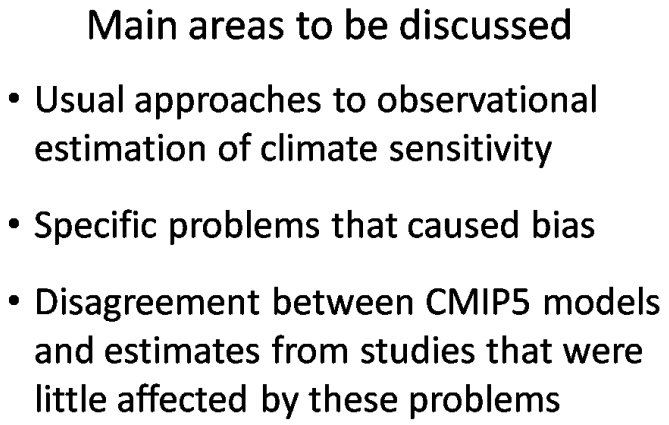
Pitfalls in climate sensitivity estimation: Part 1

A three part article by Nicholas Lewis

April 2015

As many readers will be aware, I attended the WCRP Grand Challenge Workshop: Earth's Climate Sensitivities at Schloss Ringberg in late March. Ringberg 2015 was a very interesting event, attended by many of the best known scientists involved in this field and in areas of research closely related to it – such as the behaviour of clouds, aerosols and heat in the ocean. Many talks were given at Ringberg 2015; presentation slides are available here. It is often difficult to follow presentations just from the slides, so I thought it was worth posting an annotated version of the slides relating to my own talk, "Pitfalls in climate sensitivity estimation". To make it more digestible and focus discussion, I am splitting my presentation into three parts. I've omitted the title slide and reinstated some slides that I cut out of my talk due to the 15 minute time constraint.

Slide 2



In this part I will cover the first bullet point and one of the major problems that cause bias in climate sensitivity estimates. In the second part I will deal with one or two other major problems and summarize the current position regarding observationally-based climate sensitivity estimation. In the final part I will deal with the third bullet point.

In a nutshell, I will argue that:

- Climate sensitivity is most reliably estimated from observed warming over the last ~150 years
- Most of the sensitivity estimates cited in the latest IPCC report had identifiable, severe problems
- Estimates from observational studies that are little affected by such problems indicate that climate sensitivity is substantially lower than in most global climate models
- Claims that the differences are due to substantial downwards bias in estimate from these observational studies have little support in observations.

Slide 3

What observations best constrain ECS?

- Paleoclimate: big uncertainties; dependence on assumptions; ECS may differ from now
- Instrumental: estimates based on short time-scales or non-GHG forcings unreliable (AR5)
- ECS (+ TCR) best constrained by multidecadal warming during instrumental period
- Attribution scaling method: best use of GCMs?
- Combination: instrumental dominates paleo; subjective probability – but objective possible

3

ECS refers to equilibrium climate sensitivity: the increase in global mean surface temperature (GMST) that a doubling of atmospheric CO_2 concentration leads to, once the ocean has fully equilibrated. The ECS of a coupled atmosphere-ocean general circulation climate model (AOGCM, or just GCM) can be determined by running it to equilibrium in a 2x CO_2 experiment, but that takes thousands of simulation years. All non-palaeoclimate real-world ECS estimates reflect effective climate sensitivity, which depends on the strength of climate feedbacks over the analysis period involved (although in a few cases estimates are calibrated to ECS in one or more AOGCMs).

In many but not all current generation (CMIP5) AOGCMs, effective climate sensitivity estimates based on transient forced warming fall short of ECS, to an extent depending on the model, the estimation period, the forcing profile and the method used. It is unknown whether effective and equilibrium climate sensitivity differ much in the real world.

A shorter term measure of sensitivity, transient climate response (TCR) represents the increase in GMST over a 70 year period during which CO_2 concentration increases at 1% p.a., thereby doubling. The focus at Ringberg 2015 was mainly on ECS, which can be related, at least approximately, to the physical concepts of (a) the effective radiative forcing (ERF) that a doubling of CO_2 concentration produces and (b) the sum of the climate feedbacks to surface warming. Although TCR depends also on ocean heat uptake characteristics and is thus does not have a simple physical interpretation, it is more relevant than ECS to warming over this century..

The first three bullet points reiterate what the IPCC fifth assessment WG1 report (AR5) said in Chapters 10 and 12 about palaeoclimate ECS estimates and those based on short timescales or non-greenhouse gas (GHG) forcings, and its implicit conclusion that estimates based on multidecadal warming during the instrumental period (since about 1850) were likely to prove most reliable and provide the narrowest uncertainty bounds.

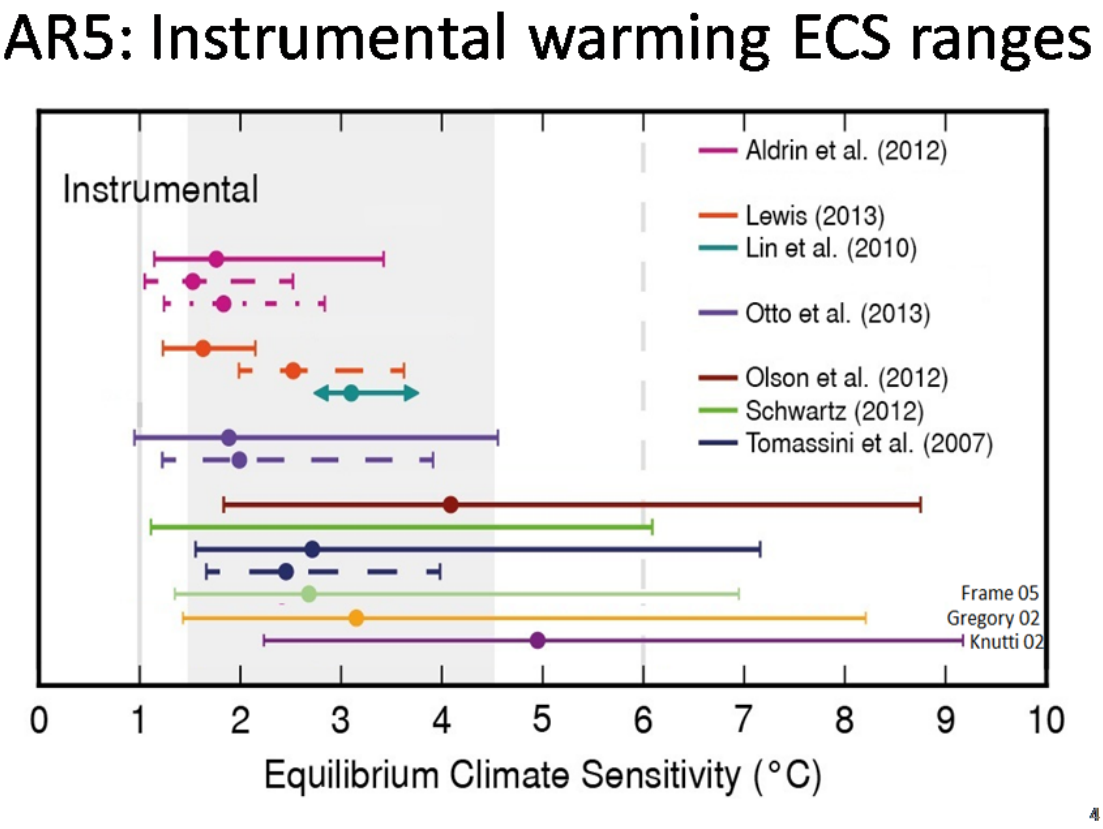
ECS estimates based on multidecadal warming typically use simple or intermediate complexity climate models driven by estimated forcing timeseries, and measure how well simulated surface temperatures and ocean heat uptake compare with observations as model parameters are adjusted. Energy budget methods are also used. These involve deriving ECS and/or TCR directly from estimates changes in forcing and measured changes in GMST and ocean heat content, usually between decadal or longer periods at the start and end of the multidecadal analysis period. Alternatively, regression-based slope estimates are sometimes used.

Attribution scaling methods refers to the use of scaling factors (multiple regression coefficients) derived from detection and attribution analyses that match observed warming to the sum of scaled

AOGCM responses to different categories of forcing, based on their differing spatiotemporal fingerprints. The derived scaling factor for warming attributable to GHG can then be used to estimate ECS and/or TCR, using a simple model. This hybrid method of observationally-estimating ECS and TCR appears to work better than the PPE approach, which involves varying AOGCM model parameters.

Various studies have combined palaeoclimate and instrumental ECS estimates using subjective Bayesian methods. I do not believe that such methods are appropriate for ECS estimation, as the results are sensitive to the subjective choice of prior distribution. It is however possible to use objective methods – both Bayesian and frequentist – to combine probabilistic ECS estimates, provided that the estimates are independent – which palaeo and instrumental estimates are usually assumed to be. Since palaeo ECS estimates are normally less precise than good instrumental ones, combination estimates are usually dominated by the underlying instrumental estimate.

Slide 4



My talk concentrates on ECS estimates based on observed warming observed during the instrumental period, as they are thought to be able to provide the most reliable, best constrained observational estimates. Slide 4 shows a version of Box 12.2, Figure 1 from AR5 with all other types of ECS estimate removed. The bars represent 5–95% uncertainty ranges, with blobs showing the best (median) estimates.

For the Lewis (2013) study, the dashed range should be ignored and the solid range widened to 1.0–3.0°C (with unchanged median) to reflect non-aerosol forcing uncertainty, as discussed in that paper.

Although the underlying forcing and temperature data should be quite similar in all these studies, the estimates vary greatly.

Slide 5

Why do warming-based ECS/TCR ranges differ?

- All relate global warming to forcing changes
- Also need heating rate estimate Q (not for TCR)
- Estimates directly or indirectly reflect energy budget equation: ECS = F_{2xCO2} x $\Delta T/(\Delta F - \Delta Q)$
- Multidecadal variability (AMO): may badly bias

The relevant forcing concept here is ERF, denoted here by ΔF. AR5 defines it as follows: "ERF is the change in net TOA [top of atmosphere] downward radiative flux after allowing for atmospheric temperatures, water vapour and clouds to adjust, but with surface temperature or a portion of surface conditions unchanged."

ΔT refers to the change in GMST resulting from a change in ERF, and ΔQ to a change in the planetary heating rate, mainly (>90%) reflected in ocean heat uptake.

Without using ocean heat content (OHC) data to estimate ΔQ, ECS tends to be ill-constrained.

ΔQ is not relevant to estimating TCR: the equivalent equation for generic TCR given in AR5 Chapter 10 is TCR = $F_{2xCO2} \times \Delta T / \Delta F$.

Pre-2006 ECS studies almost all used the Levitus (2000) OHC data, which – due apparently to an uncorrected arithmetic error – gave substantially excessive values for ΔQ. Lin 2010 also used an excessive estimate for ΔQ, taken from the primarily model-based Hansen et al (2005) study. Moreover, many of the studies make no allowance for ΔQ being non-negligibly positive at the start of the instrumental period, as the Earth continued its recovery from the Little Ice Age. Gregory et al (2013) gives estimates of steric sea-level rise from 1860 on, derived from a naturally-forced model simulation starting in 850. Converting these to the planetary energy imbalance, and scaling down by 40% to allow for the model ECS of 3°C being high, gives ΔQ values of 0.15 W/m2 over 1860-1880 and 0.2 W/m2 from 1915-1960 (ΔQ being small in the intervening period due to high volcanism).

Multidecadal variability, represented by the quasi-periodic Atlantic Multidecadal Oscillation (AMO) in particular, means that the analysis period chosen is important. The AMO seems to be a genuine internal mode of variability, not as has been argued a forced pattern caused by anthropogenic aerosols.

The NOAA AMO index exhibits 60–70 year cycles over the instrumental period, peaking in the 1870s, around 1940, and in the 2000s. The AMO affects GMST, with a stronger influence in the northern hemisphere. As well as altering heat exchange between the ocean and atmosphere, the AMO also appears to modulate internal forcing through changing clouds – a little recognised point. As I will explain, the AMO can distort ECS estimation more seriously than its influence on GMST – of maybe ~0.2°C peak-to-peak – suggests.

Slide 6

Why do warming-based ECS/TCR ranges differ?

- All relate global warming to forcing changes
- Also need heating rate estimate Q (not for TCR)
- Estimates directly or indirectly reflect energy budget equation: ECS = F_{2xCO2} x $\Delta T/(\Delta F - \Delta Q)$
- Multidecadal variability (AMO): may badly bias
- Aerosol forcing biggest source of uncertainty
- F_{aero} can be well-constrained by zonal ΔT data – but not by GMST (Tomassini 2007; Olson 2012)
- Two TCR estimates based on AR4 ECS estimate

6

Taking the last point first, the Meinshausen 2009 and Rogelj 2012 TCR distributions featured in AR5 Figure 10.20(a) as estimated from observational constraints were actually based on ECS distributions selected simply to match the AR4 2-4.5°C ECS range and 3°C best estimate. They should be regarded as estimates based primarily on expert opinion, not observations.

Uncertainty as to the change in aerosol forcing occurring during the instrumental period, ΔF_{aero}, is the most important source of uncertainty in most ECS and TCR estimates based on multidecadal warming. Chapter 8 of AR5 gives a 1.8 W/m^2 wide 5-95% range for ΔF_{aero} over 1750–2011, about as large as the best estimate for ($\Delta F - \Delta Q$). The Lewis and Curry (2014) energy budget based study used the AR5 best estimate and uncertainty range for aerosol forcing (as well as other forcings), and hence its ECS and TCR estimates have 95% bounds that are much higher than their median values.

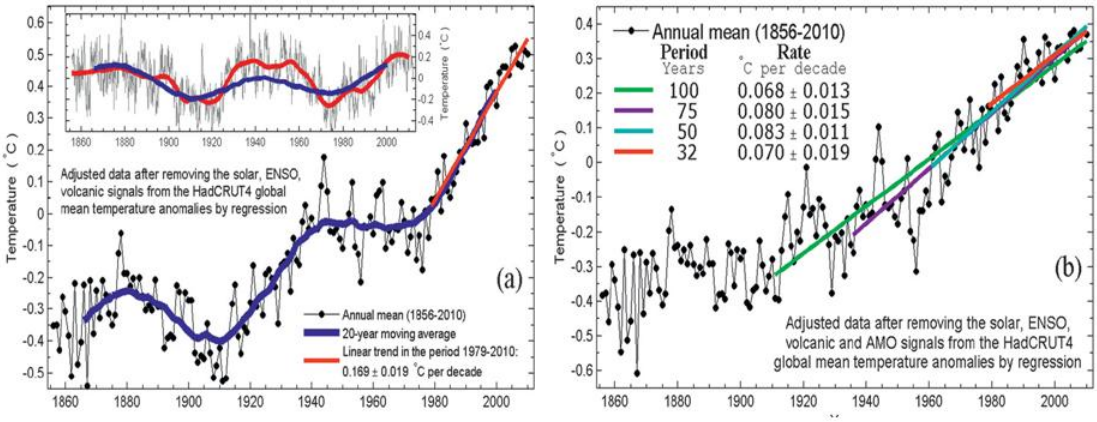
Otto et al (2013), although likewise using an energy budget method, used estimated forcings in CMIP5 AOGCMs (Forster et al 2013), which exhibit a narrower uncertainty range than AR5 gives, and adjusted their central value to reflect the difference between AOGCM and AR5 aerosol forcing estimates. Its resulting median estimate for TCR was accordingly almost identical to that in Lewis and Curry (2014), but its 95% bound based on the most recent data was lower.

It is possible to estimate ΔF_{aero} with considerably less uncertainty than that stated in AR5, using "inverse methods" that infer ΔF_{aero} from hemispherically- or zonally-resolved surface temperature data. This takes advantage of the latitudinally-inhomogeneous, northern hemisphere dominated, distribution of anthropogenic aerosol emissions, using a latitudinally-resolving model to estimate the spatial pattern of temperature changes at varying ΔF_{aero} levels. Of the ECS studies featured in AR4 and AR5, Andronova and Schlesinger (2001), Forest et al (2002 and 2006), Liberdoni and Forest (2011/13), Ring, Schlesinger et al (2012), Aldrin et al (2012) and Lewis (2013) used this approach; so did Skeie et al (2014).

However, inverse estimates of ΔF_{aero} are very unreliable if only GMST data is used. At a global level the evolution of ΔF_{aero} and ΔF_{GHG} is very highly correlated (r = 0.98 for the AR5 best estimate timeseries). Moreover, the diverge of the growth rates of the two series post the 1970s, when aerosol emissions flattened out, coincides and gets conflated with the AMO upswing.

Slide 7

Zhou & Tung 2013 JAS: Effect of AMO



Blue line in (a) inset is that in main plot, but linearly detrended. Matches AMO index

Upswing in AMO from late 1970s occurs when aerosol and GHG rises start to diverge

Zhou finds ~0.3 K AMO GMST range and Anthro warming ~ constant over 100 years

Even if < 0.15 K, AMO (i) biases ECS/TCR; (ii) => hopeless estimating F_{aer} from GMST

7

The AMO index smoothed pattern is shown by the red curve in the inset at the top LH of Slide 7, and can be seen to resemble the detrended GMST with shorter term natural signals removed (blue curve). Zhou and Tung may be overestimating the influence of the AMO on GMST; Delsole et al (2011) estimate it to be about half as strong. However, even at one-quarter of the level shown it is enough to bias estimation of ΔF_{aero} up by a factor of two or more, with an accompanying upwards bias of 20% or more in the estimate of warming attributable to GHG (and hence in TCR estimates; ECS estimates are even worse affected).

The problem is that a combination of a strongly negative estimate for ΔF_{aero}, and a high estimate for ECS is able to mimic the effect on GMST caused by a factor (the AMO) not represented in the estimation model used. The slight fall in GMST between the 1940s and the early 1970s is matched by selecting a strongly increasing negative ΔF_{aero} that counters increasingly positive ΔF_{GHG}, whilst the fast rise in GMST from the late 1970s on is matched by a high ECS (and hence high TCR), operating on a strong rise in ΔF_{GHG} that is no longer countered by strengthening ΔF_{aero}.

The ECS studies that use only the evolution of GMST (along with data pertaining to ΔQ) to estimate ΔF_{aero} jointly with ECS therefore usually reach a much more negative estimate for ΔF_{aero}, and a higher estimate for ECS, than studies that are able to estimate ΔF_{aero} from the differential evolution of hemispherically- or zonally-resolved surface temperature data. Studies affected by this problem include, for ECS, Knutti et al (2002), Tomassini et al (2007) and Olson et al (2012) and, for TCR, Knutti & Tomassini 2008 and Padilla et al (2011).

Using hemispherically- or zonally-resolved temperature data to estimate aerosol forcing fails to avoid contamination by the AMO when the analysis period is unsuitable or insufficiently long. Many AR4 era ECS and TCR studies used the 20th century as their analysis period. The1900s started with the AMO low and ended with the AMO high. Gillett et al (2012) found that, despite its uses of spatiotemporal patterns, their detection and attribution study's estimate of warming attributable to GHG was biased ~40% high when based on 1900s data compared with when the longer 1851-2010

period was used. ECS studies affected by this problem include Gregory et al (2002), Frame et al (2005) and Allen et al (2009).The Stott and Forest (2007) TCR estimate is also affected.

The Gregory and Forster (2008) TCR estimate, while avoiding the AMO's influence on aerosol forcing estimation, is significantly biased up by the AMO's direct enhancement of the GMST trend over the short 1970–2006 analysis period used.

I will leave it there for Part 1; in Part 2 I will move on to problems with Bayesian approaches to climate sensitivity estimation.

References

Aldrin, M., M. Holden, P. Guttorp, R.B. Skeie, G. Myhre, and T.K. Berntsen, 2012. Bayesian estimation of climate sensitivity based on a simple climate model fitted to observations of hemispheric temperatures and global ocean heat content. *Environmetrics*;23: 253–271.

Andronova, N.G. and M.E. Schlesinger, 2001. Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res*.,106 (D19): 22605–22611.

DelSole, T., M. K. Tippett, and J. Shukla, 2011: A significant component of unforced multidecadal variability in the recent acceleration of global warming. *J. Clim.*, 24, 909–926.

Forest, C.E., P.H. Stone, A.P. Sokolov, M.R. Allen and M.D. Webster, 2002.Quantifyinguncertainties in climate system properties with the use of recent climate observations. *Science*; 295: 113–117

Forest, C.E., P.H. Stone and A.P. Sokolov, 2006.EstimatedPDFs of climate system properties including natural and anthropogenic forcings. *Geophys. Res. Lett*., 33: L01705

Forster, P.M., T. Andrews, P. Good, J.M. Gregory, L.S. Jackson, and M. Zelinka, 2013. Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res*., 118: 1139–1150.

Frame D.J., B.B..B. Booth, J.A. Kettleborough, D.A. Stainforth, J.M. Gregory, M. Collins and M.R. Allen, 2005.Constrainingclimateforecasts: The role of prior assumptions. *Geophys. Res. Lett.*, 32, L09702Fyfe, J.C., N.P. Gillett, and F.W. Zwiers, 2013.Overestimatedglobal warming over the past 20 years. *Nature Clim.Ch.*; 3.9: 767–769.

Gillett NP, Arora VK, Flato GM, Scinocca JF, von Salzen K (2012) Improved constraints on 21st-century warming derived using 160 years of temperature observations. *Geophys. Res. Lett.*, 39, L01704, doi:10.1029/2011GL050226.

Gregory, J.M., R.J. Stouffer, S.C.B. Raper, P.A. Stott, and N.A. Rayner, 2002. An observationally based estimate of the climate sensitivity. *J. Clim.*,15: 3117–3121.

Gregory, J. M., and P. M. Forster, 2008: Transient climate response estimated from radiative forcing and observed temperature change. *J. Geophys. Res. Atmos.*, 113, D23105.

Knutti, R., T.F. Stocker, F. Joos, and G.-K. Plattner, 2002. Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature*, 416: 719–723.

Knutti, R., and L. Tomassini, 2008: Constraints on the transient climate response from observed global temperature and ocean heat uptake. *Geophys. Res. Lett*., 35, L09701.

Levitus, S., J. Antonov, T. Boyer, and C Stephens, 2000. Warming of the world ocean, *Science*; 287: 5641.2225–2229.

Lewis, N., 2013. An objective Bayesian, improved approach for applying optimal fingerprint techniques to estimate climate sensitivity. *J. Clim.*,26: 7414–7429.

Lewis N, Curry JA (2014) The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Clim. Dyn*. DOI 10.1007/s00382-014-2342-y

Libardoni, A.G. and C. E. Forest, 2011. Sensitivity of distributions of climate system properties to the surface temperature dataset. *Geophys. Res. Lett.*; 38, L22705. Correction, 2013: doi:10.1002/grl.50480.

Lin, B., et al., 2010: Estimations of climate sensitivity based on top-of-atmosphere radiation imbalance. *Atmos. Chem. Phys.*, 10: 1923–1930.

Meinshausen, M., et al., 2009: Greenhouse-gas emission targets for limiting global warming to 2 °C. *Nature*, 458, 1158–1162.

Olson, R., R. Sriver, M. Goes, N.M. Urban, H.D. Matthews, M. Haran, and K. Keller, 2012. A climate sensitivity estimate using Bayesian fusion of instrumental observations and an Earth System model. *J. Geophys. Res. Atmos.*,117: D04103.

Otto, A., et al., 2013. Energy budget constraints on climate response. *Nature Geoscience*, 6: 415–416.

Padilla, L. E., G. K. Vallis, and C. W. Rowley, 2011: Probabilistic estimates of transient climate sensitivity subject to uncertainty in forcing and natural variability. *J. Clim.*, 24, 5521–5537.

Ring, M.J., D. Lindner, E.F. Cross, and M.E. Schlesinger, 2012. Causes of the global warming observed since the 19th century. *Atmos. Clim. Sci.*, 2: 401–415.

Rogelj, J., M. Meinshausen, and R. Knutti, 2012: Global warming under old and new scenarios using IPCC climate sensitivity range estimates. *Nature Clim. Change*, 2, 248–253.

Schwartz, S.E., 2012. Determination of Earth's transient and equilibrium climate sensitivities from observations over the twentieth century: Strong dependence on assumed forcing. *Surv.Geophys*., 33: 745–777.

Tomassini, L., P. Reichert, R. Knutti, T.F. Stocker, and M.E. Borsuk, 2007. Robust Bayesian uncertainty analysis of climate system properties using Markov chain Monte Carlo methods. *J. Clim.*, 20: 1239–1254.

Zhou, J., and K.-K. Tung, 2013. Deducing multidecadal anthropogenic global warming trends using multiple regression analysis. *J. Atmos. Sci*., 70, 3–8.

Pitfalls in climate sensitivity estimation: Part 2

Nicholas Lewis

In Part 1 I introduced the talk I gave at Ringberg 2015, explained why it focussed on estimation based on warming over the instrumental period, and covered problems relating to aerosol forcing and bias caused by the influence of the AMO. I now move on to problems arising when Bayesian probabilistic approaches are used, and then summarize the state of instrumental period warming, observationally-based climate sensitivity estimation as I see it. I explained in Part 1 why other approaches to estimating ECS appear to be less reliable.

Slide 8

Bias from informative priors

- Most ECS estimates based on Subjective Bayes
- Prior reflects view about what is being estimated
- Uniform ECS & K_v priors greatly bias CrI for ECS
- 'Expert' ECS prior usually dominates over data

The AR4 report gave probability density functions (PDFs) for all the ECS estimates it presented, and AR5 did so for most of them. PDFs for unknown parameters are a Bayesian probabilistic concept. Under Bayes' theorem – a variant on the conditional probability lemma – one starts by choosing a prior PDF for the unknown parameter, multiplies it by the relative probability of having obtained the actual observations at each value of the parameter (the likelihood function) and obtains the posterior PDF by normalising the result to unit total probability.

The posterior PDF melds any existing information about the parameter from the prior with information provided by the observations. If multiple parameters are being estimated, a joint prior and a joint likelihood function are required, and marginal posterior PDFs for individual parameters are obtained by integrating out the other parameters from the joint posterior PDF.

An uncertainty range derived from percentage points of integral, the cumulative probability distribution (CDF) are known as credible intervals (CrI) the frequentist statistical approach instead gives confidence intervals (CIs), which are conceptually different. In general, a Bayesian CrI cannot be exactly equivalent to a frequentist CI no matter what prior is selected. However, for some standard cases they can be, and it is typically possible to derive a prior (a probability matching prior) which results in CrIs close to the corresponding CIs. That is critical if assertions based on a Bayesian CrI are to be true with the promised reliability.

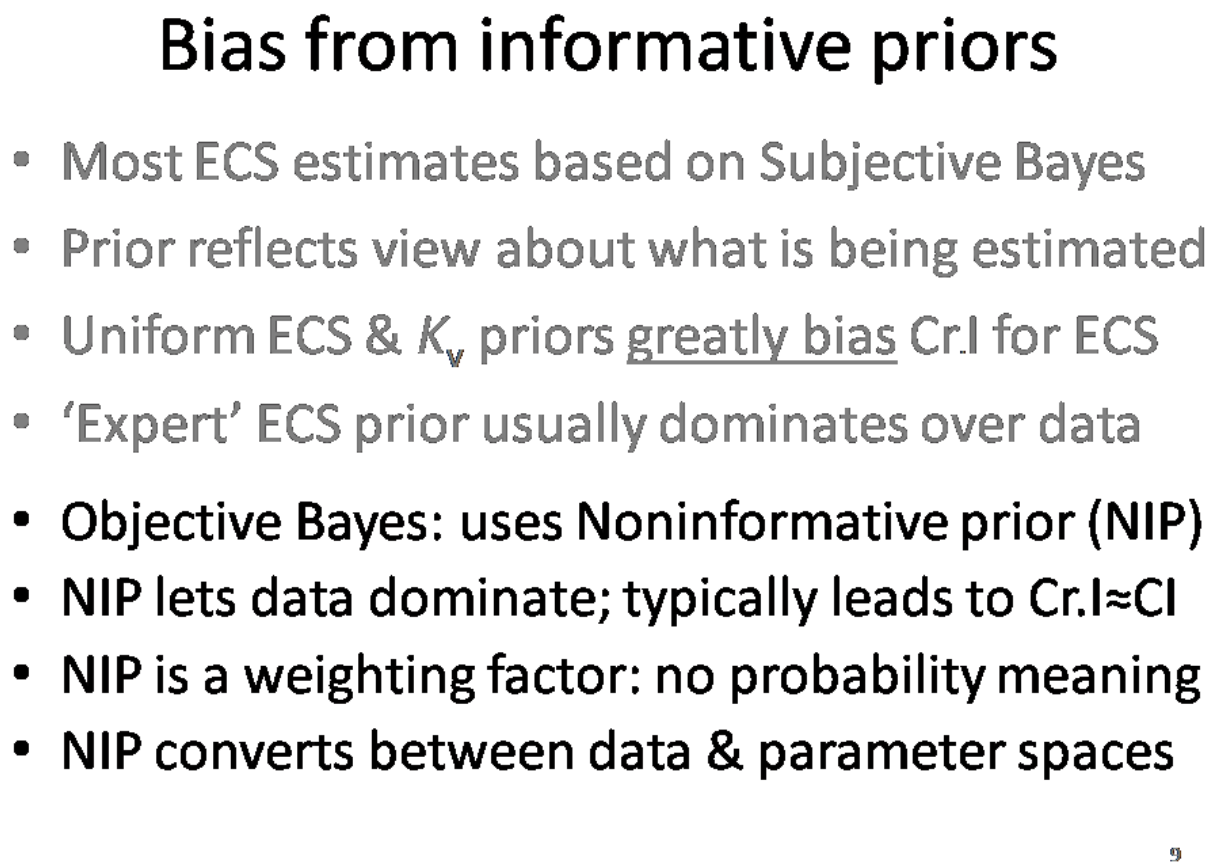
Almost all the PDFs for ECS presented in AR4 and AR5 used a 'subjective Bayesian' approach, under which the prior is selected to represent the investigator's views on how likely it is the parameter has each possible value. That may be a judgemental or elicited 'expert prior' that typically

has a peaked distribution indicating a most likely value, or a diffuse, typically uniform, distribution spread over a wide range, intended to convey ignorance and/or view a view to letting the data dominate the posterior PDF. AR4 stated that all its PDFs for ECS were presented on a uniform-in-ECS prior basis, although the AR4 authors were mistaken in two cases. In AR5, most ECS PDFs were derived using either uniform or expert priors for ECS (and for other key unknown parameters being estimated alongside ECS).

When the data is weak (is limited and uncertainty is high) the prior can have a major influence on the posterior PDF. Unlike in many areas of physics, that is the situation in climate science, certainly so far as ECS and TCR estimation is concerned. Moreover, the relationships between the principal observable variables (changes in atmospheric and ocean temperatures) and the parameters being estimated – which typically also include ocean effective vertical diffusivity (K_v) when ECS is the target parameter – are highly non-linear.

In these circumstances, use of uniform priors for ECS and K_v (or its square root) greatly biases posterior PDFs for ECS, raising their medians and fattening their upper tails. On the other hand, use of an expert prior typically results in the posterior PDF resembling the prior more than reflecting the data.

Slide 9

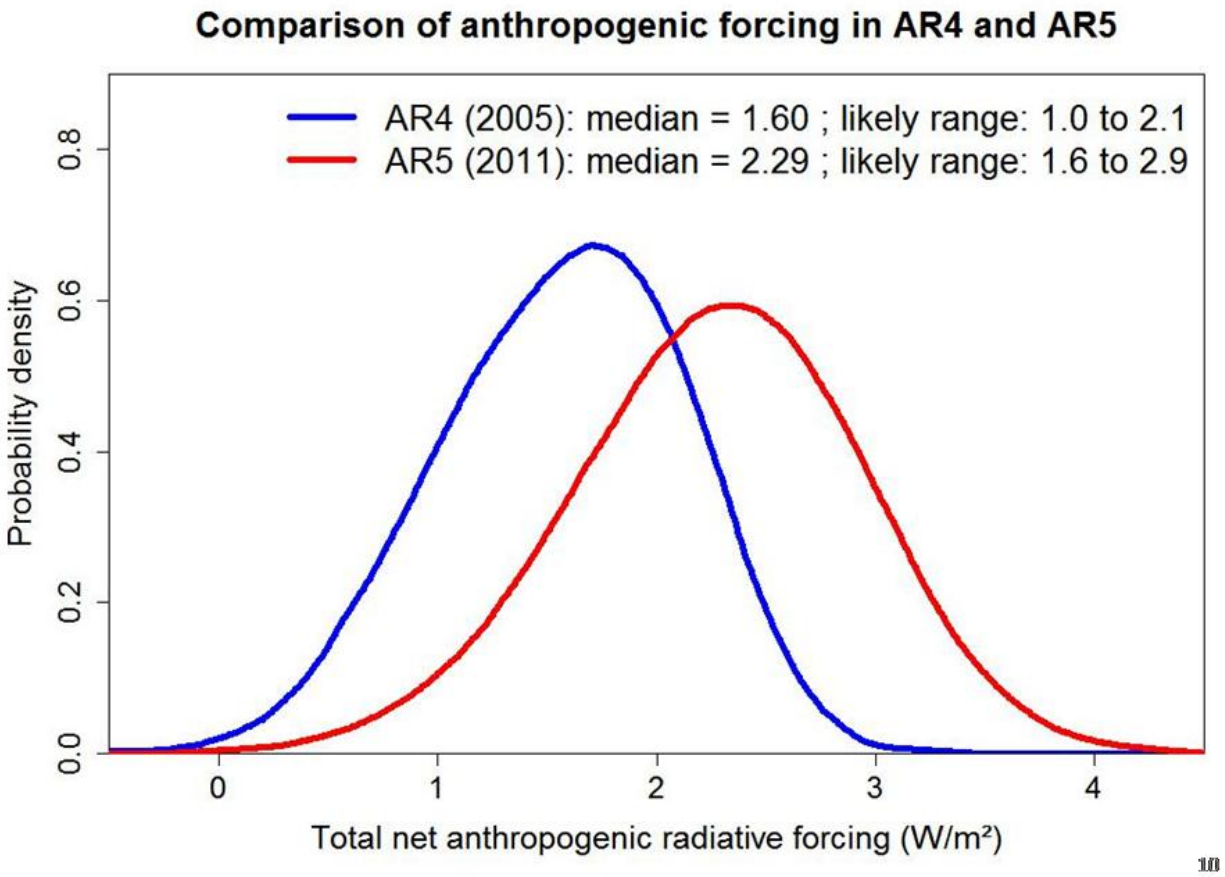


Some studies used the alternative 'objective Bayesian' approach, under which a mathematically-derived noninformative prior is used. Noninformative priors are designed to allow even weak data to dominate the posterior PDF for the parameter being estimated. They are typically judged by how good the probability matching properties of the resulting posterior PDFs are.

Noninformative priors do not represent how likely the parameter is to take any particular value and have no probabilistic interpretation. Noninformative priors are simply weight functions that convert data-based likelihoods into parameter PDFs with desirable characteristics, typically as regards probability matching. This is heresy so far as the currently-dominant Subjective Bayesian school is concerned. At least in typical cases of ECS and TCR estimation, noninformative priors are best regarded as conversion factors between data and parameter spaces. [For readers wanting insight as to why noninformative priors have no probability meaning, contrary to the standard interpretation of Bayes' theorem, and regarding problems with Bayesian methods generally, I recommend Professor Don Fraser's writings, perhaps starting with this paper.].

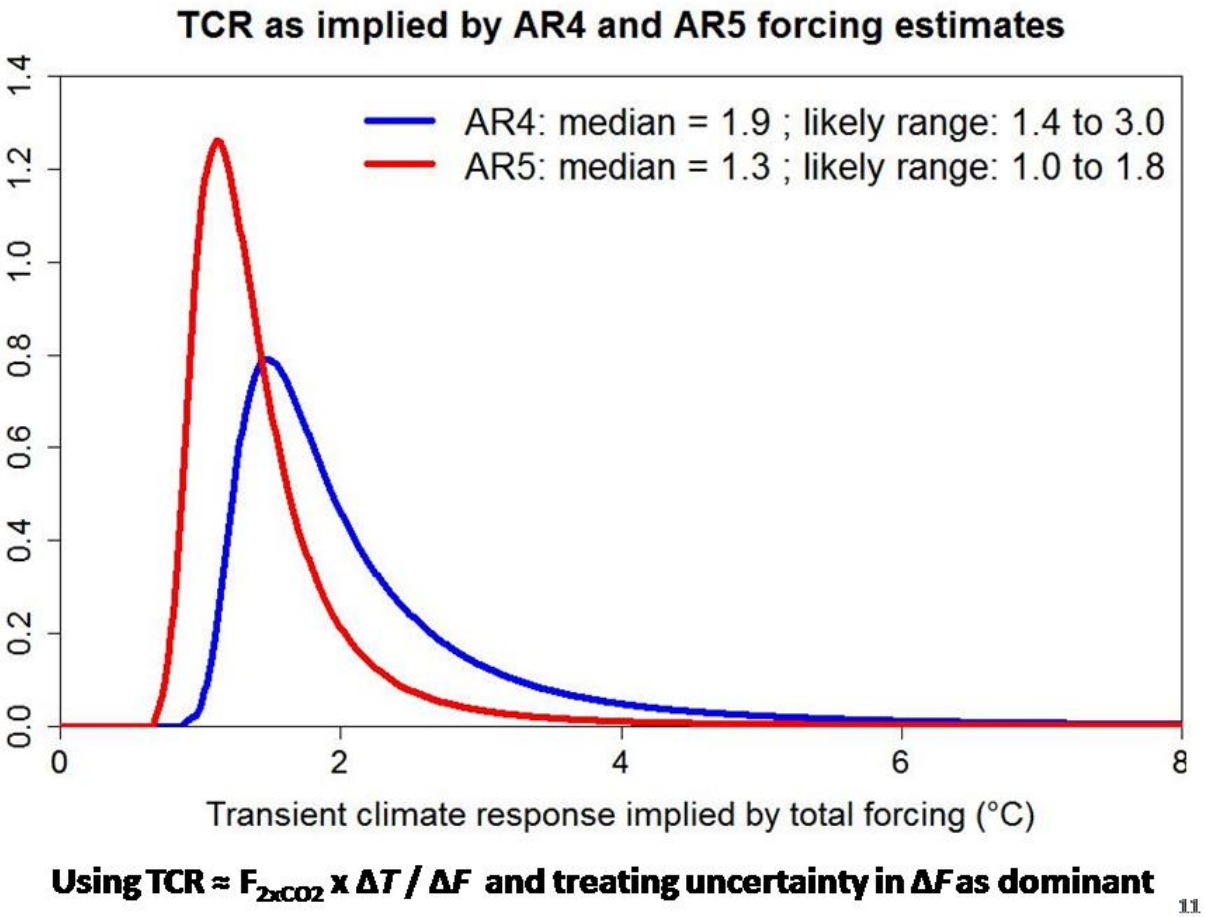
The Lewis (2013) and Lewis (2014) studies employed avowedly objective Bayesian approaches, involving noninformative priors. The Andronova and Schlesinger (2001), Gregory et al (2002), Otto et al (2013), and Lewis & Curry (2014) studies all used sampling methods that equated to an objective Bayesian approach. Studies using profile likelihood methods, a frequentist approach that yields approximate CIs, also achieve objective estimation (Allen et al 2009, Lewis 2014).

Slide 10



I will illustrate the effect of using a uniform prior for TCR estimation, that being a simpler case than for ECS. Slide 10 shows estimated distributions from AR4 and AR5 for anthropogenic forcing up to respectively 2005 and 2011. These are Bayesian posterior PDFs. They are derived by sampling from estimated uncertainty distributions for each forcing component, and I will assume for the present purposes that they can be considered to be objective.

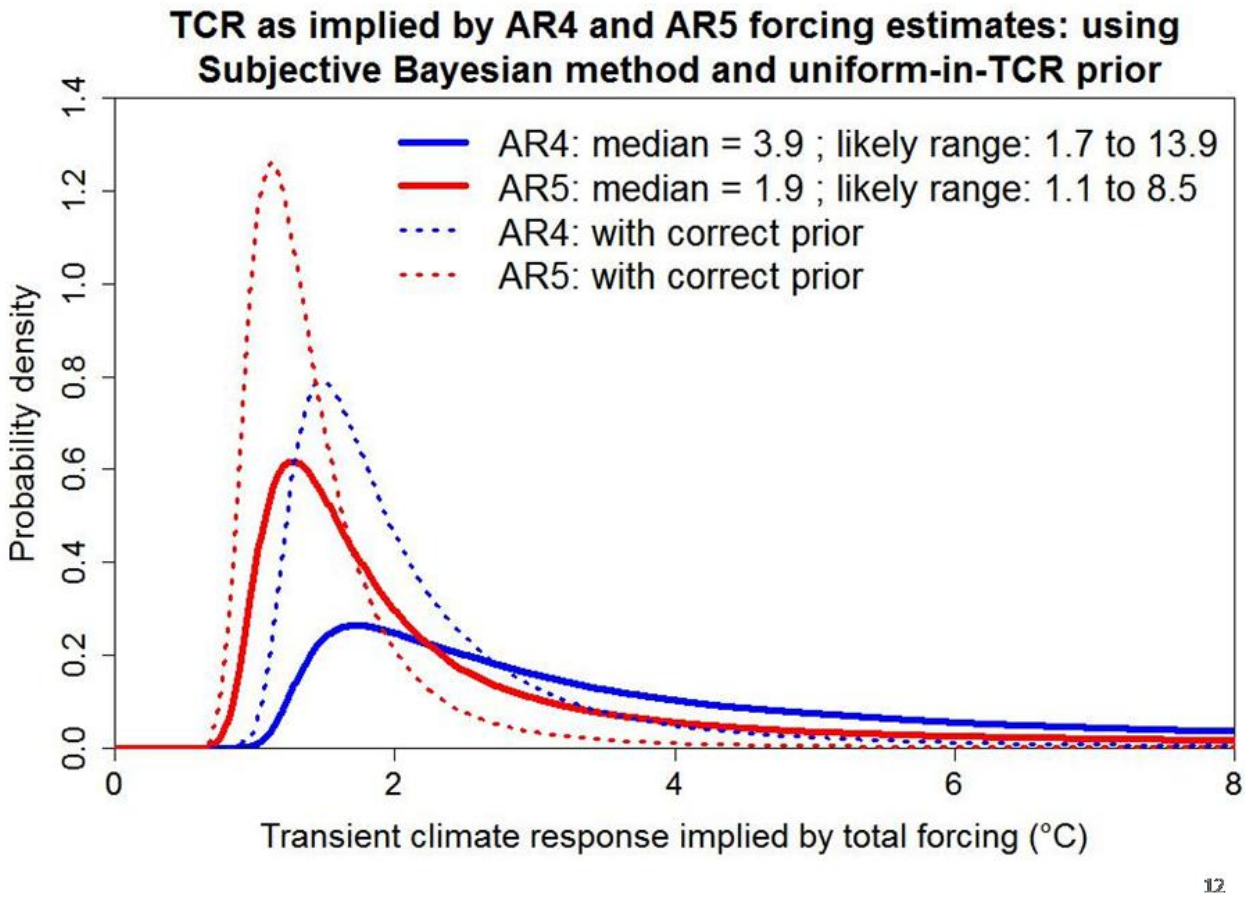
Slide 11



Slide 11 shows posterior PDFs for TCR derived from the PDFs for anthropogenic forcing, ΔF, by making certain simplifying approximations. I have assumed that the generic-TCR formula given in AR5 holds; that uncertainty in the GMST rise attributable to anthropogenic forcing, ΔT , and in F_{2xCO2}, the forcing from a doubling of CO_2, is sufficiently small relative to uncertainty in ΔF to be ignored; and that ΔT = 0.8°C and F_{2xCO2} = 3.71 W/m^2.

On this basis, PDFs for TCR follows from a transformation of variables approach. One simply changes variable from ΔF to TCR (the other factors in the equation being assumed constant). The PDF for TCR at any value TCR_a therefore equals the PDF for ΔF at $\Delta F = F_{2xCO2} \times \Delta T / TCR_a$ multiplied by the standard Jacobian factor, being the absolute derivative of ΔF with respect to TCR at TCR_a. That factor equals, up to proportionality, $1/TCR^2$.

Slide 12

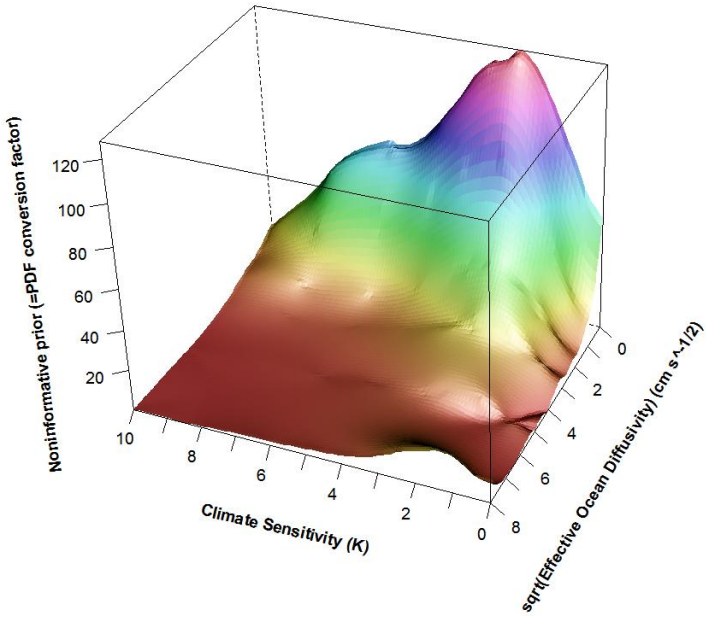


Suppose one regards the posterior PDFs for ΔF as having been derived using uniform priors. This is correct so far as components of ΔF have symmetrical uncertainty distributions, but is only an approximation since the most uncertain component, aerosol forcing, is assumed to have an asymmetrical distribution. However, the AR4 and AR5 PDFs for ΔF are not greatly asymmetrical.

On the basis that the posterior PDFs for ΔF correspond to the normalised product of a uniform prior for ΔF and a likelihood function, the PDFs for TCR derived in slide 11 correspond to the normalised product of the same likelihood function (now expressed in terms of TCR) and a prior having the form $1/\mathrm{TCR}^2$. Unlike PDFs, likelihood functions do not depend on the variable that the expressed as a function of. That is because, unlike a PDF, a likelihood function represents a density for the observed data, not for the variable that it is expressed in terms of.

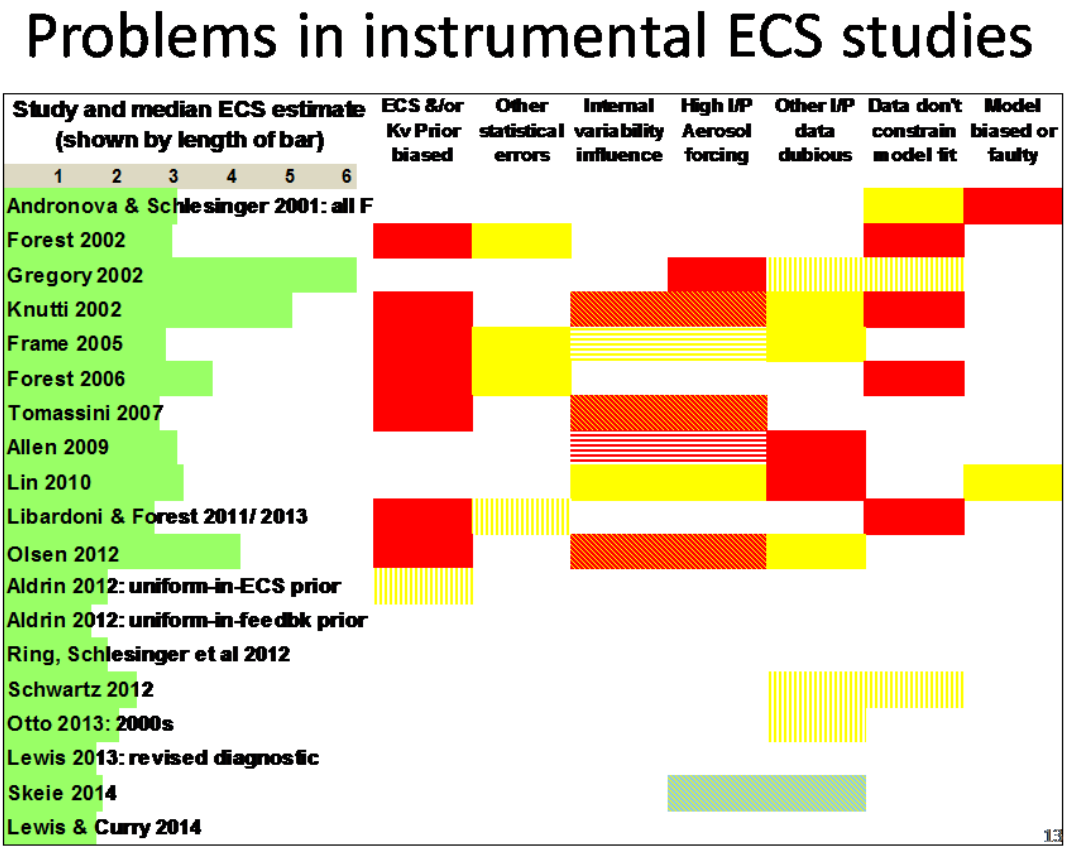
Slide 12 shows, on the foregoing basis, what the effect on the posterior PDF is of substituting a uniform-in-TCR prior for the mathematically correct $1/\mathrm{TCR}^2$ prior applying in slide 11. The median (50% probability point), which is the appropriate best estimate to give for a skewed distribution, increases substantially, doubling in the AR4 case. The top of the 17–83% 'likely' range more than quadruples. The distortion for ECS estimates would be even larger.

Slide 12a



I cut slide 12a out of my talk to shorten it. It shows the computed joint noninformative prior for ECS and sqrt(K_v) from Lewis (2013). Ignore the irregularities and the rise in the front RH corner, which are caused by model noise. Note how steeply the prior falls with sqrt(K_v), which is lowest at the rear, particularly at high ECS levels (towards the left) . The plot is probability-averaged over all values for aerosol forcing, which was also being estimated, accounting for the turndown in the prior at low ECS values. Noninformative priors can be quite complex when multiple parameters are involved!

Slide 13



Slide 13 summarises serious problems in instrumental period warming based ECS studies, ordered by year of publication, breaking problems down between seven factors. Median ECS estimates are shown by the green bars at the left.

Blank rectangles imply no significant problem in an area; solid yellow or red ones respectively a significant and a serious problem; vertical yellow bars, which looks like solid pale yellow, indicates a minor problem.

Red/yellow diagonal bars (looks like an orange shade of red) across 'Internal variability influence' and 'High input Aerosol forcing' mean that, due to use of global-only data, internal variability (the AMO) has led to a misestimate for aerosol forcing within the study concerned and hence to an overestimate of ECS. Yellow or red horizontal bars across those factors for the Frame et al (2005) and Allen et al (2009) studies mean that internal variability appears to have caused respectively significant or serious misestimation of aerosol forcing in the detection and attribution study that was the source of the (GHG-attributable) warming estimate used by the ECS study involved, and hence to bias in that estimate (reflected in the yellow or red rectangle for 'Other input data dubious').

The blue/yellow horizontal bar across 'High input Aerosol forcing' and 'Other input data dubious' for the Skeie et al (2014) study mean that problems in these two areas largely cancelled. Skeie's method estimated aerosol forcing using hemispherically-resolved model-simulation and observational data. An extremely negative prior for aerosol forcing was used, overlapping so little with the observational data-based likelihood function that the posterior estimate was biased significantly negative. However, the simultaneous use of three ocean heat content observational datasets appears to have led to the negatively biased aerosol forcing being reflected in lower modelled than observed NH warming rather than a higher ECS estimate.

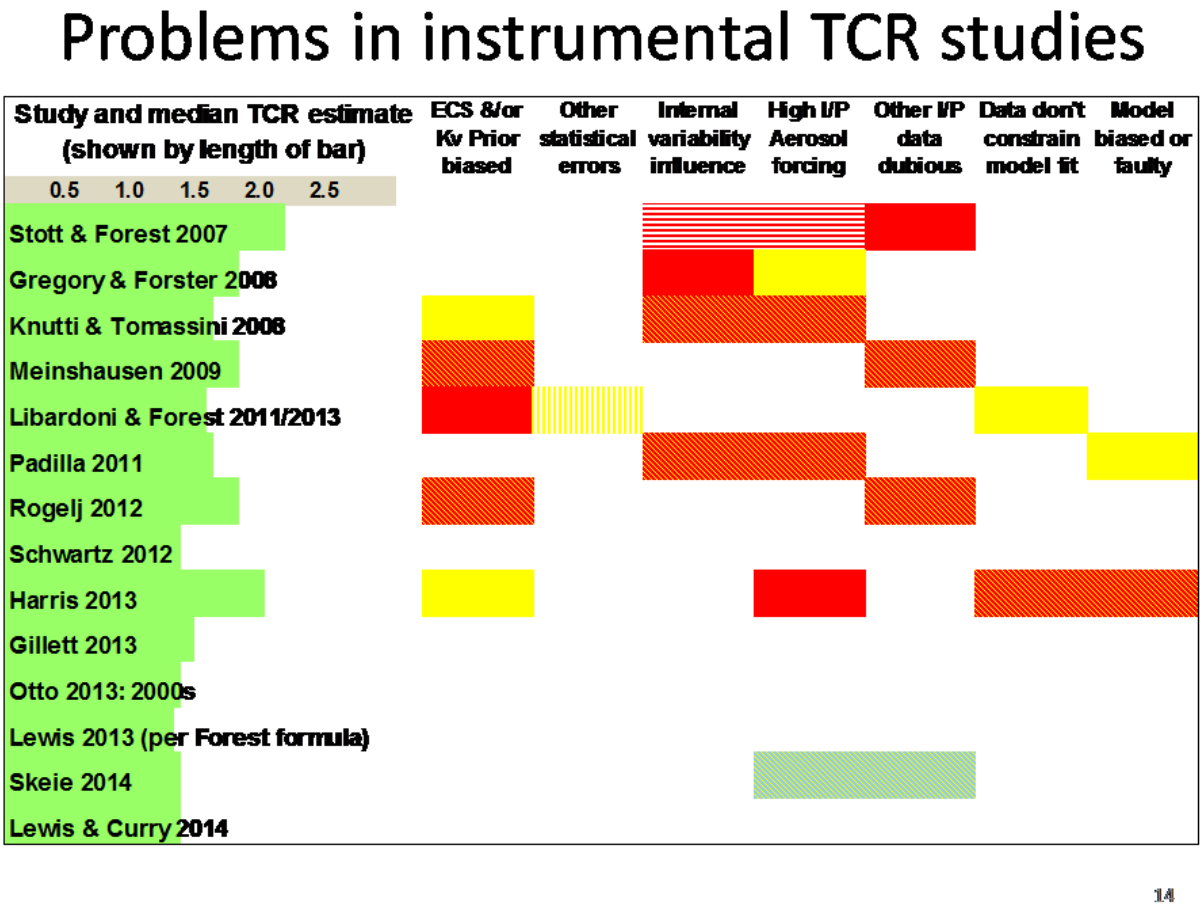
The 'Data don't constrain model fit' red entries for the Forest studies are because, from my experience, warming over the model-simulation run at the claimed best-fit parameter values is substantially greater than has been observed. The same entry for Knutti et al (2002) is because a very weak, pass/fail, statistical test was used.

The 'Model biased or faulty' red rectangle for Andronova and Schlesinger (2012) reflects a simple coding error that significantly biased ECS estimation: see Table 3 in Ring et al (2012).

A more detailed analysis of problems with individual ECS studies is available here.

To summarise: all pre-2012 instrumental-period-warming studies had one or more serious problems, and their median ECS estimates varied widely. Most studies from 2012 on do not have serious problems, and their estimates agree quite closely. (The Schwartz 2012 study's estimate was a composite of five estimates based on different forcing series, the highest ECS estimate comes from a poor quality regression obtained from one of the series.)

Slide 14



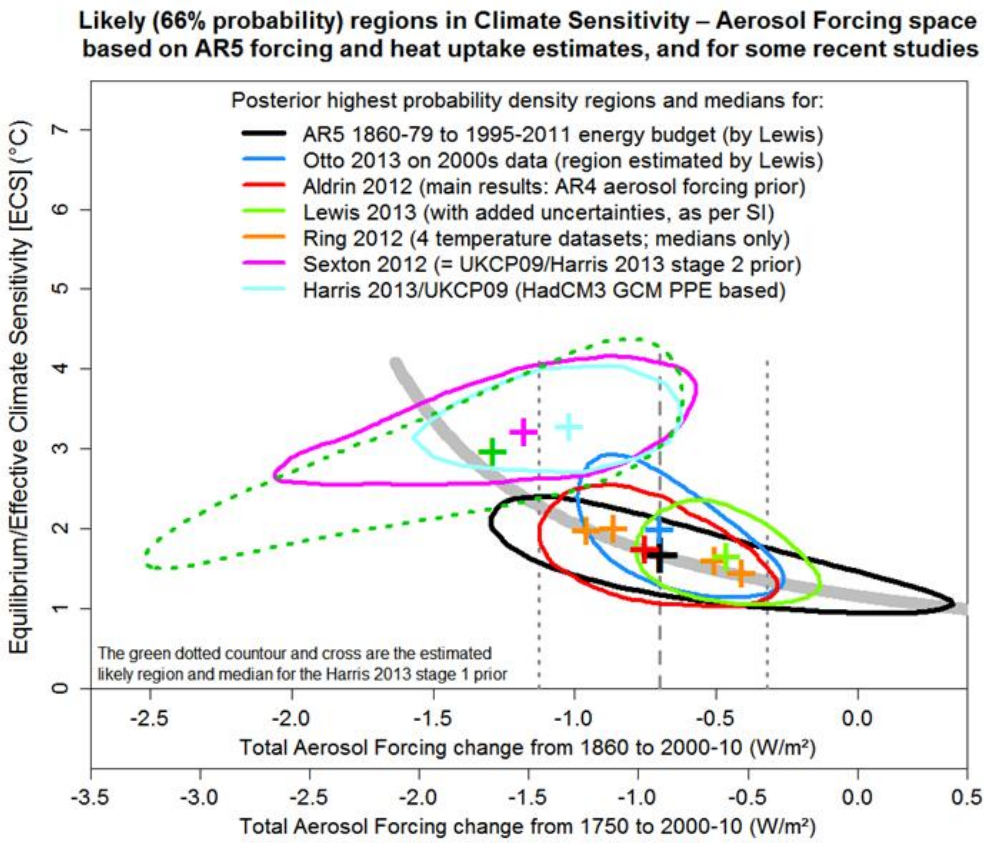
Slide 14 gives similar information to slide 13, but for TCR rather than ECS studies. As for ECS, all pre-2012 studies had one or more serious problems that make their TCR estimates unreliable, whilst most later studies do not have serious problems and their median TCR estimates are quite close.

Rogelj et al (2012)'s high TCR estimate is not genuinely observationally-based; it is derived from an ECS distribution chosen to match the AR4 best estimate and 'likely' range for ECS; the same goes for the Meinshausen et al (2009) estimate. The reason for the high TCR estimate from Harris et al (2013) is shown in the next slide.

A more detailed analysis of problems with individual TCR studies is available here.

Slide 14a

HadSM3 can't explore ECS–F_{aero} space



25

This slide came later in my talk, but rather than defer it to Part 3 I have moved it here as it relates to a PPE (perturbed physics/parameter ensemble) study, Harris et al (2013), mentioned in the previous slide about TCR studies. Although this slide considers ECS estimates, the conclusions reached imply that the Harris TCR estimate is seriously biased up relative to what observations imply.

The plot is of joint distributions for aerosol forcing and ECS; the solid contours enclose 'likely' regions of highest posterior probability density, containing 66% of total probability. Median estimates are shown by crosses; the four Ring et al (2012) estimates based on different surface temperature datasets are shown separately. The black contour is very close to that for Lewis and Curry (2014).

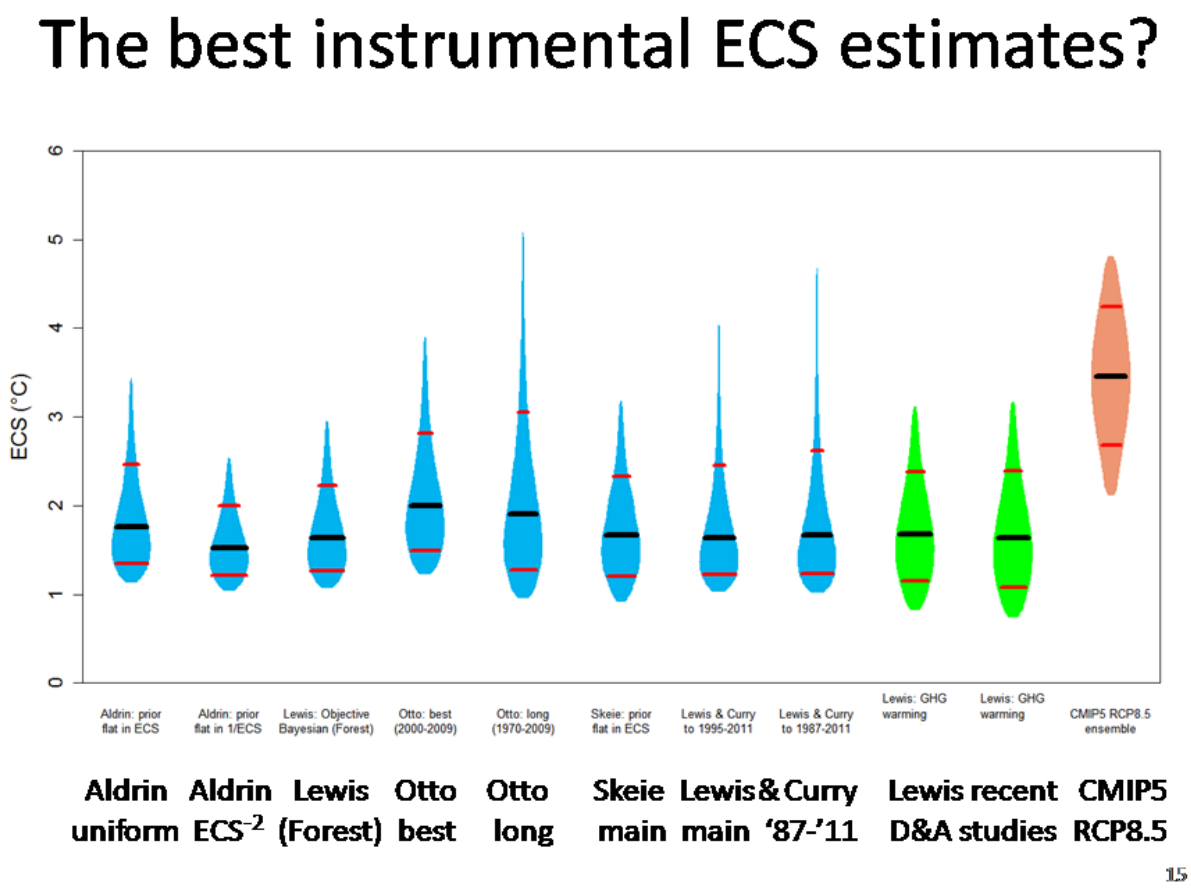
The grey dashed (dotted) vertical lines show the AR5 median estimate and 'likely' range for aerosol forcing, expressed both from 1750 (preindustrial) and from 1860; aerosol forcing in GCMs is normally estimated as the change between 1850 or 1860 and 2000 or 2005. The thick grey curve shows how one would expect the median estimate for ECS using an energy budget approach, based on AR5 non-aerosol forcing best estimates and a realistic estimate for ocean heat uptake, to vary with the estimate used for aerosol forcing.

The median estimates from the studies not using GCMs cluster around the thick grey line, and their likely regions are orientated along its curve: under an energy budget or similar model, high ECS estimates are associated with strongly negative aerosol forcing estimates. But the likely regions for the Harris study are orientated very differently, with less negative aerosol forcing being associated with higher, not lower, ECS. Its estimated prior distribution 'likely' region (dotted green contour) barely overlaps the posterior regions of the other studies: the study does not explore the region of low

to moderately negative aerosol forcing, low to moderate ECS which the other studies indicate observations support. It appears that the HadCM3/SM3 model has structural rigidities that make it unable to explore this region no matter how its key parameters are varied. So it is unsurprising that the Harris et al (2013) estimate for ECS (and hence TCR) is high: they cannot be regarded as genuinely observationally-based.

Further information on the problems with the Harris et al (2013) study is available here: see Box 1.

Slide 15



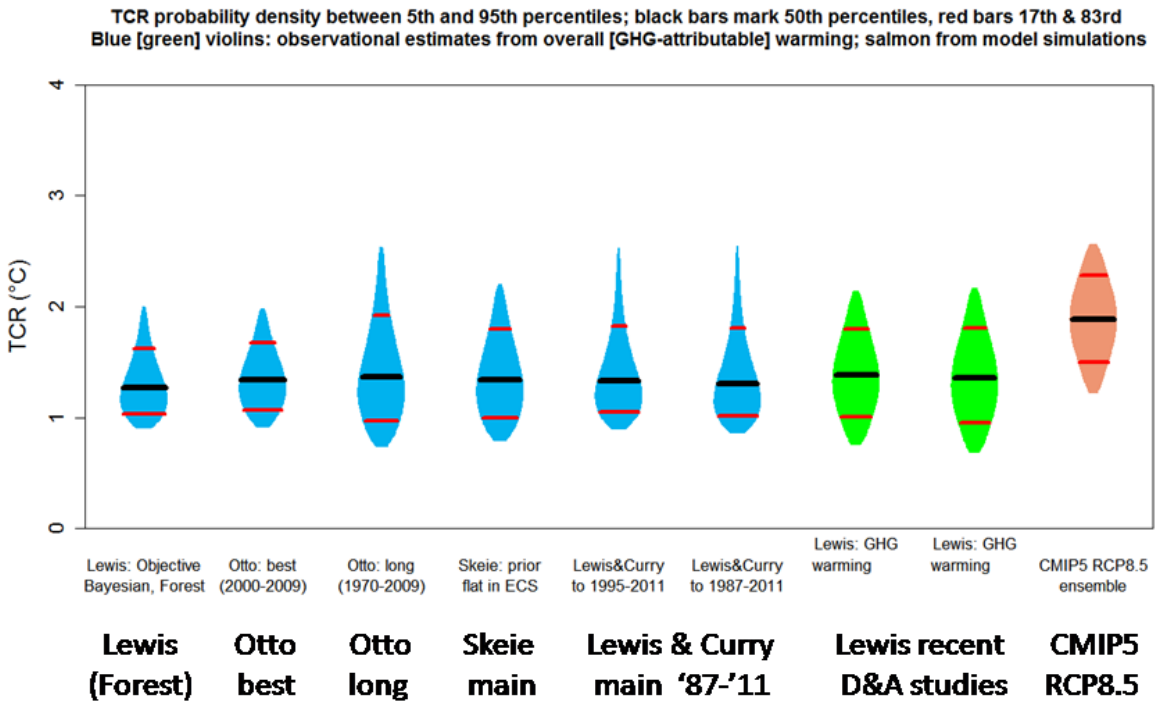
This slide shows what I regard as the least-flawed ECS estimates based on observed warming over the instrumental period, and compares them with ECS values exhibited by the RCP8.5 simulation ensemble of CMIP5 models. I should arguably have included the Schwartz (2012) and Masters (2014) estimates, but I have some concerns about the GCM-derived forcing estimates they use.

The violins span 5–95% ranges. Black lines show medians, red lines 17% and 83% bounds. Published estimates based directly on observed warming are shown in blue. Unpublished estimates of mine based on warming attributable to greenhouse gases inferred by two recent detection and attribution studies are shown in green. CMIP5 models are shown in salmon.

The 'Aldrin ECS^{-2}' violin is for its estimate that uses a uniform prior for 1/ECS, which equates to a ECS^{-2} prior for ECS. I believe that to be much closer to a noninformative prior than is the uniform-in-ECS prior used for the main Aldrin et al (2012) results. The Lewis (Forest) estimate is based on the Lewis (2013) preferred main ECS estimate with added non-aerosol forcing uncertainty, shown in the study's Supplementary Information.

Slide 16

The best TCR estimates?



16

This slide is like the previous one, but relates to TCR not ECS.

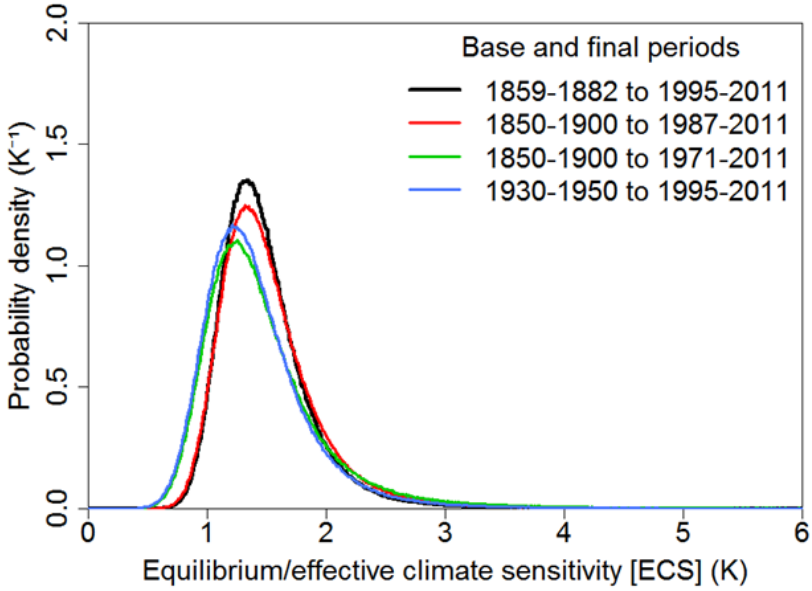
The Schwartz (2012) TCR estimate, which has been omitted for no good reason, has a median of 1.33°C and a 5–95% range of 0.83–2.0°C.

The Lewis (Forest) estimate uses the same formula as in Libardoni and Forest (2011), which also uses the MIT 2D GCM, to derive TCR for combinations of model ECS and Kv values.

Slide 17

Implications of lower aerosol forcing: ECS

Energy budget estimates using Stevens 2015 not AR5 F_{aer}



Lewis & Curry 2014: median ECS estimate 1.45 K; 5–95% 1.05–2.2 K

17

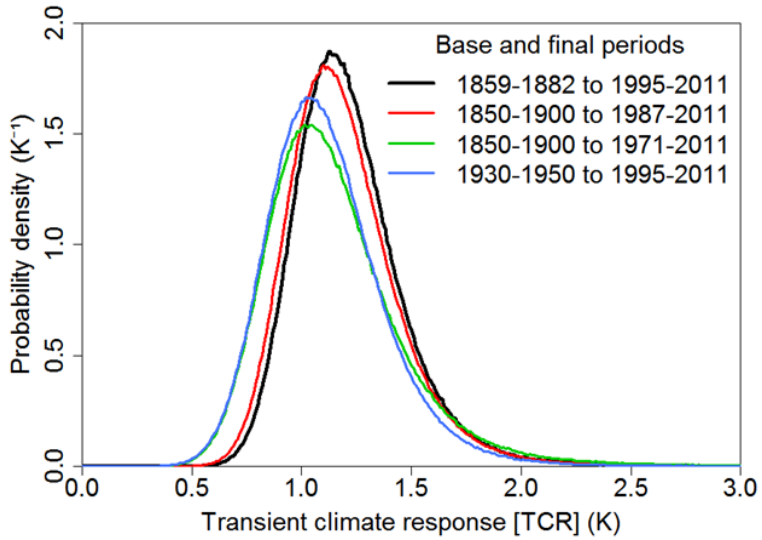
The main cause of long tails in ECS and TCR studies based on observed multidecadal warming is uncertainty as to the strength of aerosol forcing (F_{aer}). I'll end this part with a pair of slides that show how well constrained the Lewis and Curry (2014) energy-budget main ECS and TCR estimates would be if they were recalculated using the distribution for aerosol forcing implicit in Bjorn Stevens' recent study instead of the wide AR5 aerosol forcing distribution. (For some reason these slides appear much later, out of order, in the PDF version of my slides on the Ringberg 2015 website.)

The median ECS estimate reduces modestly from 1.64°C to 1.45°C, but the 95% uncertainty bound falls dramatically, from 4.05°C to 2.2°C.

Slide 18

Implications of lower aerosol forcing: TCR

Energy budget estimates using Stevens 2015 not AR5 F_{aer}



Lewis & Curry 2014: median TCR estimate 1.21 K; 5–95% 0.9–1.65 K

18

The picture is similar for TCR, although somewhat less dramatic. The median TCR estimate reduces modestly from 1.33°C to 1.21°C, but the 95% uncertainty bound falls much more, from 2.50°C to 1.65°C.

Additional references

Allen MR, Frame DJ, Huntingford C, Jones CD, Lowe JA, Meinshausen M, Meinshausen N (2009) Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, 458, 1163–6.

Frame DJ, Booth BBB, Kettleborough JA, Stainforth DA, Gregory JM, Collins M, Allen MR (2005) Constraining climate forecasts: The role of prior assumptions. *Geophys. Res. Lett.*, 32, L09702, doi:10.1029/2004GL022241.

Harris, G.R., D.M.H. Sexton, B.B.B. Booth, M. Collins, and J.M. Murphy, 2013. Probabilistic projections of transient climate change. *Clim. Dynam.*, doi:10.1007/s00382–012–1647-y.

Lewis N (2014) Objective Inference for Climate Parameters: Bayesian, Transformation-of-Variables, and Profile Likelihood Approaches. *J. Climate*, 27, 7270-7284.

Masters T (2014) Observational estimate of climate sensitivity from changes in the rate of ocean heat uptake and comparison to CMIP5 models. Clim Dyn 42:2173-2181 DOI 101007/s00382-013-1770-4

Sexton, D.M. H., J.M. Murphy, M. Collins, and M.J. Webb, 2012. Multivariate probabilistic rojections using imperfect climate models part I: outline of methodology. *Clim. Dynam.*, 38: 2513–2542.

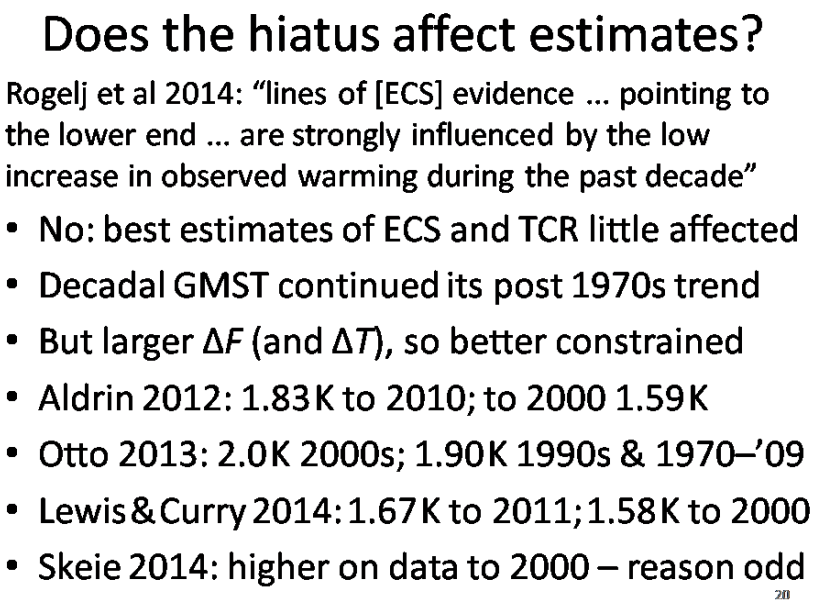
Stevens, B. Rethinking the lower bound on aerosol radiative forcing. In press, J.Clim (2015) doi: http://dx.doi.org/10.1175/JCLI-D-14-00656.1

Pitfalls in climate sensitivity estimation: Part 3

Nicholas Lewis

In Part 1 I introduced the talk I gave at Ringberg 2015, explained why it focussed on estimation based on warming over the instrumental period, and covered problems relating to aerosol forcing and bias caused by the influence of the AMO. In Part 2 I dealt with poor Bayesian probabilistic estimation and summarized the state of observational, instrumental period warming based climate sensitivity estimation. In this third and final part I discuss arguments that estimates from that approach are biased low, and that GCM simulations imply ECS is higher. I've incorporated some extra material here in the light of the content of some of the other Ringberg talks.

Slide 19



I'll start with an easy target: claims that reduced instrumental period warming based ECS estimates that have been published over the last few years reflected the hiatus in warming over the last decade. Such claims are demonstrably false. The main effect of using data extending past 2000 is to provide better constrained ECS estimates, as the anthropogenic signal rose further above background noise.

Most of recent studies that give results using data for different periods actually show lower, not higher, ECS median estimates when data extending only to circa 2000 is used. Skeie 2014 is an exception. I attribute this to the less strong observational constraints available from such data being unable to counteract the excessively negative aerosol forcing prior distribution.

Slide 20

Other possible sources of bias

- Earlier effect of F_{aer} (Shindell): <5%, from obs.
- HadCRUT4 coverage: up 14%>UAH; 29%>RSS LT
- Land sfc & night T_{min} biases likely > Artic deficit
- Land surface forcing: AR5 -0.15; total est is ~0
- Time-varying climate feedback?

21

Now for some genuine issues. First, in a 2014 paper Drew Shindell argued that inhomogeneous forcing, principally by aerosols, that had a greater concentration in the northern hemisphere, particularly in the extratropics, than homogenous GHG forcing would have a greater effect on transient global warming. That is principally because the northern hemisphere has more land and warms rapidly. That aerosol forcing reached a peak level some time ago, unlike for GHGs, also contributes. The result would be that TCR, and hence ECS, estimates based on observed global warming were biased down.

I think there is something in Shindell's argument, but I regard his GCM-based estimate of the magnitude of the bias as absurdly high. Based on a simple model and observational constraints as to the ratios of transient warming for various latitude zones, I estimate a bias of no more than about 5%. It would be difficult to reconcile a significant bias with estimates from the non-energy budget 'good' studies, which should be unaffected by this issue due to the use of models that resolve forcing and temperature by hemisphere and within each hemisphere by latitude zone and/or land vs ocean being in line with energy-budget based estimates.

In his Ringberg talk, Gavin Schmidt stated that in the GISS-E2-R AOGCM, the transient responses (over ten years) to aerosol forcing and land use were respectively 1.33x and 3.65x as large as that to GHG forcing. From this he deduced that TCR and ECS estimated from the model's historical run were biased low by about 20% and 30% respectively. Picking (over high) median estimates based on historical period unadjusted forcing, of 1.6°C for TCR and 1.9°C for ECS, he claims that these go up by respectively 35% and 60% when adjusted for forcing-specific 'transient efficacy'. I am at a loss to understand how the diagnosed increases of 20% and 30% magically turned into increases of 37% and 63% – maybe this was achieved by using uniform priors. The resulting estimates greatly exceed the GISS-E2-R model TCR of 1.4°C and 2.3°C for ECS. The very large estimated land use forcing

transient efficacy is based on an unphysical regression line that implies a very large GMST increase with zero land use forcing.

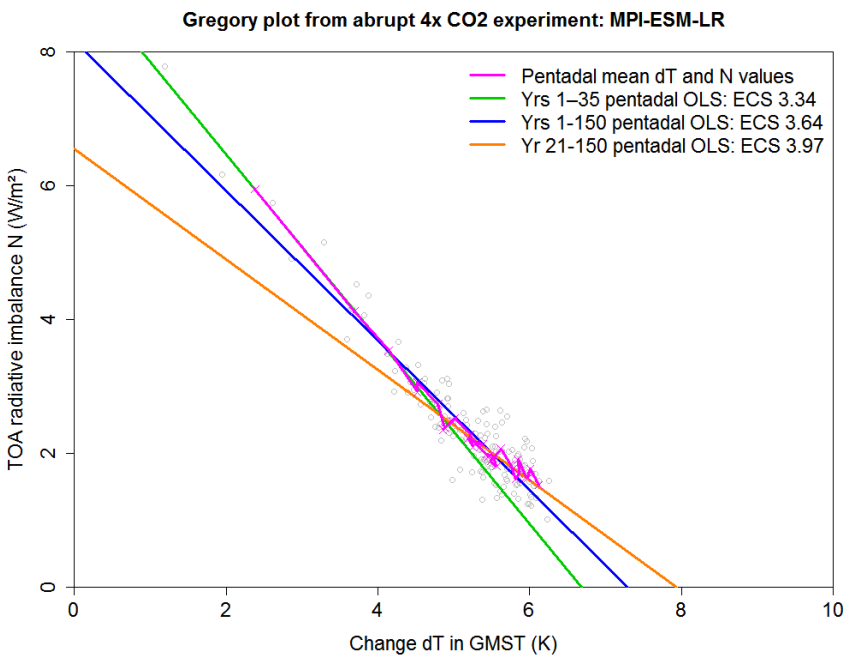
If, despite my doubts, the results Gavin Schmidt presented are correct for the GISS-E2-R model, they would support Drew Shindell's argument in relation to that model. But it would not follow that similar biases arise in other models or in the real world. I am aware of only two other AOGCMs for which transient efficacies have been likewise been diagnosed using single-forcing simulations (Shindell used the much less satisfactory standard CMIP5 simulations). One of those models shows a significantly lower transient efficacy for aerosol forcing than for GHG (Ocko et al 2014), behaviour that implies TCR and ECS estimates based on historical warming would be biased up, not down. The other model also appears to show that behaviour, albeit based only on preliminary analysis.

The next two bullet points concern arguments that the widely-used HadCRUT4 surface temperature dataset understates the historical rise in GMST. However, over the satellite era, which provide lower troposphere temperature estimates with virtually complete coverage, HadCRUT4 shows a larger global mean increase than does UAH and, even more so, RSS. It seems quite likely that upward biases arising from land surface changes (UHI, etc.) and the destabilisation of the nocturnal boundary layer (McNider et al 2012) exceed any downwards bias resulting from a deficit of coverage in the Arctic.

For land surface changes, AR5 gives a negative best estimate for albedo forcing but states that overall forcing is as likely positive as negative. On that basis it is inappropriate to include negative land surface forcing values when estimating TCR and ECS from historical warming. Those studies (probably the majority) which include that forcing will therefore tend to slightly overestimate TCR and ECS.

The final point in this slide concerns the argument, put quite strongly at Ringberg (e.g., see here) that climate feedback strength declines over time, so that ECS – equilibrium climate sensitivity – exceeds the effective climate sensitivity approximation to it estimated by comparing transient changes in GMST, forcing and radiative imbalance (or its counterpart, ocean etc. heat uptake). As explained in Part 1, in many but not all CMIP5 models global climate feedback strength declines over time, usually about 20 years after the (GHG) forcing is imposed. I address this issue in the next slide.

Slide 21



As running AOGCMs to equilibrium takes so long, their ECS values are generally diagnosed by regressing the top of atmosphere (TOA) radiative imbalance N – the planetary heat absorption rate – on dT, the rise in GMST during a period, typically 150 years, following a simulated abrupt quadrupling (300% increase) in CO_2 concentration. The regression line in such a 'Gregory plot' is extrapolated to N = 0, indicating an equilibrium state. ECS is given by half the dT value at the N = 0 intercept, since CO2 forcing increases logarithmically with concentration (that attributable to a 300% increase being twice that for a 100% increase).

Slide 21, not included in my Ringberg talk, illustrates the potential bias in estimating ECS from observed warming over the instrumental period. It is a Gregory plot for the MPI-ESM-LR model (chosen in honour of the Ringberg hosts). The faint grey open circles show annual mean data, that closest to the top LH corner being for year 1. The magenta blobs and line show pentadal mean data, which I have used to derive linear fits (using ordinary least squares regression). The curvature in the magenta line (a kink after about year 30) indicates that climate feedback strength (given by the slope of the line) is decreasing over time.

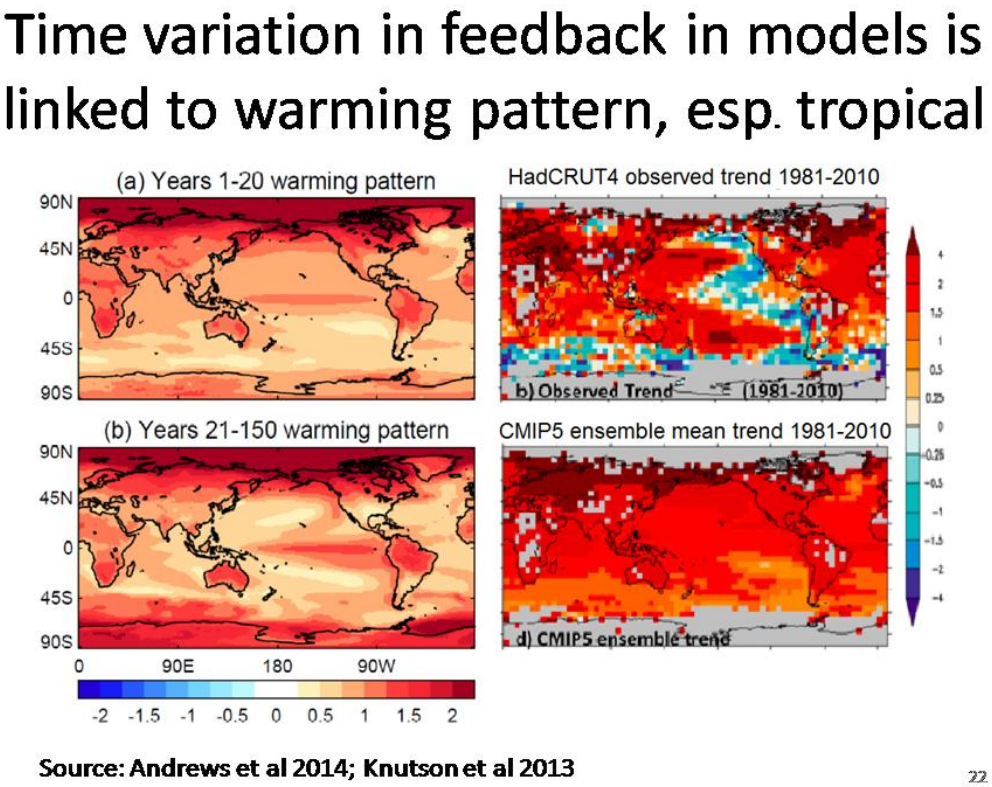
CMIP5 model ECS values given in AR5 were based on regressions over all 150 years of data available, as for the blue line in the slide. I have compared ECS values estimated by regressing over years 21-150 (orange line), as in Andrews et al (2014), with ECS values estimated from the first 35 years (green line). Since the growth in forcing to date approximates to a 70-year linear ramp, and the average period since each year's increase at the end of a ramp is half the ramp period, 35 years from an abrupt forcing increase is fairly representative of the observable data.

On average, ECS diagnosed for CMIP5 models by regressing over years 21-150 of their abrupt 4x CO_2 Gregory plots exceeds that diagnosed from years 1-35 data by 19%. Excluding models with a year 21-150 based ECS exceeding 4°C reduces the difference to 12%. This is fairly minor. The difference is not nearly large enough to reconcile the best estimates of ECS from observed warming over the instrumental period with most CMIP5 model ECS values. And it is not relevant to differences between observationally-based TCE estimates and generally higher AOGCM TCR values.

It is moreover unclear that higher AOGCM ECS values diagnosed by Gregory plot regression over years 21-150 are more realistic than those starting from year one. Andrews et al (2014) showed, by running the HadGEM2-ES simulation for 1290 years (to fairly near equilibrium), that the 36% higher ECS diagnosed for it from regressing over years 21-150 was more excessive than ECS diagnosed from regressing over years 1-35 was insufficient.

Importantly, an increase in effective climate sensitivity over time, if it exists, is almost entirely irrelevant when considering warming from now until the final decades of this century. The extent of such warming, for a given increase in GHG levels, is closely dependent on TCR. Even if effective climate sensitivity does increase over time, that would not bias estimation of TCR from observed historical warming. The projected effect on warming would be small even over 300 years – only about 5% for a ramp increase in forcing, if one excludes HadGEM2-ES and the Australian models, two of which are closely related to it and the third is an outlier.

Slide 22



Although the increase in effective climate sensitivity, or reduction in climate feedback strength, with time in many CMIP5 models appears to have little practical importance, at least on a timescale of up to a few centuries, finding out why it occurs is relevant to gaining a better scientific understanding of the climate system.

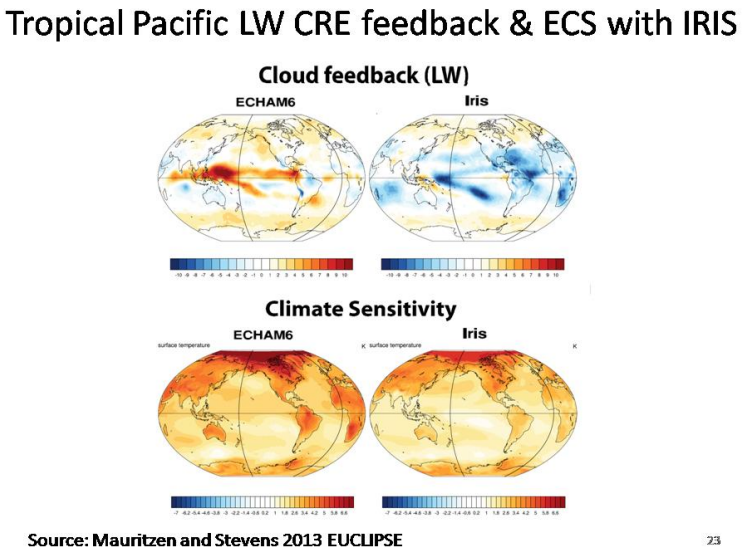
In a model-based study, Andrews et al (2014) linked the time-variation to changing patterns of sea-surface temperature (SST), principally involving the tropical Pacific. In current AOGCMs, after an initial delay of a few years, on a multidecadal timescale the eastern tropical Pacific warms significantly more than the western part and the tropical warming pattern becomes more El-Nino like, affecting cloud feedback.

The two LH panels in slide 22, from Tim Andrews' paper and talk, show the CMIP5 model ensemble mean patterns of surface warming during the first 20 and the subsequent 130 years after an abrupt quadrupling of CO_2. The colours show the rate of local increase relative to that in GMST. It can be seen that even during the first 20 years, warming is strongly enhanced across the equatorial Pacific.

The RH panels, taken from a different paper, show observed and modelled patterns of warming over 1981–2010. The CMIP5 ensemble mean trend (bottom RH panel) shows a pattern in the tropical Pacific fairly consistent with that over the first 20 years of the abrupt 4x CO_2 experiment, as one might expect. But the observed trend pattern (top RH panel) is very different, with cooling over most of the eastern tropical Pacific, including the equatorial part.

So observations to date do not appear consistent with the mean evolution of eastern tropical Pacific SST predicted by CMIP5 models. Given Tim Andrew's finding that weakening of climate feedback strength over time in CMIP5 models is strongly linked to evolving eastern tropical Pacific SST patterns, that must cast considerable doubt on whether effective climate sensitivity increases over time in the real world.

Slide 23



There are other reasons for doubting the realism of the changing SST patterns in CMIP5 models that Andrews et al (2014) found to be linked to increasing climate sensitivity.

The strong warming in the deep tropics across the Pacific over years 21–130 is linked to longwave (LW) cloud feedback, which in CMIP5 models strengthens and spreads further after years 1–20. But is this behaviour realistic? As well as their main new CMIP6 model MPI-ESM2 (ECHAM6 plus an ocean module), Thorsten Mauritzen has been developing a version with a LW iris, an effect posited by Dick Lindzen some years ago (Lindzen et al 2001). The slides for Thorsten Mauritzen's Ringberg talk, which explained the Iris version and compared it with the main model, are not available, but slide 23 comes from a previous talk he gave about this work. It shows the equilibrium position; only a slab-ocean version, which equilibriates after a short simulation period, has been run so far.

As the top panels show, unlike the main ECHAM6/MPI-ESM2 model the Iris version exhibits no positive LW cloud feedback in the deep tropical Pacific. And the bottom panels show that, as expected, warming in the central and eastern tropical Pacific remains modest. This suggests that, if the Iris effect is real, any increase in effective climate sensitivity over time would likely be much lower than CMIP5 model ensemble mean behaviour suggests. The Iris version also has a lower ECS than the main model, although not as low as might be expected from the difference in LW cloud feedback as this is partially offset by a more positive SW cloud feedback.

Slide 24

Other ECS/TCR estimation methods

- Short term solar/volcanic/TOA: ECS unreliable
- Paleo: large uncertainty so not much impact
- GCM PPEs constrained by climate mean/change
- Raw GCMs (+ feedbacks in GCMs): ECS and TCR
- Informed but not determined by observations
- CMIP5 RCP8.5 GCMs: mean ECS 3.4 K, TCR 1.9 K

26

Slide 24 lists methods of estimating ECS other than those based on observed multidecadal warming. I explained in Part 1that I concurred with AR5's conclusions that estimating ECS from short term responses involving solar or volcanic forcing or TOA radiation changes was unreliable, and that true uncertainty in paleoclimate estimates was larger than for instrumental period warming based estimates. That implies that combining paleo ECS estimates with those based on instrumental period warming would not change the latter very much. I also showed, in Part 2, that the model most widely used for Perturbed Physics/Parameter Ensemble studies, HadCM3/SM3, could not successfully be constrained by observations of mean climate and/or climate change, and so was unsuitable for use in estimating ECS or TCR. (Such use nevertheless underlies UKCP09, the official UK 21st century climate change projections.)

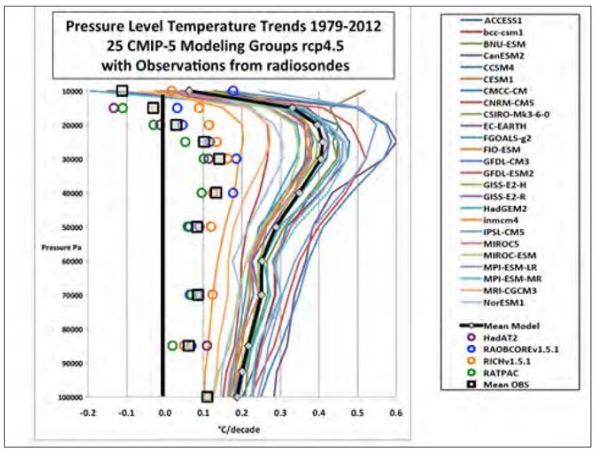
The other main source of ECS estimates involves GCMs more directly. Distributions for ECS and TCR can be and are derived from estimated model ECS and actual model TCR values. A 5-95% ECS range for CMIP5 models, of 2–4.5°C, was given in Figure 1, Box 12.2 of AR5. Feedbacks exhibited by GCMs can also be analysed. But although development of GCMs is informed by observations, their characteristics are not determined by observational constraints. If the climate system were thoroughly understood and AOGCMs accurately modelled its physical processes on all scales that mattered, one would expect all aspects of their behaviour to be similar, and the ECS and TCR values they exhibited might be regarded as reliable estimates. However, those requirements are far from being satisified.

Since AOGCMs tend to be similar in many respects, it is moreover highly doubtful that a statistically-valid uncertainty range for ECS or TCR can be derived from CMIP5 model ECS and TCR values. If some key aspect of climate system behaviour is misrepresented in (or unrepresented by) one CMIP5 model, the same problem is likely to be common to many if not all CMIP5 models.

I'll finish by highlighting two areas that are relevant to climate sensitivity where model behaviour seems unsatisfactory across almost all CMIP5 models.

Slide 25

Tropical warming by pressure level



1979-2012 mean model trends 2x to 4x reality Source: Christy 2014 presentation

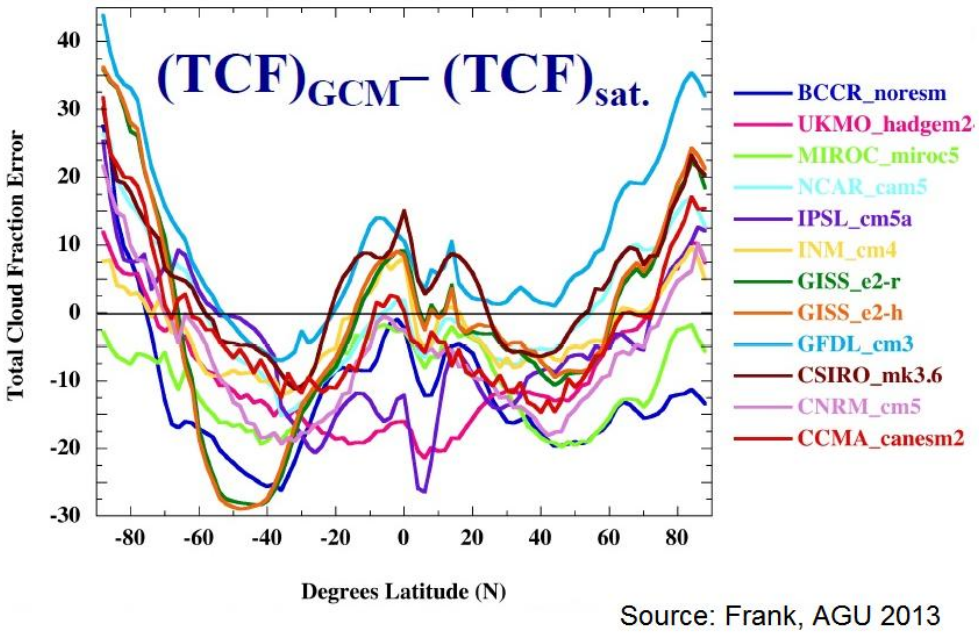
29

Slide 25 compares tropical warming by pressure level (altitude) in CMIP5 models and radiosonde observations over 1979-2012. Most models not only show excessive near-surface warming, by a factor of about two on average, but a much greater increase with height than observations indicate. This is the 'missing hot-spot' problem. The ratio of tropical mid-troposphere to surface warming

would be expected to be smaller in a model with a LW iris than in one that does not, a point in favour of such a feature.

Figure 9.9 in AR5 showed much the same discrepancy – an average factor of about 3x –between observed and modelled temperature trends in the tropical lower troposphere, over 1988-2012. Observations in that case were based on satellite MSU datasets and reanalyses that used models to assimilate data.

Slide 26



Also, models get low cloud LWP, O depth and albedo very wrong (Stephens 2010)

A lot of the discussion at Ringberg 2015 concerned clouds, one of the most important and least well understood elements of the climate system. Their behaviour significantly affects climate sensitivity.

Slice 26 shows errors in cloud fraction by latitude for twelve CMIP5 GCMs. $(TCF)_{sat}$. is the average per MODIS and ISCCP2 observations. It can be seen that most models have too little cloud cover in the tropics and, particularly southern, mid latitudes, and too much at high latitudes. Errors of this magnitude imply, IMO, that reliance should not be placed on cloud feedbacks exhibited by current climate models.

Models also appear to have excessive liquid water path, optical depth and albedo, which may result in negative optical depth climate feedback being greatly underestimated in models (Stephens 2010).

Slide 27

ECS & TCR estimation: Conclusions

- Use observational data, not raw GCMs (or PPEs)
- All AR4 and pre-2012 AR5 studies badly flawed
- Overstrong aerosols, bad priors, AMO influence
- Sound recent long term warming studies agree
- Estimates narrowed but unaltered over hiatus
- Low aerosol (Stevens): 95% ECS<2.2; TCR<1.65K
- Most CMIP5 models have excessive TCR & so ECS

Thank you for listening, and forgive me
for being so frank in discussing studies

My concluding slide reiterates some of the main points in my talk. Assuming Bjorn Stevens' revised estimate of aerosol forcing is correct, then the 95% uncertainty bounds on ECS and TCR from observed multidecadal warming are well below the mean ECS and TCR of CMIP5 models. It will be very interesting to see how these discrepancies between models and observations are resolved, as I think is likely to occur within the next decade.

Additional references

Timothy Andrews, Jonathan M. Gregory, and Mark J. Webb, 2015: The Dependence of Radiative Forcing and Feedback on Evolving Patterns of Surface Temperature Change in Climate Models. *J. Climate*, **28**, 1630–1648

Lindzen, RS, M-D Chou, AY Hou (2001) . Does the Earth have an adaptive infrared iris? Bull. Amer. Meteor. Soc. 82, 417-432, 2001

Mauritsen, T and B Stevens (2015) Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models. *Nature Geoscience* doi:10.1038/ngeo2414

McNider, R. T., et al. (2012) Response and sensitivity of the nocturnal boundary layer over land to added longwave radiative forcing. J. Geophys. Res., 117, D14106.

Ocko IB, V Ramaswamy and Y Ming (2014) Contrasting Climate Responses to the Scattering and Absorbing Features of Anthropogenic Aerosol Forcings *J. Climate*, **27**, 5329–5345

Rogelj J, Meinshausen M, Sedlácek J, Knutti R (2014) Implications of potentially lower climate sensitivity on climate projections and policy. *Environ Res Lett* 9. doi:10.1088/1748-9326/9/3/031003

Shindell, DT (2014) Inhomogeneous forcing and transient climate sensitivity. Nature Clim Chg: DOI: 10.1038/NCLIMATE2136

Stephens, GL (2010) Is there a missing low cloud feedback in current climate models? GEWEX News, 20, 1, 5-7.