Extended Summary of the

Climate Dialogue

on

Climate Sensitivity and Transient Climate Response

Bart Strengers (PBL, Netherlands Environmental Assessment Agency)

December 2014

Content

1. Introduction

The Climate Dialogue on Climate Sensitivity and Transient Climate Response took place from 12 May 2014 and lasted almost four months, although almost 90% of the expert discussion took place in the first two. Both Equilibrium Climate Sensitivity (ECS) and Transient Climate Response (TCR) summarize the global climate system's temperature response to an externally imposed radiative forcing (RF), expressed in W/m2. ECS is defined as the equilibrium change in annual global mean surface temperature following a doubling of the atmospheric CO_2 concentration. TCR is the expected transient change in temperature over a period of 70 years assuming a linear doubling of the atmospheric CO_2 concentration in this period, i.e. before equilibrium has been reached. Both metrics have a broader application than these definitions imply: ECS determines the eventual warming in response to stabilization of atmospheric composition on multi-century time scales, while TCR determines the warming expected at a given time following any steady (and linear) increase in forcing over a 50- to 100-year time scale. TCR is a useful metric next to ECS because it can be estimated more easily than ECS, and is more relevant to projections of warming over the rest of this century.

Climate sensitivity is at the heart of the scientific debate on anthropogenic climate change. In the fifth assessment report of the IPCC (AR5) it is indicated that the peer-reviewed literature provides no consensus on a formal statistical method to combine different lines of evidence (i.e. different methods to estimate ECS, see Chapter 2). Therefore, in AR5 the range of ECS (and TCR) is expert-assessed and they conclude that the Equilibrium Climate Sensitivity (ECS) is likely in the range from 1.5°C to 4.5°C. Obviously, this expert judgement in AR5 has been performed deliberately, but it is not a straightforward procedure. The discussion on how to weigh the different lines of evidence is very old, not only in the scientific literature but also in the blogosphere and in reports and is still going on.

We invited three experts: John Fasullo, James Annan and Nic Lewis. Fasullo is a project scientist at the National Centre for Atmospheric Research (NCAR) in Boulder, Colorado, studying processes involved in climate variability and change using both observations and models. He has published extensively on the topic and was co-author of the assessment reports of the IPCC. James Annan has worked as senior scientist at the Japanese Research Institute for Global Change, JAMSTEC- perhaps better known as the home of the Earth Simulator – for the past 13 years. He published many papers and his work has been heavily cited in the recent IPCC AR5. Nic Lewis is an independent climate scientist who studied mathematics and physics at Cambridge University. He published two key papers on ECS and TCR (Otto, 2013 and Lewis, 2013), one of them together with prominent IPCC lead authors. Both are cited and discussed in AR5.

	# comments	Average Length in words	Contribution in comments	Contribution in words
Nic Lewis	34	750	39%	53%
John Fasullo	12	450	14%	11%
James Annan	10	175	11%	4%
Moderator	14	420	16%	12%
Public	18	550	20%	20%
Total	**88**	**550**	**100%**	**100%**

Table 1. *Statistics of the Climate Dialogue on CS and TCR, lasting from 12 May to 9 September 2014.*

The contribution of the discussants to the dialogue was rather imbalanced. In terms of comments, almost 40% of the discussion was covered by Nic Lewis and in terms of words this was even more than halve (see Table 1). James Annan contributed 11% in terms of comments and 4% in terms of words. This was partly because of time limitations, but maybe also because, after writing his guest blog and his reaction to the blog of Nic Lewis and to the blog of John Fasullo, he indicated: "*I'm not sure that I have a lot to add to my previous comments*". Also, the involvement of the public was limited to 20% of the discussion, much less than we experienced in other discussions.

The experts' guest blogs dealt with all questions raised in our introduction, but due to the broadness of the subject and time limitations of the participating experts, we managed to cover the questions on ECS only and not those on TCR, i.e. questions 1 to 4 as formulated in the introduction, which are discussed in chapters 2 to 5.

2. What are the pros and cons of the different lines of evidence?

Figure 1 in the introduction shows the ranges and best estimates of ECS in AR5[1] based on studies that support different lines of evidence, which are: 1) the observed or instrumental surface, ocean and/or atmospheric temperature trends since pre-industrial time, 2) observed and modelled short-term perturbations of the energy balance such as those caused by volcanic eruptions, included under instrumental in figure 1, 3) climatological constraints by comparing patterns of mean climate and variability in models to observations, 4) climate models, 5) temperature fluctuations as reconstructed from paleoclimate archives and 6) studies that combine two or more lines of evidence.

2.1. Instrumental period warming-based studies

In this climate dialogue Nic Lewis was outspoken in his conviction that instrumental period warming-based studies are superior to all other studies that estimate ECS. Of this type of study cited in AR5, he only considered four studies, which agree with an energy budget analysis based on AR5 forcing and

[1] AR5 = The fifth assessment report of working group I the IPCC(2013).

heat uptake estimates, to be satisfactory for estimating ECS. In his blog he wrote that, studies that are "*both reliable and able to usefully constrain ECS*[2]" are "*Aldrin (2012), Ring (2012), Lewis (2013)*[i] *and Otto (2013), in accordance with the conclusions of AR5.*" With respect to all other instrumental based studies (including combination studies) Lewis indicated they should be rejected for different reasons (see also Table 2):

1. Instrumental studies using the aerosol forcing of AR4 are useless because in AR5 it is substantially higher (less negative), based on better scientific understanding and observational data. Link. More details can be found in paragraph 2.4.2 and 5.2.
2. Studies that use values for aerosol forcing that are estimated along with ECS using *global* mean temperature data are useless. Because the time-evolution of global aerosol forcing is almost identical to that from GHGs, it is impossible to estimate both aerosol forcing – which largely affects the northern hemisphere – and ECS (or TCR) with any accuracy. The Northern Hemisphere (NH) and Southern Hemisphere (SH) must be separated. Blog
3. Most ECS estimates in AR4 and most of the instrumental studies cited in AR5 use uniform priors in ECS (i.e. the starting assumption that all climate sensitivities are, over a very wide range, equally likely). A uniform prior prevents a fair estimate of ECS because ECS has a non-linear relationship with the observational data from which it is being estimated. This greatly inflates the upper uncertainty bounds for ECS. Blog. More details can be found in paragraph 5.4 on the use of priors.
4. Observational studies using expert priors, which a number of the AR5 studies did, produce ECS estimates that reflect the prior, with the observational data having limited influence. Blog and paragraph 5.4.
5. The ocean heat uptake, if used, is in some case either incorrectly calculated, or erroneous Ocean Heat Content (OHC) data is used or treated in a non-standard way. Blog
6. Studies that use biased General Circulation Model (GCM)-based anthropogenic warming estimates instead of directly observed temperatures. Blog and paragraph 2.4.
7. Studies based on short-term changes such as volcanic eruptions and satellite measured variations in top-of-atmosphere (TOA) radiation (see 10.8.2.2 in AR5) are deprecated by AR5. Blog and paragraph 2.2.
8. In some cases other data was unsatisfactory or badly used. For example Libardoni & Forest (2013) used surface temperature data extended only to 1995; Lin(2010) ignored strong initial volcanic forcing.

Based on his guest blog James Annan seemed to **agree** with Nic Lewis to some extent when he writes that "*Transient 20th century warming-based estimates (Aldrin et al 2012, Ring et al 2012, Otto et al 2013) are more trustworthy than other approaches, as they are more-or-less directly based on the long-term response of the climate system to anthropogenic forcing. They point at the low end of the IPCC range* [3] *due to better quality and quantity of data and better understanding of aerosol effects.*" However, later in the discussion he was more critical as he writes that "*These estimates rely on models of the climate system, which are so simple and linear (and thus certainly imperfect) that they may not be recognized as such.*"

[2] ECS = Equilibrium Climate Sensitivity as opposed to the Effective Climate Sensitivity. As Lewis indicated the Effective CS is slightly lower than ECS, but these terms are largely used synonymously in AR5.

[3] This refers to the likely range of the ECS as reported in AR5 which is 1.5 to 4.5 ^{0}C.

Reasons to reject	Knutti 2002	Frame 2005	Annan & Hargraves 2006	Hegerl 2006	Forster & Gregory 2006	Tomassini 2007	Murphy 2009	Bender 2010	Lin 2010	Lindzen & Choi 2011	Olson 2012	Libardoni & Forest 2013
Poor aerosol forcing (1,2)	X					X			X		X	
Unsuitable use of priors (3,4)	X	X		X		X					X	X
Ocean data incorrect or mishandled (5)	X	X							X			
Temperatures from GCMs instead of observations (6)		X		X								
Based on short term changes (7)			X		X		X	X		X		
Other data unsatisfactory (8)									X			X

Table 2. *Lewis' reasons to reject instrumental period warming-based studies.*[4]

In his guest blog, John Fasullo pointed at some positive aspects of observational-based studies, by writing: "*Simple models rooted in statistics can be powerful tools for interpreting complex systems*" and in one comment he wrote that "*These studies hold the promise of saving the countless CPU-hours of computation involved in estimating ECS from a fully coupled simulation*." But the main conclusion in his guest blog was that "*Assessing ECS solely with statistical approaches using simple models that capture little of the climate system's physical complexity, trained on a limited subset of questionably relevant surface observations, and based on largely untested physical assumptions is impossible.*" And also in the remainder of the discussion he was very critical on the current achievements of instrumental studies. For example, he indicated these studies are severely limited by the absence of a unique "correct" prior, (Trenberth & Fasullo, 2013 and Shindell, 2014), see paragraph 5.4 for more details on the prior-discussion. He added that "*Without a physical understanding of the climate system, based on robust observations of key processes, […], there cannot be high confidence in climate projections"* and also pointed out that "*GCM's incorporate several orders of magnitude more observational information in their development and testing than do the typical "instrumental" approaches. The dataset I am using is NOAA AVHRR OLR. The achievement in constructing this record [since 1974] is both remarkable and unprecedented, and lessons learned have contributed to numerous follow-on efforts (e.g. CALIPSO, CERES, CLOUDSAT, ERBE, GPCP, GRACE, ISCCP, QUIKSCAT, SSM/I, TOPEX, TRMM, …). Given this era of such remarkable observations, accompanied by similar achievements across a realm of disciplines (e.g. ocean and atmospheric observations, operational models, reanalysis methods, supercomputing, …) I cannot help but be struck by the fact that there are those advocating for assessing climate solely with statistical approaches using simple models that*

[4] More details can be found in Lewis' blog and here

capture little of the climate system's physical complexity, trained on a limited subset of questionably relevant surface observations, and based on largely untested physical assumptions."

Lewis replied that finding the appropriate 'prior' is far more of a problem with a GCM "*because of the much higher dimensionality of the parameter space – a GCM has hundreds of parameters. […] The number of degrees of freedom available in relevant observations is limited.[…] It is therefore more practicable to constrain a smaller number of parameters using observations."* Furthermore, he indicated he did not claim that observations alone can be used to generate a useful estimate of ECS and that he **agrees** that observationally-based ECS estimates also involve use of climate models. But he added that although "*GCMs or similar climate models are needed for observationally-based ECS estimates, their ECS values have very little effect on those estimates.*"

More in general John Fasullo agreed that **"***the statistical approach [as also applied in the four studies of Lewis, red] has the potential to play an important role in constraining ECS"* but only "*once the strengths and weaknesses are broadly understood by exploring such methods in a framework that is tightly constrained. Using a GCM whose sensitivity is known and whose variability is thoroughly vetted provides such an opportunity. […] Colleagues and I at NCAR are currently collaborating in an effort to do just this*." Fasullo indicated both in his blog and in a later comment that this effort will show that uncertainty in observations and the need to disentangle the response of the system to CO_2 from the convoluting influences of internal variability and responses to other forcings (aerosols, solar, etc) entails considerable uncertainty in ECS as also shown by Schwartz (2012) who determined an ECS range from 1.16 ± 0.09 to 4.9 ± 1.2 K, more than spanning the IPCC estimated "likely" uncertainty range. In other words, these studies "*do not resolve individual feedbacks and thus how to incorporate them in the approach for future progress remains unclear*".

With respect to the latter, Lewis held the opposite opinion by stating that "*the climate system may be too complex and current understanding of it too incomplete for strong constraints on ECS or TCR to be achieved in the near future from just narrowing constraints on individual feedbacks"* and also on Schwartz he disagreed because "*The upper part of the range, 3.0-6.1 in Schwartz (2012) derives from a poor quality regression using one of six alternative forcings datasets: MIROC. Its regression of temperature change on forcing has an R^2 of only 0.29, far lower than for the remaining four datasets (R^2 from 0.54 to 0.78) and is thus rejected by me.*"

Fasullo replied that Lewis'reasoning to reject MIROC is an example of "*the same lack of questioning of basic assumptions that I've identified in your other work*" and "*is based on what you 'believe' is the right value of R^2 for the relationship between forcing and temperature, and that it should be high. In fact, we know very well from both GCMs and observations that surface temperature and forcing need not correlate strongly at all, and that their degree of correlation over any finite and therefore transient record can be strongly positive, weak, or even negative as a consequence of internal variability."* and therefore "*there is no basis for narrowing the range of uncertainty presented by Schwartz et al. 2012*".

Nic Lewis replied that if "*you reject the simple proportionality model underlying Steve's study [i.e. Schwartz et al, 2012, red] then you should certainly not conclude – as you do – that his uncertainty range stands."* Fasullo disagreed: "*…on very long timescales (a century and longer) one would expect fairly good coherence. On shorter timescales, the expectation is that the coherence would degrade considerably due to internal variability. On decadal timescales, we find in GCM simulations that variability in global mean temperature arising from forcing can easily be swamped by internal variability. Variance on this and shorter timescales is likely to be key to the criterion you use for rejecting the MIROC forcing dataset used in Steve's paper. So my answer is no, I see no need to reject* *the approach of Schwartz (which is centered primarily on lower frequency variability) while having misgivings regarding your approach for rejecting Steve's uncertainty range."*

From Steve Schwartz: "*The exclusion of the data sets [by Lewis, red] from my further analysis was based on the fact that they did not fit the model relating forcing and observation, but I would not use even that to exclude such forcing histories from the realm of possibility; we need to evaluate forcing independently from its implications on response. Try to maintain a firewall. Otherwise it becomes circular reasoning*."

Lewis and Fasullo did agree , however, with the conclusion in Schwartz (2012) that "*The main drawback [of instrumental period warming-based studies, red] is the large uncertainty as to changes in total radiative forcing, resulting principally from uncertainty in aerosol forcing."* Where Fasullo added that *"…this point is fundamental to Steve's uncertainty range.*"

2.2. Short-term perturbations of the energy balance

There was no discussion on studies based on observed and modelled short-term perturbations of the energy balance such as those caused by volcanic eruptions. In his blog, Nic Lewis indicated that these studies have "*timescales different from those relevant for climate stabilization*", that "*Some of these studies have non-overlapping uncertainty ranges*." and that it "*is unclear whether the estimates they arrive at really represent ECS, or something different. The case for simply disregarding all such estimates as unreliable is stronger there*." Although John Fasullo indicated in his blog that it is "*inappropriate to place high confidence in any single approach.",* he was also critical with respect to these perturbation studies when he wrote that "*The relevance of perturbation studies are limited by the degree to which they can serve as analogues to climate change, the certainty with which their forcing can be known, and the potentially complex and poorly understood interactions between that forcing and nature (e.g. clouds)."*

2.3. Climatological constraint studies

No in depth discussion took place. Nic Lewis mentioned them a few times during the discussion, but from the different blogs it can be concluded that all three experts more or less agreed that these studies are of limited value:

- In his blog, Nic Lewis wrote that "*All Perturbed Physics Ensemble (PPE) studies in AR5 are based on the UK HadCM3/SM3 GCM and thus only reflect the characteristics of this GCM. It has a structural link – probably via clouds – between ECS and aerosol radiative forcing. At parameter settings that produce even (moderately) low ECS values, aerosol cooling becomes so high that the model climate becomes inconsistent with observations.*" In a later comment he added that in PPE studies "*it may prove impracticable to explore all combinations of climate system properties that are compatible with the observations. That was the problem with the Sexton et al (2012) and Harris et al (2013) PPE studies.*"
- James Annan agreed as he wrote in his blog that the result of PPE studies, including his, are "*highly dependent on the underlying GCM , as was first shown by Yokohata et al 2010 and has also been confirmed by others (Klocke et al 2011)"* and he therefore no longer considers "*such methods to be of much use*" and adds that "*the underlying problem here appears to be that changing parameters within a given GCM structure does not adequately represent our uncertainty regarding the climate system.*"

- John Fasullo indicated in his blog that “*Useful insight has been gained for some fields (e.g. snow cover and water vapor, Hall & Qu 2006 or Soden et al 2002)”, but added that “The relevance of perturbation studies are limited by the degree to which they can serve as analogues to climate change, the certainty with which their forcing can be known, and the potentially complex and poorly understood interactions between that forcing and nature (e.g. cloud, Dessler, 2010).*” and therefore it is “*Difficult to establish statistical confidence in identified relationships, due to a lack of independence across GCMs, and the need to firmly establish a physical basis for why a climatological constraint should act as an indicator of future change.*”

2.4. State of the art General Circulation Models (GCMs)

A substantial part of the dialogue was devoted to GCMs. Nic Lewis indicated that GCMs “*Offer the most accurate way of constraining ECS once they are known to represent all significant climate system processes sufficiently accurately”.* However, he adds that the current generation of GCMs are useless to estimate ECS because “*Many combinations of GCM parameters can produce good simulations of the current climate but with substantially different ECSs (Forest, Stone & Sokolov, 2008)*”. or, in other words: “*Even if reasonably low ECS values can be achieved, they may always be accompanied by values for other climate system properties that make the simulated climate unrealistic.*”

In his blog, John Fasullo wrote that “*GCMs offer a uniquely physical approach for estimating ECS and TCR and readily allow for controlled experimentation.*” He also mentioned the shortcomings: “*Representations of key processes is often lacking - such as the interaction of aerosols with clouds – and some processes particularly those acting on low frequency timescales or for which observations are generally unavailable, contain additional uncertainty.*” However, during the discussion he strongly emphasized that the model-approach inevitably leads to the conclusion that ECS is on the high end of the IPCC-uncertainty range (see chapter 5 for more details).

Four “*climate system properties*” of GCMs were discussed in more detail, i.e. clouds, aerosols, cross-equatorial heat transport and the discrepancy in warming between models and observations, and will be summarized below.

2.4.1. Discussion on clouds in GCMs

In his blog Lewis claimed that “*much of the Earth's surface cloudiness is too low in most models.*” Therefore he concludes that too much sunlight reaches the model-surface leading to a too high ECS: “*the excess of model ECS over 2°C comes primarily from positive cloud feedbacks and adjustments*”. And in a later comment he added that “*Cloud characteristics are 'parameterised' in GCMs rather than derived directly from basic physics*”. Fasullo disagreed when he stated that “*There is a considerable body of work […] that addresses the gap between GCMs and the microphysical scale. This work excludes strong negative cloud feedbacks and thus a low ECS.*” and “*I wonder what large negative feedback you might envision that is “not included in any model physics”? To me, it seems more like wishful thinking than informed conjecture. I too wish it were so – but I see no evidence that it is.”* Lewis replied that “*If the modellers claim that their model is correct and observational evidence is at* *fault then it is incumbent on them to prove so. It is not up to someone who accepts the observational evidence that the model is not a good representation of the real world to show where and how it misrepresents the real world.*"

More details on the discussion on clouds in models can be found in paragraph 5.1.

2.4.2. Discussion on aerosols and the discrepancy between GCMs and observations

In his blog Lewis wrote that: 1) since 1979 global surface temperatures in GCMs warmed by 50% too much on average. A period where aerosol forcing changed only little; 2) since 1988 Tropical Lower Tropospheric (TLT) temperatures over the oceans in GCMs warmed three times that of the average of observational satellite datasets (UAH and RSS); and 3) GCMs use a post 1850 aerosol forcing that is 0.4 to 0.5 W/m^2 more negative than the best estimate in AR5 and thus GCMs must be excessively sensitive to match 20th century warming. For example "*the NCAR CESM1-CAM5 model [Fasullo is from NCAR, red] matched global actual warming reasonably well"* because the diagnosed aerosol forcing (Shindell et al, 2013) was -0.7 W/m² more negative from 1850 to 2000 than the AR5's best estimate [-0.75 W/m2 over this period] and the other NCAR model, CCSM4, simulates 1) over the 1988-2012 period, four times faster warming in the tropical troposphere than the average of two satellite-observation based datasets (UAH and RSS vs CCSM4 (blue circle) in Figure 9.9 of AR5); 2) global surface warming over 1979-2013 is more than 50% higher than the observational datasets, in particular HadCrut4; 3) over the period 1950-2013 nearly 85% higher than per HadCRUT4; 4) a mean Top Of the Atmosphere (TOA) radiative imbalance of 1.1 W/m² over 2002-2011, twice the estimate calculated above from observational data. That is the greatest overestimation of any CMIP5 model.

Fasullo replied that *"the total effective aerosol forcing from AR5 [from 1750-2000, red] is –0.9 (–1.9 to –0.1) W/m^2 […]. The value of CESM1-CAM5 is about -1.5 W/m^2* so […] *I don't view it as a basis for discrediting the model's sensitivity.*" and he explained that CCSM4 does not "*include any aerosol indirect effect and so they obviously shouldn't be expected to replicate the observed temperature or energy imbalance records. It is an error in framing to suggest they should.*"

Lewis disagreed on the latter since CCSM4-simulations are "*used by the IPCC for projecting future temperatures, which is used for many purposes. Although CCSM4 does not include indirect aerosol forcing, according to Lamarque et al (2011) its change in direct aerosol forcing from 1850-2000 was -0.81 W/m², in itself slightly higher that AR5's best estimate of the change in total (direct + indirect) aerosol forcing over that period of -0.74 W/m²*."

But Fasullo replied this is not correct because "*the aerosol direct effects in CCSM4 are -0.45 W/m^2 for sulfate and +0.14 W/m^2 for the black carbon direct effect. For details, see Meehl et al 2012.*"

Fasullo also stressed the point that CCSM4 and CESM1-CAM5 are very different, as described in Gettelman et al. (2013): *"Most of the cloud/convective schemes were rebuilt from the ground up and so aerosol forcing from one cannot be assumed to be the same as the other. The contribution from clouds in the midlatitudes to the increase in climate sensitivity from CCSM4 to CESM1-CAM5 was one of the surprising aspects of that study.*"

Then Nic Lewis showed a figure of the Historical warming vs Aerosol ERF (figure 1) of all the CMIP5 models analysed in Forster et al 2013 for which Shindell et al 2013 gives Aerosol ERF estimates and concluded that the correlation is "*almost 0.9*", and "*The Aerosol ERF for the best fit line through the points in the figure that corresponds to Historical warming of 0.75°C is about -1.1 W/m². By contrast, the AR5 best estimate for the increase in Aerosol ERF over the same period as that diagnosed in Shindell et al 2013 (1850 to 2000) is -0.75 W/m², some 0.35 W/m² less negative.*"

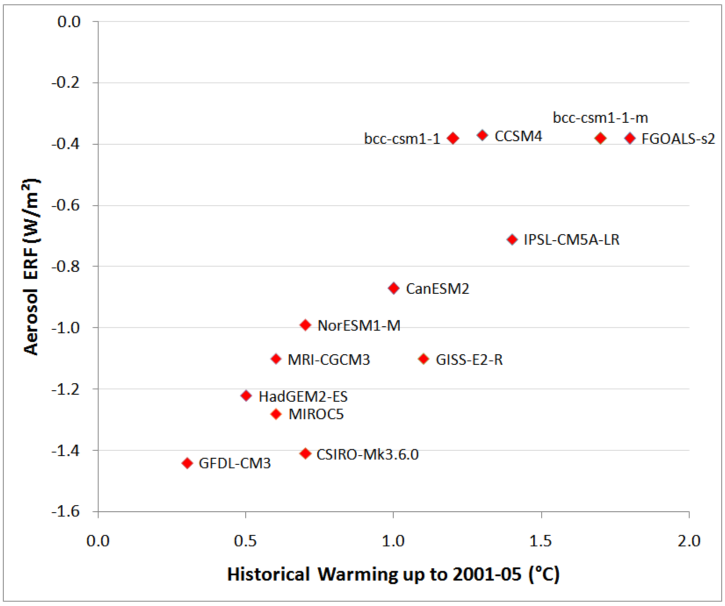

Figure 1 *Historical warming versus Aerosol Effective Radiative Forcing (ERF). The marker for CESM1(CAM5), which was not analyzed in Forster et al 2013, would be almost identical to that for CSIRO-Mk3.6.0.There is a large spread in simulated Historical warming to 2001-2005, but by then the models analyzed were on average simulating a significantly greater rise in surface temperature than observed.*

So, in *summary*, Nic Lewis argues that GCMs which reproduce the observed warming have an aerosol forcing on the more negative side of the abovementioned range. This is why until the last few decades GCMs did not simulate excessive surface warming or ocean heat uptake despite having high ECS and TCR values, but have done so since then. Their large aerosol forcing may have indirectly led to GCMs having high sensitivities, as the model developers chose model variants and tuned them with an eye on matching the historical temperature record. Fasullo replied that the aerosol forcing values in models fall well within the wide uncertainty range of AR5, which is -0.1 to -1.9 W/m2 and therefore the conclusion of Lewis is, according to him, unjustified.

2.4.3. Discussion on cross-equatorial heat transport

Nic Lewis brought up the point that "*The average cross-equatorial ocean heat transport is 0.2 PW northward. Figure 9.21 of AR5 shows that the CMIP5 models have a mean northward cross-equatorial ocean heat transport four times higher. There are only four models with heat transports substantially below 0.8 PW. The 0.6 PW excess of the CMIP5 multimodel mean northwards ocean heat transport over the average of the observationally-based estimates is equivalent to an excess of forcing in the northern hemisphere over the southern hemisphere of 4.8 W/m². That excess is greater than estimated total anthropogenic forcing, so this is a major issue.*" John Fasullo replied that "*…the magnitude of the bias [in the cross-equatorial heat transport] doesn't relate in any systematic way to*

simulated climate sensitivity […]. But perhaps the lack of any relationship is unsurprising, given that the ocean heat transport is not a forcing as Nic's comments might lead one to believe." Nic Lewis explained that he wasn't suggesting that ocean heat transport is a forcing, but that it is equivalent, in terms of the rate of energy input, to an excessive inter-hemispheric forcing differential of 4.8 W/m².
No further discussion on this topic followed.

2.5. Paleo Climatic studies

In his guest blog, James Annan wrote that "*When averaged over a sufficiently long period of time, the earth must be in radiative balance or else it would warm or cool massively. This enables us to use paleoclimatic evidence to estimate ECS.*" And although "*We cannot add up the forcings and use the temperature response to determine the ECS as non-linearity occurs due to multiple changes: ice sheets, vegetation cover, continental drift, mountain ridges, opening or closing ocean passages, etc. (e.g. Yoshimori et al 2011)".* Annan concludes that "*The global mean temperature has varied by several degrees over long time scales in tandem with substantial changes to radiative forcings. This can only be reconciled with a sensitivity around 2 - 4.5 °C (Rohling et al 2012)."*

John Fasullo agrees and highlights that "*Paleo record provides a vital perspective for evaluating the slowest climate feedbacks.*" But he also stresses that due to "*sensitivity to nonlinearities, major uncertainty in proxy records (Rohling 2012), data problems, and uncertainty in forcing […] it is unclear whether progress on these fronts presents an immediate opportunity for reducing uncertainty in ECS in the near future.*"

Nic Lewis indicated that AR5 *"concludes in the final sentence of section 10.8.2.4 that paleoclimate estimates support only a wide 10–90% range for ECS of 1.0–6°C.*" where he added that he would be inclined to treat it as "*a 17–83% likely range rather than a 10–90% range." Because "The uncertainties are too great to support the narrower ~2–4.5°C range."* as mentioned by Annans. With respect to the median he added that "*the overall PDF for ECS from paleoclimate studies should have a rather similar skew to that derived from instrumental period warming based studies, implying a median estimate far below the midpoint of the 1–6°C (or whatever) range*." and therefore "*the paleo estimate should not greatly affect the overall median and likely range derived from warming over the instrumental period.*" Both John Fasullo and James Annan did not reply to these claims.

2.6. Combination studies

There was no separate discussion on combination studies. In his guest blog, John Fasullo wrote that "*These studies incorporate two or more of the other methods in an attempt to leverage the strengths of each, but in doing so are also susceptible to their weaknesses.*" Nic Lewis indicated that he considers studies in the Combination category together with those in the Instrumental category, since the combination estimates all include an instrumental estimate and the instrumental estimate tends to dominate the result.

3. What weight should be assigned to the different lines of evidence?

In AR5 it is indicated that the peer-reviewed literature provides no consensus on a formal statistical method to combine different lines of evidence. Therefore, in AR5 the range of ECS and TCR is expert-assessed, supported by, as indicated above, several different and partly independent lines of evidence, each based on multiple studies, models and data sets. Obviously, this expert judgement in AR5 has been performed deliberately, but it is not a straightforward procedure. The discussion on how to weigh the different lines of evidence is very old, not only in the scientific literature but also in the blogosphere and in reports and is still going on.

According to James Annan "*Weighting different methods higher or lower should not really apply, so long as the methods are valid and correctly applied.[…] The value of ECS probably lies in the intersection of the various high-probability ranges.*" However, based on his claims that "*Transient 20th century warming-based estimates (Aldrin et al 2012, Ring et al 2012, Otto et al 2013) are more trustworthy than other approaches*" (see Chapter 1) and he seems to have some confidence in Paleo Climatic studies (see paragraph 2.5), we conclude that he has more confidence in these two lines of evidence than the others.

In his guest blog John Fasullo more or less agrees, writing that: "*The distinctions between ECS estimation methods are artificial.*" and it is "*inappropriate to place high confidence in any single approach.*" However, based on the key arguments raised in the discussion on the most likely value of ECS (see chapter 5), it can be concluded that John Fasullo has most confidence in the model-based estimates (see Table 3). He disagreed with Annan that the likely value of ECS probably lies at the intersection of the high-probability ranges and he adds that *"There is also a need to more fully consider the sensitivity of any method to observations, particularly when using ocean heat content. As we show in a paper earlier this year, the choice of an ocean heat content dataset can change the conclusions of such an analysis from being a critique of the IPCC range to being consistent with it."*

As indicated in paragraph 2.1, Nic Lewis believes that instrumental period warming-based studies are superior to all other studies cited in AR5 that estimate ECS. Of this type of study, he only considers four of those to be free of serious shortcomings. Consequently, his confidence in these type of studies is high as compared to other types thus being low (see Table 3).

Study Type	**Lewis**	**Annan**	**Fasullo**
Instrumental – Observations	High	Middle?	Unknown
Instrumental – Perturbations	Low	n.a.	Unknown
Climatological Constraints	Very Low	Low	Unknown
Climate Models	Low	n.a.	Highest?
Paleo Climatic studies	Lowish	Middle?	Unknown
Combination	As Instrumental	n.a.	Unknown

Table 3. *Weights of the different lines of evidence according to the experts.*

4. Why would a lack of agreement between the lines of evidence not allow for a best estimate for ECS?

With respect to the best estimate it was reported in AR5 that: "*No best estimate for equilibrium climate sensitivity can now be given because of a lack of agreement on values across assessed lines of evidence.*" So IPCC did not choose between the different lines of evidence with respect to the best estimate, but it was not discussed in much detail why.

During this Climate Dialogue, there was not much discussion on this topic. Only Nic Lewis and John Fasullo gave a speculative answer:

According to Nic Lewis it was "*Maybe because of the major disagreement between ECS best estimates based on the energy budget method, of no more than about 2°C, and the average ECS value of GCMs of about 3°C.*" and since "*All the projections of future climate change in AR5 are based on GCMs and giving a best estimate below their average could have destroyed the credibility of the Working Group 2 and 3 reports.*"

John Fasullo did not agree and argued that in his view "*This would have required a firmer understanding of the uncertainties inherent to each approach than is presently available." and therefore Improved assessment of these uncertainties is a high priority that is achievable in the not-so-distant future.*"

5 What do you consider as a range and best estimate of ECS, if any?

Table 4 summarizes the answer to this question given by the three experts and the most important considerations underlying these estimates. The discussions surrounding these key arguments are summarized in this chapter.

ECS	**Lewis**	**Annan**	**Fasullo**
Likely Range	1.2 – 3.0 Link	2.0 – 3.0 Link	2.7 – 4.5 Link
Key arguments	Evidence for ECS being lower than in AR4 is ***not*** because of the hiatus since instrumental studies are mainly based on data up to 2001. Link and paragraph 5.3. All instrumental period warming based observational studies that have no evident serious flaws arrive at best estimates for ECS in the 1.5–2.0°C range. Blog, link and chapter 2. The excess of model ECS over 2°C comes primarily from positive cloud feedbacks and adjustments and reliable observational evidence for cloud feedback being positive rather than negative is lacking. Blog and paragraph 5.2. Studies using expert priors produce ECS estimates that primarily reflect the prior. Blog and paragraph 5.4.2.	Paleo studies can only be reconciled with a sensitivity around 2 - 4.5 °C. Blog. State of the art GCMs such as CMIP3/5, give a reasonable estimate of ECS in the range of 2-5 °C. Blog and paragraph 2.5. Global warming combined with Ocean Heat Uptake and knowledge of climate forcings points to an ECS at the lower end of the IPCC range, consistent with other analyses. Blog.	AR5 wrongly concluded to lower its lower bound from 2.0 to 1.5 °C because the instrumental approach would suggest that the hiatus argues for such a reduction. Blog and paragraph 5.3. There is no credible GCM with an ECS of less than 2.7 °C. Link Low sensitivity models have difficulty in simulating even the basic features of observed variability in both clouds and radiation. Link Key processes that drive ECS (clouds, radiation) are better represented in many of the high sensitivity GCMs. Link Poorest performing models in CMIP3 have been improved in CMIP5 and produce higher ECSs now. Link Forcing of aerosols are more effective than forcings of CO_2. Paragraph 5.2. There exist no valid studies supporting the strong negative cloud feedback needed to arrive at a sensitivity well below 2 °C. link and link and paragraph 5.1. CESM1-CAM5 ensemble shows no obvious bias in its reproduction of the surface temp record. Yet its ECS is 4.1 °C. Blog.
Best Estimate	1.7 Link	2.5	3.4 Link
Remarks	Same as above.	Number of 2.5 based on e-mail conversation held after the actual discussion.	Slow feedbacks could raise the likely range. Blog and paragraph 2.5.

Table 4. *Likely ranges and best estimates.*

5.1. Cloud feedbacks

In his blog, Lewis wrote that the excess of model ECS over 2°C comes primarily from positive cloud feedbacks and adjustments." and in a comment he adds that "*Reliable observational evidence for cloud feedback being positive rather than negative is lacking."* He emphasized that AR5 comes to the same conclusion in section 7.2.5.7.

In the first public comment by Andy Dessler he stated that: "*Doubling carbon dioxide by itself gives you about 1.2°C of warming. Add in the water vapor and lapse-rate feedbacks, which we have pretty high confidence in, and you get close to 2°C. Then add in the ice-albedo feedback and you get into the low 2s. To get back down to 1.5-ish, the cloud feedback needs to be large and negative. Is that possible? Yes, but essentially none of the evidence supports that. Instead, most evidence suggests a small positive cloud feedback, which would push the ECS to closer to 3°C."*

In a comment Steven Sherwood added that Lewis has ignored "*the multiple lines of evidence for positive cloud feedbacks articulated in Chapter 7 of AR5.*", but Lewis adds to his previous statement that "*The first approach Section 7.2.5.7 discusses is to seek observable aspects of present-day cloud behaviour that reveal cloud feedback or some component thereof. Its conclusion: 'In summary, there is no evidence of a robust link between any of the noted observables and the global feedback'; all it can point to is some apparent connections that are being studied further. Section 7.2.5.7 then discusses attempts to derive global climate sensitivity from interannual relationships between global mean observations of TOA radiation and surface temperature, but notes studies contradicting the basic assumption of these attempts. It goes on to note all sorts of problems in finding acceptable cloud-response derived observational constraints on climate sensitivity, ending by stating 'These sensitivities highlight the challenges facing any attempt to infer long-term cloud feedbacks from simple data analyses.'*"

The moderator then confronted Lewis with the overall conclusion in the summary of Chapter 7 of AR5, where it is written that the overall cloud feedback is "likely positive" and quantified as +0.6 (−0.2 to +2.0) $W/m^2/°C$.

Lewis replied that this range "*is based on the mean from GCMs and a widened version of the distribution of cloud feedback in GCMs. I do consider this conclusion to be wrong. In my view, it is not good scientific practice to assign a range for overall cloud feedback based on models when there is no solid observational evidence as to its value and models are known to be very far from perfect.*"

Sherwood, who was a co-author of AR5 Chapter 7, strongly disagreed when he stated that the quote from chapter 7.2.5.7 given by Lewis "*was taken out of context and does not imply there is no evidence for positive feedback. It applied only to one particular strategy that has been used.." and "The multiple lines of evidence were not only based on GCMs. […] We explicitly required observational evidence or back-up from detailed cloud simulations. The two feedback mechanisms we identified as having such support, are both positive (relating to the rise of the tropopause and the poleward shifting of cloud bands) and have support both from observations and explicit models of the relevant processes. And as Andy Dessler points out in a comment, to get ECS < 2C you need very strong* *negative cloud feedbacks to come from somewhere in order to cancel out the known positive ones. We have no evidence for such a thing after decades of searching.*"

Lewis repeated that Section 7.2.5.7 of AR5 "*Observational constraints on Global Cloud Feedback' deals with "the global level of overall cloud feedback and the observational evidence relating to it […] discussing various approaches and citing many studies. […] Of course almost all GCMs show overall positive cloud feedback – that is why they have high climate sensitivity! I never claimed that Ch.7 does not cite observationally-based evidence for some specific positive cloud feedbacks, just that it concludes – as it does – that robust observational evidence for positive OVERALL cloud feedback is lacking.*" With respect to Andy Dessler's comment (that you need strong negative cloud feedbacks to get an ECS<2) Lewis adds that "*Lindzen & Choi (2011) show such evidence, and Spencer & Braswell (2011) show the difficulty in estimating cloud feedbacks. Counterarguments were made in Dessler (2011) but have been challenged. Clearly, the separation of internal cloud fluctuations from feedbacks is difficult and represents an ongoing research problem.*"

Fasullo continued that from recent work - he mentions 15 authors - that "*examine the issue across observations, cloud resolving models, and GCM archives of various sorts, there is persuasive evidence that the feedback is not strongly negative but rather is likely to be positive, perhaps strongly so.*"

Lewis replied by referring to a study from Jiang et al (2012) that shows that the modelled mean CWCs [cloud water contents] over tropical oceans range from ~0.03 to ~15 times the observations in the Upper Troposphere (UT) and from 0.4 to 2 times the observations in the lower/mid-troposphere (L/MT). Modelled water vapour over tropical oceans was within 10% of the observations in the L/MT, but mean values ranged from ~0.01x to 2x the observations in the UT. Additionally, as Figure 3 of Lewis' guest blog showed, GCMs have severe biases as to cloud extent. The discussion on clouds ended with a remark from Lewis: "*If ultimately it proves to be the case that cloud feedback is positive rather than negative, then so be it. But there is a long way to go before cloud feedbacks are fully understood and correctly represented in GCMs.*"

5.2. Efficacy

A related discussion was on the so-called 'efficacy', i.e. the hypothesis that the transient climate response (TCR and thus also ECS) to historical aerosols and ozone is substantially greater than the transient response to CO_2. According to Shindell et al (2014) this is primarily caused by more of the short-lived aerosol and ozone forcing being limited to the places of emission, which are predominantly in the Northern Hemisphere continental regions. Since land temperatures respond stronger to a change in forcing than ocean temperatures do, this triggers a stronger temperature response, relative to the magnitude of the forcing, than the more evenly distributed CO_2 does. Annan and Fasullo indicate that estimates of ECS based on 20th-century observations have assumed that the efficacy is unity, i.e. that the forcing of aerosols is as effective as the forcing of CO_2. Kummer and Dessler (2014) showed that increasing "*the efficacy to 1.33 increases the ECS to 3.0 ^{0}C (1.9 – 6.8) a value in excellent agreement with other estimates*" and thus bridges the gap between the instrumental-based approach and the model-approach.

Lewis disagrees as he rejects both studies: "*Shindell […] never claims that these inhomogeneous forcings [mainly aerosols, red] have a efficacy of greater than one. He never refers to efficacy at all in his paper*." and Kummer & Dessler confuse "*forcing efficacy with transient climate sensitivity*" and therefore "*their calculations make no physical sense. […] Troy Masters has an excellent blog explaining this problem here.*" Furthermore, "*Kummer & Dessler state that their forcing time series is referenced to the late 19th century and accordingly use a reference (base) period to measure changes in global surface temperature from of 1880-1900. That would be fine were it true, but it is not. Their forcing time series actually come from AR5 and are referenced to 1750. The mean total forcing during 1880-1900 was substantially negative relative to 1750 due to high volcanic activity. Referencing the forcing change to a base period of 1880-1900, as necessary to match their temperature change, reduces their non-efficacy-adjusted ECS estimate to 1.5°C. And their headline 3.0°C best ECS estimate, based on an aerosol and ozone 'efficacy' of 1.33 and their faulty adjustment method, become 1.7°C.*"

Andrew Dessler replied that Lewis' statement about the referencing period of the forcing is right and that it "*will be corrected in the galleys*" but he continues that "*Assuming that the climate in the late 19th century is warmer than that in the mid 18th century (probable since radiative forcing is +0.3 W/m2 in the late 19th century), then referencing both time series to 1750 will increase the calculated climate sensitivity (I can explain why if it's not clear). Thus, it does **not** affect our conclusion that incorporating efficacy has a significant effect on the inferred climate sensitivity.*"

5.3. Hiatus

In his guest blog John Fasullo raised the issue that "*IPCC AR5 lowered its lower bound estimate on the likely range for ECS from 2.0 to 1.5 °C because simple models (i.e. the instrumental approach) suggest that the hiatus argues for such a reduction due to negative feedbacks (Collins 2013). However, warming during the hiatus has been driven by the vertical redistribution of heat in the ocean (Meehl 2011; Levitus 2012), confirmed by persistence in the rate of thermal expansion since 1993 (Cazenave 2014).*"

James Annan disagreed as he argued that one should consider ECS "*on the merits of the available literature rather than considering the previous AR4 estimate and/or the GCM model range as some sort of prior or null hypothesis to only be changed if and when the observational data become overwhelming.*"

Nic Lewis disagreed even more: "*Evidence for ECS being lower than in AR4 has been piling up, but that is **not** because of the hiatus since instrumental studies are mainly based on data up to 2001.*" More in detail, in Aldrin (2012) "*the median ECS estimate using data up to 2000 was lower, not higher, than the one using data to 2007.*" With respect to Otto (2013), the "*median ECS estimate using 2000s data was the highest; the estimates using data from the 1970s, 1980s or 1990s were all lower.*" and Lewis (2013) "*used data ending in August 2001*."Furthermore, Lewis emphasized that in Box 9.2 of AR5 it is stated that: "*there are no apparent incorrect or missing global mean forcings in the CMIP5 models over the last 15 years that could explain the model–observations difference during the warming hiatus*" and box 12.2 of AR5 writes: "*This change [of the lower bound of ECS] reflects the evidence from new studies of observed temperature change, using the extended records in atmosphere and ocean. These studies suggest a best fit to the observed surface and ocean warming for ECS values in the lower part of the likely range*" and does not mention the slowdown in warming this century. According to Lewis "*Internal climate system variability made a significant contribution to the fast warming over 1970-1995 due to the warm phase of the AMO.*"

5.4. The use of priors

A prior distribution in Bayesian theory is intended to tell how likely different values, of ECS in this case, are without considering some given data and in the 'non-informative' case, without considering any data at all. When you introduce the data, the prior probability distribution is updated and gives rise to the posterior distribution or the Probability Density Function (PDF).

Uniform vs non-informative priors

In his blog, Nic Lewis indicated that the priors of many of the observational instrumental-period warming based ECS estimates cited in AR5 start from a 'uniform prior' in ECS, and that all of those shown in AR4 were stated to be on that basis. A uniform prior means that the starting position is that "*all climate sensitivities are, over a very wide range, equally likely*." According to Lewis, the problem is that this biases ECS estimates substantially upwards. Lewis therefore claims that a '*non-informative prior*' should always be used. He asserted that the observable variables have a much more linear relationship to the reciprocal of ECS, the climate feedback parameter (lambda), than to ECS itself, and that it follows that a uniform prior in lambda is fairly uninformative. On that basis, he wrote "*It follows mathematically that a prior of the form* $1/ECS^2$ *will be non-informative for estimating ECS.*"

An interesting but rather technical discussion evolved with the expert Salvador Pueyo who emphasized that many experts including him and Lewis agree that "*the uniform prior vastly overestimates climate sensitivity*", but he adds that "*The overestimation resulting from this prior is so obvious that, in practice, the uniform is assumed only between S=0 and some Smax, and a zero probability is assumed above Smax, with no explicit criterion to choose Smax (discussed in Annan & Hargreaves 2016). With this correction,* ***it is not so obvious that this method should overestimate sensitivity****, but it is obvious that it is inappropriate.*" Pueyo published a paper (Pueyo, 2012) - and a very informative comic version of it - that concludes that a non-informative prior of climate sensitivity should be proportional to 1/ECS, the so-called sensitivity parameter lambda, resulting in a log-uniform distribution that should be refined with "*a limited use of expert elicitation or other methods*."

non-informative vs reference priors

Pueyo claimed that Lewis does not use a non-informative prior but rather a 'reference prior'. In his opinion, "it is intended just as a convention, as something that everybody is supposed to use when they don't know what to use, so that everybody's results are comparable (and, since the reference prior has several good statistical properties, you avoid some types of "accident"). This is a practical option when the posterior distribution is strongly constrained by the data. However, this is not the case of climate sensitivity. In the case of sensitivity, small differences in the prior can have a visible impact on the posterior [i.e., the PDF of ECS, red]. Since the reference prior cannot be given the strict meaning of a prior probability distribution, what you obtain by updating it cannot either be given the meaning of a posterior probability distribution. In fact, it is meaningless."

Lewis fully disagreed and replied that the distinction between 'reference priors' and 'non-informative priors' makes no sense. "*The whole point about a non-informative prior is that it is constructed so that only weak constraints by the data are required in order for the resulting posterior PDF for the parameter(s) to be dominated by (correctly-reflected) information from the data rather than* *information from the prior. Indeed, Berger and Bernardo show that reference priors have a minimal influence on inference, in the sense of maximising the missing information about the parameters.*"

According to Pueyo the point is that using a reference prior as if it were non-informative "*can cause serious trouble unless the amount of data makes the result quite insensitive to the prior, which is rarely the case with climate sensitivity.*" and the method of Nic Lewis "*results into a **vast underestimation** of climate sensitivity*."

Claimed failure of non-informative prior

During this discussion James Annan remarked that Nic Lewis "*provides a good example of a catastrophic failure of his approach in […] this climate audit blog post. The topic is carbon dating, but the point is a general one: his "objective" algorithm returns a probability distribution that assigns essentially zero probability to the interval 1200-1300 AD. That is, it asserts with great confidence that the object being dated does not date from that interval even in the case that the object does in fact date from that interval, and despite the observation indicating high likelihood (in the Bayesian sense) over that interval."* and "*there is nothing in Nic's approach that provides for any testing of the method, i.e. to identify in which cases it might give useful results, and when it fails abysmally*"

Nic Lewis responded that this zero probability "*simply reflects that over the interval concerned the data is very uninformative about the parameter of interest, because the interval corresponds to a small fraction of the data error distribution. If some non-radiocarbon data […] between 1200 and 1300 AD is obtained, then the non-informative prior for inference from the combined data would cease to be low in that region, and the posterior PDF would become substantial in the calendar region consistent with the new data, resulting in a much tighter uncertainty range*." He adds that "*the non-informative Jeffreys' prior provided uncertainty ranges that in all cases gave exact probability matching*" and that "*most statisticians (and scientists) would regard the accuracy of probability matching as a very useful – and widely used – way of identifying when a statistical method gives useful results*."

Expert priors

Some climate sensitivity estimates use 'expert priors'. This is, according to Lewis, "*a particular type of informative prior – one might say it is an intentionally informative prior that is derived from subjective opinions rather than only from data.*" and in his blog: "*These are mainly representations of pre-AR5 'consensus' views of climate sensitivity, which largely reflect estimates of ECS derived from GCMs. Studies using expert priors typically produce ECS estimates that primarily reflect the prior, with the observational data having limited influence.*".

Unfortunately, there was no in depth discussion on expert priors except that Salvador Pueyo indicated that the expert prior is "*no less problematic"* than the reference prior. So, he agrees that expert priors are problematic, but comparable to the use of non-informative priors (or reference priors) as used by Lewis (see paragraph 5.4.1).

6. What is key to narrowing uncertainty ranges in ECS and TCR?

This question was not explicitly raised in our introduction, but during the discussion both Nic Lewis and John Fasullo mentioned several ways to narrow the uncertainty ranges in ECS and, to a lesser extent, TCR. Table 5 summarizes their main points.

	Key to narrow down uncertainty in ECS and TCR
Lewis	• Better constraining aerosol forcing is the key to narrowing uncertainty in all ECS and TCR estimates based on observed multidecadal warming during the instrumental period. Link • Better observations of clouds (to reduce the uncertainty in cloud feedback, see paragraph 5.1) and their interactions with aerosols. Link • Less of the available resources should be put into model development and more into observations because, for example, modern ocean "reanalysis" methods are no substitute for good observations. Link
Fasullo	The statements of Fasullo come down to understanding the individual processes by a smart combination of observations, models and theory: • Further progress in estimating both ECS and TCR can best be made by a combined consideration of the individual approaches and the adoption of a physically-based perspective rooted in narrowing uncertainty in the individual feedbacks that govern sensitivity across a broad range of timescales. Blog. • A combined effort that makes use of various approaches for constraining ECS with an emphasis on evaluating individual climate feedbacks with targeted observations (7). Blog • Improved estimates of ocean heat content such as ARGO and improved ocean reanalysis methods and other climate indices (e.g. sea level, terrestrial storage) are fundamental in providing improved metrics of climate variability and change, evaluating models, and narrowing remaining uncertainties. Blog. • For hypothesis testing, there is a need for well-understood, well-calibrated, global-scale observations of the energy and water cycles and reanalyses. Blog • Improved assessment of the uncertainties inherent to each approach is a high priority that is achievable in the not-so-distant future. link • Testing GCMs with paleoclimate archives, where uncertainties in proxy data and forcings are adequately small. • Explore strengths and weaknesses of observationally based studies in a framework that is tightly constrained using a model whose sensitivity is known and whose variability is thoroughly vetted. See paragraph 2.1. • Weighting or culling model archives based on various physically-based rationales is likely to play a key role in constraining GCM estimates of sensitivity in the near future. link • Potential improvements in so-called "20th Century" approaches include a more thorough consideration of the adequacy of any "prior" and the uncertainty in both forcings and their efficacy. link • There is a need to more fully consider the sensitivity of any method to observations, particularly when using ocean heat content. link

Table 5. *How to narrow down uncertainty ranges in ECS and TCR?*

7. Relevance of the debate to policy makers

An additional question was raised on the relevance of the scientific debate on Climate Sensitivity to climate policy and policy makers. Fasullo wrote "*In the US, the two are largely disconnected.*" . Lewis agreed as he wrote that "*the debate on climate sensitivity and TCR should still be very pertinent in the political context, even though it currently is not.*" and also Annan joined this opinion writing that there are no " *politicians in the UK, or for that matter Japan (where I lived and worked until recently) [that] is paying much attention to the arcane debate in the literature about climate sensitivity.*"

Lewis continued that "*Very few politicians have any real understanding of the science*" and suggested that if they did, they would understand that: "*lowish sensitivity/TCR estimates (in line with what AR5 forcing and heat uptake best estimates imply) point to global warming from now to 2081-2100 of little more than 1 K on a business-as-usual scenario*". Lewis argued that one of the reasons why they don't understand is that "*The highbrow media (BBC, Guardian newspaper) are committed believers in dangerous anthropogenic climate change and present only that viewpoint. There are also various pressure groups warning of dangerous climate change and pushing for strong actions to reduce emissions. These include renewable energy groups and other subsidy farmers with vested interests, environmental NGOs and radical politico-environmental campaigning/protesting groups.*"

Fasullo, however, held the opposite opinion and wrote that those politicians (in the US) who "*often insist that sensitivity is low (or zero)*" don't do this because they understand the science, but "*largely out of convenience rather than being based on specific tidbits of convincing evidence.*" Fasullo suggested this is quite a large group as he wrote that "*half of Congress essentially deems the issue unworthy of further study – an opinion clearly voiced by members with vested interests that would change little if the scientific consensus were to become firmer.*" and thus "*Given the reluctance to embrace even the broadly accepted facts on the issue (e.g. that humans have contributed substantially to Earth's warming since the mid-20th C), any strong connection between the scientific debate and policy seems thus far to be elusive. It is both my hope and expectation that this will change.*"

Annan held the opinion that the scientific debate is not important to policy makers because "the remaining debate concerning the precision of our estimates [of ECS] is not, or at least rationally should not be, so directly pertinent for policy decisions. We already know with great confidence that human activity is significantly changing the global climate, and will continue to do so as long as emissions continue to be substantial."

References

Annan, J.D. and J.C. Hargreaves, 2006. Using multiple observationally-based constraints to estimate climate sensitivity. Geophys. Res. Lett., 33: L06704.

Annan, J. D., & Hargreaves, J. C. (2011). On the generation and interpretation of probabilistic estimates of climate sensitivity. Climatic Change, 104(3-4), 423–436. doi:10.1007/s10584-009-9715-y

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.

Bender, M. A., Knutson, T. R., Tuleya, R. E., Sirutis, J. J., Vecchi, G. A., Garner, S. T., and Held, I. M., 2010. Modeled Impact of Anthropogenic Warming on the Frequency of Intense Atlantic Hurricanes. Science, 327(5964), 454-458.

Cazenave, A., Dieng, H. B., Meyssignac, B., von Schuckmann, K., Decharme, B., & Berthier, E. (2014). The rate of sea-level rise. Nature Climate Change.

Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W.J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver and M. Wehner, 2013: Long-term Climate Change: Projections, Com- mitments and Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Dessler, A. E., 2010. A determination of the cloud feedback from climate variations over the past decade. Science, 330(6010), 1523-1527.

Dessler, A.E., 2011. Cloud variations and the Earth's energy budget. Geophysical Research Letters. http://dx.doi.org/10.1029/2011GL049236

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

Forster, P.M.D., and J.M. Gregory, 2006. The climate sensitivity and its components diagnosed from Earth radiation budget data. J.Clim.,19: 39–52.

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. Atmos., 118, 1139–1150, doi:10.1002/jgrd.50174.

Frame D.J., B.B..B. Booth, J.A. Kettleborough, D.A. Stainforth, J.M. Gregory, M. Collins andM.R. Allen, 2005.Constraining climate forecasts: The role of prior assumptions. Geophys. Res. Lett., 32, L09702

Gettelman, A., H. and Coauthors, 2010: Global simulations of ice nucleation and ice supersaturation with an improved cloud scheme in the community atmosphere model. J. Geophys. Res., 115, D18216, doi:10.1029/2009JD013797.

Hall, A., & Qu, X., 2006. Using the current seasonal cycle to constrain snow albedo feedback in future climate change. Geophysical Research Letters, 33(3).

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.

Hegerl, G.C., T.J. Crowley, W.T. Hyde, and D.J. Frame, 2006. Climate sensitivity constrained by temperature reconstructions over the past seven centuries. Nature;440: 1029–1032.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Jiang, J. H., et al., 2012. Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations, J. Geophys. Res., 117, D14105, doi:10.1029/2011JD017237.

Klocke, D., Pincus, R., & Quaas, J. (2011). On constraining estimates of climate sensitivity with present-day observations through model weighting. Journal of Climate, 110525130752016. doi:10.1175/2011JCLI4193.1

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.

Kummer JR and AE Dessler, 2014. The impact of forcing efficacy on the equilibrium climate sensitivity. Geophysical Research Letters.

Levitus, S., et al. (2012), World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010, Geophys. Res. Lett., 39, L10603, doi:10.1029/ 2012GL051106.

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

Libardoni, A.G. and C.E.Forest, 2013. Correction to 'Sensitivity of distributions of climate system properties to the surface temperature dataset'. Geophys. Res. Lett.; 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.

Lindzen, R.S. and Y.S. Choi, 2011. On the observational determination of climate sensitivity and its implications. Asia-Pacific J. Atmos. Sci.;47: 377–390.

Meehl, G. A., Arblaster, J. M., Fasullo, J. T., Hu, A., & Trenberth, K. E. (2011). Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nature Climate Change*, *1*(7), 360-364.

Murphy, D.M., S. Solomon, R.W. Portmann, K.H. Rosenlof, P.M. Forster, and T. Wong, 2009. An observationally based energy balance for the Earth since 1950. J. Geophys. Res. Atmos.,114: D17107.

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., F. E. L. Otto, O. Boucher, J. Church, G. Hegerl, P. M. Forster, N. P. Gillett, J. Gregory, G. C. Johnson, R. Knutti, N. Lewis, U. Lohmann, J. Marotzke, G. Myhre, D. Shindell, B Stevens and M. R. Allen, 2013: Energy budget constraints on climate response. Nature Geoscience, 6, 415–416.

Pueyo, S. 2012. Solution to the paradox of climate sensitivity. Climatic Change 113: 163-179

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.

Rohling, E. J., Sluijs, A., Dijkstra, H. A., Köhler, P., van de Wal, R. S. W., et al. (2012). Making sense of palaeoclimate sensitivity. Nature, 491(7426), 683–691. doi:10.1038/nature11574

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.

Shindell, D. T. et al, 2013. Radiative forcing in the ACCMIP historical and future climate simulations. Atmos. Chem. Phys. 13, 2939–2974

Shindell D.T., 2014. Inhomogeneous forcing and transient climate sensitivity, Nature Climate Change, vol. 4, pp. 274-277.

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

Spencer, R.W., and Braswell, W. D. (2010) On the diagnosis of radiative feedback in the face of unknown radiative forcing, Journal of Geophysical Research, 114, D16109.

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.

Trenberth, K. E., & Fasullo, J. T. (2013). An apparent hiatus in global warming?. Earth's Future.

Yoshimori, M., Hargreaves, J. C., Annan, J. D., Yokohata, T., & Abe-Ouchi, A. (2011). Dependency of Feedbacks on Forcing and Climate State in Physics Parameter Ensembles. Journal of Climate, 24(24), 6440–6455. doi:10.1175/2011JCLI3954.1

[i] In a comment on May 16, Steven Sherwood was very critical on Lewis(2013). Lewis responded in detail in a quite technical manner and rejected all criticisms raised by Sherwood. Sherwood replied one month later, fully disagreeing with the response of Lewis. He ends his comment saying that: "Nic challenges me to defend the studies he wishes to dismiss. All I can say is that one could dismiss every single study, including his, by cherry-picking some random imperfection in the methods or models used. These studies all passed peer review, which does not prove they are valid, but means that if Nic wishes to dismiss them the burden is on him to identify the key flaw and explain why it would have led to an overestimate of ECS rather than an underestimate." A few days later Lewis responded by mainly referring to his earlier comments and concludes by saying that he gives "specific reasons for dismissing each model. If Steven thinks any of them are wrong, I invite him to say so and to explain why. Steven has failed to do so. Passing peer review means little. I have identified the key flaw in each study and shown why it leads to an overestimate of ECS – it doesn't look to me as if Steven has even read my critiques of the studies."