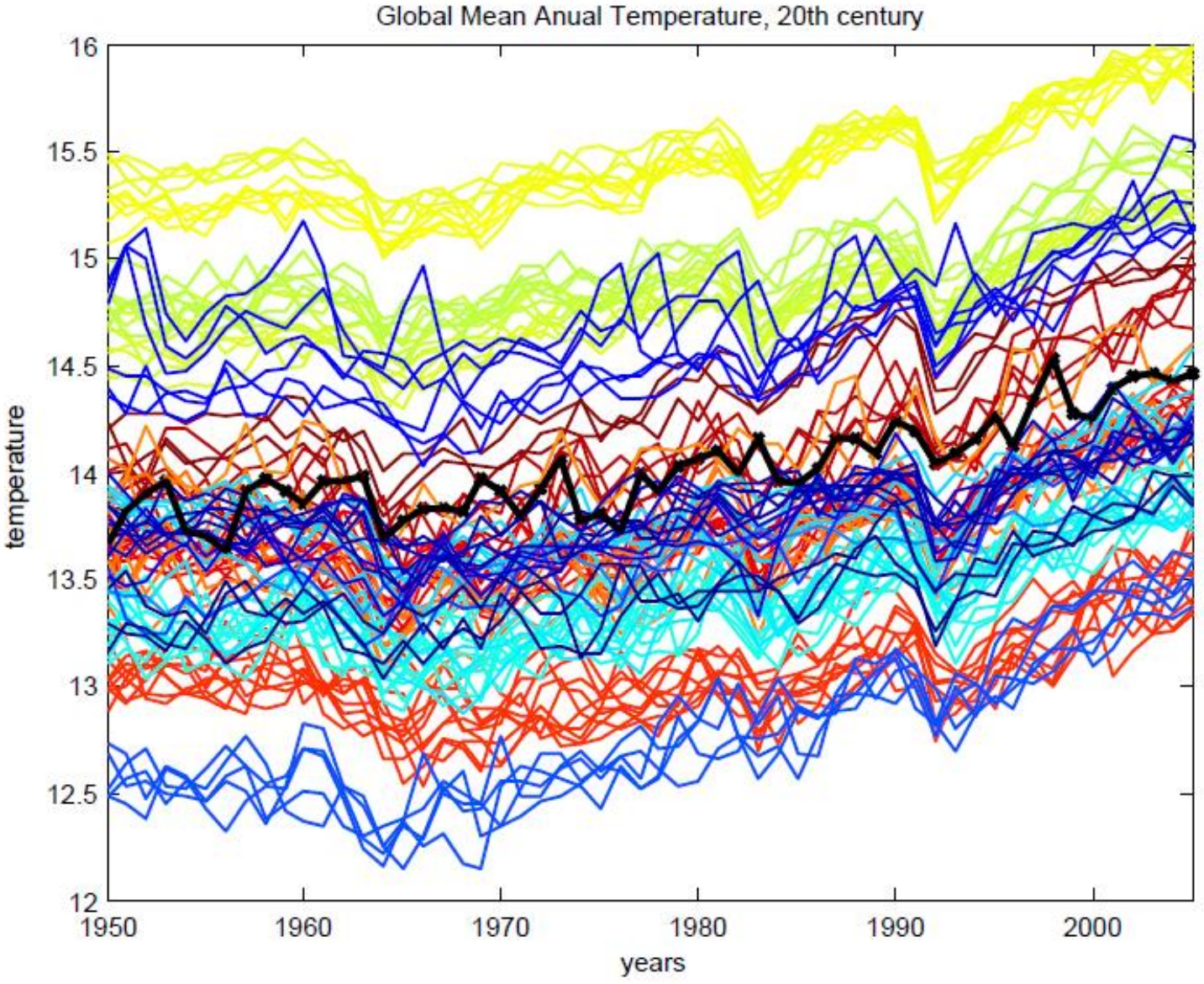
Systematic errors are larger than the observed effect



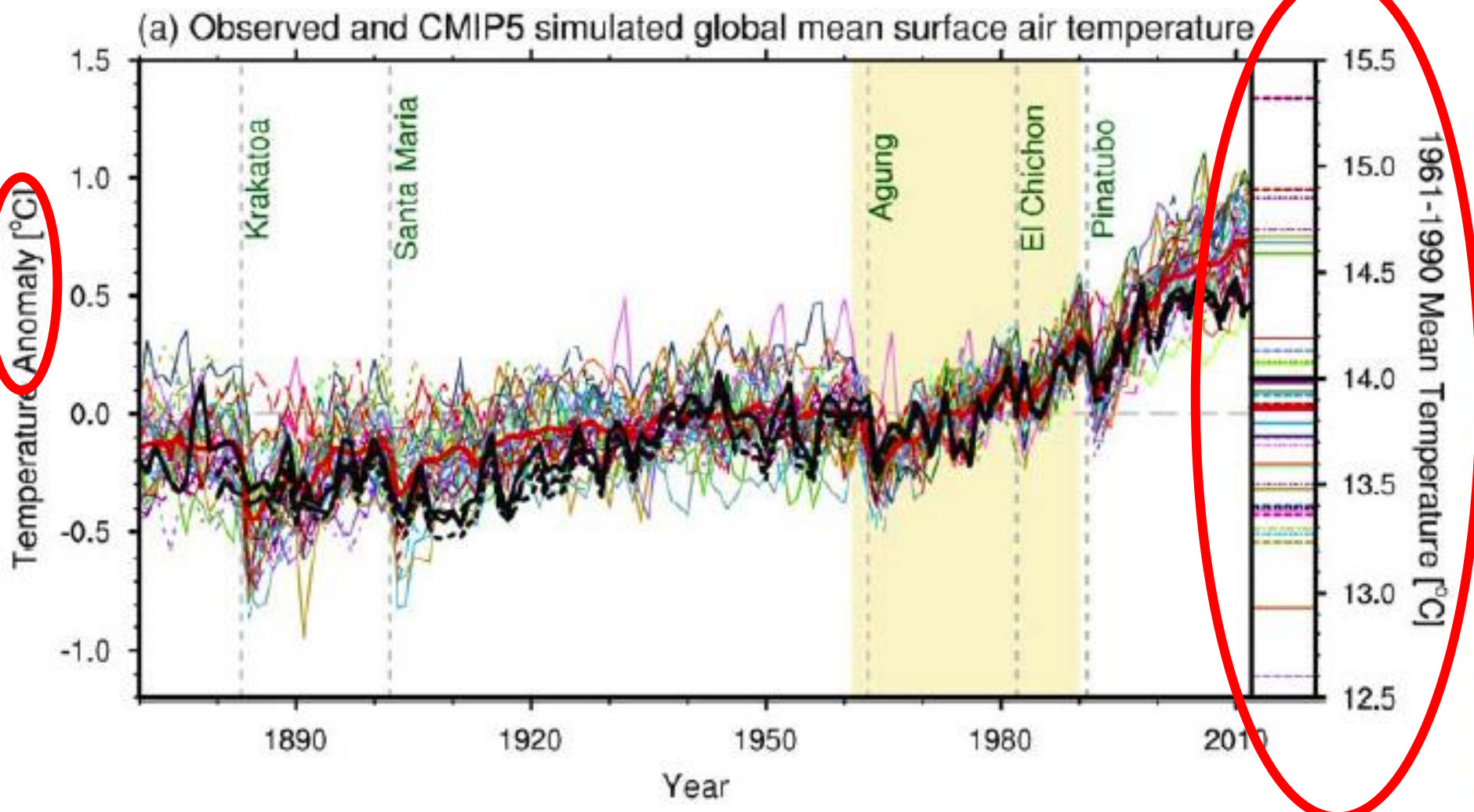
While systematic errors are larger than the observed effect



CMIP5


Global Mean Anual Temperature, 20th century
temperature
years
16
15.5
15
14.5
14
13.5
13
12.5
12
1950
1960
1970
1980
1990
2000

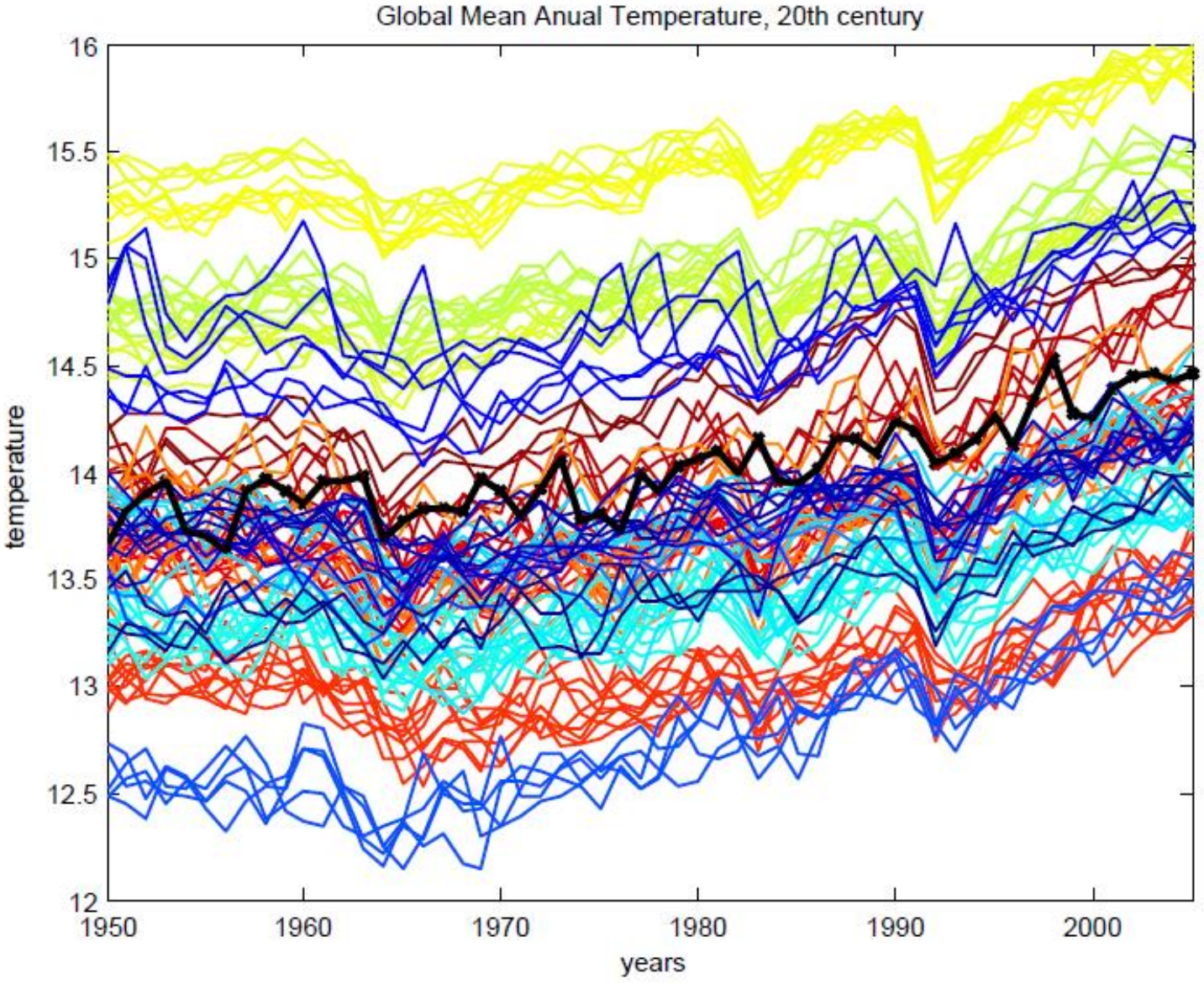
The AR5 is a bit more forthcoming



CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

And why was the anomaly period shifted?

CMIP5


Global Mean Anual Temperature, 20th century
temperature
years
16
15.5
15
14.5
14
13.5
13
12.5
12
1950
1960
1970
1980
1990
2000

Limits to Transparency: Anomalies 2013

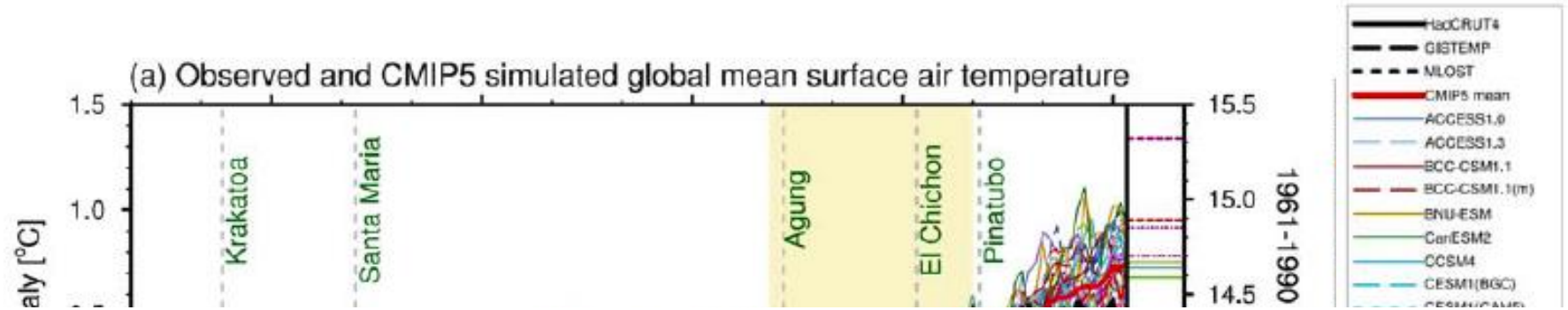
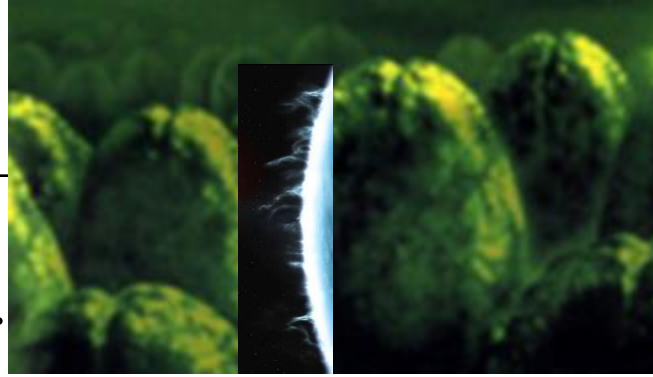


Figure 9.8: Observed and simulated time series of the anomalies in annual- and global-mean surface temperature. All anomalies are differences from the 1961–1990 time-mean of each individual time series. The reference period 1961–1990 is indicated by yellow shading; vertical dashed grey lines represent 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.

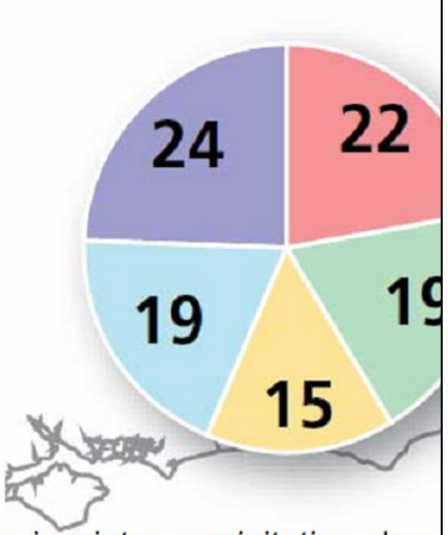
I've got, that sinking feeling…

When lower bounds, identified as such in the peer reviewed literature, are said to be estimates.



How important are different sources of uncertainty?

- Varies, but typically no single sour



Uncertainties in winter precipitation chan
1961-90, at a 25km box in SE England

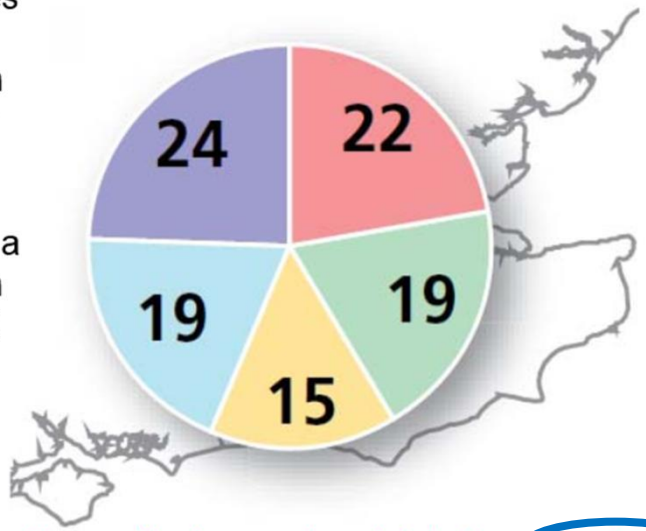
2008

UK CLIMATE PROJECTIONS

Met Office Hadley Centre

Reducing different sources of uncertainty?

Uncertainties in winter precipitation changes for the 2080s relative to 1961-90, at a 25km box in SE England



- Internal variability
- Carbon cycle
- Structural uncertainty
- Parameter uncertainty
- Downscaling

New information, methods, experimental design can reduce uncertainty so projections will change in future and decision makers need to consider this

2011

Do we have a firm estimate of each slice of the pie?

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,

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

From the AR4

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

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

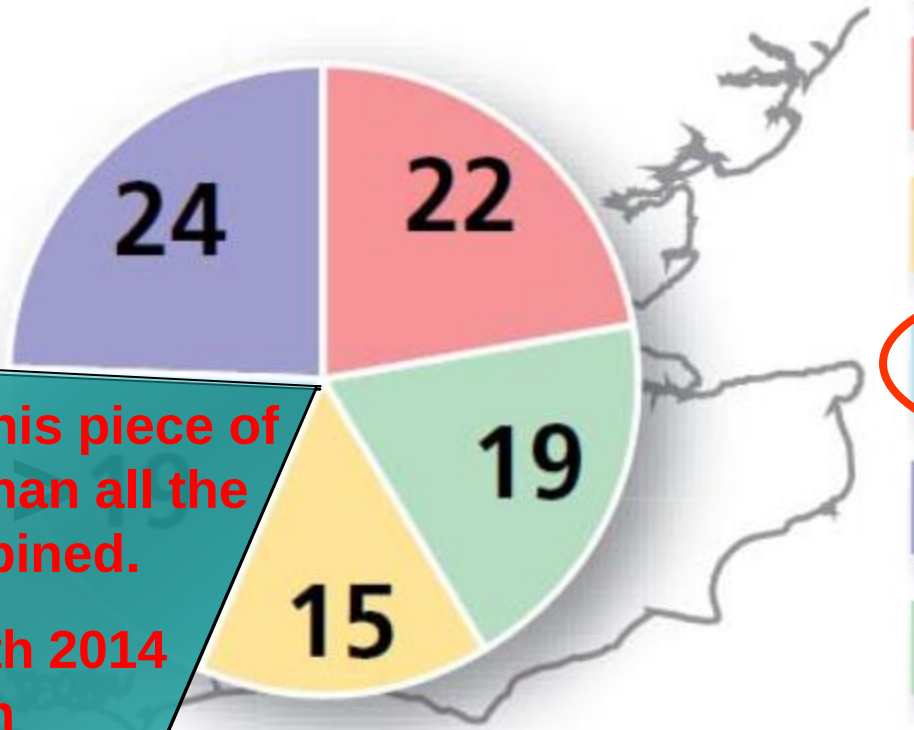
J.M Murphy, B.B.B Booth, M Collins, G.R Harris, D.M.H Sexton and M.J Webb

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

Reducing different sources of uncertainty?

Model diversity is only a lower bound on structural uncertainty. It may well be by far the biggest piece of the pie.

Uncertainties in winter precipitation changes for the 2080s relative to 1961-90, at a 25km box in GB England



Internal variability

Carbon cycle

Structural uncertainty

Parameter uncertainty

Downscaling

For all we know, this piece of the pie is bigger than all the other pieces combined.

A lower bound with 2014 hardware is not an estimated value.

What to do? Say?

New information, methods, experimental ... change in future and decis...

We know today our model-based probabilities are not mature.

Sexton 2010 http://www.exeter.ac.uk/media/universityofexeter/research/inspiringresearch/sciencestrategy/ccsf/docs/Making_probabilistic_climate_projections_for_the_UK_presentation.pdf

Is a sea ice anomaly unphysical?!?

Regardless, my point here is not to contrast structural differences with simulation differences; but rather to note that sea ice vs no sea ice has immediate nontrivial local impacts.

Coast guard stations in the North.

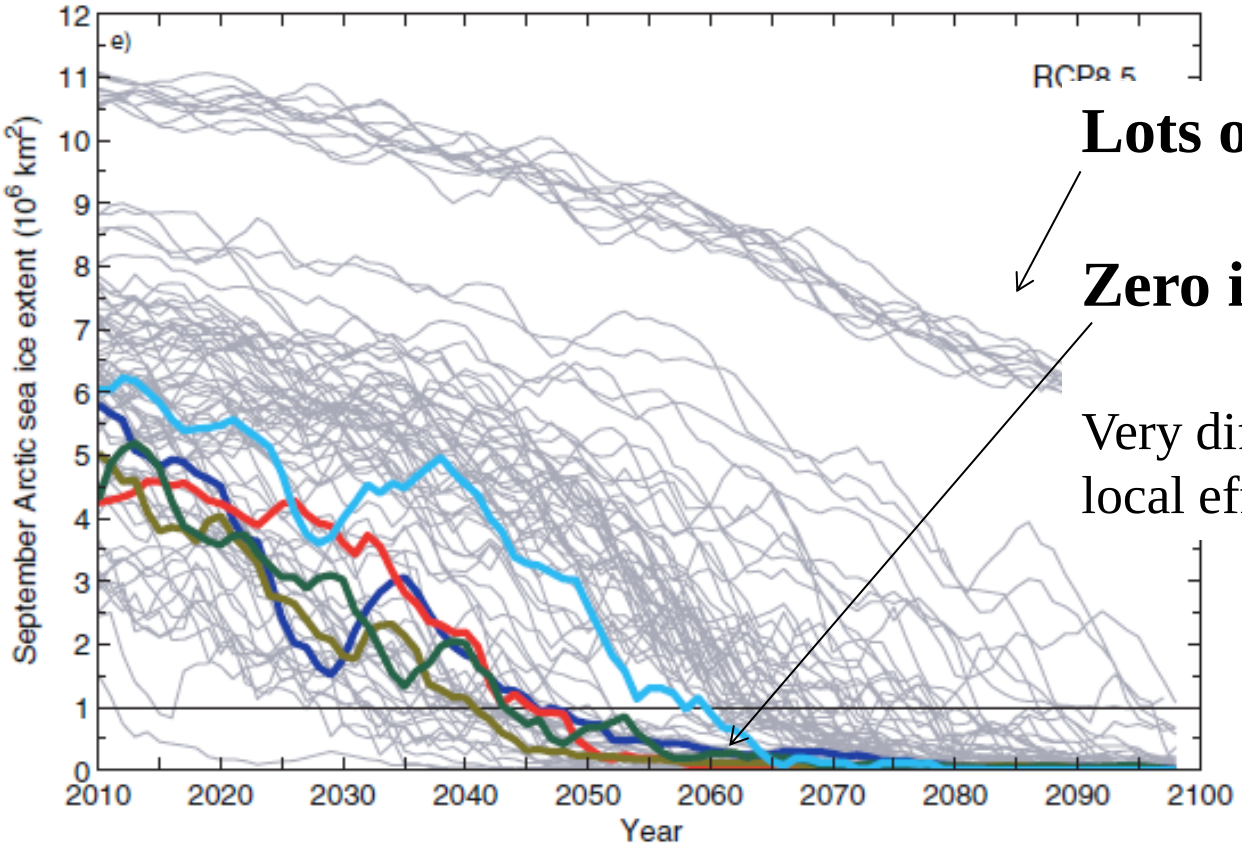


Figure 12.31 | (a–d) First year during which the September Arctic sea ice extent falls below 1 × 10^6 km² in CMIP5 climate projections (37 models, RCP8.5) as a function of (a) the September Arctic sea ice extent averaged over 1986–2005, (b) the annual mean Arctic sea ice volume averaged over 1986–2005, (c) the amplitude of the 1986–2005 mean seasonal cycle of Arctic sea ice extent and (d) the trend in September Arctic sea ice extent over 1979–2012. The sea ice diagnostics displayed are calculated on the original model grids. The correlations and one-tailed *p*-values are computed from the multi-member means for models with several ensemble members (coloured crosses), but the ensemble members of individual models are also depicted (coloured dots). The vertical solid and dashed lines show the corresponding observations or bias-adjusted PIOMAS (Pan-Arctic Ice-Ocean Modelling and Assimilation System) reanalysis data (a, c and d: Comiso and Nishio, 2008, updated 2012; b: Schweiger et al., 2011) and the ±20% interval around these data, respectively. (e) Time series of September Arctic sea ice extent (5-year running mean) as simulated by all CMIP5 models and their ensemble members under RCP8.5 (thin curves). The thick, coloured curves correspond to a subset of five CMIP5 models selected on the basis of panels a–d following Massonnet et al. (2012) (see text for details). Note that each of these models provides only one ensemble member for RCP8.5.

CATS CENTR THE AN OF TIM

"If you think it is 5 your are stupid"

Working in User's Words/Thought Space

Pedagogically, if someone asks a weird question, it is sometimes useful to try and answer the question as asked.

But that is not what we did!

Yes, but that's not the way it was done!

Sure. I understand that. She knows that.
But why resist looking at the problem from another angle?

This is an example of trying to engage and have a discussion (BBC).
I am not disputing science, but illustrating what (might have been) better communication.

That's not my job!

Fine. It is just an example: my talk is on Good Karma Communication.

AR5

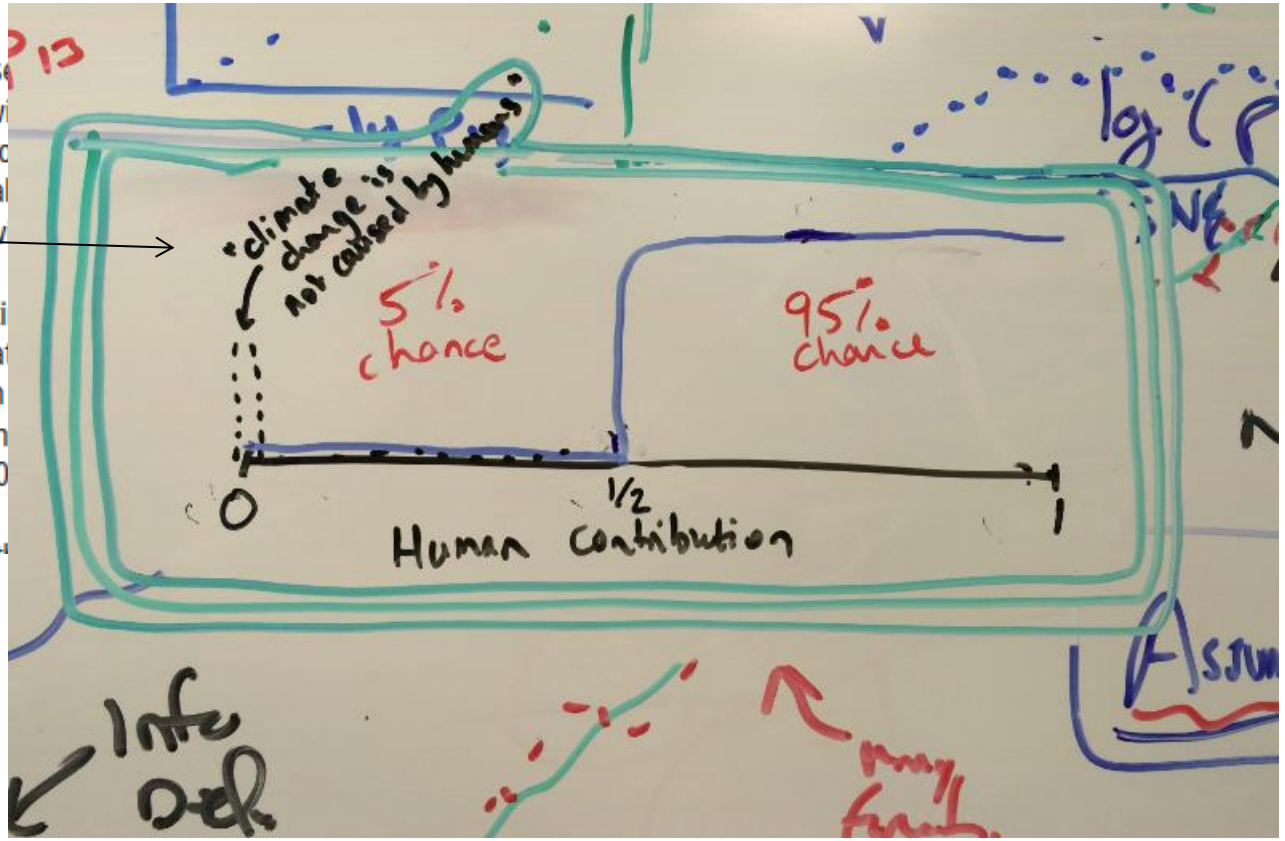
D.3 Detection and Attribution of Climate Change

Human influence has been detected in warming of the atmosphere and the ocean, in changes in the global water cycle, in reductions in snow and ice, in global mean sea level rise, and in changes in some climate extremes (see Figure SPM.6 and Table SPM.1). This evidence for human influence has grown since AR4. It is *extremely likely* that human influence has been the dominant cause of the observed warming since the mid-20th century. {10.3–10.6, 10.9}

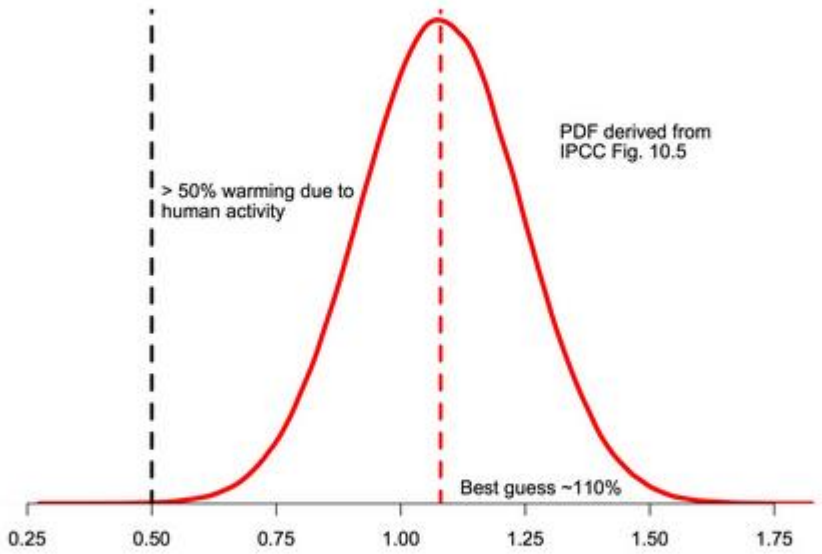
- It is *extremely likely* that more than half of the observed increase in global average surface temperature from 1951 to 2010 was caused by the anthropogenic increase in greenhouse gas concentrations and other anthropogenic forcings together. The best estimate of the human-induced contribution to warming is similar to the observed warming over this period. {10.3}

- Greenhouse gase
  1951 to 2010, wi
  be in the range c
  and from natural
  are consistent w

- Over every conti
  surface temperat
  tainties result in
  available station
  since the mid-20

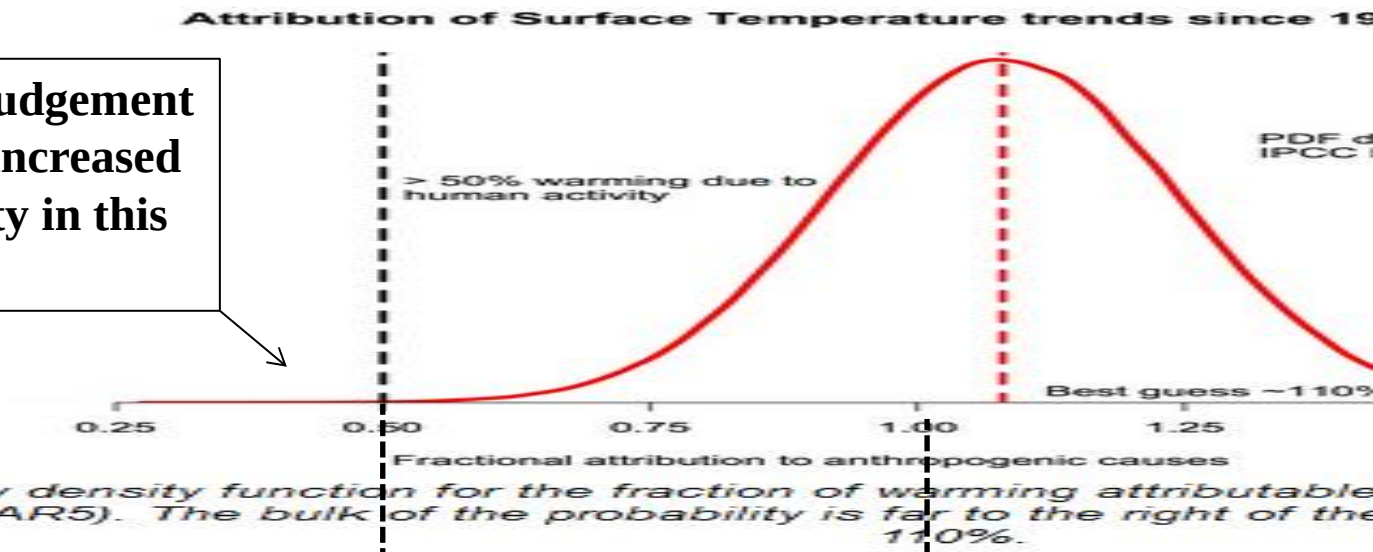


Attribution of Surface Temperature trends since 1950

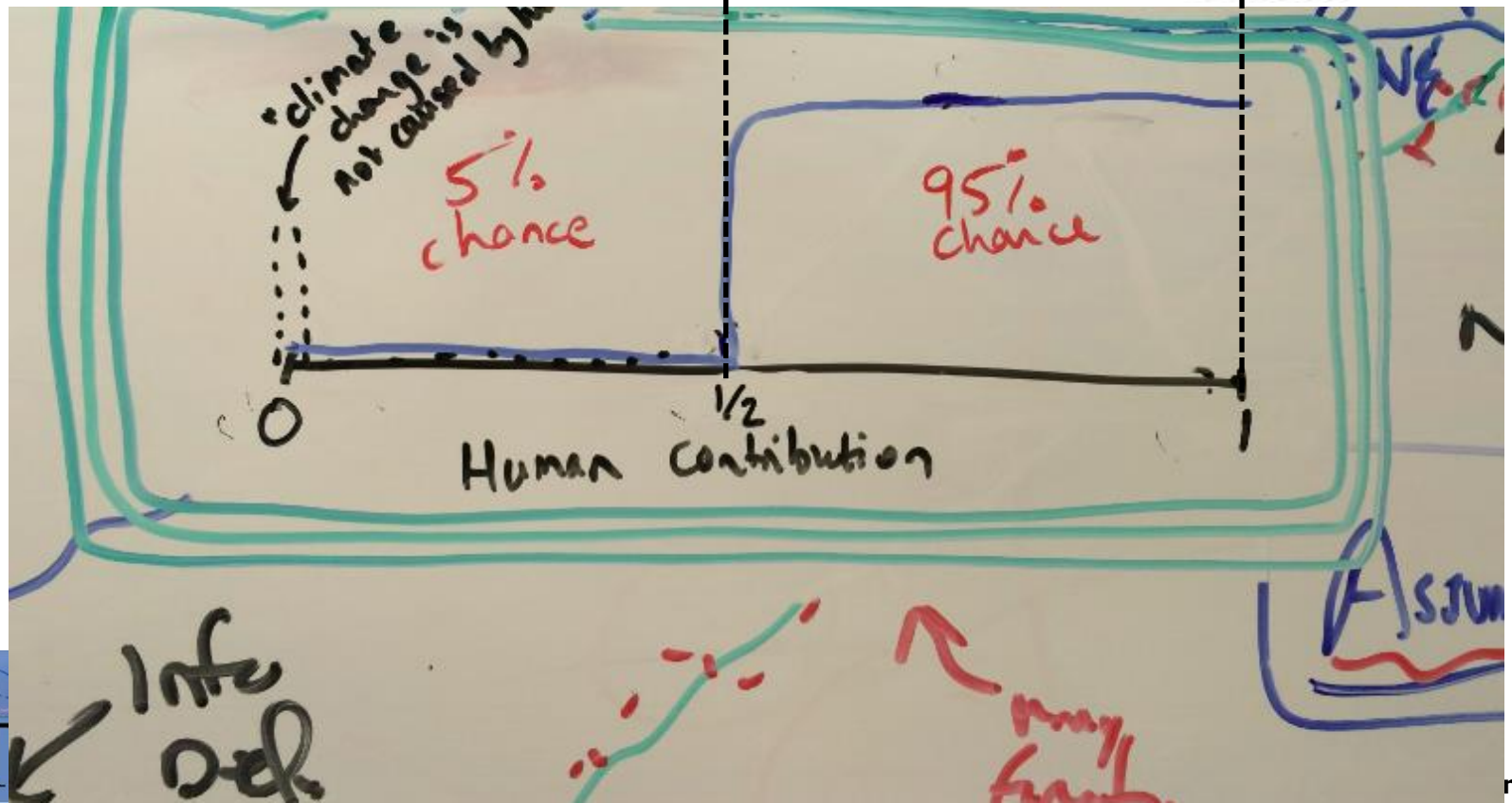


The probability density function for the fraction of warming attributable to human activity (derived from Fig. 10.5 in IPCC AR5). The bulk of the probability is far to the right of the "50%" line, and the peak is around 110%.

Note expert judgement significantly increased the probability in this tail area.



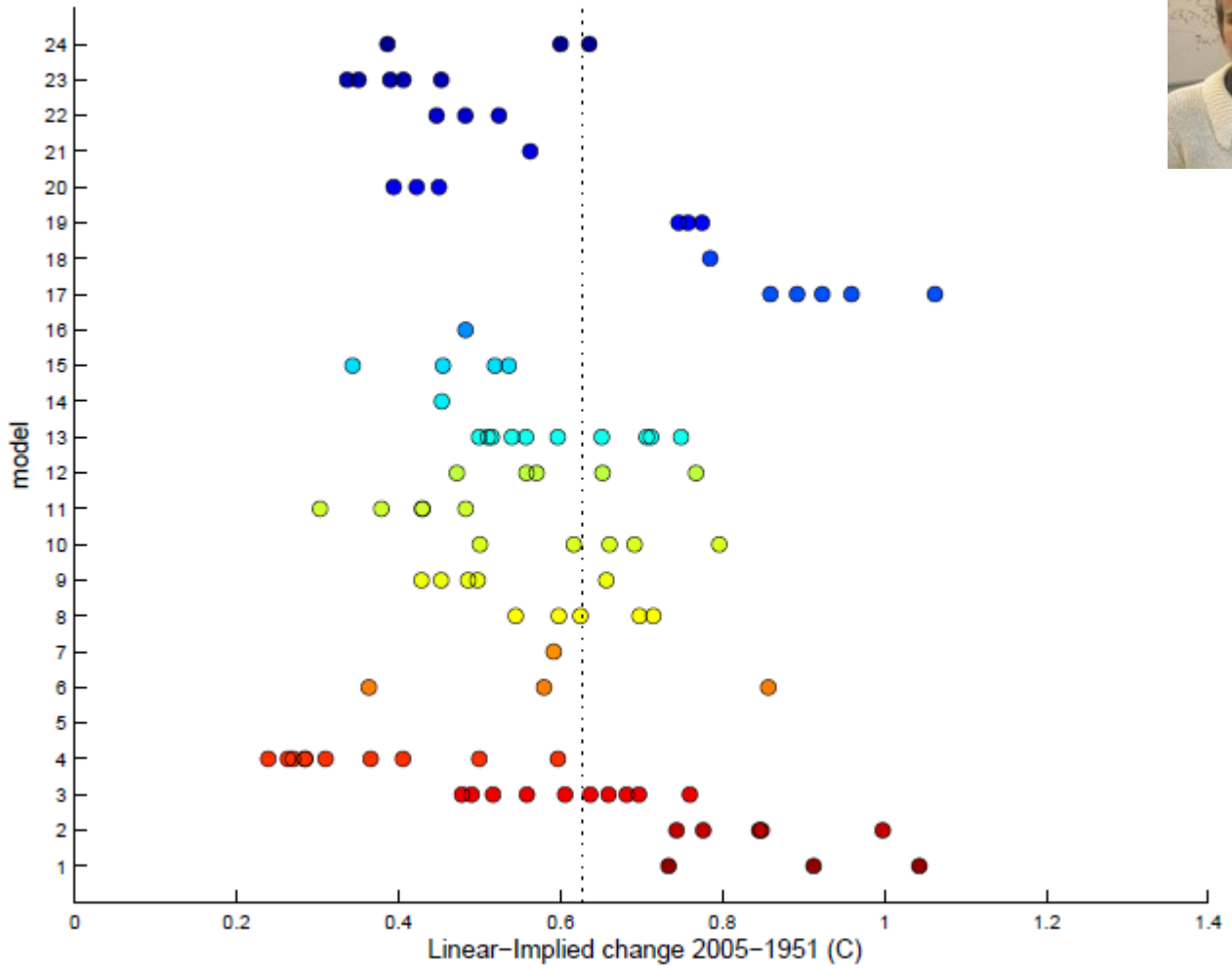
The probability density function for the fraction of warming attributable
10.5 in IPCC AR5). The bulk of the probability is far to the right of the
110%.

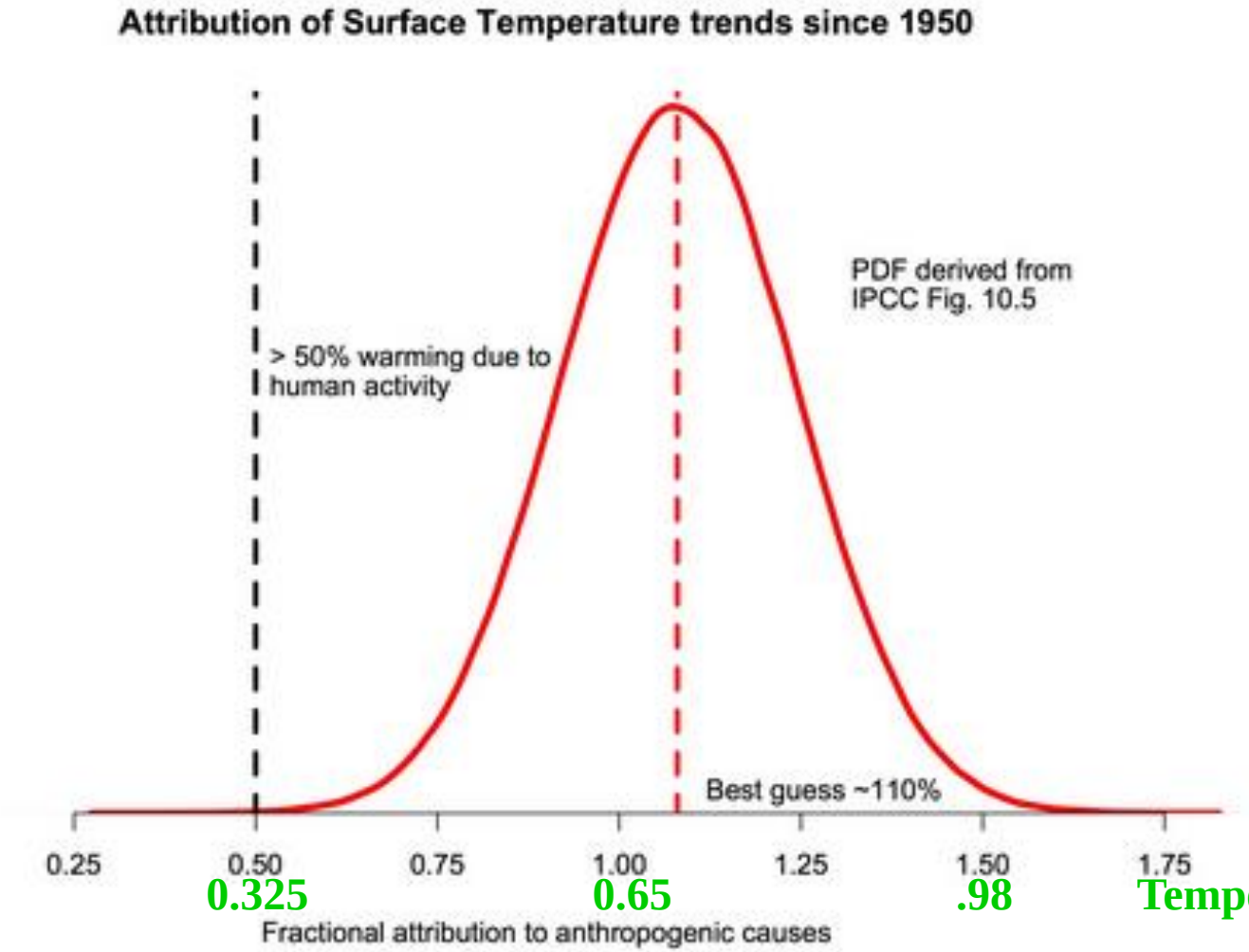


CAT

mith

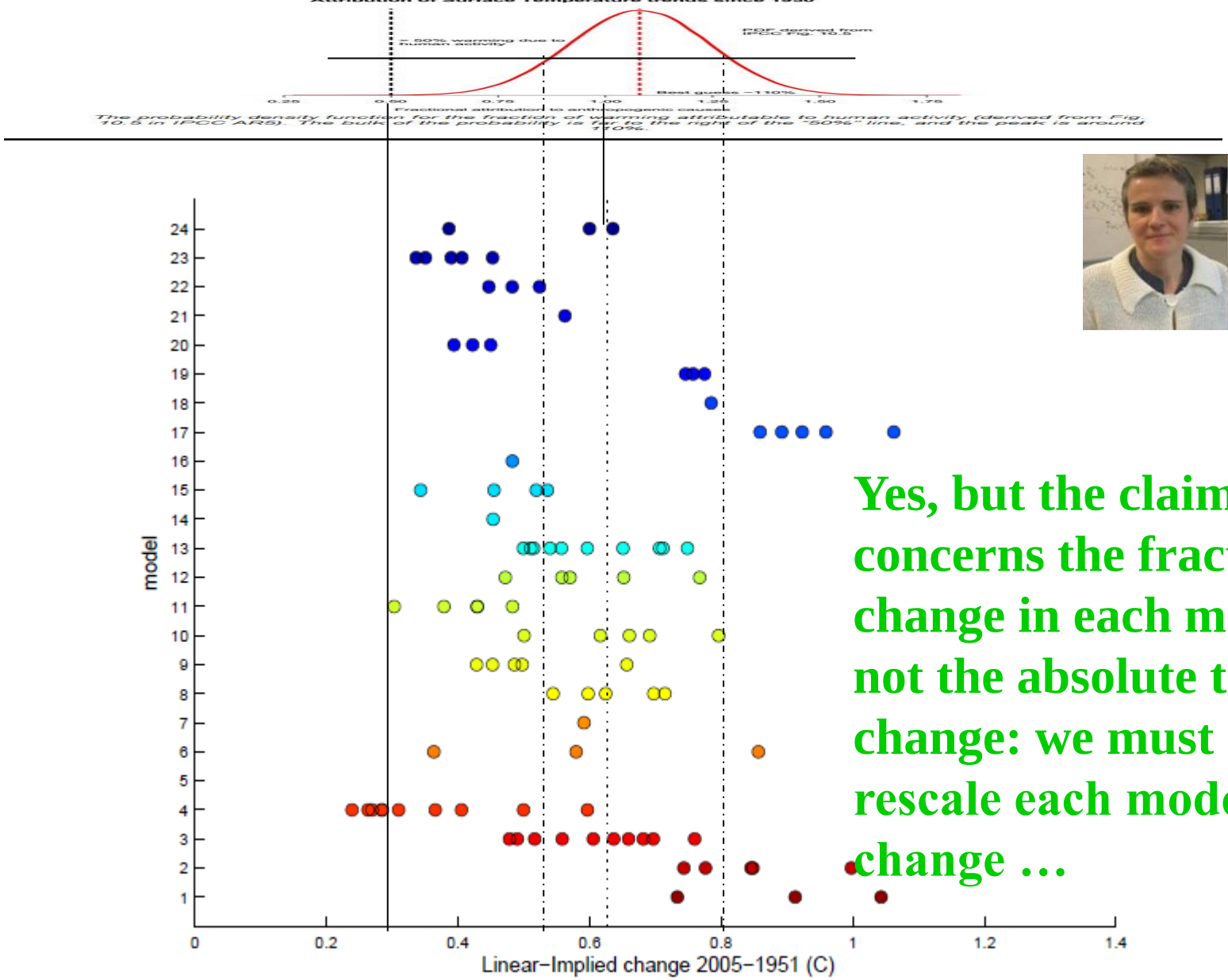
But this is what the model's did, it looks different





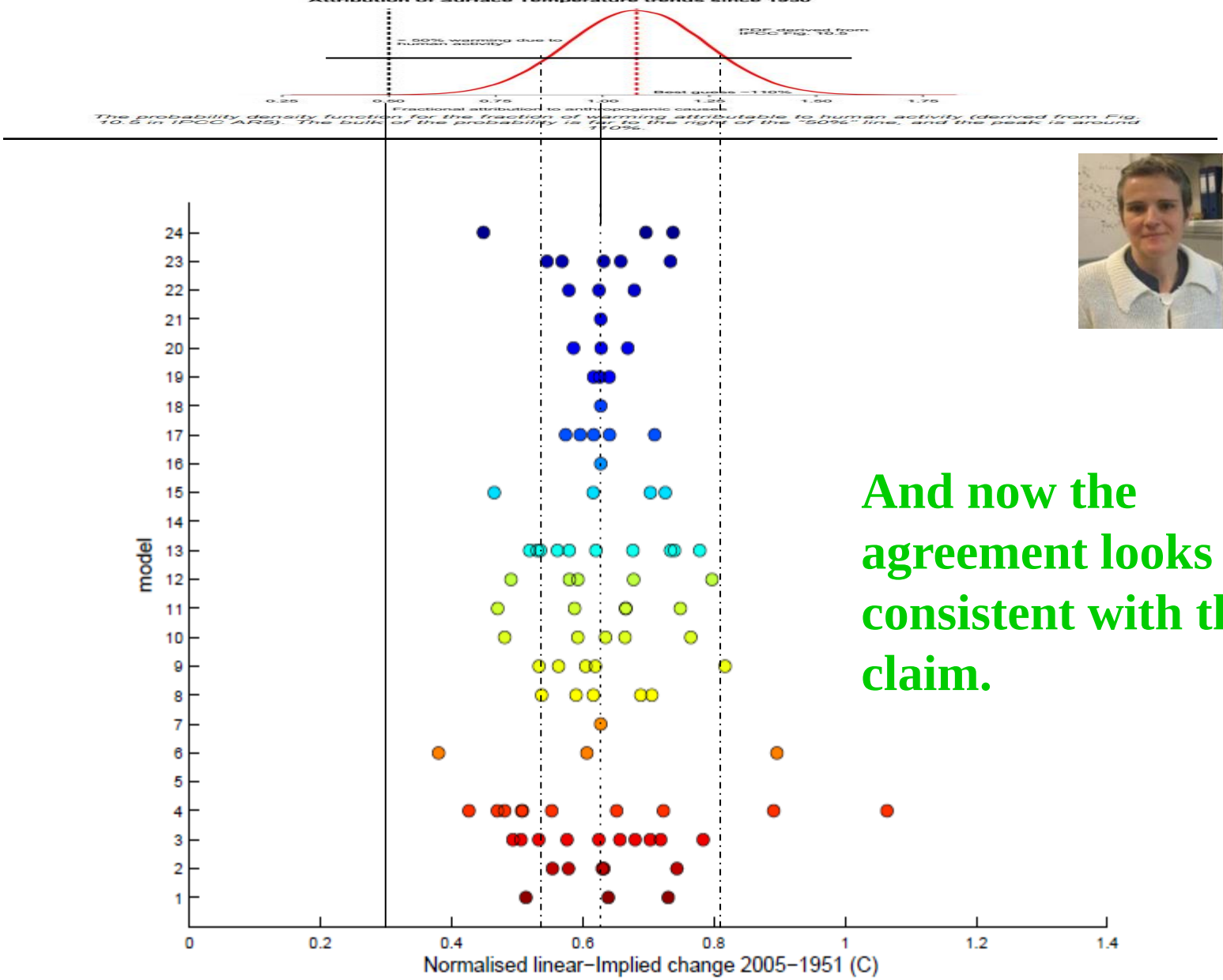
The probability density function for the fraction of warming attributable to human activity (derived from Fig. 10.5 in IPCC AR5). The bulk of the probability is far to the right of the "50%" line, and the peak is around 110%.

OK, so plot the models on this graph.



Yes, but the claim concerns the fraction of change in each model, not the absolute temp change: we must rescale each model's change …

See: Still Completely Different and not Exchangeable!

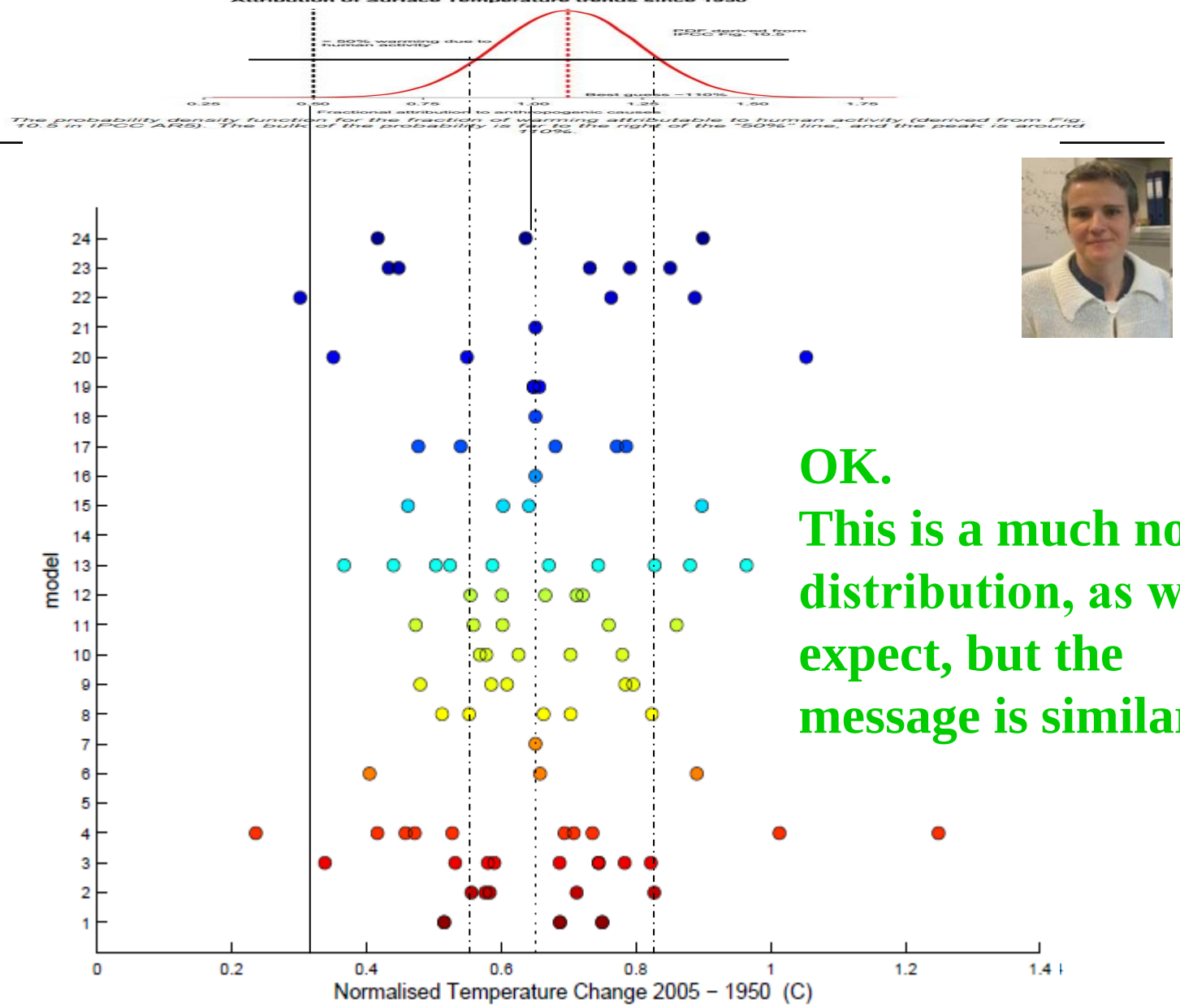


And now the agreement looks consistent with the claim.

What is this "linear-implied change"?

Just show me 2005 - 1950!

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

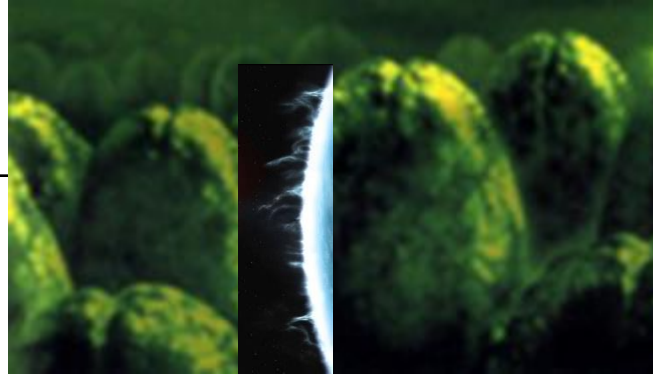


OK.
This is a much noisier distribution, as we'd expect, but the message is similar.

….

OK. But why are you using 2005?

I've got, that sinking feeling…



Under the CMIP5 experimental design, there is a break-point at 2005.

Shirley, this does not matter.

(But why introduce a hook for the anti-science lobby, and annoy the statisticians?)

Careful, coherent experimental design is basic good practice. Are we not clear on the questions before we look at the output? It is the fifth assessment: have we not frozen most of the policy questions to allow comparability?

Actionable Credible Transparent Applied Science
Can we avoid talking past each other?

I find this both fun and disquieting.

Rhetorical Sword Crossing

Stephen Koonin: WSJ piece.

Ian Foster: Identified "Strictly true but misleading factoids."

Ray Pierrehumbert: Slams relevant and "weak" claims in one go.

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Actionable Credible Transparent Applied Science
Can we avoid talking past each other?

Consider a "Chatham House Rule" discussion to refine simple clear statements. I believe:

There are many things we'd all admit at the 1 in 200 (0.005) level.
And not many things we'd each insist on at the 0.995 level.
And that if worded by a neutral party (so as to be inclusive, not so as to be woolly), we might find people agree on things they would never reach agreement on in conversation. P(x|I) P changes with I.

Speaking probabilistically, we can "drop" discussion of things that we agree are << 0.005, or >> 0.10, and refine and report the diversity of views on those things near the threshold.

Then demonstrate the level of unity.
With coherent "minority reports" as needed.

The One in 200 Threshold and Risk Management

http://riskfriends.wordpress.com/2009/03/10/solvency-ii-dealing-with-operational-risk/

Solvency II is a set of regulatory requirements for insurance firms that operate in the EU designed to prevent insurance company failures by unbundling "operational risk".

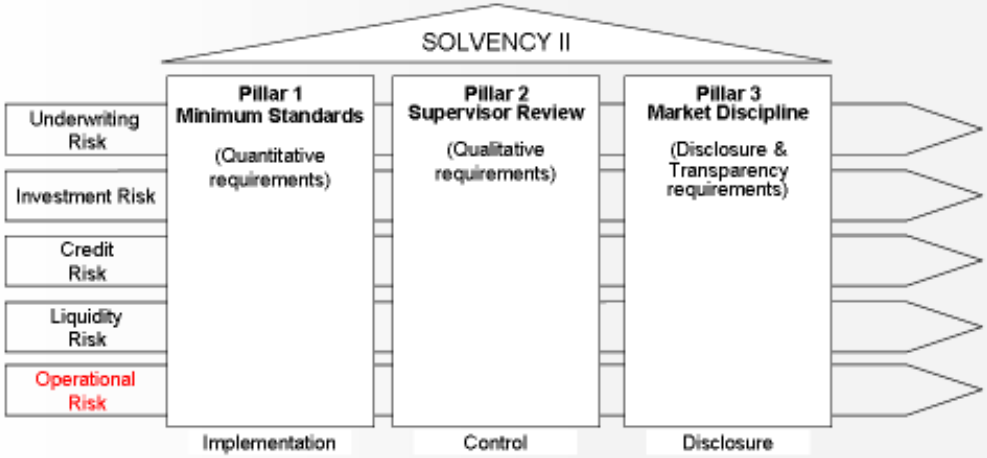
The aim here is not to integrate over all risks and opportunities to estimate the PDF of expected annual income but simply to ensure that insurers have sufficient "regulatory capital" to survive any (every) adverse event which has more than a 1 in 200 chance of occurring.

Question: Can climate science ascertain whether the probability of an outcome is

a) >> 1 in 200
b) ~ 1 in 200
c) << 1 in 200

Clearly identify risks without the investigative distraction posed by the whole shebang of a PDF.

The Solvency II framework consists of three pillars, each covering a different aspect of the economic risks facing insurers, see figure 1. This three-pillar approach aims to align risk measurement and risk management. The first pillar relates to the quantitative requirement for insurers to understand the nature of their risk exposure. As such, insurers need to hold sufficient regulatory capital to ensure that (with a 99.5% probability over a one-year period) they are protected against adverse events. The second pillar deals with the qualitative aspects and sets out requirements for the governance and risk management of insurers. The third pillar focuses on disclosure and transparency requirements by seeking to harmonise reporting and provide insight into insurers' risk and return profiles.

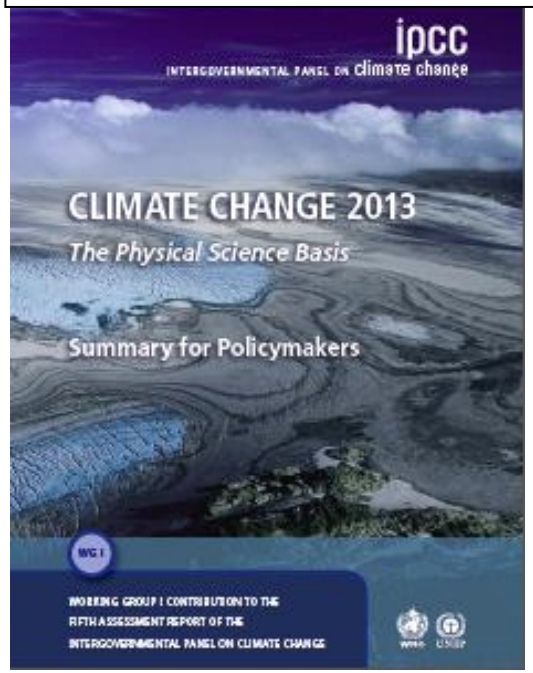


Solvency II (SII) is the updated set of regulatory requirements for insurance companies operating in the European Union. It revises the existing capital adequacy regime and is expected to come into force in 2012. It has a number of expected benefits, both for insurers and consumers. Although the most obvious benefit seems to be preventing catastrophic losses, other less obvious benefits which are considered to be important are summarised in table 1.

https://www.ipcc.ch/report/ar5/wg1/docs/WGIAR5_SPM_brochure_en.pdf

What does the IPCC AR5 Say?

- Increase of global mean surface temperatures for 2081–2100 relative to 1986–2005 is projected to *likely* be in the ranges derived from the concentration-driven CMIP5 model simulations, that is, 0.3°C to 1.7°C (RCP2.6), 1.1°C to 2.6°C (RCP4.5), 1.4°C to 3.1°C (RCP6.0), 2.6°C to 4.8°C (RCP8.5). The Arctic region will warm more rapidly than the global mean, and mean warming over land will be larger than over the ocean (*very high confidence*) (see Figures SPM.7 and SPM.8, and Table SPM.2). {12.4, 14.8}



Real-world GMT is "likely" (66% chance) to be in "the range" of model-land GMT.

That suggests there is a significant chance the real-word will be outside the range of the models.

If your downscaling model was perfect, there remains a huge chance you could not catch the relevant pathway (as none of today's models do).

I think it is fair say the IPCC implies that the Probability of a Big Surprise (GMT in 2100) is about one in ~ four to ~ten.

By law, we require banks and insurance companies hold reserves to cover one in 200 year events.

Regarding Prob(Big Surprise) in GMT, the AR5 is much improved.

- Increase of global mean surface temperatures for 2081–2100 relative to 1986–2005 is projected to *likely* be in the ranges derived from the concentration-driven CMIP5 model simulations, that is, 0.3°C to 1.7°C (RCP2.6), 1.1°C to 2.6°C (RCP4.5), 1.4°C to 3.1°C (RCP6.0), 2.6°C to 4.8°C (RCP8.5). The Arctic region will warm more rapidly than the global mean, and mean warming over land will be larger than over the ocean (*very high confidence*) (see Figures SPM.7 and SPM.8, and Table SPM.2). {12.4, 14.8}

AR5 SPM E.1

This is an explicit statement that nontrivial probability density lies outside the range of the CMIP5 model runs.

It would be very nice to know the distribution of belief on where that extra probability lies!
Mostly above? Symmetrically above and below?...

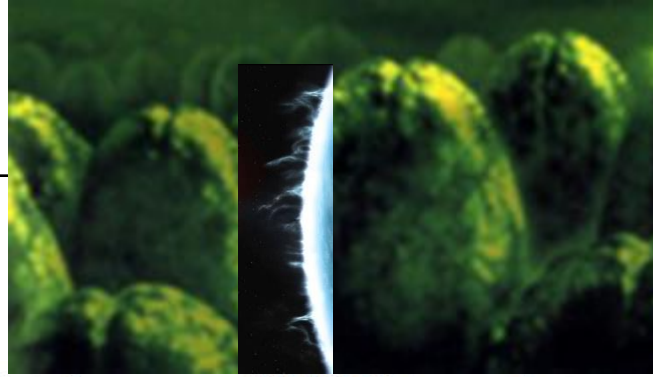
But in a public panel at the AAAS it was said:

~"An adjustment was needed in the AR4 (the 60:40 rule) but not this time (AR5)"~

Physicists on the Higgs-like at the AAAS

I've got, that sinking feeling…

Rank order beauty contests, without comparison with some absolute measure of quality, are misleading in several ways.



Limits to Transparency:

Equidismality

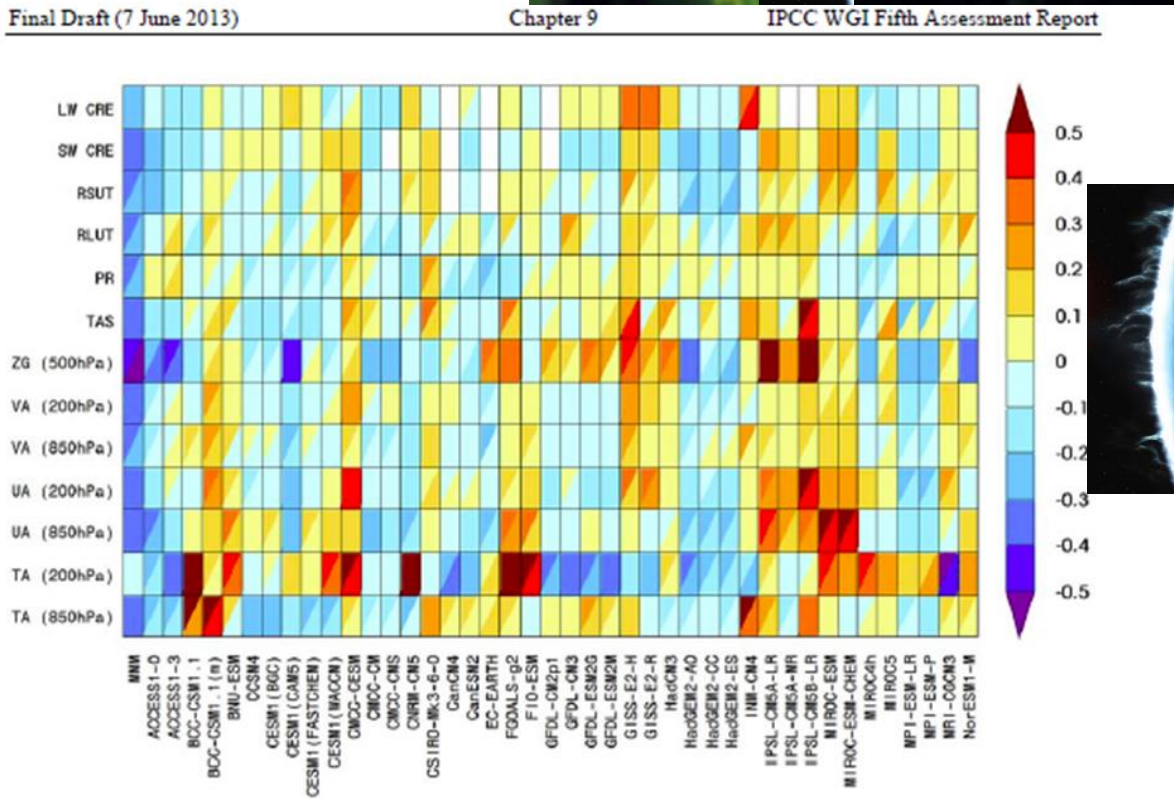


Figure 9.7: Relative error measures of CMIP5 model performance, based on the global seasonal-cycle climatology (1980–2005) computed from the historical experiments. Rows and columns represent individual variables and models, respectively. The error measure is a space–time root-mean-square error (RMSE), 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

Skill in Decadal Forecasting (ENSEMBLES)

(The empirical model "wins")

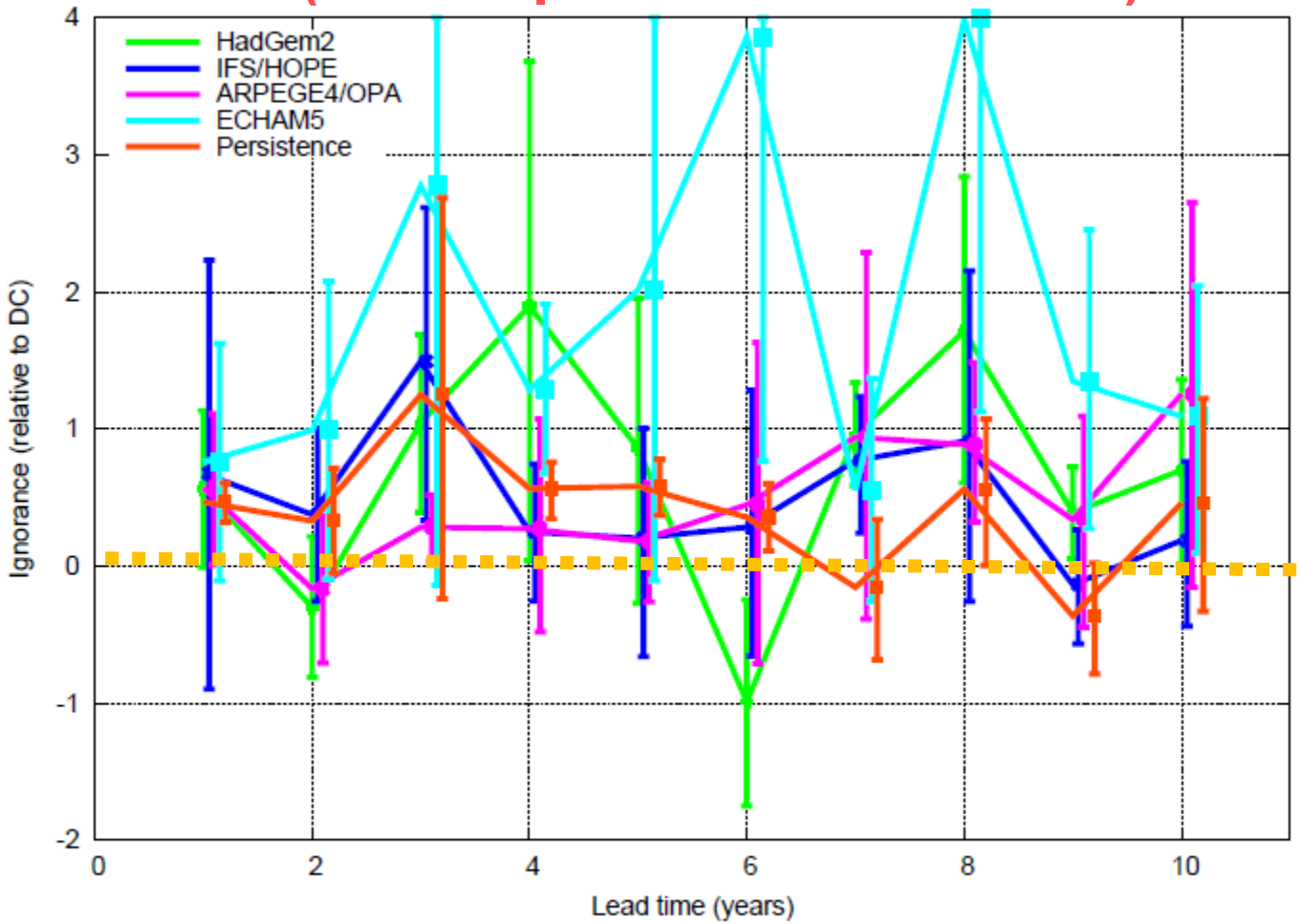
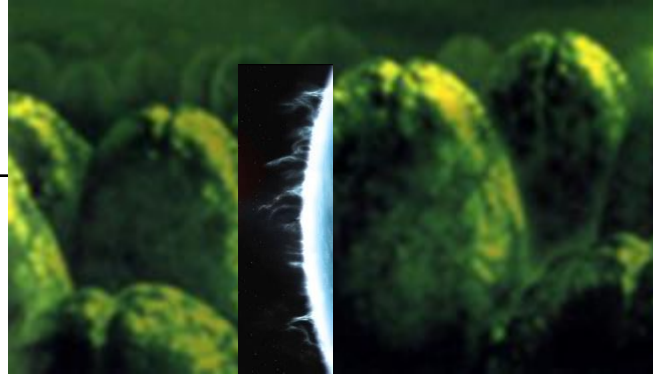


FIG. 9. Ignorance of the ENSEMBLES models relative to DC as a function of lead time. Note that the simulation models tend to have positive scores (less skill) than the DC model at every lead time.

UKCP09 Worked Examples…



A "charitable reading" of the UKCP09 documentation is not possible.

An ensemble a "charitable reading" is possible, and leads to vastly different implications for intended interpretation and use, and implied adequacy.

The intention is made rather more clear by the "worked examples" in the main report, and the description of the designers said what would will allow.

UKCP09 Guidance to perform a sustainability assessment

Objectives: To determine how many times in the future a known threshold might be crossed to ensure energy use sustainability in a school and perform a cost-benefit analysis on adaptation measures. The school is new-build so if operation of the school requires more energy than specified in the building contract then the contracting developer is held accountable.

- **How they used UKCP09 dummy data**
  - It is known that a persistently warm minimum temperature (exceeding 16° for three nights in a row) makes the building uncomfortably warm during the day.
  - This was used as a threshold for analysis in the threshold detector and was defined, in this case, as a *heatwave*. This

I do not have time to discuss the UKCP worked examples…
But each of these makes naïve realist assumptions

UKCP09 in practice: acclimatise

UKCP09 Guidance
UKCP09 in practice
Good practice
Table of worked examples
Inappropriate uses

Keywords for this example:

Energy, sustainability, weather generator, threshold detector

UKCP09 Guidance to perform a sustainability assessment

Objectives: To determine how many times in the future a known threshold might be crossed to ensure energy use sustainability in a school and perform a cost-benefit analysis on adaptation measures. The school is new-build so if operation of the school requires more energy than specified in the building contract then the contracting developer is held accountable.

- **How they used UKCP09 dummy data**
  - It is known that a persistently warm minimum temperature (exceeding 16° for three nights in a row) makes the building uncomfortably warm during the day.
  - This was used as a threshold for analysis in the threshold detector and was defined, in this case, as a *heatwave*. This is a user-defined heatwave and not the same as the pre-defined heatwave in the threshold detector.
  - By analysing the weather generator output with the threshold detector, an increase in frequency in comparison to the baseline climate was presented in a table showing the range of error in the threshold detector output and the average value, for each season (this is standard output from the threshold detector).
  - A number of adaptation measures were suggested to manage these climate risks.

Selections made:

Data source: Weather generator

UKCP09 Product: Weather generator output, Threshold detector

Other products: None

Climate variables: Temperature

Emission scenario: Medium

Time period: 2080s

Temporal averaging: Seasonal

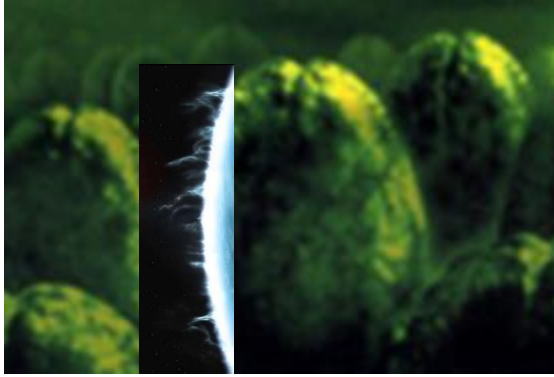
Spatial averaging: 5 km grid squares

Location: Gordon Heights School (fictitious)

Probability level: N/A

Simulate and Count

What is a "scientifically sound preview"?
What does that mean exactly?

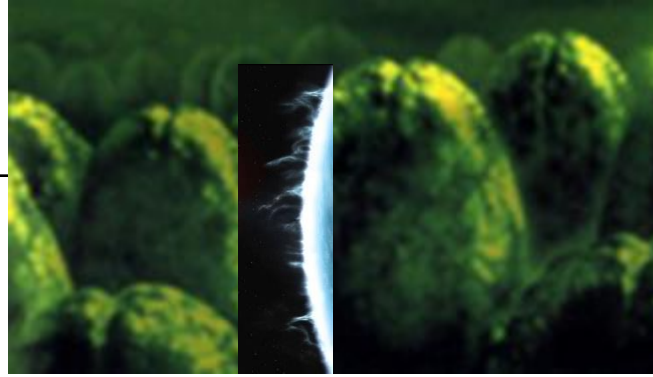


FAQ 9.1 AR5

…climate models are based, to a large extent, on verifiable physical principles and are able to reproduce many important aspects of past response to external forcing. In this way, they provide **a scientifically sound preview** of the climate response to different scenarios of anthropogenic forcing.

825

Lift the carpets



Realistic,
Competitive with other models,
Research challenges remain…,
“Might it be the case that …?”

We do not want to throw the baby out with the bathwater!

The baby here is the "as good as it gets" science claim that increasing CO_2 levels will increase the temperature of the Earth.

We wish to protect this baby!

Regardless of whether or not it is the fear of throwing the baby out with the bath water that has stopped climate science from criticising it's own: we could use a better mode of thinking…

Perhaps an additional set of models focusing on empirically confirmed skill only?

As new (unsupportable) claims of over-detailed knowledge appear; will they be openly criticised?

Once the bathwater is becoming rancid, we have to save the baby!

There is going to be another sequel

How do we communicate better next time?

More clarity regarding model fidelity.

Better experimental design.

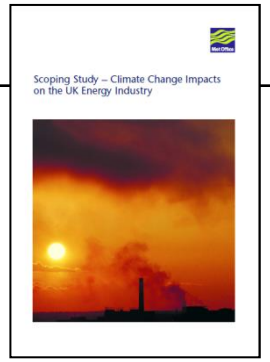
Kinder, more pedagogical discussion styles.
(More understanding, if we cannot reach "nicer".)

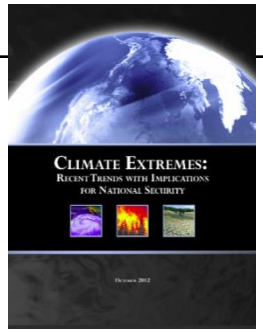
Clarity on where we (almost all) agree
(conditioned on I).

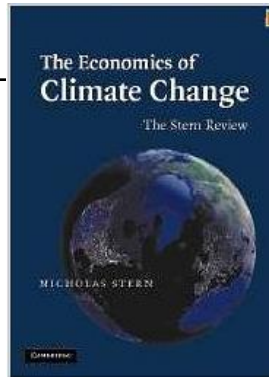
Clarity on exactly what is in dispute.

Lifting all the carpets, revealing what's underneath.

I worry more about the Scepticism growing here than the anti-science lobby.







HiH Moment

2006

Focused on RDU

Asked for more
? After checking?

EPA

eDF ENERGY

e·on

nationalgrid

CENTRAL INTELLIGENCE AGENCY

HM TREASURY

Department of Energy & Climate Change

Face AR5 quandary given "misreading" of AR4 statement

Embraced "Credible and Transparent"

ASA AMERICAN STATISTICAL ASSOCIATION

APS physics

AGU American Geophysical Union

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Trust is easily lost:
How do we protect the credibility of science over climate time scales?
How do we ease the pushback when current oversell becomes clear?


Thank you


Thank you
Watch the livestream now - Knowledge and Models in Climate Science - Oct 24-26

Probability in **Weather** and Climate

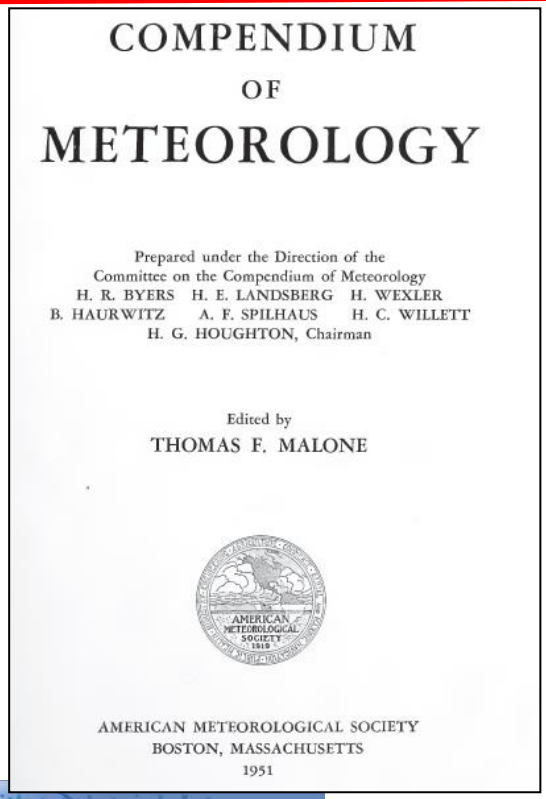
THE FORECAST PROBLEM

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 identify where today's science is not actionable today harms both science and policy.

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Transparency, Credibility & Realism in CFD

Where is climate modelling in relation to other computational fluid dynamics (CFD) groups in terms of: **Credibility, Robustness of Results, Validation, Verification, Uncertainty Quantification and Transparency?**

Credibility, robustness and UQ are not a new question to me:

Smith, L.A. (1992) Identification and Prediction of Low-Dimensional Dynamics. *Physica D,* 58 (1-4): 50.

Nor is this new to CFD. In the late 80's I was told why climate modelling was different.
I do not believe those reasons hold today.

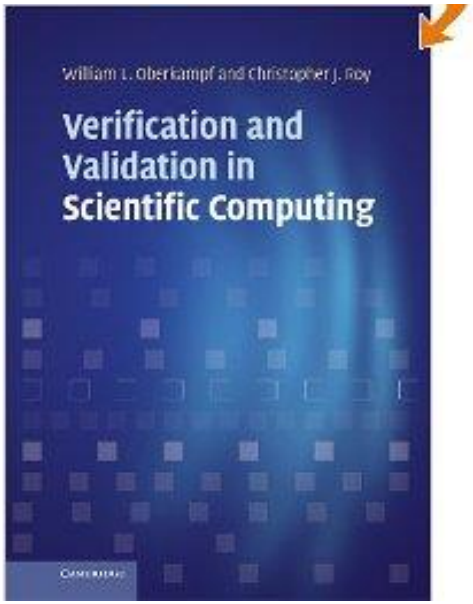
"By credibility of computational results we mean that the results of an analysis are worthy of belief or confidence".

"Two key factors are the users depth of understanding of the information produced and the appropriateness of the information for its intended use."

"The perspective of V&V is distinctly on the side of skepticism, sometimes to the degree of being radical."

These guys, like me, love models. They consider me rather mild mannered and trusting of simulation.

Oberkampf, & Roy (2010) Verification and Validation in Scientific Computing. CUP.



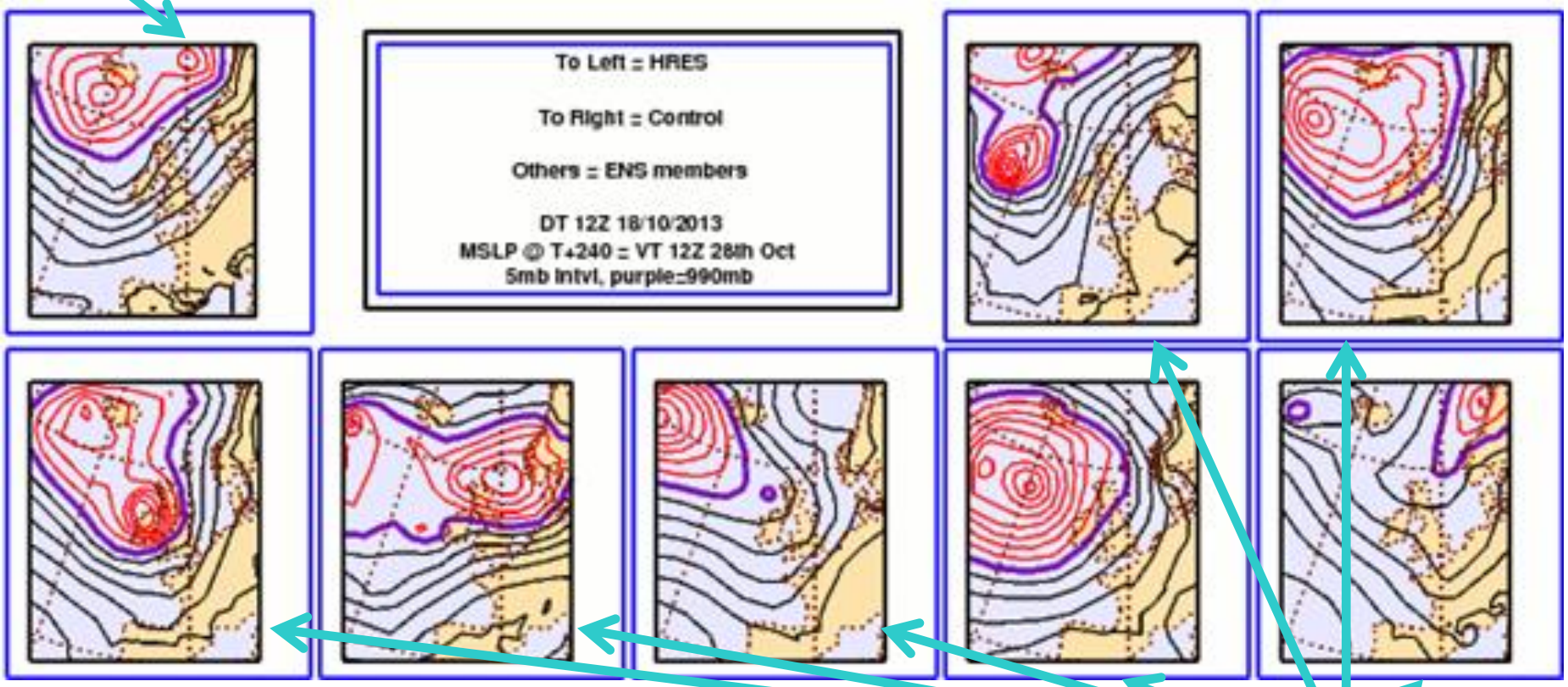
Trust can trump Uncertainty

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Foreseeing the St Jude Storm

(golf balls)

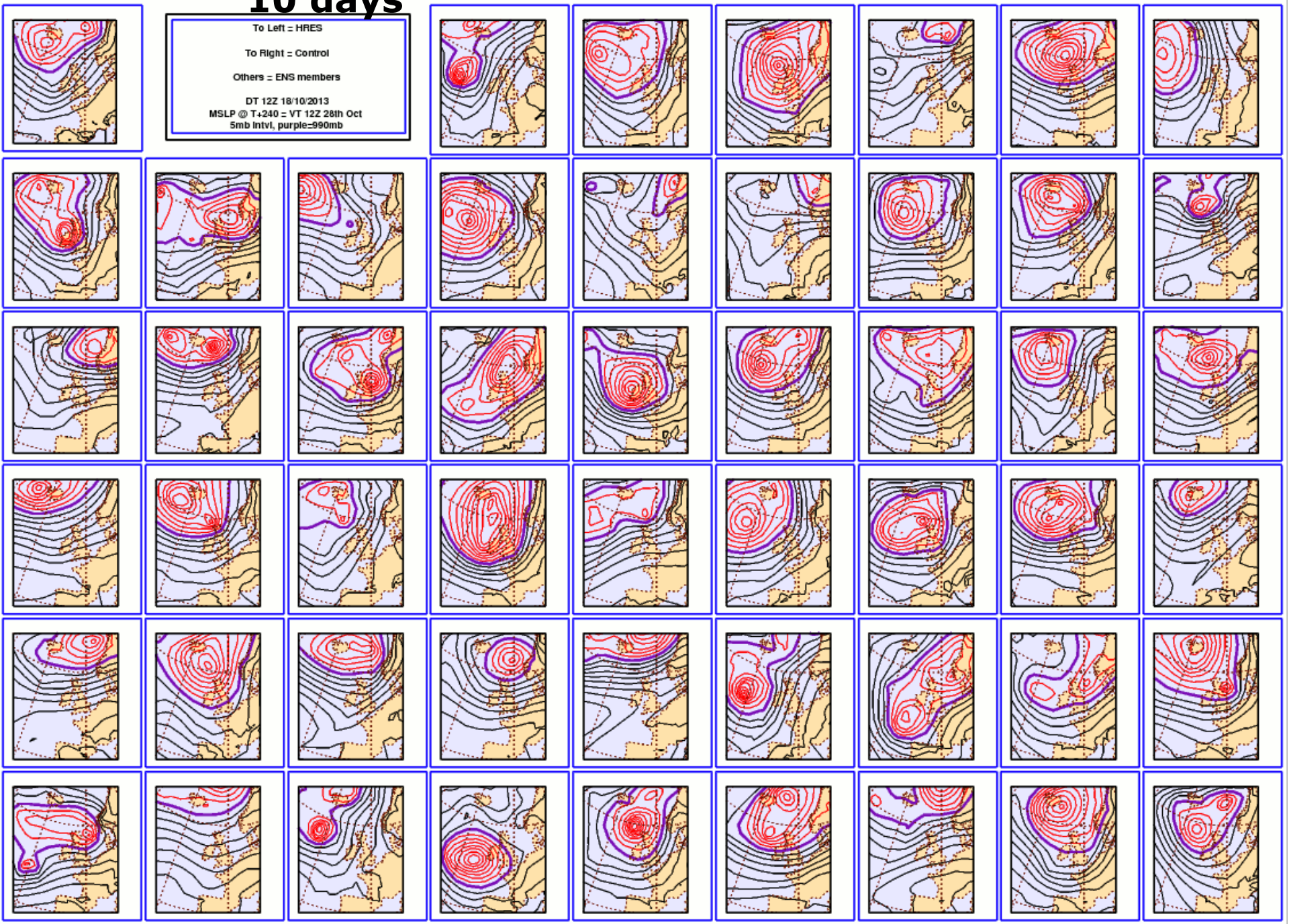
"Quality" Golf ball



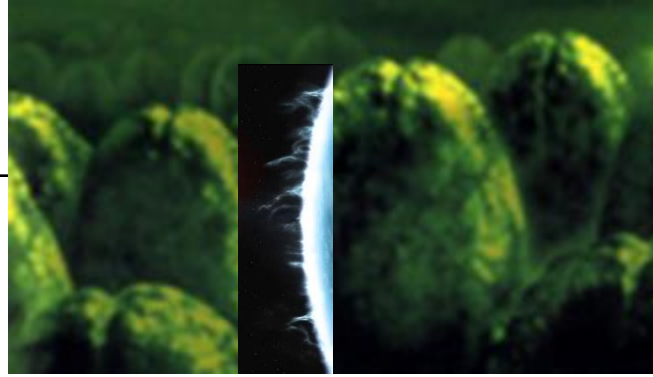
Forecasts from 18 Oct 2013
(10 days in advance)
- for 06 and 12UTC on 28 Oct 2013
(same contour colour scheme)

"Cheap" Golf balls

10 days


To Left = HRES
To Right = Control
Others = ENS members
DT 12Z 18/10/2013
MSLP @ T+240 = VT 12Z 28th Oct
5mb intvl, purple=990mb

I've got, that sinking feeling…



When footnotes and subtle boxes appear to aim for "plausible deniability".

Probability in Weather and Climate

There are many different kinds of probability.

Two Properties of a Probability Forecast:

- **Actionable**: useful in decision making using the tools taught in DT101
- **Mature**: encapsulates all the information implied by our knowledge. More compute power is not expected to alter a mature probability, but of course new observations or new insights (knowledge) are.

Kelvin's Gambit: "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."

William Thomson (1862)
Macmillan's Magazine **5** 388

This clarification (conditioning on **I**) is a source of strength, not of weakness.

It is nothing more than covering the possibility that P(data | **I**) = 0, that is, that "**I** is False" in some important manner.

Should you fear Senator Inhofe's List?

U. S. Senate Report
Over 400 Prominent Scientists Disputed Man-Made Global Warming Claims in 2007
Scientists Debunk "Consensus"

U.S. Senate Environment and Public Works Committee
Minority Staff Report (Inhofe)
www.epw.senate.gov/minority

Released: December 20, 2007

cyclopedia

nhofe (/ˈɪnhɒf/; born
senior United States
nd a member of the
ected to the Senate in

n Environment
vities cause
ce, Chris
the science of

reviously served as
Representative and

[hide]
business career

2.1 State legislature
2.2 1974 gubernatorial election

Jim Inhofe

United States Senator
from Oklahoma

Do you fear things like "Senator Inhofe's List"?

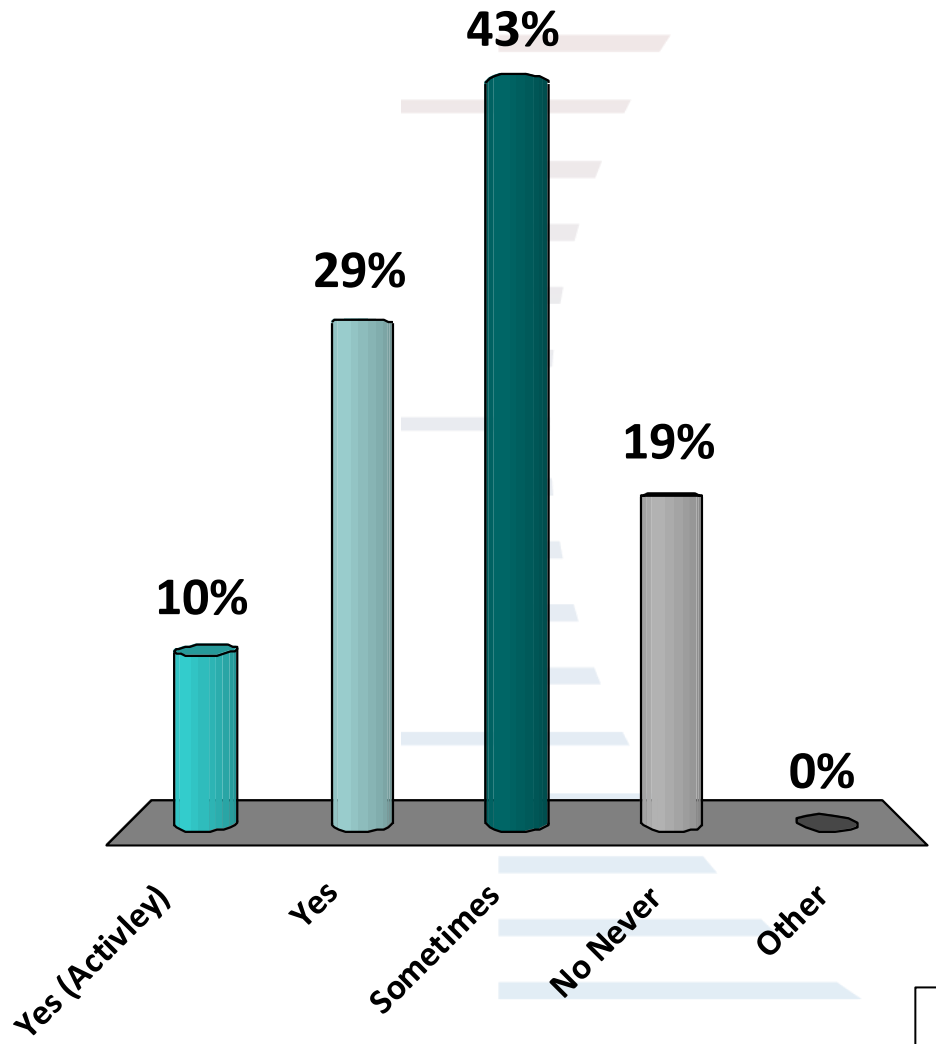
A. Yes (Actively)

B. Yes

C. Sometimes

D. No Never

E. Other



Q9.1

How can we know our simulation models are inadequate?
Science is more than simulations

When does "Sit and Think" trump "Simulate and Count"?

Example: When we know moist air must go over or around in (and only in) the real world!

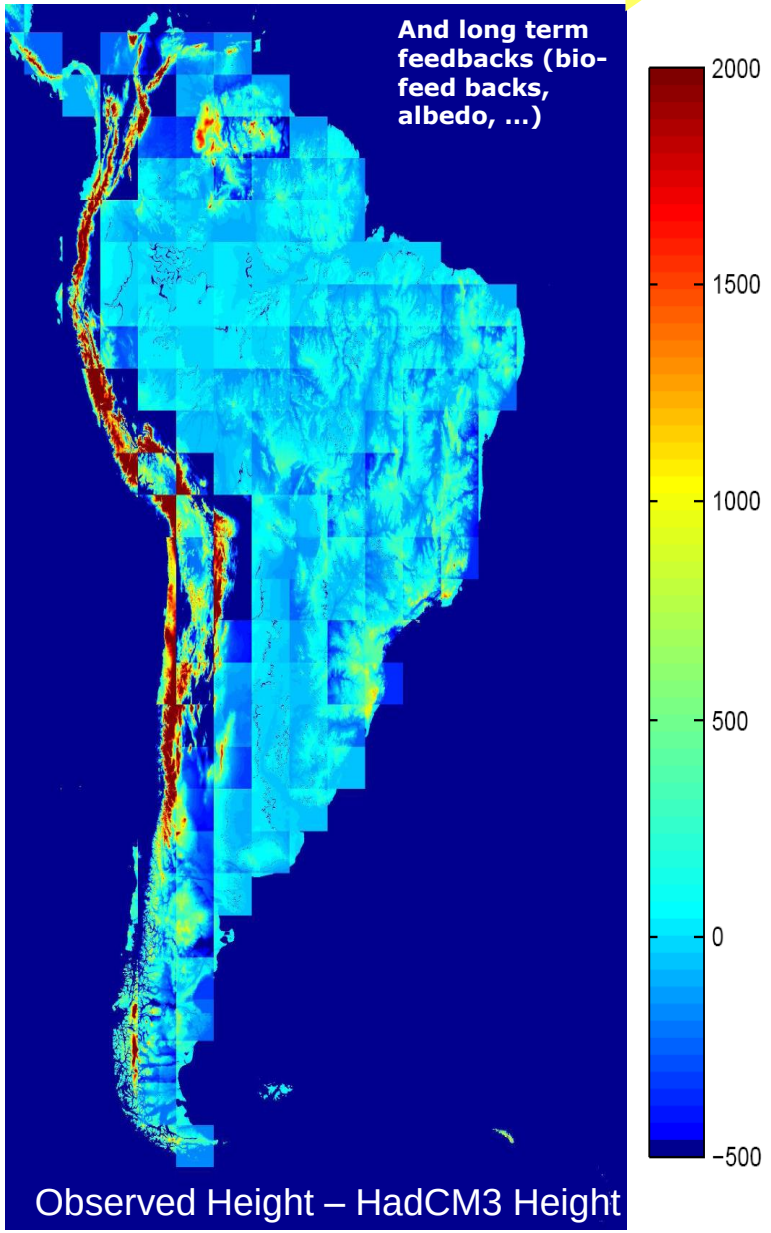
If our models cannot reproduce today's driving meteorological phenomena, can we expect them to get second order feedbacks "well enough"?

One-way coupled regional models *cannot* account for missing physics or inactive feedbacks.

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

Why not provide Prob(Big Surprise) with lead time?

Missing 2km tall walls of rock.



A genuine expert can always foretell a thing that is 500 years away easier than he can a thing that's only 500 seconds off.
- *A Connecticut Yankee in King Arthur's Court*

What is a "Big Surprise"?

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

Often we can identify 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 I in $P(x|I)$ changes. (Arguably **"Bayes" does not apply**: this is not a question of probability theory.)

In weather forecasting, we can see when our models become silly, but in climate forecasting we are in the dark.

If our models agreed (in distribution) would we have more confidence in their simulations?

Decadal Forecasting Skill

Take last years GMT.

Distribute the last one hundred years amongst yourselves (each of you will get about three years to care for)

Compute the one-year first-differences in GMT for your years.

Take everyones first-difference(s) and add each to this years GMT to form an ensemble of values.

Form a probability forecast for next year's GMT from this ensemble

Compute the two year differences in GMT for your years.

… Form a probability forecast for GMT in 2016

And so on…

Suckling, E.B. and Smith, L.A. (2013) 'An evaluation of decadal probability forecasts from state-of-the-art climate models', Journal of Climate, 26 (23): 9334-9347.

http://www.cru.uea.ac.uk/cru/data/temperature/CRUTEM4-gl.dat

Probability Forecasts in Climate Science

Uncertainty and the IPCC Sixty-Forty Rule

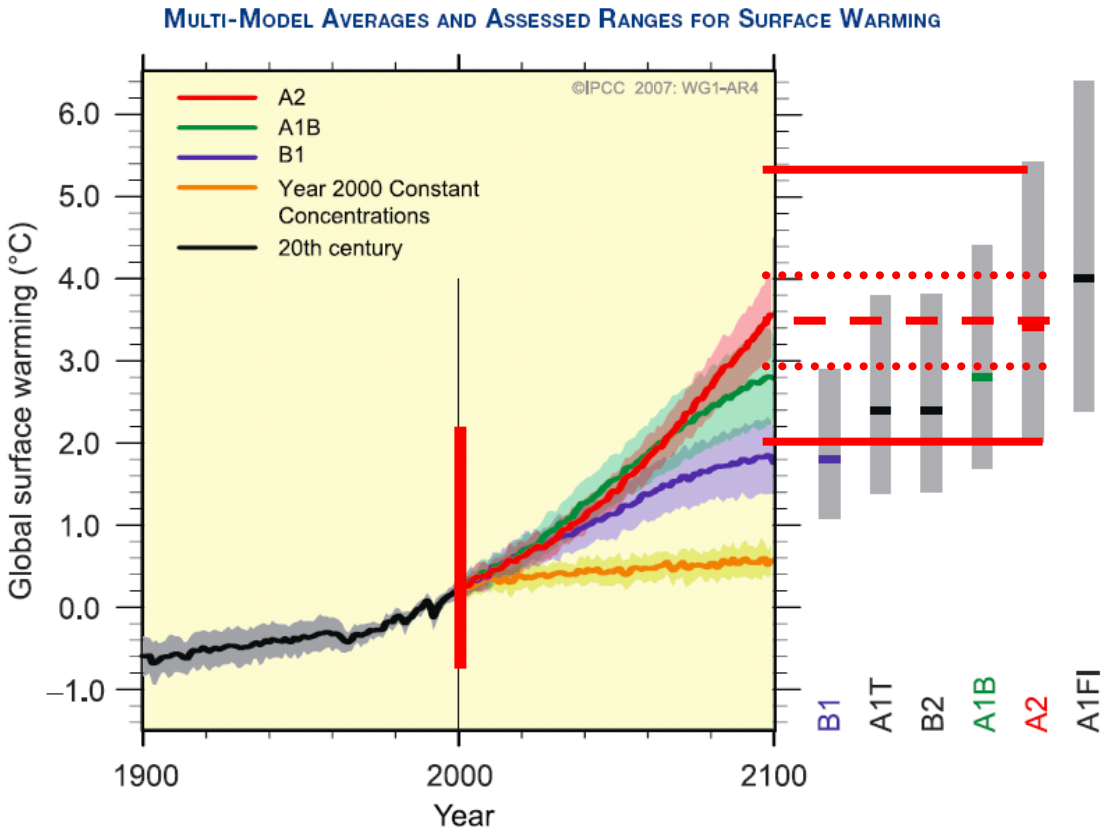
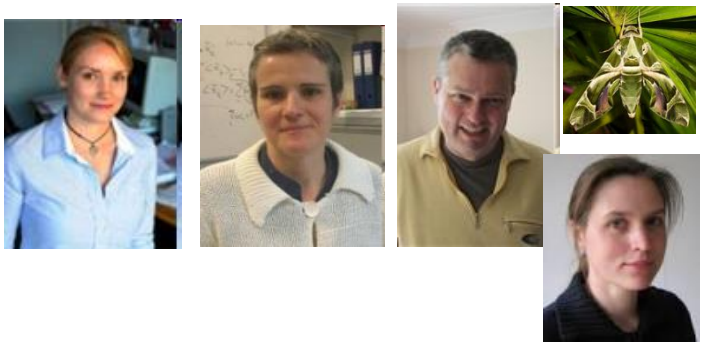


Figure SPM.5. *Solid lines are multi-model global averages of surface warming (relative to 1980–1999) for the scenarios A2, A1B and B1, shown as continuations of the 20th century simulations. Shading denotes the ±1 standard deviation range of individual model annual averages. The orange line is for the 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 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 have rational confidence in model-based probabilities?

The uncertainty here is not in "x" but in the structure of the model.



The IPCC rejects the diversity of ensembles directly reflecting the pdf of GMT, it follows that "downscaling" alone cannot provide local probabilities.

Skill in Decadal Forecasting

Suckling, E.B. and Smith, L.A. (2013) 'An evaluation of decadal probability forecasts from state-of-the-art climate models', Journal of Climate, 26 (23): 9334-9347.

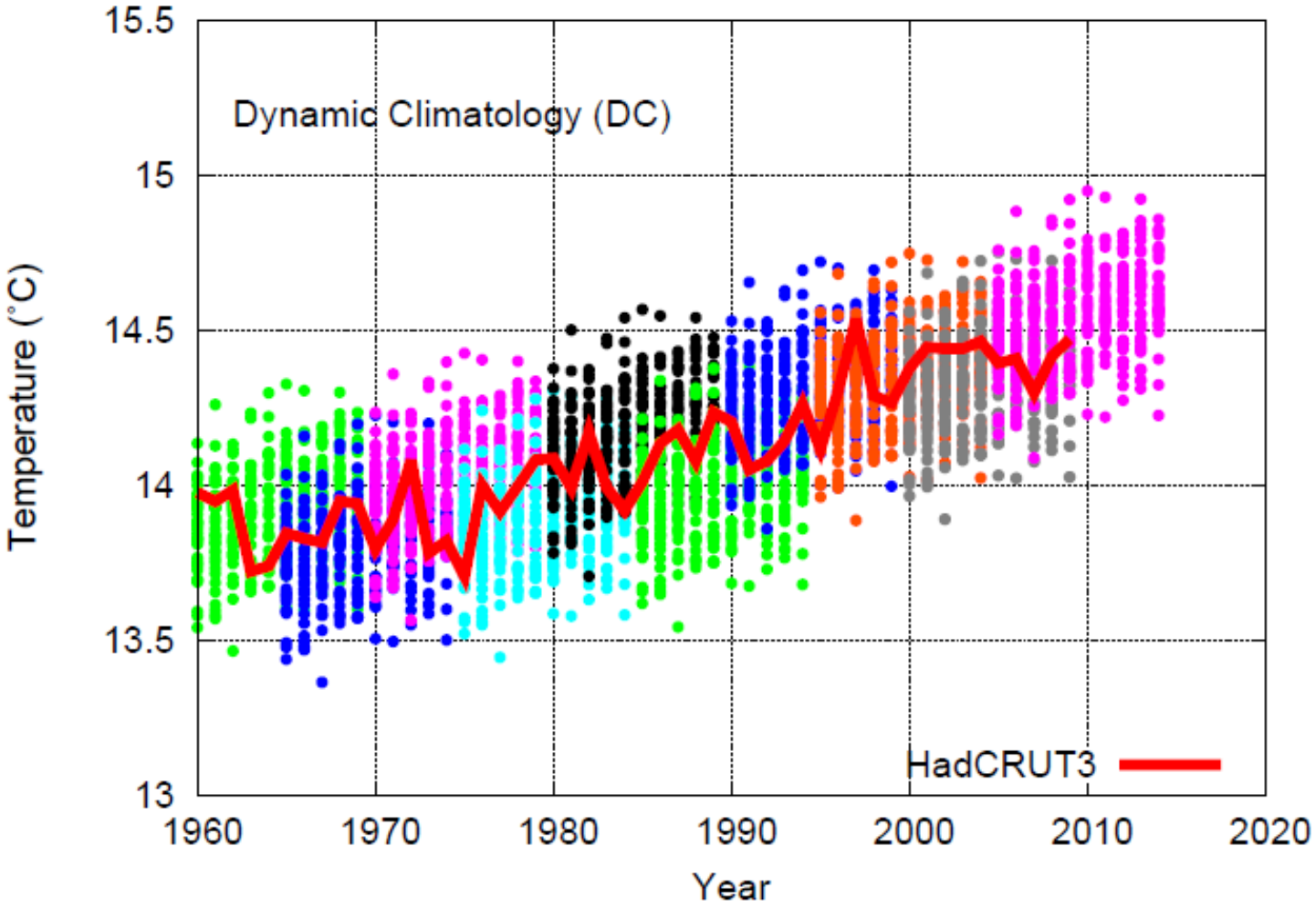
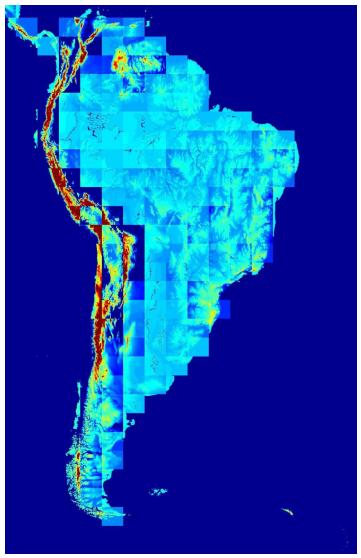


FIG. 3. Dynamic climatology (DC) over the period of the ENSEMBLES hindcasts (figure 1). HadCRUT3 (from which the DC model is constructed) is shown for comparison.

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



Spatial Scales

metres

km

1000km

If precip over the Amazon (or Okeefenokee) is poorly simulated, then a missing feedback may eventually lead to model irrelevance…
First local, then global…

The timescales for such things can be estimated using sound science!

Prob(Big Surprise)

Target Lead-time

hours weeks years decades centuries

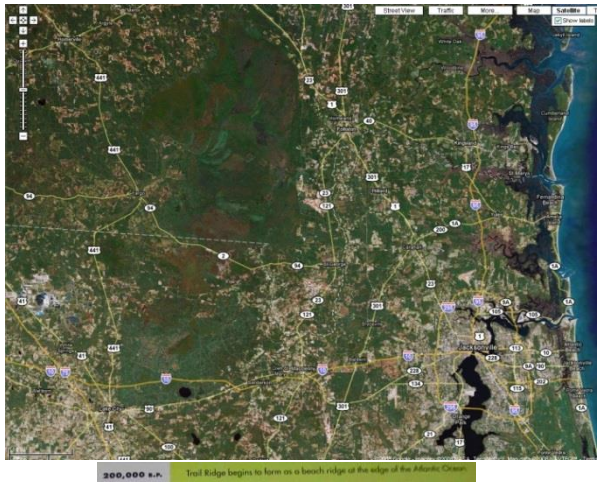
years

weeks

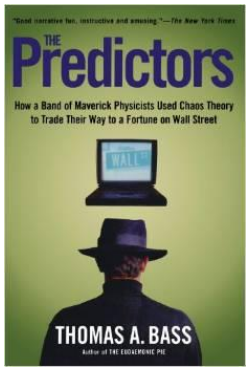
day

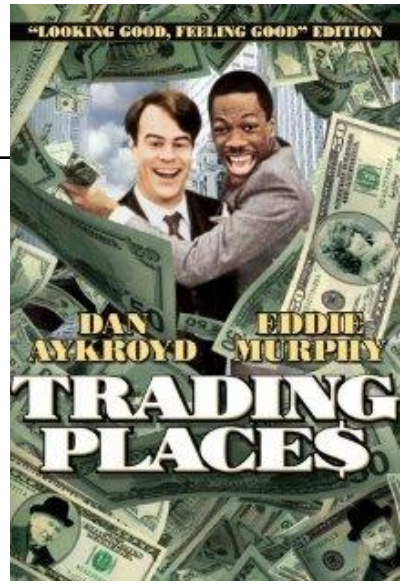
Temporal Average Scale

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



CATS CENTRE FOR THE ANALYSIS OF TIME SERIES





Knowing or Feeling

… 8 min early morning discussion on day-trading energy future given model forecasts…
Gary: "I don't see how it helps when the model does that."
Lenny: "It means that/" Gary: "Yes, you told me what it means."
Gary: "I don't think you have a good feel for it."
Gary: "Here, push speed dial 7, when he answers do 100 either way."
("100": that would be 100,000 barrels of brent sweet crude oil, "either way": buy or sell)

I am not sure what I did, but I clearly recall how it felt. (I think I bought.)
And I didn't even think of not taking the phone (not even until I made this slide)
instinctively it was clear that "not deciding" would have been the highest cost outcome.

Theiler, Crutchfield, Shaw, …

http://www.minyanville.com/trading-and-investing/stocks/articles/Insiders-Buy-at-Seattle-Genetics-CIT/5/19/2014/id/55020
http://www.amazon.co.uk/The-Predictors-Maverick-Physicists-Fortune/dp/0805057579

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.

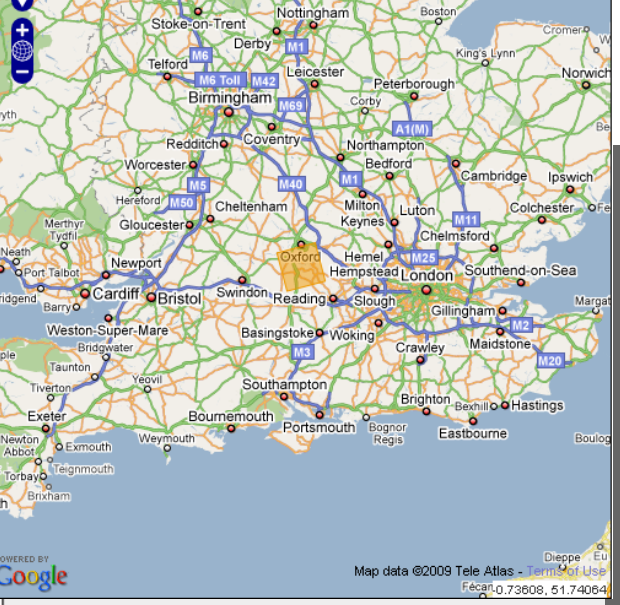
Climate-like forecasting tasks:

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 no true out-of-sample observations.

Best practice principles of forecasting **differ** in these two settings.

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



- Future Climate Change Only
- Future Absolute Climate Values

Variable

- Change in mean temperature (°C)
- Change in mean daily maximum temperature (°C)
- Change in mean daily minimum temperature (°C)
- Change in temperature of the coolest day (°C)
- Change in temperature of the warmest day (°C)
- Change in temperature of the coldest night (°C)
- Change in temperature of the warmest night (°C)
- Change in precipitation (%)
- Change in precipitation on the wettest day (%)
- Change in mean sea level pressure (hPa)
- Change in total cloud (%)
- Change in relative humidity (%)
- Change in specific humidity (%)
- Change in net surface longwave flux (W m-2)
- Change in net surface shortwave flux (W m-2)
- Change in total downward surface shortwave flux (W m-2)

UK CLIMATE PROJECTIONS USER INTERFACE

http://www.ukcip.org.uk/

Start Page | My Jobs | My Details | Using UKCP09 | UI Manual | Need help?

Logged in as: lenny@maths.ox... Logout

Logged in users: 2

You have no pending jobs. See **My Jobs** for previously run jobs.

Request Status:

Request Summary:

Selecting your UK location first

This page is intended for novice users of the UI who know what location they are interested in. This page should be used as follows:

Step 1: Click on a point on the map (or type in the latitude/longitude coordinates and click "Select".
Step 2: Select a data source of interest from the list that appears on the right.
Step 3: Select the variable you are interested in and click the "Next" button.

You can search by place name or postcode using the box on the right-hand side. Note that clicking a result re-centres and zooms the map to the new location but does make a selection.

Selections on this page are restricted in that only a single location may be selected. Weather Generator simulations and Marine Model Simulations are not available from this start point.

Read about starting your request by making spatial selections in the UI Manual.

Next

Search place name or postcode to re-centre map:

ox1 1dw Search Clear

Postcode: OX1 1DW

Select by Latitude / Longitude by:

Latitude: 52.0018
Longitude: -0.1044

Select

Step 2: Select a data source

At your chosen location, there is data for following data sources (clicking an option will highlight the selected location on the map adjacent):

- UK Probabilistic Projections of Climate Change over Land for the 25km Grid Box with the ID: **1551**
- UK Probabilistic Projections of Climate Change over Land for the Administrative Region: **East of England**
- UK Probabilistic Projections of Climate Change over Land for the River Basin: **Anglian**

Step 3: Select a variable

Please choose one of the following variables:

Next

Funded by: defra | Department of Energy & Climate Change | Department of the Environment | The Scottish Government | Llywodraeth Cymru Welsh Assembly Government

Provided by: Met Office Hadley Centre

UK Climate Impacts Programme

Service hosted at: Science & Technology Facilities Council, Rutherford Appleton Laboratory.

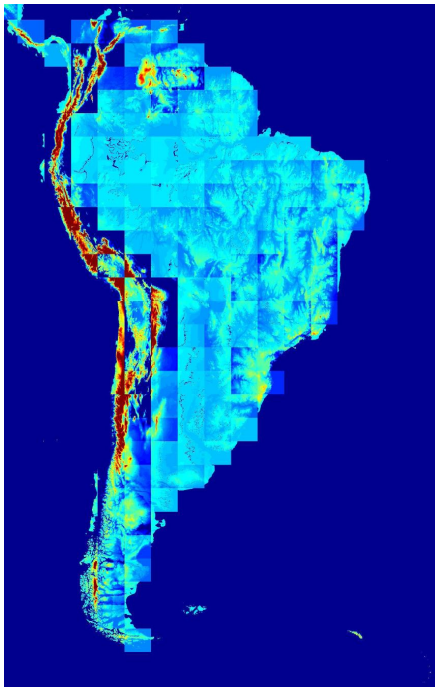
Communicating the Relevant Dominate Uncertainty

"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 well modelled or not***, would aid the use of projections in decision support.

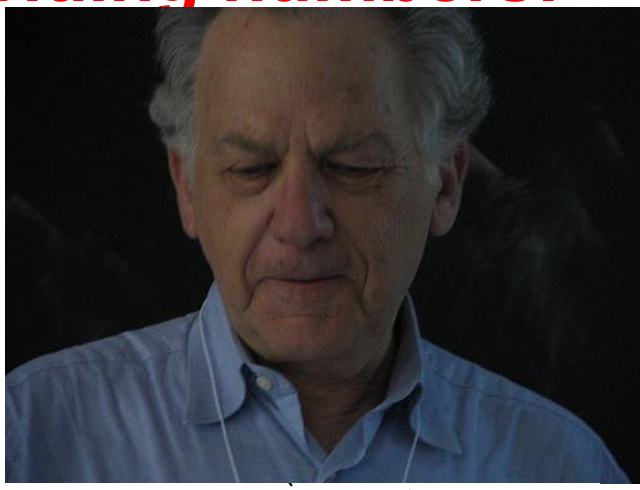
Alternatives better than this probability of a big surprise would be very welcome.

Thank you

Models can aid insight, without providing numbers!

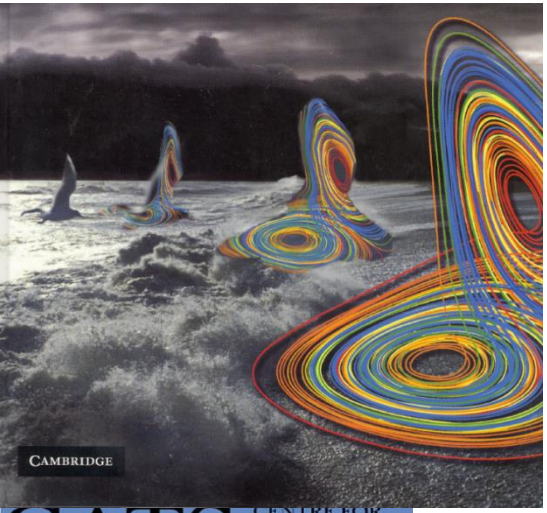
The events are connected with one another. Remember the charming, if somewhat irrelevant, analogy used by explicators of chaos theory: A butterfly flutters its wings in Beijing and next month there is an earthquake in Chile.

The Western Australian (1994)



We are therefore led to conclude that a great number of phenomena observed in variable stars can be explained by the instability mechanism discussed by Eddington, once non-linear, non-adiabatic solutions can be found. We feel, however, that progress in this direction can be made through the study of elementary prototype equations perhaps more closely related to the stellar model than ours, but hopefully not more complicated.

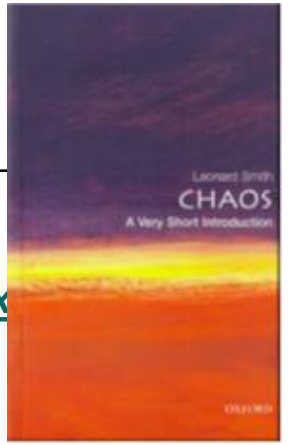
There remains the question as to whether our results really apply to the atmosphere. One does not usually regard the atmosphere as either deterministic or finite, and the lack of periodicity is not a mathematical certainty, since the atmosphere has not been observed forever.



Decision makers often act on insight.

(Remember: Du got a seat on the that train…)

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES



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

Page contents > *Papers | Conference & Workshop Proceedings | Recent Abstracts | In Production and/or*

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

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.

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

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.

Brian Hoskins fears throwing out the baby with the bathwater.
So did I, as a grad student in 1988 and then again in 2002.
But *if* the bathwater becomes rancid we need to save the baby.

So what do Decision makers want?

In my experience their main aim is achieve a better outcome, ideally by making better decisions. You really should ask yours, but I expect they would like:

a) to trust their level of trust in their source(s) of information

b) to learn from their mistakes (in weather-like situations)

c) to cover their posteriors

(That is a rank ordered list, but only to this point!)

d) not get bogged down in the details between 9 & 5 , unless it aids execution.

e) to recognize where they are naïve. (Monte Hall Problem)

f) to know tell-tale early-warning signals; the costs/benefits of delay/steping back.

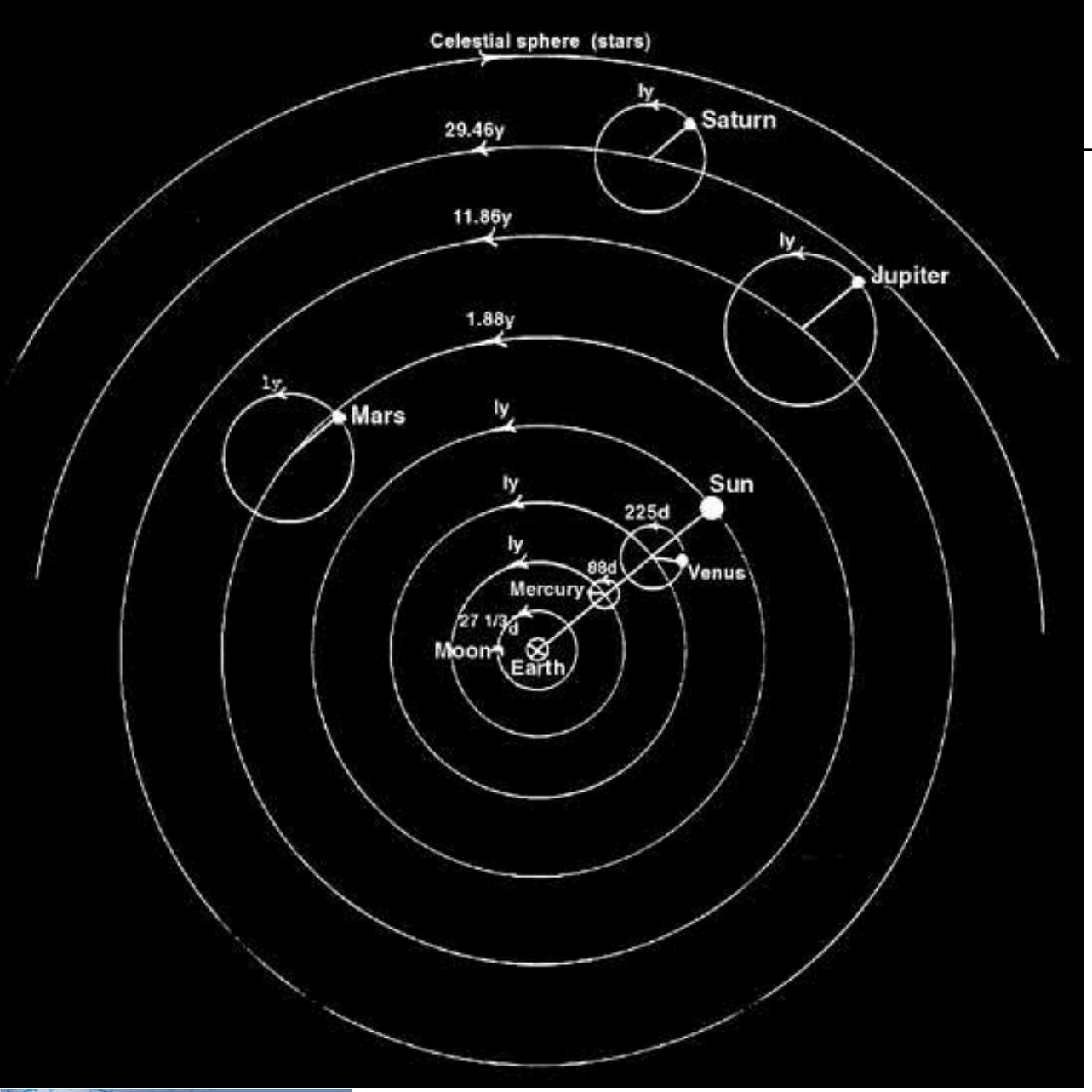
g) to know where the exit is (this is more than "plan B") .

h) to understand those decisions which lead to poor outcomes.

i) tools to do what they do better (with out fear of being replaced by a machine)

What else belongs in this list?

On Wed: What can environmental modellers provide?

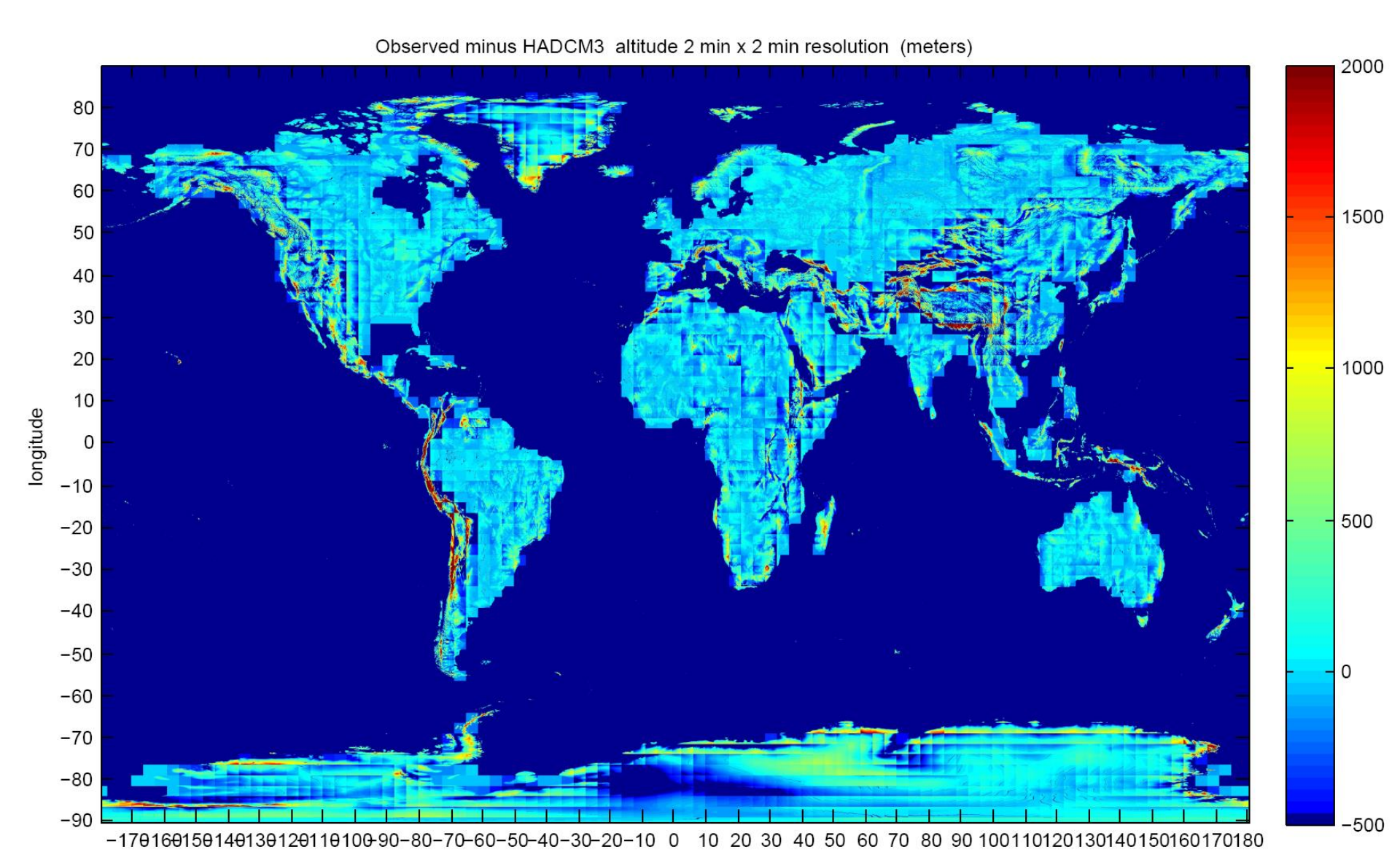


What is the correct diameter/number of epicycles for Mars?

What is the correct entrainment rate (ice fall velocity) for HadAM3?

Imprecision of an unknown quantity or indeterminacy of an empirically vacuous fiction?

What is your question?
Really:
Your question?



This map shows what is missing in HadCM3
Including 2km walls of rock.
All 2012's GCMs suffer from related inadequacies

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

Bayesian's Burden

Rev. T. Bayes

Take up the Bayesian's burden,
Your best students send out,
To give each and every science,
It's PDF of quantified doubt.

Sacrifice theoretical advances
In maths, your career may cease,
To help doubters in the darkness
Find their distributions and peace.

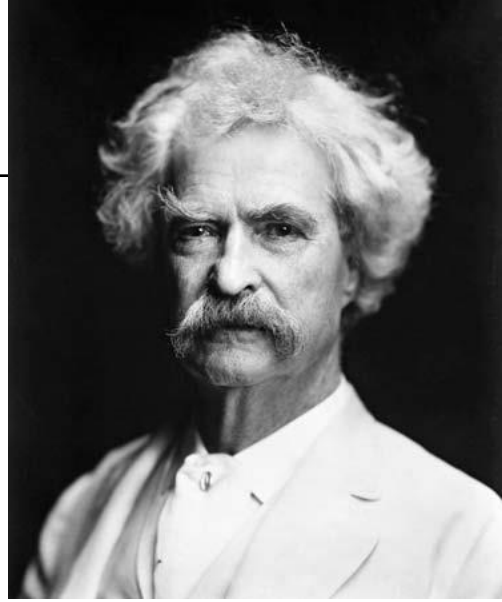
In the dreary halls of physics,
Encapsulate their beliefs,
Their model's empirically inadequate,
Still only B's way gives coherent release.

Extract priors without mercy,
It is the only way,
The numbers must mean something,
Whatever the captives say!

Allow him his posterior only
Not his heart, certainly not his head;
Constrain the result with priors,
Before the data's been read.

Then free him to act blindly,
As his posterior says he should,
Once he finds a utility function,
All will be well and good.

(with apologies to sam)



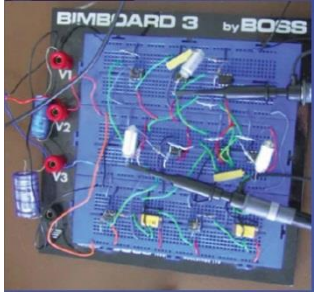
To the Physicist Sitting in Darkness

Probabilities are all well and good. And it is a fine thing to get in touch with your beliefs and feelings. Shall we bang ahead in our old-time, loud pious way, and commit new sciences to the game; or shall we sober up, sit down, and think it over first?

The Blessings-of-Subjective-Probability Trust, wisely and cautiously administered, is a Blue Chip. But Bayesians have been playing it badly of late, and must certainly suffer from it, in my opinion; they have been eager to solve every problem, especially the poorly posed ones, and the Physicists who sit in Darkness have begun to notice it – they have noticed it and have begun to show alarm. They have become suspicious of posteriors on empirically vacuous reals, not to mention function spaces; they have begun to resist the kindly extraction of priors. More – they have begun to examine them! This is not well. The Blessings of Bayesianism are all right, and a good NSF commercial property; there could not be better, in a dim light. In the right kind of light, and at the proper distance, with the goods a little out of focus, they are a desirable enticement to the Physicists who sit in darkness.

Probability theory eases the stress of decision making. And improves the outcome, but not if we adulterate it. For the Empirically Adequate and the Large Number Statistic, it is pie. But in cutting edge science, and in extrapolation, here the Physicist sitting in darkness is (almost) sure to say: “These is something curious about this – curious and unaccountable.” … There have been lies yes, but told in a good cause, it might have worked; yet we have passed on a Shadow from one who hadn’t it to sell, and long term infrastructure investments are being made.

Take home questions



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

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?

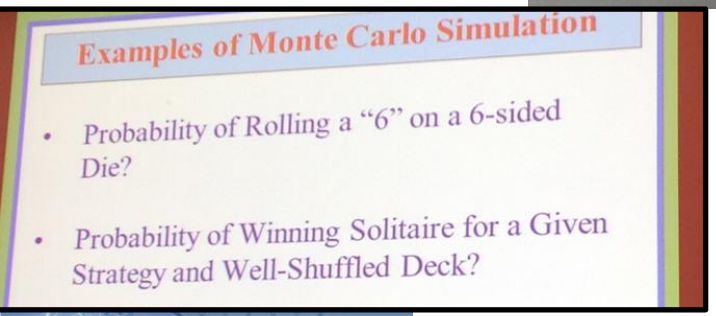
How to distinguish the value of improvement from the utility of prediction?

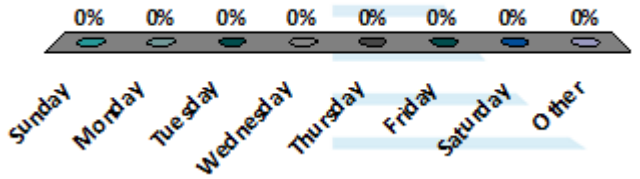
Might the provision of probability be maladaptive?

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?

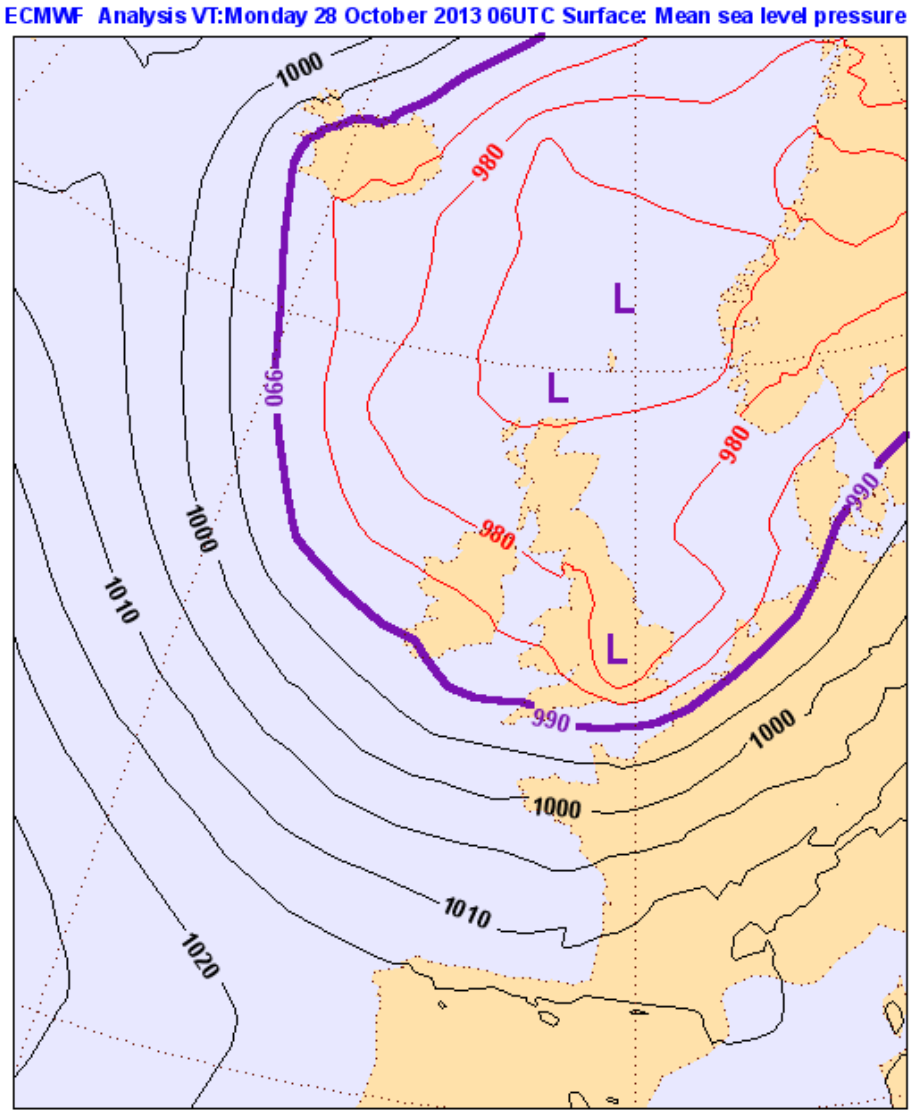
What is the Probability that "Eight 6-sided fair die sum to an even number"

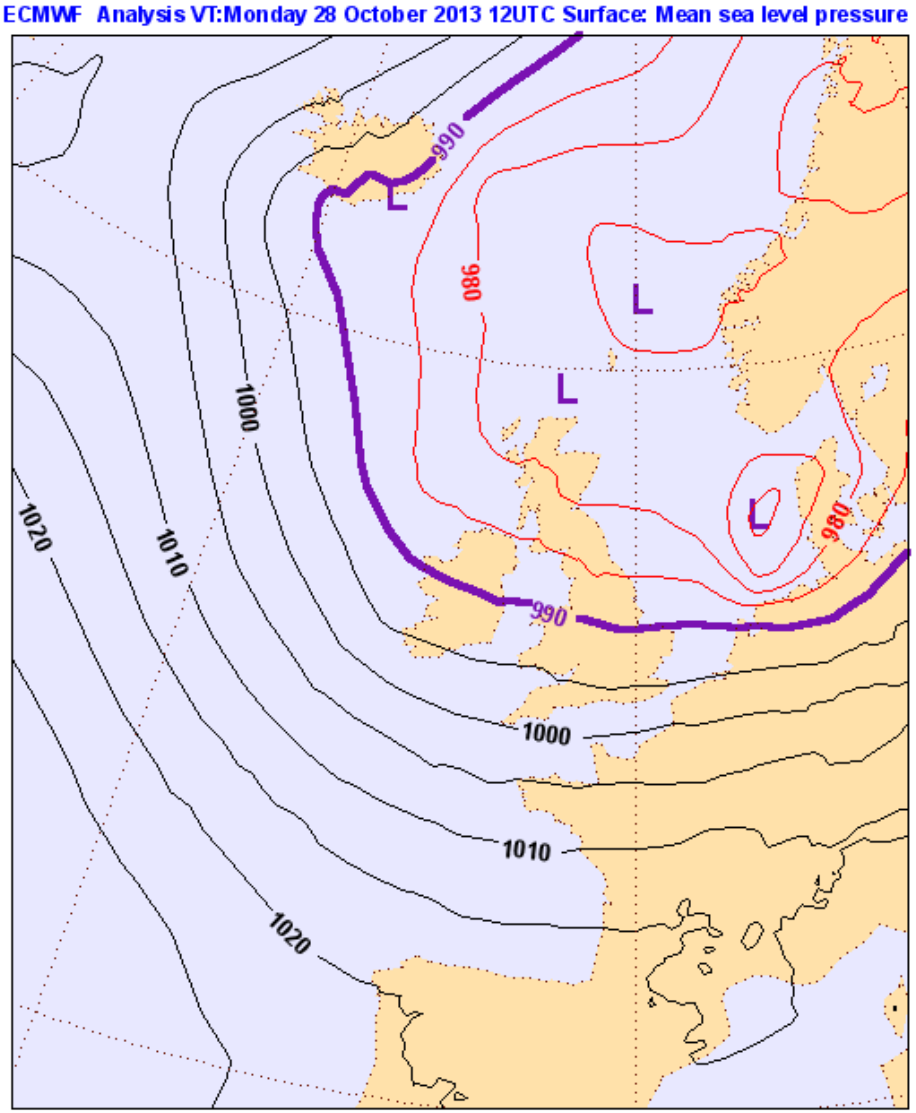




Mistaking an immature probability for a tautological probability can be costly.

< 970 : 4 red contours below the purple 990

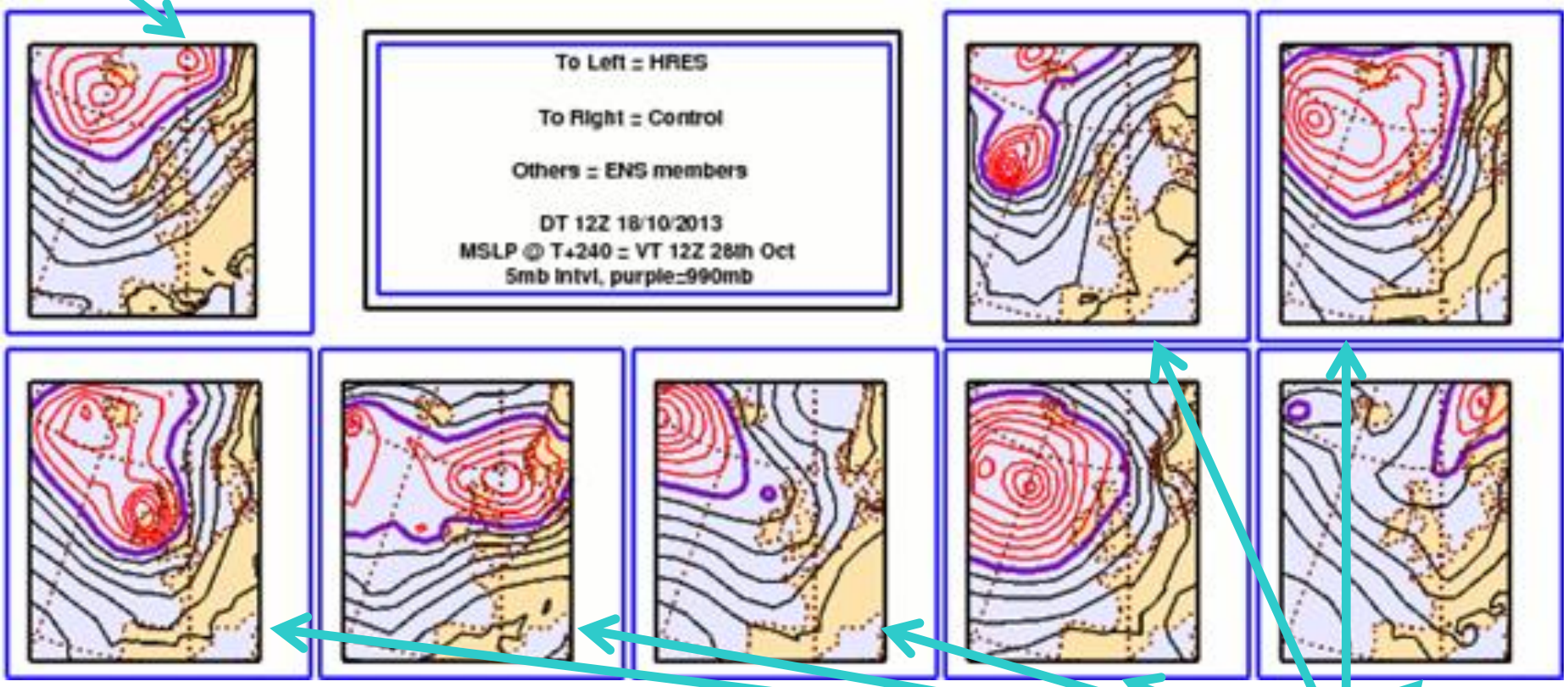




Foreseeing the St Jude Storm

(golf balls)

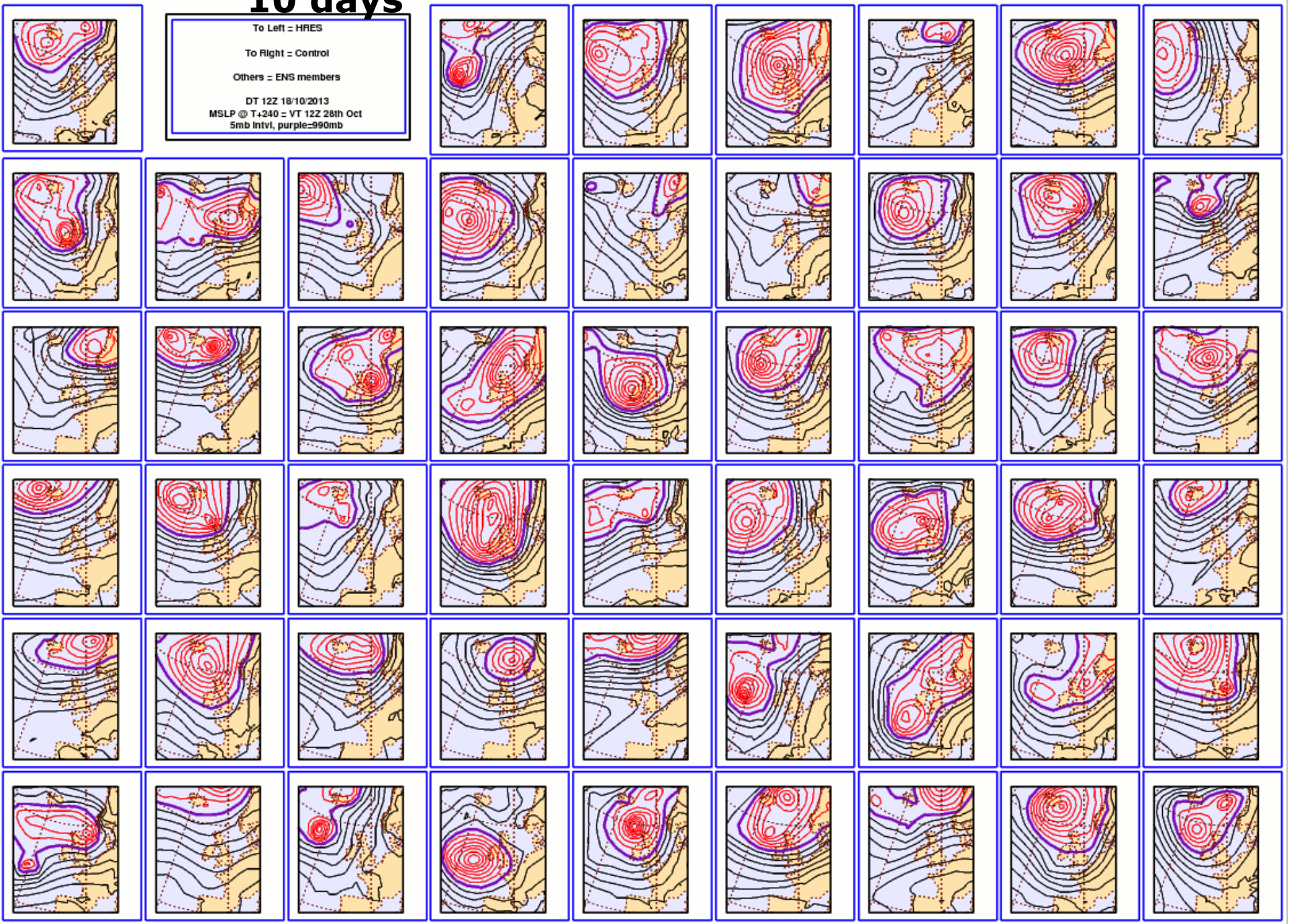
"Quality" Golf ball



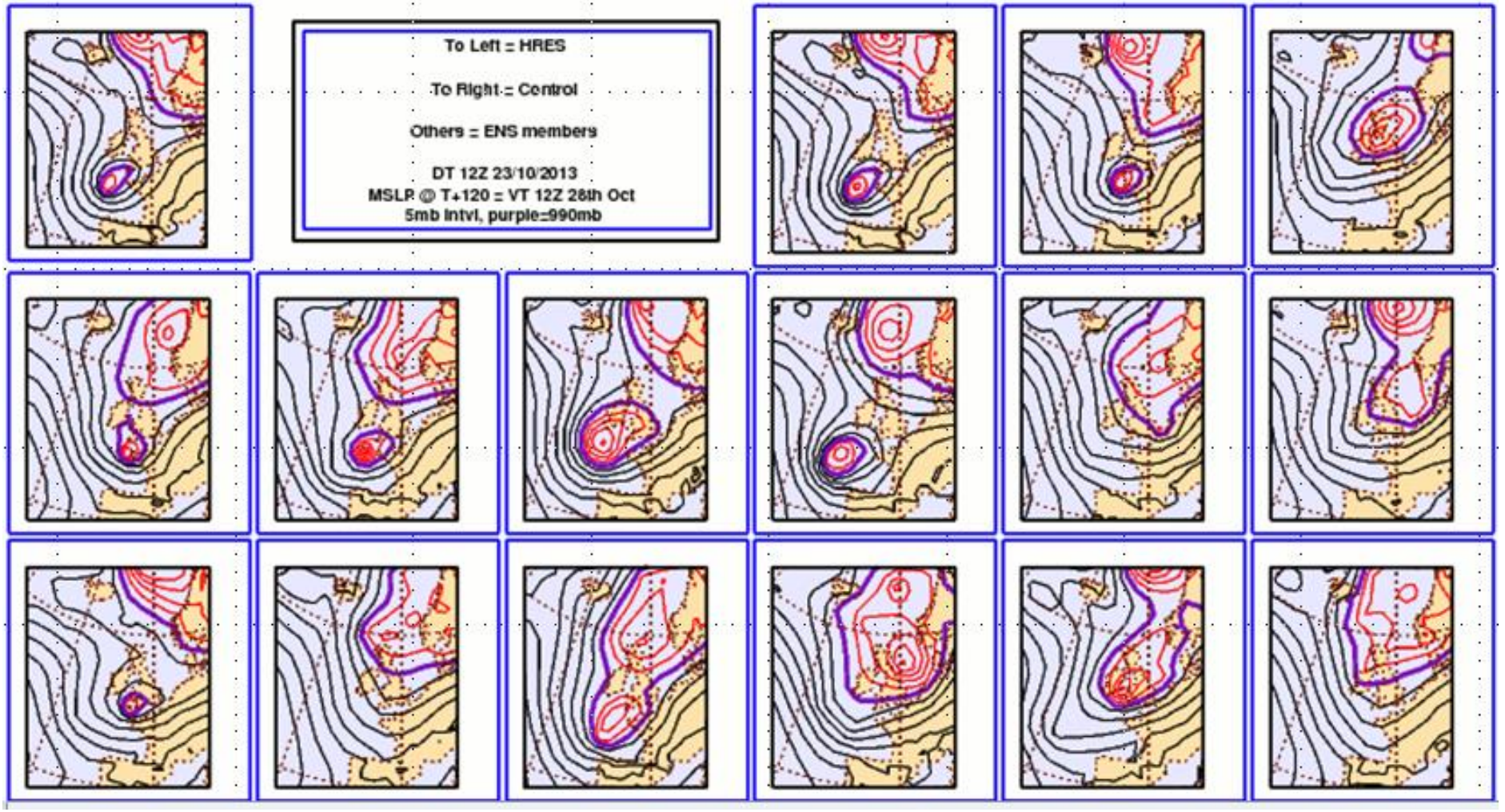
Forecasts from 18 Oct 2013
(10 days in advance)
- for 06 and 12UTC on 28 Oct 2013
(same contour colour scheme)

"Cheap" Golf balls

10 days


To Left = HRES
To Right = Control
Others = ENS members
DT 12Z 18/10/2013
MSLP @ T+240 = VT 12Z 28th Oct
5mb intvl, purple=990mb

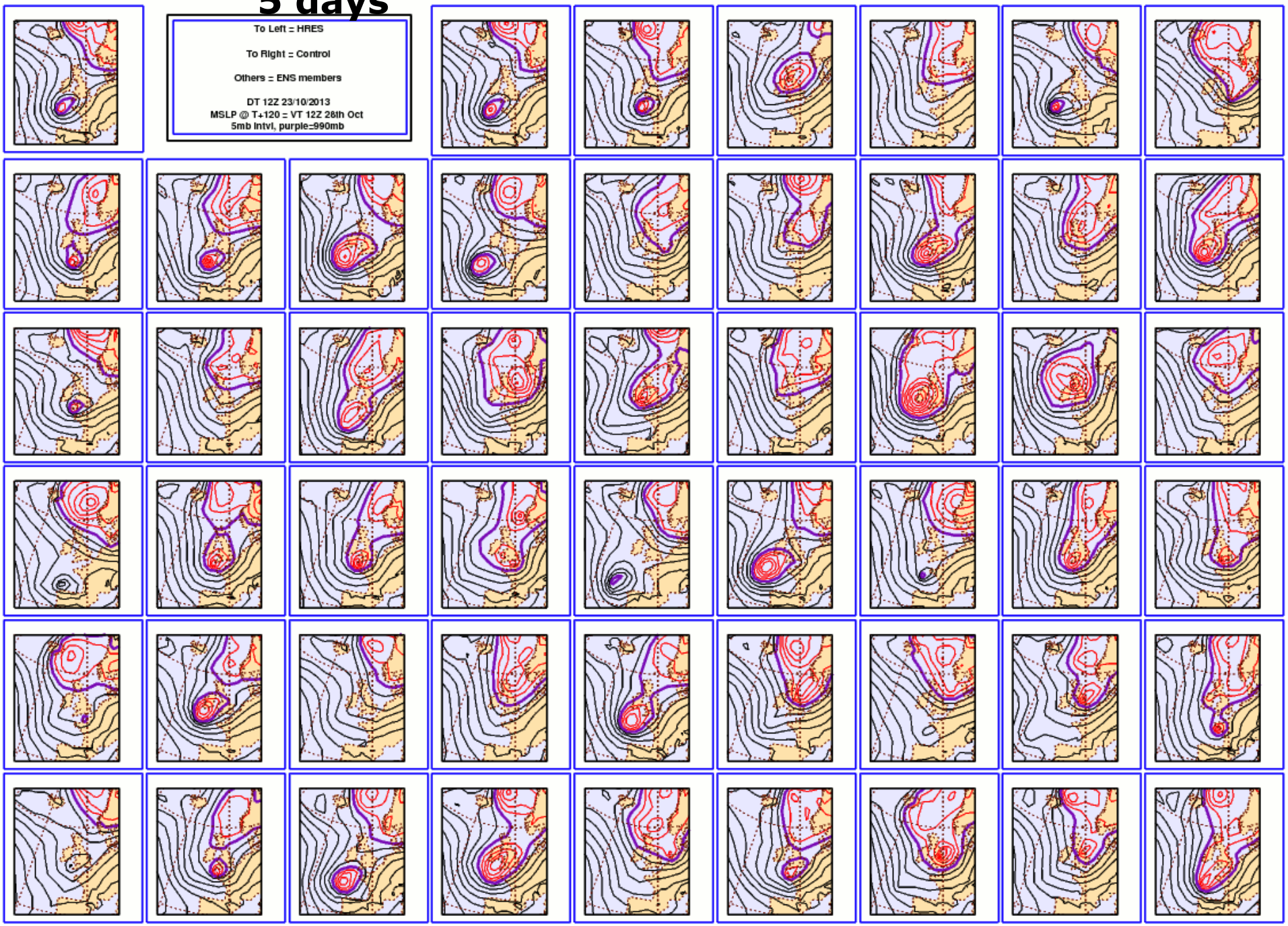
Foreseeing the St Jude Storm



Forecasts from 12UTC 23 Oct 2013
(5 days in advance)

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

5 days


To Left = HRES
To Right = Control
Others = ENS members
DT 12Z 23/10/2013
MSLP @ T+120 = VT 12Z 28th Oct
5mb Intvl, purple=990mb

END

Watch out for the Penguin Effect

The challenge of climate change will be with us for some time.

Can we maintain parallel streams: pure research to apply in 2050, and applied research to improve the modelling position we are in when we get there?

When selecting a thesis problem: do you suggest something important, like understanding cloud dynamics (better), blocking, circulation change?

Or to be the first person in the world to include the penguin effect in a global model? (and thereby all but assured a job at a rival modelling centre?)

(Similar effects plague economics and statistics)

THERE IS NO PENGUIN EFFECT

(My prior on this effect is zero)

It is a joke regarding climate,

but sadly not career paths!

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

Herman Rubin (1970)

Varieties of **I** include

G : True, complete knowledge of the world **Actionable**

g_t : a subset of G, incomplete/imprecise but wholly True **Actionable**

g : a set of useful beliefs based on G seen through a glass, darkly **???**

Probability in Weather and Climate

There are many different kinds of probability.

Two Properties of a Probability Forecast:

- **Actionable**: useful in decision making using the tools taught in DT101
- **Mature**: encapsulates all the information implied by our knowledge. More compute power is not expected to alter a mature probability, but of course new observations or new insights (knowledge) are.

Kelvin's Gambit: "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."

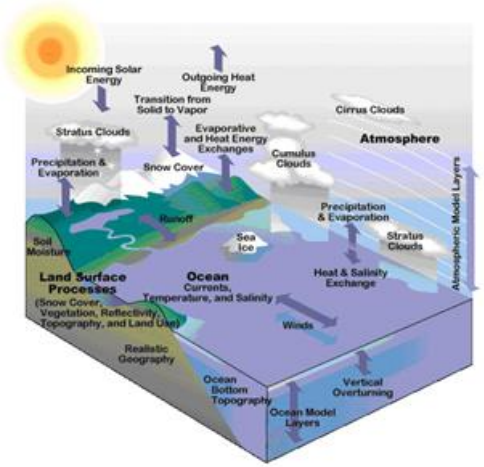
William Thomson (1862)
Macmillan's Magazine **5** 388

This clarification (conditioning on **I**) is a source of strength, not of weakness.

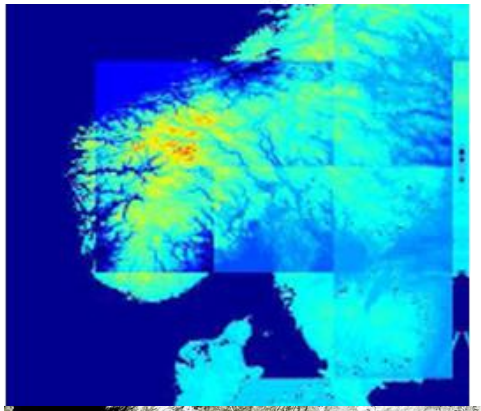
It is nothing more than covering the possibility that P(data | **I**) = 0,
that is, that "**I** is False" in some important manner.

Plausible Planets or Implausible Earths?

How can we best develop our models as the available computational power increases?

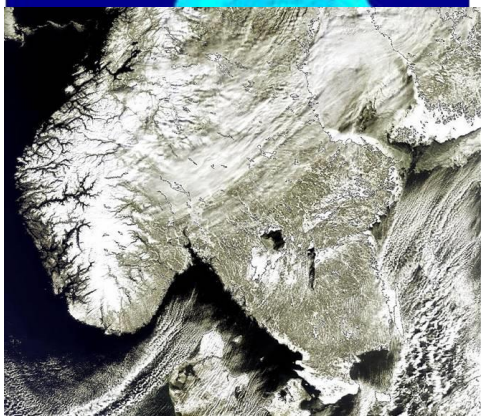


The kitchen sink approach "includes" a hodgepodge everything we can think of that might be important.
At best, this yields an implausible Earths in a sample of unphysical, unbiological, uninteresting & irrelevant model diversity which all expected to suddenly become Earth-like at some resolution in an ill-defined higgledy-piggledy way.



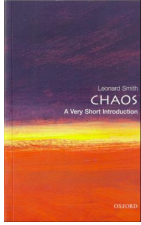
An alternative is to simulate (and evaluate empirically) potentially real planets that get more and more Earth-like while omitting any Earth-relevant process for which the model cannot provide coherent physical drivers on Earth-like scales.
(No suggestion of linear superposition!)

Does water vapour come after mountains?
Does vegetation come after water vapour?
Do we avoid the penguin effect?
(until it is simulated realistically)
?Termites?



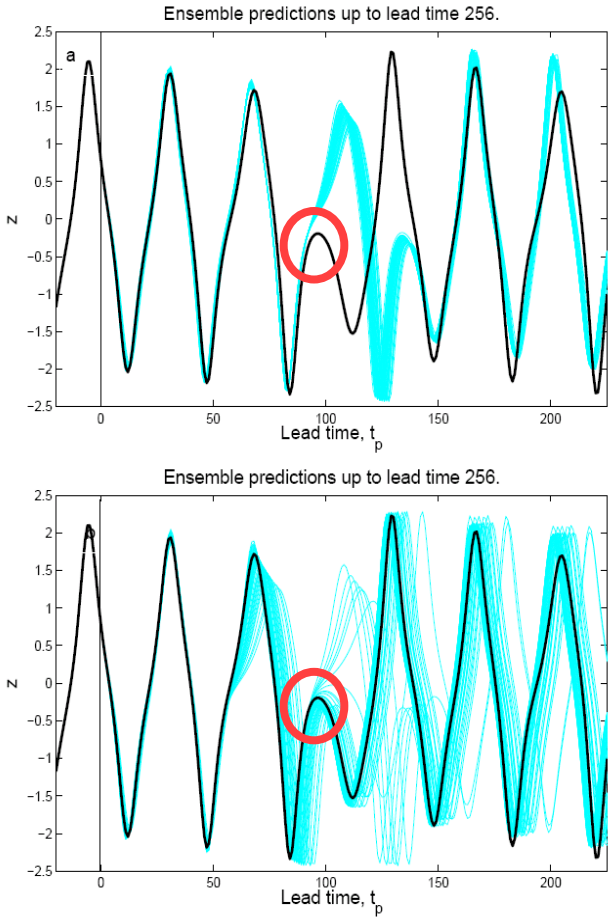
One might argue physical intuition is more effective in evaluating plausible planets, as there is physics to intuit in that case.
(and at least a few examples.)

But should be even be aiming at Probabilities?



Model 1

Model 2



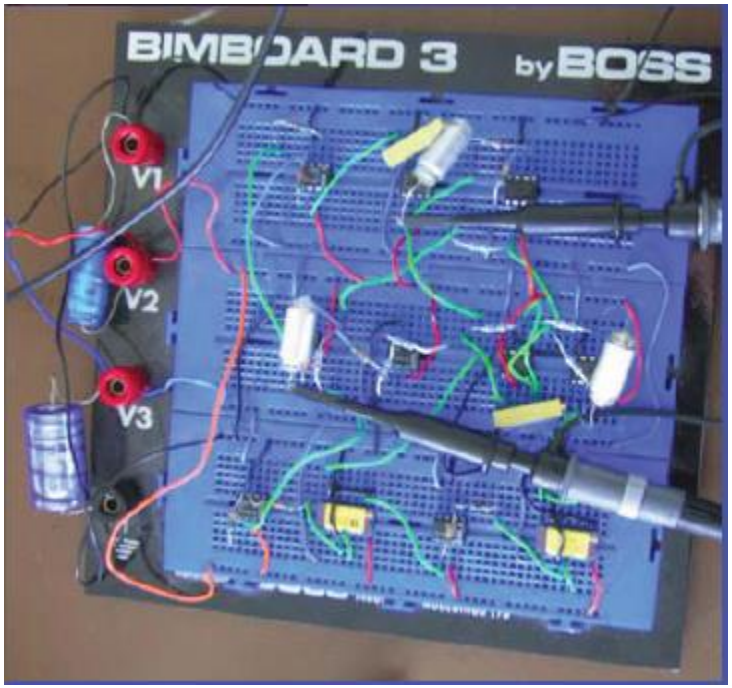
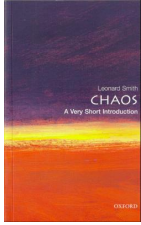


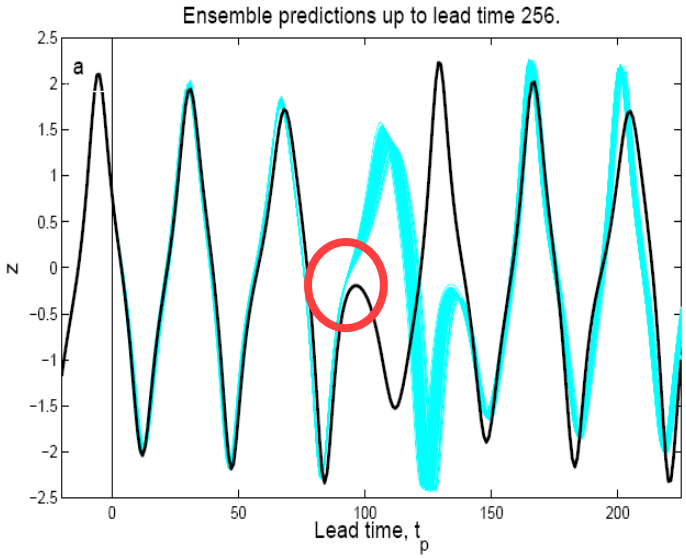
Figure 7: Ensemble predictions using (a) model 1 and (b) model 2. The

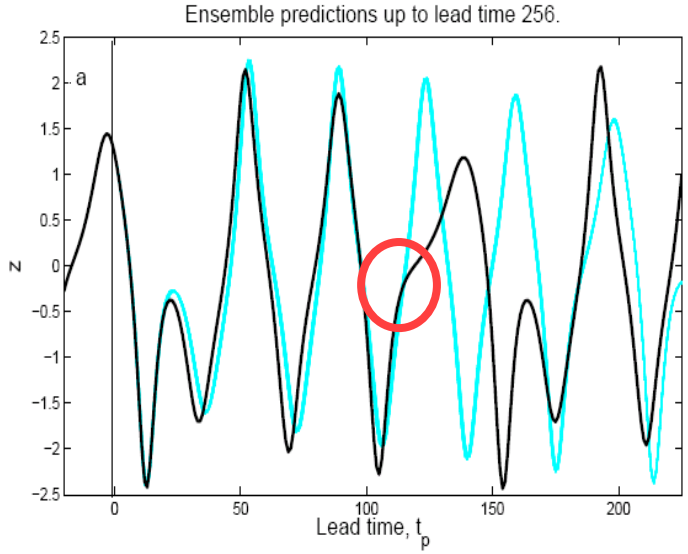
Moore-Spiegel Circuit (by Reason Machette)



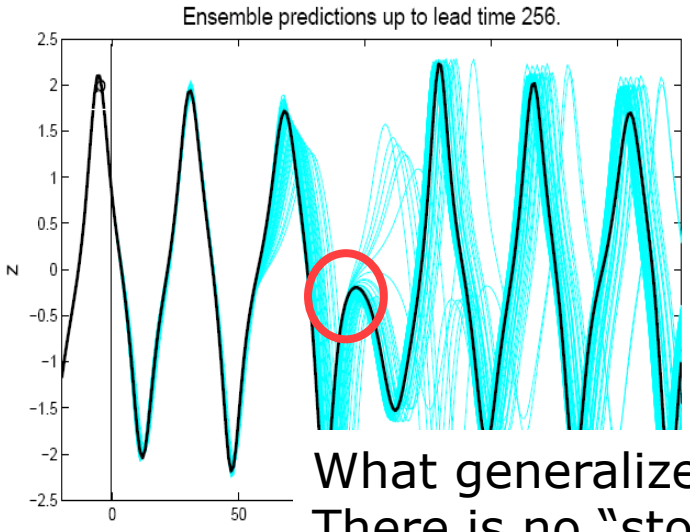
One Initial State —— Another Initial State –

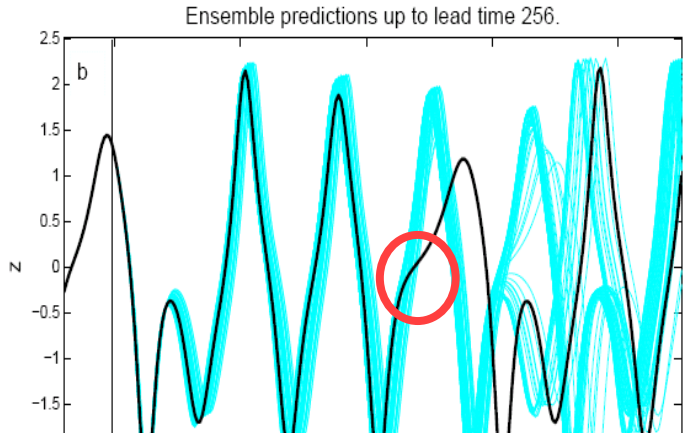
Model 1





Model 2





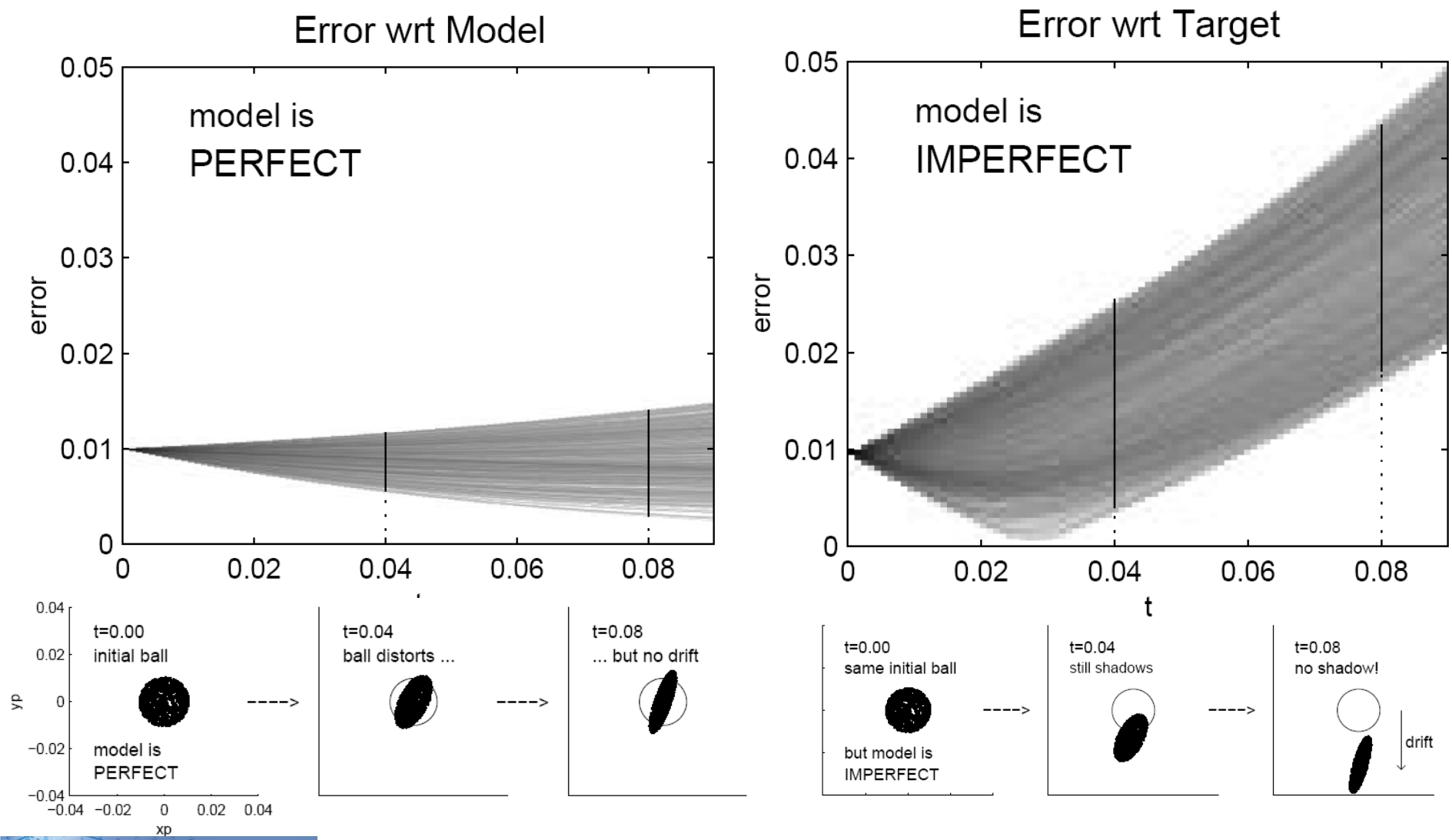
What generalizes:

There is no "stochastic fix" for an inadequate model (cla

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

Model Imperfections

D Orrell, LA Smith, T Palmer & J Barkmeijer (2001) Model Error in Weather Forecasting, Nonlinear Processes in Geophysics 8: 357-371.



Types of Probability (Forecasts): P(x| data, I)

(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)

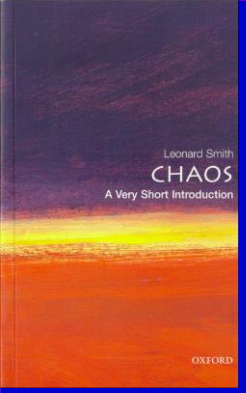
(ii) Psychological Probability: "Personal probability inferred from one's behaviour."

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

(iv) 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).

(v) **Mature Probability**: P(x| g<G) when G is known (not) to be encapsulated in g.
Mature probability is not expected to change without some additional insight or additional empirical observation (even given vast computational power).

Rational Decisions I. J. Good (1952) Journal of the Royal Statistical Society. Series B (Methodological) Vol. 14, No. 1 , pp. 107-114
Good Thinking I.J. Good (1983) Dover.



Background Reading:

LA Smith(2002) *What might we learn from climate forecasts?* P. Nat. Acad. Sci (99)

LA Smith (2003) *Predictability Past Predictability Present. Predictability and Weather Forecasting* (ed. Tim Palmer, CUP).

LA Smith (2000) *Disentangling Uncertainty and Error*, in Nonlinear Dynamics and Statistics (ed A.Mees) Birkhauser.

Stainforth et al (2005) *Uncertainties in Prediction of Climate response*. Nature.

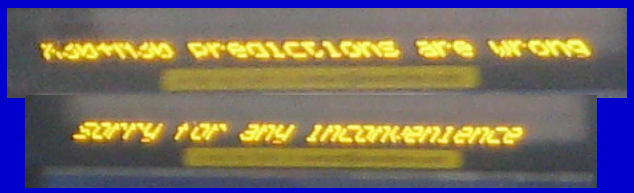
Stainforth et al (2007) *Uncertainty & Decision Support*. Phil Trans Roy. Soc. A,1098

LA Smith (2007) *A Very Short Introduction to Chaos*. OUP

Nancy Cartwright (1983) *How the Laws of Physics Lie*. OUP

www.cccep.ac.uk

L.Smith@lse.ac.uk



> *When in doubt*, distrusting the indications, or inferences from them (duly considered on purely scientific principles, and checked by experience), the words "*Uncertain*," or "*Doubtful*," may be used, without hesitation.
>
> Fitzroy, 1862

CATS CENTRE FOR THE ANALYSIS OF TIME SERIES

communication

Being more clear

Maths -> Physics

"chaos" need not limit predictability

Being more pedagogical

Being more designing (as in experimental)