AGU PUBLICATIONS

Earth's Future

RESEARCH ARTICLE

10.1002/2016EF000376

Key Points:
- Current understanding of bounds on Earth's Equilibrium Climate Sensitivity (ECS)
- The refutation of physical storylines (or hypotheses) can more effectively bound the ECS range
- Reasons to be optimistic as to our ability to reduce uncertainty in our assessment of ECS in the future

Supporting Information:
- Supporting Information S1

Corresponding author:
B. Stevens,
bjorn.stevens@mpimet.mpg.de

Citation:
Stevens, B., S. C. Sherwood, S. Bony, and M. J. Webb (2016), Prospects for narrowing bounds on Earth's equilibrium climate sensitivity, *Earth's Future*, *4*, doi:10.1002/2016EF000376.

Received 18 APR 2016
Accepted 12 OCT 2016
Accepted article online 31 OCT 2016

© 2016 The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Prospects for narrowing bounds on Earth's equilibrium climate sensitivity

Bjorn Stevens[1], Steven C. Sherwood[2], Sandrine Bony[3], and Mark J. Webb[4]

[1]Max Planck Institute for Meteorology, Hamburg, Germany, [2]CCRC and Centre of Excellence for Climate System Science, University of New South Wales, Sydney, NSW, Australia, [3]LMD/IPSL, CNRS, Université Marie et Pierre Curie, Paris, France, [4]Met Office/Hadley Centre, Exeter, UK

Abstract The concept of Earth's Equilibrium Climate Sensitivity (ECS) is reviewed. A particular problem in quantifying plausible bounds for ECS has been how to account for all of the diverse lines of relevant scientific evidence. It is argued that developing and refuting physical storylines (hypotheses) for values outside any proposed range has the potential to better constrain these bounds and to help articulate the science needed to narrow the range further. A careful reassessment of all important lines of evidence supporting these storylines, their limitations, and the assumptions required to combine them is therefore required urgently.

1. Introduction

Nearly 40 years have passed since the U.S. National Academies issued the "Charney Report." This landmark assessment popularized the concept of the "equilibrium climate sensitivity" (ECS), the increase of Earth's globally and annually averaged near surface temperature that would follow a sustained doubling of atmospheric carbon dioxide relative to its preindustrial value. Through the application of physical reasoning applied to the analysis of output from a handful of relatively simple models of the climate system, Julge G. Charney and his co-authors estimated a range of 1.5 –4.5 K for the ECS [*Charney et al.*, 1979]. Since then, and as summarized by *Stocker et al.* [2013], physical understanding of the climate system has improved, more sophisticated models have been developed, global temperatures have risen by more than 0.5 K, and evidence of many past climate changes and the forces that drove them have been uncovered.

Reconciling the implications of these scientific advances for the ECS constitutes an open problem. But at a recent gathering of experts, participants were optimistic that new ways of looking at the problem could appreciably narrow the ECS range from that in 1979 [*Stevens et al.*, 2015]. Because the ECS remains such a fundamental measure of the sensitivity of Earth's climate to atmospheric CO_2, appreciably narrowing the ECS range would be a triumph for climate science. To achieve more rapid progress, however, requires a shift in focus: a move away from arguments in support of particular ECS values, and a move toward efforts to rule values out through a careful process of hypothesis testing.

2. The Equilibrium Climate Sensitivity—Some Fundamentals

ECS is an idealized but central measure of climate change, which gives specificity to the more general idea of Earth's radiative response to warming. This specificity makes ECS something that is easy to grasp, if not to realize. For instance, the high heat capacity and vast carbon stores of the deep ocean mean that a new climate equilibrium would only be fully attained a few millennia after an applied forcing [*Held et al.*, 2010; *Winton et al.*, 2010; *Li et al.*, 2012]; and uncertainties in the carbon cycle make it difficult to know what level of emissions is compatible with a doubling of the atmospheric CO_2 concentration in the first place. Concepts such as the "transient climate response" or the "transient climate response to cumulative carbon emissions" have been introduced to account for these effects and may be a better index of the warming that will occur within a century or two [*Allen and Frame*, 2007; *Knutti and Hegerl*, 2008; *Collins et al.*, 2013; *MacDougall*, 2016]. But the ECS is strongly related and conceptually simpler, so it endures as the central measure of Earth's susceptibility to forcing [*Flato et al.*, 2013].

The socioeconomic value of better understanding the ECS is well documented. If the ECS were well below 1.5 K, climate change would be a less serious problem. The stakes are much higher for the upper bound. If the ECS were above 4.5 K, immediate and severe reductions of greenhouse gas emissions would be imperative to avoid dangerous climate changes within a few human generations. From a mitigation point of view, the difference between an ECS of 1.5 K and 4.5 K corresponds to about a factor of two in the allowable CO_2 emissions for a given temperature target [*Stocker et al.*, 2013] and it explains why the value of learning more about the ECS has been appraised so highly [*Cooke et al.*, 2013; *Neubersch et al.*, 2014]. The ECS also gains importance because it conditions many other impacts of greenhouse gases, such as regional temperature and rainfall [*Bony et al.*, 2013; *Tebaldi and Arblaster*, 2014], and even extremes [*Seneviratne et al.*, 2016], knowledge of which is required for developing effective adaptation strategies. Being an important and simple measure of climate change, the ECS is something that climate science should and must be able to better understand and quantify more precisely.

Physically, the ECS depends on how the outward directed radiation at the top of the atmosphere responds to warming. By increasing the infrared opacity of the atmosphere, greenhouse gases reduce the rate at which Earth loses energy to space. This results in an imbalance of power at the top of the atmosphere ("radiative forcing"), which causes Earth's surface to warm until a new balance is achieved. The more responsive the net (infrared plus reflected solar) outgoing radiation to temperature, the less the temperature needs to rise, and the lower the ECS.

Feedbacks within the system determine how responsive the net outgoing radiation is. Fundamentally, the stability of Earth's climate is due to the principle that warmer bodies radiate more energy, thereby balancing the forcing. The net feedback depends on how this response is modified by other processes [*Dufresne and Saint-Lu*, 2016]. As was already appreciated in the first attempt to calculate the sensitivity of surface temperature to the amount of atmospheric CO_2, surface ice disappears as the temperature rises, so less sunlight reflects to space, yielding a positive contribution to the net feedback and higher ECS [*Arrhenius*, 1896]. Water vapor, clouds, and the thermal structure of the atmosphere change with warming in ways that introduce additional feedbacks. In an analysis of climate models that contributed output to the fifth phase of the coupled model intercomparison study, CMIP5, *Vial et al.* [2013] show that in nearly all models, changes in water vapor, clouds and surface ice, each contribute a positive feedback, with water vapor feedback correlated with, and partially offset by, a negative feedback from changes in the atmospheric thermal structure. Feedbacks from clouds are by far the most inconsistent across models [*Boucher et al.*, 2013].

Evidence of planetary scale glaciations during the Cryogenian period—635–720 million years ago (Mya)—makes the case for climate states that may be particularly sensitive to perturbations [*Pierrehumbert et al.*, 2011]. Hence, a generalization of the concept of climate sensitivity to different eras may need to account for differences that arise from the different base state of the climate system, increasingly so for large perturbations. Even for small perturbations, there is mounting evidence that the outward radiation may be sensitive to the geographic pattern of surface temperature changes. *Senior and Mitchell* [2000] argued that if warming is greater over land, or at high latitudes, different feedbacks may occur than for the case where the same amount of warming is instead concentrated over tropical oceans. These effects appear to be present in a range of models [*Armour et al.*, 2013; *Andrews et al.*, 2015]. Physically they can be understood because clouds—and their impact on radiation—are sensitive to changes in the atmospheric circulation, which responds to geographic differences in warming [*Kang et al.*, 2013], or simply because an evolving pattern of surface warming weights local responses differently at different times [*Armour et al.*, 2013]. Hence different patterns of warming, occurring on different timescales, may be associated with stronger or weaker radiative responses. This introduces an additional state dependence, one that is not encapsulated by the global mean temperature. We call this a "pattern effect." Pattern effects are thought to be important for interpreting changes over the instrumental period [*Gregory and Andrews*, 2016], and may contribute to the state dependence of generalized measures of Earth's climate sensitivity as inferred from the geological record.

3. From Consensus Building to Hypothesis Testing

The authors of the Charney Report obtained their 1.5–4.5 K range of ECS on the basis of physical arguments, informed by evidence from modeling. In more recent assessments, this uncertainty range has increasingly

become identified with a confidence interval, determined from the consensus of an ever larger mix of modeling and observational studies [e.g., *Knutti and Hegerl*, 2008; *Collins et al.*, 2013]. In emphasizing consensus, the physical underpinnings of each study are easily lost and there is little to reward the effort of judging the merits of individual studies. Though the process seems objective, its results (especially near the upper and lower extremes where the implications are most profound) are sensitive to prior judgments and assumptions about how independent the various estimates were that went into the mix. By obscuring the way in which the assumptions and various strands of evidence are interwoven, this process neither yields wholly convincing bounds nor points the way toward further progress.

Insufficient emphasis on physical arguments also stunts the full exploitation of data, whose emergence over the last 40 years has been an underappreciated revolution. Estimates of the ECS in the Charney Report admirably combined physical arguments with model output, but observational data that could be applied to the question were extremely limited and hardly mentioned. Today, data describe the state of the past climate system over a range of time periods: techniques are improving for reconstructing the global temperature record from measurements over an instrumental period dating back more than 150 years [*Morice et al.*, 2012]; temperature proxies [*Waelbroeck et al.*, 2009; *Tripati et al.*, 2014] and ice-core based reconstructions of atmospheric composition are available throughout the Late Quaternary (last 800,000 years) [*Lüthi et al.*, 2008]; and data from a variety of rough proxies document large climate changes in earlier epochs [*Pagani et al.*, 2014]. Observations of clouds, water vapor, energy flows to space and ocean enthalpy changes now provide a wealth of data to test feedback processes and temperature responses [*Loeb et al.*, 2009; *Dessler*, 2013; *Forster*, 2016] and are increasingly being complemented by mechanistic studies using high-resolution (cloud-resolving) models more grounded in the underlying laws of fluid dynamics [*Kuang and Hartmann*, 2007; *Rieck et al.*, 2012; *Bretherton*, 2015].

In attempting to arrive at a tighter bound on the ECS, the challenge will be to combine these various sources of information in a physically consistent manner, whereby the evidence required to rule out the argument for a value of the ECS that exceeds a particular bound becomes transparent. To meet this challenge, we propose a more systematic application of what has long been a fundament of the scientific approach, namely the development and eventual *refutation* of the physically based — hence testable — "storylines" that would enable a given bound for ECS to be violated. We call this "reasoning by refutation."

The storylines at the heart of our proposal would take the form of compound propositions involving multiple lines of evidence, each of which links models and theory to data. To be convincing, an argument for a very low or very high (e.g., below 1.5 K or above 4.5 K) ECS value would have to be reconciled with *each* of these lines of evidence. We appreciate that a comprehensive assessment of ECS should consider other storylines than those we discuss, or a range in the ECS different from the conventional one (1.5–4.5 K). Nonetheless, to illustrate the power of reasoning by refutation, in the following three sections we draw on an extensive literature (more systematically assembled in the Appendix S1, Supporting Information) to adumbrate what we believe to be compelling physical storylines for values either above or below the conventional range in the ECS, and outline the evidence with which it must be reconciled. A methodology for combining these different lines of evidence is then proposed in the form of an example.

4. Reasoning by Refutation and a Physical Storyline for a Very Low ECS

Our physical storyline for a very low value of the ECS (<1.5 K) is comprised of four conditions: (i) cooling of climate associated with anthropogenic aerosols would have to have been modest (less negative than -1 W m^{-2}), and/or a historical "pattern effect" would have to be less important than indicated by models; (ii) tropical sea-surface temperatures during the last glacial maximum (LGM, 21 kya) would need to have been at the warm end of the expected range, and/or a "pattern effect" for the LGM would have to be *more* important than current models predict; (iii) climate feedbacks would have to have been much larger in past hot climates than they are at present, or else climate forcing at those times has been significantly underestimated; and (iv) cloud feedbacks from warming would have to be negative. For a very low ECS to prevail, all four of these conditions must be satisfied; exploring them individually shows why such a low ECS is unlikely, and what kinds of evidence could refute it (Figure 1).

Condition (i) arises from the interpretation of the instrumental record using simple models whose value of the ECS is constrained by the data [e.g., *Aldrin et al.*, 2012; *Otto et al.*, 2013; *Lewis and Curry*, 2014]. These

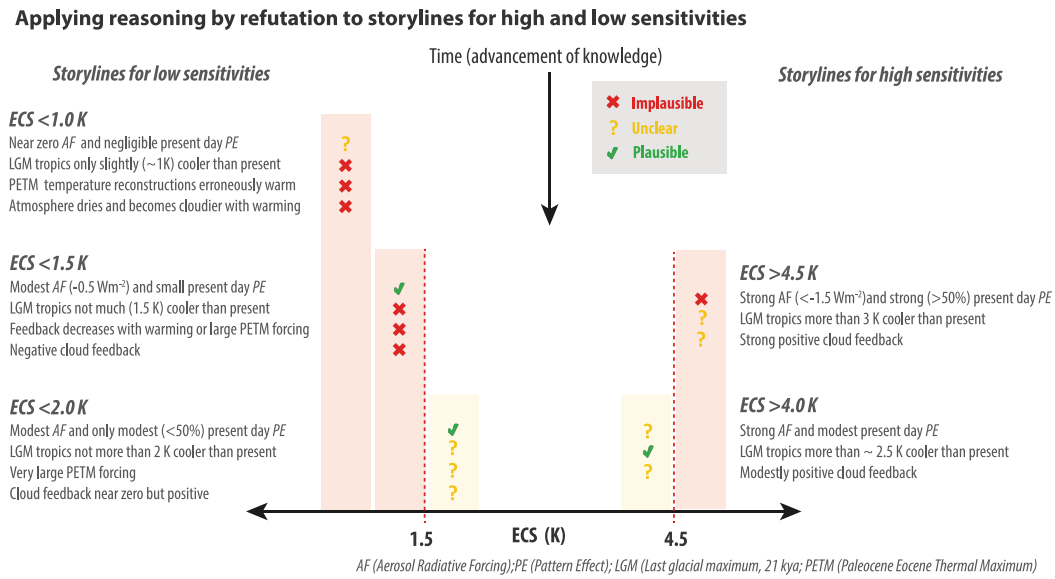

Figure 1. The application of refutational reasoning to storylines for successively more constraining bounds on ECS. For a particular bound on ECS to hold it must be consistent with all the lines of evidence, here formulated in terms of physical statements related to some interpretation of data or some specific process, for example, cloud response to warming. The hypotheses most fruitful to test progressively narrower bounds are those that are marked by a yellow question mark. This illustration is meant to encourage, rather than substitute, a more thorough assessment.

studies suggest that a value of the ECS less than 1.5 K would be consistent with changes in the energy budget and surface temperatures over the course of the instrumental record provided that the aerosol forcing and/or the historical "pattern effect"—which is neglected by the simple models—are modest (less than -1 W m^{-2} and 50%, respectively). A weak to modest aerosol forcing is consistent with a broad range of evidence [*Stevens*, 2015] and although studies using comprehensive climate models suggest that the historical "pattern effect" is not negligible [*Andrews*, 2014; *Huber et al.*, 2014], its magnitude is uncertain, making it difficult to refute condition (i).

Condition (ii) expresses the expectation that the temperature difference between the present and the LGM should increase with ECS. This expectation is not straightforward, because many details regarding the shape and height of great ice sheets at that time are poorly known, and uncertainties related to the response of the climate system to changing patterns of insolation confound predictions, particularly at middle and high latitudes [e.g., *Braconnot et al.*, 2012; *Jochum et al.*, 2012]. Aerosol forcing is also an uncertain factor. Nonetheless, model studies suggest that the decrease of tropical temperatures during the LGM offers some constraint on ECS [*Schmittner et al.*, 2011; *Hargreaves et al.*, 2012; *Rohling et al.*, 2012; *Schmidt et al.*, 2014]. This decrease, as evaluated on the basis of multiple proxies [*Waelbroeck et al.*, 2009; *Tripati et al.*, 2014; *Annan and Hargreaves*, 2015], appears too large to be reconciled with an ECS less than 1.5 K.

Going back further in the geological record one finds compelling evidence of very warm "hothouse" climates. Background temperatures in the early Eocene, during the Paleocene-Eocene Thermal Maximum (about 56 Mya), and during more recent episodes appear to require climate sensitivities of 4–6 K, although the uncertainty in radiative forcing is considerable. To reconcile this with a modern ECS below 1.5 K requires either much higher greenhouse gas forcing than anticipated [*Berner and Kothavala*, 2001; *Pagani et al.*, 2006; *Alexander et al.*, 2015] or a weaker radiative response to warming at high global temperatures [*Covey et al.*, 1996; *Royer et al.*, 2007; *Bijl et al.*, 2010; *Rohling et al.*, 2012]. Models actually predict the latter to some extent [e.g., *Meraner et al.*, 2013], but still have trouble simulating enough Eocene warmth unless the ECS is larger than about 3 K—hence condition (iii) also appears difficult to meet.

Finally we consider the implications of our understanding of the strength of feedbacks from specific physical processes on the value of the ECS. A very low ECS would either require a substantially negative cloud feedback or much weaker water vapor feedback than is freely simulated by any model [*Sherwood et al.*, 2010].

In an attempt to force a model to achieve a weaker water vapor feedback, and a negative cloud feedback due to changes in high clouds [as suggested by *Lindzen et al.*, 2001], low-cloud cover also decreased, largely offsetting the intended effect [*Mauritsen and Stevens*, 2015]. A negative global cloud feedback is also very difficult to reconcile with understanding of cloud feedback mechanisms, which are overwhelmingly associated with positive feedbacks [see Appendix S1, also, *Hartmann and Larson*, 2002; *Dessler*, 2010; *Zelinka and Hartmann*, 2011; *Webb and Lock*, 2012; *Boucher et al.*, 2013; *Bretherton*, 2015; *Zhou et al.*, 2015; *Bony et al.*, 2016]. An exception is a predicted negative cloud feedback at high latitudes where warming causes clouds to become more reflective [e.g., *Betts and Harshvardhan*, 1987; *Zelinka et al.*, 2013; *Kay et al.*, 2014]. *Gordon and Klein* [2014] provide evidence that this feedback is overestimated by models and even if not, it is outweighed by clouds becoming less reflective in middle and low latitudes. Observational tests of the feedback-relevant processes in different models further identify the most realistic models as those having relatively strong positive feedback and high ECS [*Clement et al.*, 2009; *Sherwood et al.*, 2014; *Su et al.*, 2014], nowhere near the net negative cloud feedback required for ECS <1.5 K.

5. Reasoning by Refutation and a Physical Storyline for a Very High ECS

Our physical storyline for a very high ECS (>4.5 K) is comprised of three conditions: (i) the aerosol cooling influence in recent decades would have to have been strong enough to offset most of the effect of rising greenhouse gases, or else the net radiative response to warming would have to have been at least twice as large for the last century as it will be in the future; (ii) tropical sea-surface temperatures at the time of the LGM would have to have been much (3 K) cooler than at present; and (iii) cloud feedbacks from warming would have to be strong and positive.

Condition (i) arises from the relatively modest (near 1 K) warming seen in the instrumental record so far. The strong aerosol effect needed to reconcile a very high ECS with this warming is not consistent with values inferred from satellite data, and difficult to reconcile with the spatiotemporal structure of the instrumental record. Also aerosol effects on clouds saturate at high concentrations, which argues against a large effect [*Boucher et al.*, 2013; *Stevens*, 2015]. There may, however, be other cooling mechanisms which blunted 20th-century warming; for instance, the above-mentioned "pattern effect" could have significantly damped 20th-century warming compared to that which will occur over the longer term with the same forcing [*Gregory and Andrews*, 2016]. If so, this would reduce the magnitude of aerosol effect needed for this storyline to be true.

Condition (ii) likewise reflects similar reasoning as for the low-ECS case. The LGM evidence is difficult to reconcile with a very high ECS, unless one allows for either a strong LGM "pattern effect" [cf. *Hopcroft and Valdes*, 2015] or feedbacks that grow more positive in warmer climates. There is in fact some support for each from climate models, which typically show both a pattern effect on cloud feedbacks [*Senior and Mitchell*, 2000; *Armour et al.*, 2013; *Geoffroy et al.*, 2013; *Andrews et al.*, 2015; *Knutti and Rugenstein*, 2015] and an increase in the radiative forcing of successive CO_2 doublings [*Gregory et al.*, 2015] and/or strengthening of cloud or water vapor feedbacks [*Meraner et al.*, 2013], making this condition more plausible. Because evidence from "hothouse" climates seems unlikely to discount a very high ECS, the deep past does not play a role in this storyline.

Condition (iii) embodies understanding of physical processes, which imply that a very high ECS would require a strong positive feedback (>0.5 W m^{-2}) from clouds. Attempts to assess the realism of such feedbacks are necessarily indirect. Strong cloud feedbacks do occur in some climate models, especially those models where cloud-relevant processes better match observations [*Clement et al.*, 2009; *Sherwood et al.*, 2014; *Su et al.*, 2014], consistent with understanding of important cloud controlling factors [*Stevens and Brenguier*, 2009; *Rieck et al.*, 2012; *Bretherton et al.*, 2013; *Brient and Bony*, 2013]. The strength of these feedbacks depends on the balance of sometimes opposing process [*Zhang et al.*, 2013; *Bretherton*, 2015; *Brient et al.*, 2015], and are sensitive to how the models are formulated [*Gettelman et al.*, 2012; *Qu et al.*, 2013; *Zhao*, 2014; *Webb et al.*, 2015]. There is also some evidence that shallow-cloud cover predicted by climate models may be oversensitive to local atmospheric properties [*Nuijens et al.*, 2015], which could imply they exaggerate the amplitude of cloud feedbacks. These issues are discussed further in Appendix S1; until they are resolved, the evidence that understanding of cloud feedbacks are in conflict with the condition for the very high (ECS >4.5 K) storyline will remain weak, at best.

6. Combining the Different Lines of Evidence

In principle, refuting one of the conditions that must be satisfied for the ECS to exceed a given bound is sufficient to support the maintenance of that bound. A single red cross, the equivalent of a counter example in a mathematical proof, in Figure 1 is sufficient to falsify the hypothesis with which it is associated. In practice refutation is never this simple. To the extent a more nuanced, and quantitative, assessment of the confidence in a particular bound is desired, the rules of Bayesian inference provide a useful framework for combining different lines of evidence [*Annan and Hargreaves*, 2006; *Hegerl et al.*, 2006; *Annan*, 2015].

For sake of argument, let χ denote the value of ECS, and S the evidence presented as a storyline. Bayesian inference calculates a *posterior* distribution as the probability of the hypothesis (in this case a value of χ) given the evidence. It is proportional to the likelihood of the evidence for a given value of χ times the prior distribution:

$$P(\chi|S) = \frac{P(S|\chi)P(\chi)}{P(S)}. \tag{1}$$

The likelihood of the storyline,

$$P(S) = \int_{-\infty}^{\infty} P(S|\chi)P(\chi)\,\mathrm{d}\chi > 0, \tag{2}$$

arises as a normalizing factor and ensures that the posterior integrates to unity. The probability that the ECS is within some bounds follows by integrating equation 1 over these bounds. The storyline is organized along its different conditions, which we associate with different lines of evidence. The conditions constituting a particular storyline were chosen on the basis of their differing evidentiary basis, and as such are treated as being independent.

To illustrate the method we consider the storylines for a very low, <1.5 K, and very high, >4.5 K, ECS. Let us denote evidence related to the jth condition for the storyline S by e_j. The assumed independence of N different lines of evidence yields:

$$P(S|\chi) = \prod_{j=1}^{N} P\left(e_j|\chi\right). \tag{3}$$

Reasoning by refutation is akin to saying that the probability of the evidence is low if the hypothesis is false but high if it is true. We model this by specifying each $P(e_j|\chi)$ as a function that transitions from a likelihood α_j to $1-\alpha_j$ following an error function centered at some critical ECS, χ_j, near the hypothesized bound (see Figure 2a and Appendix S1). The α_j thus express the likelihood that the evidence, which appears to refute a particular hypothesis, is misleading. A smaller value of α_j expressing more confidence that the evidence refutes the hypothesis. The form of the likelihood function we have adopted requires $\alpha_j \leq 0.5$ and thereby ensures that only storylines are considered whose refutation is more likely than not.

The α_j and the χ_j are determined by the nature of the evidence. For the very low ECS storyline, we have some confidence that conditions (ii)–(iv) contraindicate the hypothesis (cf. discussion above, Figures 1 and S1). Hence for these conditions we adopt a likelihood α_j of 0.25, 0.35, 0.20, respectively, with $\chi_j = 1.5$ K for conditions (ii) and (iii), and $\chi_j = 2$ K for condition (iv). Likewise we have some confidence that conditions (i)–(iii) contraindicate the hypothesis of a very high ECS and for these we adopt a likelihood α_j of 0.25, 0.35, 0.40, respectively, with $\chi_j = 4.5$ K for conditions (ii) and (iii), and $\chi_j = 4$ K for condition (i). All six conditions (Figure 2a) have been constructed from a reasonable consideration of the evidence (as discussed in the Appendix S1), but also with our didactic purpose in mind.

As the prior distribution should be based on knowledge independent of the evidence subsequently applied to the range of ECS, we chose a prior that arises from our most robust and basic understanding of atmospheric physics [*Stevens and Bony*, 2013; *Zelinka et al.*, 2016, and Appendix S1]. We thus choose as the prior an ECS composed by forming the ratio of an effective radiative forcing and a feedback parameter, with a 20% (1-sigma) width Gaussian distribution in the effective forcing and a 50% width Gaussian in the feedback parameter λ. This choice of prior (Figure 2b) yields a 19% chance that ECS <1.5 K and a 11% chance that ECS >4.5 K (Figure 2b). Combining this prior with the different lines of evidence following equation 1 results in a narrower posterior distribution, one with only a 3% chance of ECS being <1.5 K, and a similarly

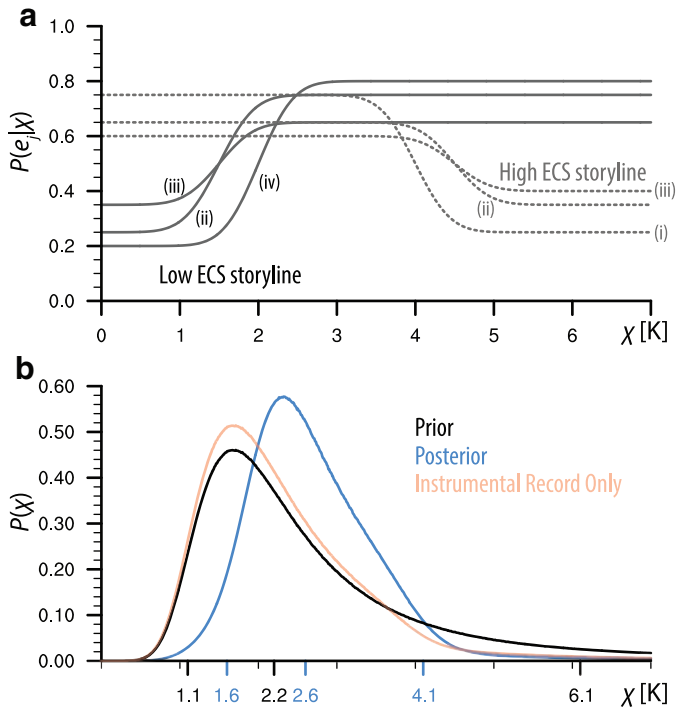

Figure 2. (a) Likelihoods of evidence refuting a particular storyline of being misleading given a value of the ECS (denoted by χ) and differentiated by the storyline conditions for a very low ECS (solid) and very high ECS (dashed); (b) the prior (black) and posterior (blue) distributions with 5, 50, and 95 percentiles of the respective distributions indicated by major tick marks. The case when only one line of evidence is considered, here the instrumental record (coral colored line), unduly emphasizes those values of ECS which it does not constrain.

small chance of ECS being >4.5 K. The posteriori estimate is relatively insensitive to modest changes in the prior, or how we assess the evidence that appears to refute the storyline for a high ECS (Appendix S1). This is less true if only a single line of evidence in isolation is considered, as reasoning from a single line of evidence is more prone to be misleading. For instance, taking only the instrumental record, which appears to rule out a very high ECS, actually ends up increasing the chance of ECS <1.5 K (Figure 2b).

We nonetheless caution against focusing too narrowly on the posterior distribution that emerges from one or the other example. As doing so risks missing the point, which is to illustrate the role of different lines of evidence in modifying a prior judgment. In evaluating how different physical storylines shape thinking as to which values of ECS seem implausible it will be important to argue about the choice of the prior, the likelihoods assigned to different lines of evidence, and their degree of independence. In this spirit, our examples help demonstrate how a physical storyline can help to quantify confidence in a particular ECS bound and make transparent the role of different lines of evidence.

7. Prospects for Narrowing Uncertainty

By elucidating the reasoning behind a given bound on ECS, the storyline approach helps define research directions that could refine and narrow the ECS range. For instance, greater confidence in the sign of cloud feedbacks would have implications for the lower bound on ECS. We estimate that increased confidence in a net positive cloud feedback may be sufficient to refute a condition of the storyline for an ECS less than 2.0 K (see Appendix S1). Better reconstructions of the tropical temperatures at the time of the LGM, or greenhouse gas levels in past "hothouse" climates, could also lift this lower bound.

Critical tests for the upperbound could be provided by a better understanding of pattern effects, or by testing recent arguments for a more modest aerosol forcing [*Stevens*, 2015]. They could also come from glacial cycle evidence, either via better understanding of pattern effects or improved proxy evidence—especially temperature reconstructions. Importantly, a more specific understanding of low-cloud feedback mechanisms [e.g., *Sherwood et al.*, 2014] now makes it possible to design field experiments in ways that can better address cloud feedback mechanisms. Such field studies, together with high-resolution modeling, may offer the best opportunity to refine the upper bound of ECS by ruling out very strong positive cloud feedbacks.

Attempts to tighten the bounds on ECS may also need to account for recent findings that the radiative forcing from rising greenhouse gases is less tightly constrained by ab initio calculations than is commonly thought [e.g., *Geoffroy et al.*, 2013; *Vial et al.*, 2013]. Recent work makes a compelling case that the radiative impact of a change in the atmosphere's composition can be modified by poorly understood changes in

cloudiness (or surface reflectance) that arise in response to a perturbing agent in the absence of global surface warming [*Gregory and Webb*, 2008; *Sherwood et al.*, 2015]. Present evidence does not suggest that this is a large effect, but the uncertainty it introduces into forcing estimates will become more important as uncertainty in ECS is reduced.

8. Refutational Reasoning: Linking Assessments to Storylines

Explicitly laying out the storylines for exceeding a particular ECS bound achieves two important things. First, it illustrates the constraining power of different lines of evidence; refutational reasoning then reveals why experts broadly familiar with this evidence are so skeptical about very high or very low ECS values, even though individual lines of evidence considered in isolation would not merit such skepticism and may sometimes even support such values. Second, it clarifies what evidence is most important, and from where improvements in our knowledge of the ECS are most likely to come. As *Bony et al.* [2015] argued in a different context, physical storylines thus support a more active and effective assessment process, one in which important gaps in understanding—and the strategies needed to fill them—are naturally identified. The physical storylines explored here are powerful, but how absolute or definitive they are, and whether our approach for combining different lines of evidence is as promising as it appears, remains to be judged. A formal and thorough assessment is needed.

Acknowledgments
The manuscript was jointly written by its four authors. We are indebted to Earth's Future and its editor, Michael Ellis, for their support in the drafting this article. Critical comments of the editor and an anonymous reviewer influenced our thinking and greatly improved the manuscript. We also wish to acknowledge all the participants of the Ringberg Workshop on Earth's Climate Sensitivities (2015): Ayako Abe-Ouchi, Myles Allen, Timothy Andrews, James Annan, Kyle Armour, Nicolas Bellouin, Lennart Bengtsson, Rodrigo Caballero, John Church, Michel Crucifix, Andrew Dessler, Tamsin Edwards, John Fasullo, Dagmar Fläschner, Piers Forster, Olivier Geoffroy, Chris Golaz, Jonathan Gregory, Julia Hargreaves, Gabi Hegerl, Reto Knutti, Mojib Latif, Nic Lewis, Thorsten Mauritsen, Gavin Schmidt, David Sexton, Graeme Stephens, Rowan Sutton, Jessica Vial, Kosaka Yu, Mark Zelinka. Their constructive and collegial contributions greatly influenced our thoughts. Timothy Andrews, James Annan, Jean-Louis Dufresne, Julia Hargreaves, Nicolas Gruber, James Murphy, Gian-Kasper Plattner, and Mark Ringer are thanked for their comments on a draft version of the manuscript, and J. Annan is additionally thanked for his insightful comments on our use of Bayesian inference. Primary data and scripts used in the analysis and other supporting information that may be useful in reproducing the author's work are archived by the Max Planck Institute for Meteorology and can be obtained by contacting publications@mpimet.mpg.de. B. S. thanks Margot Hillinger for her support and encouragement. S. S. was supported by the Australian Research Council (DP140101104). S. B. was supported by the European Research Council from the European Union (grant agreement No 694768). M. W. was supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). Support by the World Climate Research Programme's Grand Science Challenge on Clouds, Circulation and Climate Sensitivity, and the Max Planck Society, helped make this all possible.

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*(3), 253–271, doi:10.1002/env.2140.

Alexander, K., K. J. Meissner, and T. J. Bralower (2015), Sudden spreading of corrosive bottom water during the Palaeocene-Eocene Thermal Maximum, *Nat. Geosci.*, *8*(6), 458–461, doi:10.1038/ngeo2430.

Allen, M. R., and D. J. Frame (2007), Call off the quest, *Science*, *318*(5850), 582–583, doi:10.1126/science.1149988.

Andrews, T. (2014), Using an AGCM to diagnose historical effective radiative forcing and mechanisms of recent decadal climate change, *J. Clim.*, *27*(3), 1193–1209, doi:10.1175/JCLI-D-13-00336.1.

Andrews, T., J. M. Gregory, and M. J. Webb (2015), The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models, *J. Clim.*, *28*(4), 1630–1648, doi:10.1175/JCLI-D-14-00545.1.

Annan, J. D. (2015), Recent developments in Bayesian estimation of climate sensitivity, *Curr. Clim. Change Rep.*, *1*(4), 263–267, doi:10.1007/s40641-015-0023-5.

Annan, J. D., and J. C. Hargreaves (2006), Using multiple observationally-based constraints to estimate climate sensitivity, *Geophys. Res. Lett.*, *33*, L06704, doi:10.1029/2005GL025259.

Annan, J. D., and J. C. Hargreaves (2015), A perspective on model-data surface temperature comparison at the Last Glacial Maximum, *Quat. Sci. Rev.*, *107*, 1–10, doi:10.1016/j.quascirev.2014.09.019.

Armour, K. C., C. M. Bitz, and G. H. Roe (2013), Time-varying climate sensitivity from regional feedbacks, *J. Clim.*, *26*(13), 4518–4534, doi:10.1175/JCLI-D-12-00544.1.

Arrhenius, S. (1896), On the influence of carbonic acid in the air upon the temperature of the ground, *Philos. Mag. Ser.* , *41*(251), 237–276, doi:10.1080/14786449608620846.

Berner, R. A., and Z. Kothavala (2001), Geocarb III: A revised model of atmospheric CO2 over Phanerozoic time, *Am. J. Sci.*, *301*(2), 182–204, doi:10.2475/ajs.301.2.182.

Betts, A. K., and Harshvardhan (1987), Thermodynamic constraint on the cloud liquid water feedback in climate models, *J. Geophys. Res. Atmos.*, *92*(D7), 8483–8485, doi:10.1029/JD092iD07p08483.

Bijl, P. K., A. J. P. Houben, S. Schouten, S. M. Bohaty, A. Sluijs, G.-J. Reichart, J. S. S. Damste, and H. Brinkhuis (2010), Transient middle eocene atmospheric CO_2 and temperature variations, *Science*, *330*(6005), 819–821, doi:10.1126/science.1193654.

Bony, S., B. Stevens, I. H. Held, J. F. Mitchell, J.-L. Dufresne, K. A. Emanuel, P. Friedlingstein, S. Griffies, and C. Senior (2013), Carbon dioxide and climate: Perspectives on a scientific assessment, in *Climate Science for Serving Society*, edited by J. W. Hurrell and G. Asrar, pp. 391–413 , Springer Netherlands, Dordrecht, the Netherlands.

Bony, S., et al. (2015), Clouds, circulation and climate sensitivity, *Nat. Geosci.*, *8*(4), 261–268, doi:10.1038/ngeo2398.

Bony, S., B. Stevens, D. Coppin, T. Becker, K. A. Reed, A. Voigt, and B. Medeiros (2016), Thermodynamic control of anvil cloud amount, *Proc. Natl. Acad. Sci. U. S. A.*, *113*(32), 8927–8932.

Boucher, O., et al. (2013), Clouds and aerosols, 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*, edited by T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, pp. 571–657 , Cambridge Univ. Press, Cambridge, UK and New York.

Braconnot, P., S. P. Harrison, M. Kageyama, P. J. Bartlein, V. Masson-Delmotte, A. Abe-Ouchi, B. Otto-Bliesner, and Y. Zhao (2012), Evaluation of climate models using palaeoclimatic data, *Nat. Clim. Change*, *2*(6), 417–424, doi:10.1038/nclimate1456.

Bretherton, C. S. (2015), Insights into low-latitude cloud feedbacks from high-resolution models, *Philos. Trans. R. Soc. A*, *373*(2054), 20140415, doi:10.1098/rsta.2014.0415.

Bretherton, C. S., P. N. Blossey, and C. R. Jones (2013), Mechanisms of marine low cloud sensitivity to idealized climate perturbations: A single-LES exploration extending the CGILS cases, *J. Adv. Model. Earth Syst.*, *5*(2), 316–337, doi:10.1002/JAME.20019.

Brient, F., and S. Bony (2013), Interpretation of the positive low-cloud feedback predicted by a climate model under global warming, *Clim. Dyn.*, *40*(9–10), 2415–2431, doi:10.1007/s00382-011-1279-7.

Brient, F., T. Schneider, Z. Tan, S. Bony, X. Qu, and A. Hall (2015), Shallowness of tropical low clouds as a predictor of climate models' response to warming, *Clim. Dyn.*, *47*(1–2), 433–449.

Charney, J. G., A. Arakawa, D. J. Baker, B. Bolin, R. E. Dickinson, R. M. Goody , C. E. Leith, H. M. Stommel, and C. I. Wunsch (1979), Carbon dioxide and climate: A scientific assessment, National Academy of Sciences, Washington, D. C.

Clement, A. C., R. Burgman, and J. R. Norris (2009), Observational and model evidence for positive low-level cloud feedback, *Science*, *325*(5939), 460–464, doi:10.1126/science.1171255.

Collins, M., et al. (2013), Long-term climate change: Projections, commitments 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*, edited by T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley , pp. 1029–1136 , Cambridge Univ. Press, Cambridge, UK and New York.

Cooke, R., B. A. Wielicki, D. F. Young, and M. G. Mlynczak (2013), Value of information for climate observing systems, *Environ. Syst. Decis.*, *34*(1), 98–109, doi:10.1007/s10669-013-9451-8.

Covey, C., L. C. Sloan, and M. I. Hoffert (1996), Paleoclimate data constraints on climate sensitivity: The paleocalibration method, *Clim. Change*, *32*(2), 165–184, doi:10.1007/BF00143708.

Dessler, A. E. (2010), A determination of the cloud feedback from climate variations over the past decade, *Science*, *330*(6010), 1523–1527, doi:10.1126/science.1192546.

Dessler, A. E. (2013), Observations of climate feedbacks over 2000–10 and comparisons to climate models, *J. Clim.*, *26*(1), 333–342, doi:10.1175/JCLI-D-11-00640.1.

Dufresne, J.-L., and M. Saint-Lu (2016), Positive feedback in climate: Stabilization or runaway, illustrated by a simple experiment, *Bull. Am. Meteorol. Soc.*, *97*(2), 755–765, doi:10.1175/BAMS-D-14-00022.1.

Flato, G., et al. (2013), Evaluation of climate models, 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*, edited by T. F. Stocker, D. Qin, G. K. Plattner, M. Tingor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley , pp. 741–866 , Cambridge Univ. Press, Cambridge, UK and New York.

Forster, P. M. (2016), Inference of climate sensitivity from analysis of earth's energy budget, *Annu. Rev. Earth Planet. Sci.*, *44*(1), 85–106, doi:10.1146/annurev-earth-060614-105156.

Geoffroy, O., D. Saint-Martin, D. J. L. Olivié, A. Voldoire, G. Bellon, and S. Tyteca (2013), Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments, *J. Clim.*, *26*(6), 1841–1857, doi:10.1175/JCLI-D-12-00195.1.

Gettelman, A., J. E. Kay, and K. M. Shell (2012), The evolution of climate sensitivity and climate feedbacks in the community atmosphere model, *J. Clim.*, *25*(5), 1453–1469, doi:10.1175/JCLI-D-11-00197.1.

Gordon, N. D., and S. A. Klein (2014), Low-cloud optical depth feedback in climate models, *J. Geophys. Res. Atmos.*, *119*(10), 6052–6065, doi:10.1002/2013JD021052.

Gregory, J., and M. Webb (2008), Tropospheric adjustment induces a cloud component in CO2 forcing, *J. Clim.*, *21*(1), 58–71, doi:10.1175/2007JCLI1834.1.

Gregory, J. M., T. Andrews, and P. Good (2015), The inconstancy of the transient climate response parameter under increasing CO_2, *Philos. Trans. R. Soc. A*, *373*(2054), 1–25.

Gregory, J. M., and T. Andrews (2016), Variation in climate sensitivity and feedback parametersduring the historical period, *Geophys. Res. Lett.*, *43*, 3911–3920, doi:10.1002/2016GL068406.

Hargreaves, J. C., J. D. Annan, M. Yoshimori, and A. Abe-Ouchi (2012), Can the last glacial maximum constrain climate sensitivity? *Geophys. Res. Lett.*, *39*(24), L24702.

Hartmann, D. L., and K. Larson (2002), An important constraint on tropical cloud − climate feedback, *Geophys. Res. Lett.*, *29*(20), 1951, doi:10.1029/2002GL015835.

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, doi:10.1038/nature04679.

Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis (2010), Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing, *J. Clim.*, *23*(9), 2418–2427, doi:10.1175/2009JCLI3466.1.

Hopcroft, P. O., and P. J. Valdes (2015), How well do simulated last glacial maximum tropical temperatures constrain equilibrium climate sensitivity? *Geophys. Res. Lett.*, *42*(13), 5533–5539, doi:10.1002/2015GL064903.

Huber, M., U. Beyerle, and R. Knutti (2014), Estimating climate sensitivity and future temperature in the presence of natural climate variability, *Geophys. Res. Lett.*, *41*(6), 2086–2092, doi:10.1002/2013GL058532.

Jochum, M., A. Jahn, S. Peacock, D. A. Bailey, J. T. Fasullo, J. Kay, S. Levis, and B. Otto-Bliesner (2012), True to Milankovitch: Glacial inception in the new community climate system model, *J. Clim.*, *25*(7), 2226–2239, doi:10.1175/JCLI-D-11-00044.1.

Kang, S. M., I. M. Held, and S.-P. Xie (2013), Contrasting the tropical responses to zonally asymmetric extratropical and tropical thermal forcing, *Clim. Dyn.*, *42*(7–8), 2033–2043.

Kay, J. E., B. Medeiros, Y. T. Hwang, A. Gettelman, J. Perket, and M. G. Flanner (2014), Processes controlling Southern Ocean shortwave climate feedbacks in CESM, *Geophys. Res. Lett.*, *41*(2), 616–622, doi:10.1002/2013GL058315.

Knutti, R., and G. C. Hegerl (2008), The equilibrium sensitivity of the Earth's temperature to radiation changes, *Nat. Geosci.*, *1*(11), 735–743, doi:10.1038/ngeo337.

Knutti, R., and M. A. A. Rugenstein (2015), Feedbacks, climate sensitivity and the limits of linear models, *Philos. Trans. R. Soc. A*, *373*(2054), 20150146, doi:10.1098/rsta.2015.0146.

Kuang, Z., and D. L. Hartmann (2007), Testing the fixed anvil temperature hypothesis in a cloud-resolving model, *J. Clim.*, *20*(10), 2051–2057, doi:10.1175/JCLI4124.1.

Lewis, N., and J. A. Curry (2014), The implications for climate sensitivity of AR5 forcing and heat uptake estimates, *Clim. Dyn.*, *45*, 1009–1023.

Li, C., J.-S. Storch, and J. Marotzke (2012), Deep-ocean heat uptake and equilibrium climate response, *Clim. Dyn.*, *41*, 1071–1086.

Lindzen, R., M. D. Chou, and A. Y. Hou (2001), Does the earth have an adaptive infrared iris? *Bull. Am. Meteorol. Soc.*, *83*(2), 417–432, doi:10.1175/1520-0477(2001)082<0417:DTEHAA>2.3.CO;2.

Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong (2009), Toward optimal closure of the Earth's top-of-atmosphere radiation budget, *J. Clim.*, *22*(3), 748–766, doi:10.1175/2008JCLI2637.1.

Lüthi, D., et al. (2008), High-resolution carbon dioxide concentration record 650,000–800,000 years before present, *Nature*, *453*(7193), 379–382, doi:10.1038/nature06949.

MacDougall, A. H. (2016), The transient response to cumulative CO_2 emissions: A review, *Curr. Clim. Change Rep.*, *2*(1), 39–47, doi:10.1007/s40641-015-0030-6.

Mauritsen, T., and B. Stevens (2015), Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models, *Nat. Geosci.*, *8*, 346–351.

Meraner, K., T. Mauritsen, and A. Voigt (2013), Robust increase in equilibrium climate sensitivity under global warming, *Geophys. Res. Lett.*, *40*(22), 5944–5948, doi:10.1002/2013GL058118.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, *J. Geophys. Res. Atmos.*, *117*(D8), D08101.

Neubersch, D., H. Held, and A. Otto (2014), Operationalizing climate targets under learning: An application of cost-risk analysis, *Clim. Change*, *126*(3–4), 305–318, doi:10.1007/s10584-014-1223-z.

Nuijens, L., B. Medeiros, I. Sandu, and M. Ahlgrimm (2015), The behavior of trade-wind cloudiness in observations and models: The major cloud components and their variability, *J. Adv. Model. Earth Syst.*, *7*(2), 600–616, doi:10.1002/2014MS000390.

Otto, A., et al. (2013), Energy budget constraints on climate response, *Nat. Geosci.*, *6*, 415–416, doi:10.1038/ngeo1836.

Pagani, M., K. Caldeira, D. Archer, and J. C. Zachos (2006), An ancient carbon mystery, *Science*, *314*(5805), 1556–1557, doi:10.1126/science.1136110.

Pagani, M., M. Huber, B. Sageman (2014), Greenhouse Climates, In Treatise on Geochemistry (Second Edition), edited by H.D. Holland and K.K. Turekian, Elsevier, Oxford, 281–304, doi:10.1016/B978-0-08-095975-7.01314-0.

Pierrehumbert, R. T., D. S. Abbot, A. Voigt, and D. Koll (2011), Climate of the neoproterozoic, *Annu. Rev. Earth Planet. Sci.*, *39*(1), 417–460, doi:10.1146/annurev-earth-040809-152447.

Qu, X., A. Hall, S. A. Klein, and P. M. Caldwell (2013), On the spread of changes in marine low cloud cover in climate model simulations of the 21st century, *Clim. Dyn.*, *42*(9–10), 2603–2626.

Rieck, M., L. Nuijens, and B. Stevens (2012), Marine boundary layer cloud feedbacks in a constant relative humidity atmosphere, *J. Atmos. Sci.*, *69*(8), 2538–2550, doi:10.1175/JAS-D-11-0203.1.

Rohling, E. J., et al. (2012), Making sense of palaeoclimate sensitivity, *Nature*, *491*(7426), 683–691, doi:10.1038/nature11574.

Royer, D. L., R. A. Berner, and J. Park (2007), Climate sensitivity constrained by CO_2 concentrations from the last 420 million years, *Nature*, *446*, 530–532, doi:10.1038/nature05699.

Schmidt, G. A., et al. (2014), Using palaeo-climate comparisons to constrain future projections in CMIP5, *Clim. Past*, *10*(1), 221–250, doi:10.5194/cp-10-221-2014.

Schmittner, A., N. M. Urban, J. D. Shakun, N. M. Mahowald, P. U. Clark, P. J. Bartlein, A. C. Mix, and A. Rosell-Mele (2011), Climate sensitivity estimated from temperature reconstructions of the last glacial maximum, *Science*, *334*(6061), 1385–1388, doi:10.1126/science.1203513.

Seneviratne, S. I., M. G. Donat, A. J. Pitman, R. Knutti, and R. L. Wilby (2016), Allowable CO2 emissions based on regional and impact-related climate targets, *Nature*, *529*(7587), 477–483, doi:10.1038/nature16542.

Senior, C. A., and J. F. B. Mitchell (2000), The time-dependence of climate sensitivity, *Geophys. Res. Lett.*, *27*(17), 2685–2688, doi:10.1029/2000GL011373.

Sherwood, S. C., W. Ingram, Y. Tsushima, M. Satoh, M. Roberts, P. L. Vidale, and P. A. O'Gorman (2010), Relative humidity changes in a warmer climate, *J. Geophys. Res. Atmos.*, *115*(D9), D09104, doi:10.1029/2009JD012585.

Sherwood, S. C., S. Bony, and J.-L. Dufresne (2014), Spread in model climate sensitivity traced to atmospheric convective mixing, *Nature*, *505*(7481), 37–42, doi:10.1038/nature12829.

Sherwood, S. C., S. Bony, O. Boucher, C. Bretherton, P. M. Forster, J. M. Gregory, and B. Stevens (2015), Adjustments in the forcing-feedback framework for understanding climate change, *Bull. Am. Meteorol. Soc.*, *96*(2), 217–228, doi:10.1175/BAMS-D-13-00167.1.

Stevens, B. (2015), Rethinking the lower bound on aerosol radiative forcing, *J. Clim.*, *28*(12), 4794–4819, doi:10.1175/JCLI-D-14-00656.1.

Stevens, B., and S. Bony (2013), Water in the atmosphere, *Phys. Today*, *66*(6), 29, doi:10.1063/PT.3.2009.

Stevens, B., and J.-L. Brenguier (2009), Cloud-controlling factors: Low clouds, in *Clouds in the Perturbed Climate System*, edited by J. Heintzenberg and R. J. Charlson , pp. 1–24 , MIT Press, Cambridge, Mass..

Stevens, B., A. Abe-Ouchi, S. Bony, G. C. Hegerl, G. A. Schmidt, S. C. Sherwood, and M. J. Webb (2015), Ringberg15: Earth's climate sensitivities, *Tech. Rep. 11*, World Climate Research Programme, Geneva, Switzerland.

Stocker, T. F., D. Qin, G. K. Plattner, M. Tingor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (2013), *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*, vol. *413*, Cambridge Univ. Press, Berlin and Heidelberg.

Su, H., J. H. Jiang, C. Zhai, T. J. Shen, J. D. Neelin, G. L. Stephens, and Y. L. Yung (2014), Weakening and strengthening structures in the Hadley Circulation change under global warming and implications for cloud response and climate sensitivity, *J. Geophys. Res. Atmos.*, *119*(10), 5787–5805, doi:10.1002/2014JD021642.

Tebaldi, C., and J. M. Arblaster (2014), Pattern scaling: Its strengths and limitations, and an update on the latest model simulations, *Clim. Change*, *122*(3), 459–471, doi:10.1007/s10584-013-1032-9.

Tripati, A. K., S. Sahany, D. Pittman, R. A. Eagle, J. D. Neelin, J. L. Mitchell, and L. Beaufort (2014), Modern and glacial tropical snowlines controlled by sea surface temperature and atmospheric mixing, *Nat. Geosci.*, *7*(3), 205–209, doi:10.1038/ngeo2082.

Vial, J., J.-L. Dufresne, and S. Bony (2013), On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates, *Clim. Dyn.*, *41*(11–12), 3339–3362, doi:10.1007/s00382-013-1725-9.

Waelbroeck, C., et al. (2009), Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum, *Nat. Geosci.*, *2*(2), 127–132, doi:10.1038/ngeo411.

Webb, M. J., and A. P. Lock (2012), Coupling between subtropical cloud feedback and the local hydrological cycle in a climate model, *Clim. Dyn.*, *41*(7–8), 1923–1939.

Webb, M. J., et al. (2015), The impact of parametrized convection on cloud feedback, *Philos. Trans. R. Soc. A*, *373*(2054), 20140414, doi:10.1098/rsta.2014.0414.

Winton, M., K. Takahashi, and I. M. Held (2010), Importance of ocean heat uptake efficacy to transient climate change, *J. Clim.*, *23*(9), 2333–2344, doi:10.1175/2009JCLI3139.1.

Zelinka, M. D., and D. L. Hartmann (2011), The observed sensitivity of high clouds to mean surface temperature anomalies in the tropics, *J. Geophys. Res. Atmos.*, *116*(D23), D23103, doi:10.1029/2011JD016459.

Zelinka, M. D., S. A. Klein, K. E. Taylor, T. Andrews, M. J. Webb, J. M. Gregory, and P. M. Forster (2013), Contributions of different cloud types to feedbacks and rapid adjustments in CMIP5*, *J. Clim.*, *26*(14), 5007–5027, doi:10.1175/JCLI-D-12-00555.1.

Zelinka, M. D., C. Zhou, and S. A. Klein (2016), Insights from a refined decomposition of cloud feedbacks, *Geophys. Res. Lett.*, *43*, 9259–9269, doi:10.1002/2016GL069917.

Zhang, M., et al. (2013), CGILS: Results from the first phase of an international project to understand the physical mechanisms of low cloud feedbacks in single column models, *J. Adv. Model. Earth Syst.*, *5*(4), 826–842, doi:10.1002/2013MS000246.

Zhao, M. (2014), An Investigation of the connections among convection, clouds, and climate sensitivity in a global climate model, *J. Clim.*, *27*(5), 1845–1862, doi:10.1175/JCLI-D-13-00145.1.

Zhou, C., M. D. Zelinka, and A. E. Dessler (2015), The relationship between interannual and long-term cloud feedbacks, *Geophys. Res. Lett.*, *42*, 10,463–10,469, doi:10.1002/2015GL066698.

Supporting Information for "Narrowing Bounds on Earth's Climate Sensitivity"

Bjorn Sevens[1], Steven C. Sherwood[2], Sandrine Bony[3], and Mark J. Webb[4]

Corresponding author: B. Stevens, Max-Planck-Institut für Meteorologie, Bundesstrasse 53, 20146 Hamburg Germany (bjorn.stevens@mpimet.mpg.de)

[1]Max Planck Institute for Meteorology, Hamburg Germany

[2]CCRC and Centre of Excellence for Climate System Science, University of New South Wales, Sydney, Australia

[3]LMD/IPSL, CNRS, Université Marie et Pierre Curie, Paris, France

[4]MetOffice/Hadley Centre, Exeter, United Kingdom

Contents of this file

1. S1: Instrumental Record

2. S2: Last Glacial Maximum, LGM

3. S3: Eocene and late Palaeocene Thermal Maximum

4. S4: Physical understanding of climate feedbacks

5. S5: Bayesian inference

6. Figures S1: Posterior distribution with a broader prior.

7. Table S1: ECS ranges from studies fitting simple models to historical temperature record.

8. Table S2: Parameters used to construct likelihood functions.

Introduction

This Supplemental Information describes in more detail the authors' reasoning as applied to the hypotheses for a surprisingly (very) high or low Equilibrium Climate Sensitivity, ECS. It does so by elaborating on studies based primarily on interpretations of the instrumental record, palaeoclimate studies, and physical understanding, including present day observations designed to constraint specific processes relevant to feedback processes. This analysis helps rationalize the choice of likelihoods assigned to different lines of evidence in the framework of the Bayesian inference. Further details and sensitivities within this inferential framework are also described. Note that in some cases citations are less exhaustive then they would be for a stand-alone supplement (which would appear as part of a more formal study or assessment) as journal policy does not allow articles to be cited in the supporting information if they are not cited in the main article.

S1: Instrumental record.

A large number of studies [*Stocker et al.*, 2013] have attempted to infer the ECS by fitting simple models to different aspects of the instrumental record. The models are characterized as simple because they assume that temperature anomalies can be linearly related to radiation anomalies through a single parameter, often called the climate feedback parameter and denoted by λ. By convention a negative net feedback implies a stable climate so that the temperature change, ΔT, after the system returns to stationarity in response to a radiative forcing, F, is given as

$$\Delta T = -\frac{F}{\lambda} \tag{S1}$$

For sufficiently small perturbations about an arbitrary stationary reference state, λ (with units of $\mathrm{W\,m^{-2}\,K^{-1}}$) can be thought of as the radiative response to warming. The particular value of λ for the special case where the starting state of the model is taken to be the pre-industrial climate state, and F is the forcing that arises from a doubling of atmospheric CO_2, this equation can be used to define the ECS as

$$\mathrm{ECS} = -\frac{F_{2\times}}{\lambda_{2\times}}. \tag{S2}$$

So this implies that the ECS is determined by a particular value of λ, namely $\lambda_{2\times}$, which arises when the forcing is a doubling of CO_2 and the stationary starting state is taken to be the state of the climate system in the preindustrial period. In assuming that the radiative response to an arbitrary forcing, λ, is constant, simple models do not distinguish between λ and $\lambda_{2\times}$. In contrast to many simple models, comprehensive models are defined in part by the fact that their value of λ is an emergent property of the model that arises through the competition of different processes, and as such is not directly specified or fitted, and can vary with the state of the climate system.

By matching the state of the simple model to available data it is possible to derive estimates for the model parameters, and by inference the ECS. Early studies adopting this approach suggested that the instrumental record did not strongly constrain ECS [*Knutti and Hegerl*, 2008; *Collins et al.*, 2013]. More recent studies, which benefit from methodological advances and more data, conclude that the instrumental temperature record confidently rule out high (greater than about 2.8 K) values of ECS, Table S1.

This conclusion is, however, subject to some important caveats. The first is that the upper (95 %) confidence bound is sensitive to the end period chosen for the analysis as well as assumptions about the aerosol forcing. Choosing a longer end period as in the original *Lewis and Curry* [2014] and *Otto et al.* [2013] studies increases the upper bound of the 95 % confidence interval to 5.4 K and 5.0 K respectively. Studies which continue to allow for a very strong (more negative than $-1\,\mathrm{W\,m^{-}2}$) aerosol forcing yield a 95 % confidence bound close to 4 K, studies that no longer accommodate such a large aerosol forcing (TableS1), give a 95 % bound nearer 2.5 K.

A second caveat emerges from a perfect model study which tests the ability of the simple model approach to reproduce the ECS of a more comprehensive model using an ensemble of historical simulations covering a range of decadal variability realizations [*Huber et al.*, 2014]. The estimates of ECS using the simple model assumptions were shown not only to underestimate the actual ECS on average, but also to be sensitive to the particular realization of variability in the historical period, varying by up to ± 1.6 K relative to the ensemble mean prediction. This finding raises the possibility that simple model estimates of ECS could be biased because of a dependence on the particular realization of unforced

variability which happened to occur in the historical period, but does not influence the actual ECS.

More recent studies with comprehensive climate models have made these caveats more precise and call into question the idea that λ can be assumed to be constant. A number of studies argue for a "pattern effect", whereby λ depends on the pattern of the surface temperature response (or the state of the model) in a way that the interpretation of the instrumental temperature record based on simple models does not capture [e.g., *Senior and Mitchell*, 2000; *Geoffroy et al.*, 2013; *Armour et al.*, 2013; *Gregory et al.*, 2015; *Knutti and Rugenstein*, 2015; *Gregory and Andrews*, 2016]. In the great majority of comprehensive models, this pattern effect would cause estimates of ECS using a constant λ to be biased low, for reasons that are partially understood [*Senior and Mitchell*, 2000; *Geoffroy et al.*, 2013; *Armour et al.*, 2013; *Gregory et al.*, 2015]. Most studies using comprehensive models suggest a modest pattern effect between initial and long term warming in the carbon dioxide quadrupling scenarios, leading to a 20-30 % under-estimate of ECS if not accounted for [*Geoffroy et al.*, 2013; *Armour et al.*, 2013; *Knutti and Rugenstein*, 2015; *Andrews et al.*, 2015].

Correcting the simple model interpretations of the instrumental record by allowing for a 33 % pattern effect would imply that the studies cited in Table S1 can, with some (here the eighty-third percentile) confidence, rule out values of ECS greater than about 2.4 K to 3.7 K, depending on the study. We adopt a more conservative statement, and assert that an ECS greater than 4 K becomes difficult to reconcile with the instrumental temperature record. Some recent work suggests that even such a large correction to the upper bound (from 2.8 K to 4.0 K) may still insufficiently account for pattern effects [*Gregory and*

Andrews, 2016], thereby further undermining the ability of the instrumental record to refute the story line for a very high ECS.

S2: Last Glacial Maximum, LGM.

Simple Models have been applied to studies of the last glacial maximum (LGM) in a similar way as has been done for the historical temperature record [*Rohling et al.*, 2012; *Knutti and Hegerl*, 2008]. These provide estimates of ECS whose one sigma range varies from 1.7 K to 4.5 K. Although most of these studies also apply simple models whose parameter sensitivities can be fully sampled, the question as to whether the proxy record for the LGM constrains the climate sensitivity of more comprehensive models has also been explored [*Hargreaves et al.*, 2012; *Schmidt et al.*, 2014]. Comprehensive modeling studies carried out as part of the palaeoclimate modelling intercomparison project (PMIP) show a reasonably strong relationship between tropical sea-surface temperature changes in the LGM relative to the present (denoted δ_T) and the ECS [*Hargreaves et al.*, 2012]. A relationship between δ_T and ECS is evident also in studies employing simpler models [*Schmittner et al.*, 2011], and makes physical sense. Nonetheless, a recent study incorporating models which have begun to incorporate a more diverse treatment of uncertain physical processes has called this relationship into question [*Hopcroft and Valdes*, 2015].

Estimates of δ_T vary in the literature. The Multi-proxy Approach for the Reconstruction of the Glacial Ocean Surface [MARGO, *Waelbroeck et al.*, 2009] project was designed to address many previous shortcomings in earlier reconstructions of tropical SSTs during the LGM. Using estimates of δ_T to better constrain estimates of ECS, a confidence interval (5 % to 95 %) of 1 K to 4.2 K has been derived based on analysis of comprehensive models and 1.4 K to 2.8 K from a simple model whose parameters could be more formally

constrained [*Schmittner et al.*, 2011]. Studies of the LGM with both comprehensive and simple models may be subject to the pattern effect discussed above in the context of interpretations of the instrumental record, and are also influenced by any possible biases in the reconstructions of tropical SSTs. Recently the MARGO reconstruction has been re-evaluated using SST patterns derived from comprehensive modeling of the LGM, resulting in a larger δ_T [*Annan and Hargreaves*, 2015]. When applied to the analysis of comprehensive models this shifts the ECS confidence interval by about 0.5 K (1.6 K to 4.5 K). Sensitivity tests suggest that a similar change can be expected for earlier studies which would otherwise fail to constrain surprisingly low values of ECS [*Schmittner et al.*, 2011]. The idea that MARGO estimates of δ_T are biased low, and hence ECS estimates constrained by the MARGO data are also biased low has received further support from recent work using clustered isotopes to study changes in the tropical snow-line during the LGM [*Tripati et al.*, 2014].

S3: Eocene and late Palaeocene Thermal Maximum.

Large and positive values of δ_T occurred in the more distant past. Several epochs through the late Phanerozoic have been used to constrain ECS [see *Rohling et al.*, 2012, for a review], most prominently the late Palaeocene and early Eocene during which δ_T was at least 10 K, and a brief "hyper-thermal" warming spike at the boundary between these periods known as the Palaeocene-Eocene Thermal Maximum (PETM) during which δ_T shot up by another 5 K. Temperatures are not known exactly but are supported by incontrovertible evidence such as crocodilian fossils in the Arctic and tropical plants growing in (presently) cold continental interiors at high latitudes.

The Paleocene-Eocene background warmth can be attributed to the high levels of CO_2 inferred from proxy evidence and calculated based on the more active plate tectonics at the time [*Berner and Kothavala*, 2001]. CO_2 amounts are highly uncertain however (in contrast to the LGM case) and even less is known about other greenhouse gases, such that radiative forcings are not well known. Another problem is that climate feedbacks and forcings are each expected to vary under large climate changes, such that one cannot simply take a ratio of δ_T to forcing to yield ECS [*Rohling et al.*, 2012]. Nonetheless such warm climates appear to require a robust minimum ECS of at least 1.5 K, given reasonable assumptions about the carbon cycle and proxy evidence for CO_2 [*Royer et al.*, 2007; *Bijl et al.*, 2010; *Covey et al.*, 1996; *Rohling et al.*, 2012]. More recent investigations suggest a much more elevated lower bound of around 4 K on present-day ECS based on evidence from the Pliocene, late Eocene and early Oligocene periods [*Rohling et al.*, 2012].

The PETM event was caused by a rapid carbon release into the atmosphere as shown by carbon isotopic changes. Proxy greenhouse gas estimates imply ECS of 3.5 K-6.0 K [*Pagani et al.*, 2006; *Rohling et al.*, 2012], but this depends on the amount of greenhouse gases involved, which is sensitive to how records are interpreted and may be more than previously thought [*Alexander et al.*, 2015]. Uncertainty in background CO_2 is not quite as important, since it has a compensating effect on ECS estimated from the background warmth vs. from the PETM: the former is higher if there was less background CO_2, while the latter is higher if there was more (due to the logarithmic dependence of radiative forcing on CO_2 amount).

Clearly the evidence from hothouse climates as a whole does not present any real problem for a weak radiative response to warming in that era, consistent with a high ECS of

4.5 K or more, but at face value, rules out a very low ECS. Given the wide range of time periods and sources of proxy evidence that have been explored, to permit a very low ECS would seem to require a systematic (and substantial) bias having affected all studies. The main candidates that we see would be: i) increases in methane or other greenhouse gases systematically coinciding with those of CO_2, ii) systematic underestimation of past CO_2 changes, or iii) a large (factor of two or more) difference between λ in warmer climates and $\lambda_{2\times}$. The first is problematic because it would imply a positive methane feedback that would presumably also apply today. The second would require that multiple lines of evidence on CO_2 (several proxies and carbon-cycle modeling calculations of CO_2) each be incorrect. The third appears likely to some extent, and should be more actively explored to see how large the difference could be. To the extent that these possibilities can be eliminated or their effect quantified, hothouse climate changes may prove useful in establishing a lower bound on ECS, but given the very large differences in the climate system, and the difficulty in interpreting the evidence from such a distant past, caution should be exercised in attributing too much weight to this element of the story line.

S4: Physical understanding of climate feedbacks.

Climate feedbacks characterize how responsive the net outgoing radiation at the top of the atmosphere is to a global warming of the Earth's surface. The feedbacks associated with the radiative impact of clouds changes with warming are referred to as cloud feedbacks, while the other ones (associated with changes in surface and atmospheric temperatures, humidity, snow and sea-ice) are referred to as clear-sky (or non-cloud) feedbacks.

Clear-sky feedbacks are physically well understood and fairly robust across models: they arise from the temperature dependence of infrared emission (Stefan-Boltzmann law),

from water vapor changes with warming that closely follow the thermodynamic relationship of Clausius-Clapeyron (i.e. that correspond to a nearly unchanged relative humidity as climate is warming), and from the enhanced melting of snow and sea-ice as the surface temperature rises. The sum of clear-sky feedbacks (including the Planck response) diagnosed from comprehensive models is $-1.8 \pm 0.2\,\mathrm{W\,m^{-2}K^{-1}}$ for long-term anthropogenic climate change [*Vial et al.*, 2013; *Dessler*, 2013]. The nature and amplitude of these feedbacks are supported by a large and growing body of evidence, including physical reasoning tested in a hierarchy of numerical models and in observations [*Sherwood et al.*, 2010; *Stevens and Bony*, 2013] and observational tests of the model feedbacks on a variety of time scales [*Boucher et al.*, 2013; *Dessler*, 2013; *Zhou et al.*, 2015]. Given these feedbacks, an ECS lower than 2 K seems very unlikely, unless clouds provide a substantially negative feedback. On the other hand, positive cloud feedbacks have the potential to produce substantially higher ECS estimates.

A quantitative assessment of cloud feedbacks remains challenging. Nevertheless, while long considered as a mystery of climate science, the physics of cloud feedbacks is now becoming less enigmatic. Recent analyses using a wide hierarchy of models, ranging from process to comprehensive models, have pointed out a number of positive feedback mechanisms. One relates to the tendency of upper-tropospheric clouds to rise as the climate warms while maintaining roughly the same temperature at cloud-top. This mechanism, which is well explained by physical reasoning [*Hartmann and Larson*, 2002], tested by high-resolution process models [*Kuang and Hartmann*, 2007] and supported by observations [*Zelinka and Hartmann*, 2011], is considered as robust and responsible for a positive ($\approx 0.2\,\mathrm{W\,m^{-2}K^{-1}}$) feedback [*Zelinka et al.*, 2016].

Another positive feedback arises from the reduction of the low-cloud amount as the climate warms. A number of explanations have been proposed to explain this behavior, including reductions in moisture availability from surface evaporation [*Rieck et al.*, 2012; *Webb and Lock*, 2012], boundary layer drying by enhanced mixing with dry air from the free troposphere [e.g., *Brient and Bony*, 2013; *Zhang et al.*, 2013; *Sherwood et al.*, 2014; *Webb et al.*, 2015], and increased downward longwave emission from the free troposphere suppressing boundary layer turbulent mixing [*Bretherton et al.*, 2013]. Some of these arise from studies with very high resolution process models, which provide additional support for positive cloud feedbacks from low clouds [*Bretherton*, 2015]. However, the strength of the low-cloud feedback depends on the relative importance of antagonist influences [*Rieck et al.*, 2012; *Zhang et al.*, 2013; *Brient et al.*, 2015], and in models it is sensitive to the representation of turbulent and convective processes [*Gettelman et al.*, 2012; *Qu et al.*, 2013; *Zhao*, 2014; *Webb et al.*, 2015]. Consistently, its estimate varies greatly across models (ranging from $-0.1\,\mathrm{W\,m^{-2}\,K^{-1}}$ to $+0.6\,\mathrm{W\,m^{-2}\,K^{-1}}$), and it remains the primary source of uncertainty in ECS estimates [*Boucher et al.*, 2013].

A negative cloud feedback mechanism has been pointed out at high latitudes [*Zelinka et al.*, 2013], where warming causes clouds to become more reflective owing to ice changing to liquid [*Kay et al.*, 2014] and maybe additional processes [*Betts and Harshvardhan*, 1987]. Although this feedback is presumably too pronounced in climate models [*Gordon and Klein*, 2014], it is much too weak to compensate for the other positive cloud feedbacks at the global scale. It has been argued that an Iris Effect [i.e. a reduction of the upper-level cloud amount as the climate warms, e.g., *Lindzen et al.*, 2001] could produce a negative feedback. However, although an Iris effect is possible [*Bony et al.*, 2016] and even present

in many models, both observations [*Zelinka and Hartmann*, 2011] and comprehensive models [*Mauritsen and Stevens*, 2015] suggest that its ability to produce a substantial negative feedback is largely hindered by the strong compensation between the infrared and solar effects of changes in high-cloud amount.

The evidence for robust positive feedbacks from clear-sky processes, low-level clouds and upper-level clouds, plus the lack of evidence for substantially negative cloud feedbacks, constitute strong refutation arguments for an ECS less than 1.5 K, and even calls into question values of ECS less than 1.5 K. Physical understanding is however not sufficiently advanced to firmly constrain the upper bound of ECS estimates, largely because of uncertainties in the magnitude of the low-cloud feedback. If some observational constraints support ECS values in the upper range of model estimates [*Sherwood et al.*, 2014], other studies question the implications for the strength of the low-cloud feedback of model biases in the simulation of present-day low-level clouds [*Nuijens et al.*, 2015; *Brient et al.*, 2015]. A better quantification of the low-cloud feedback would greatly help assess the plausibility of a high or very high ECS.

S5: Bayesian inference.

For the section on Bayesian inference the prior arises from our most basic understanding of atmospheric physics, namely of how the atmosphere would respond to forcing. [*Stevens and Bony*, 2013]. It describes how the atmosphere would respond if the troposphere deepened with temperature consistent with the radiative definition of the tropopause (Fixed Anvil Temperature Hypothesis, or FAT), with relative humidity constant and no changes in clouds, except that their vertical extent is extended or contracted as the troposphere warms or cools, following FAT, and surface albedo increasing as ice retreats

with warming. In contrast to what was adopted by *Stevens and Bony* [2013], the feedback associated with clouds rising in a deepening troposphere is reduced in magnitude from $0.4\,\mathrm{W\,m^{-2}\,K^{-1}}$ to $0.2\,\mathrm{W\,m^{-2}\,K^{-1}}$ following more recent work to estimate this effect (Mark Zelinka personal communication, 2016). This leads to the prior being centered around a somewhat smaller value of ECS (2.2 K versus 2.7 K) as compared to what was suggested by Stevens and Bony. The uncertainty in the forcing, F, and feedback, λ, used to generate the prior are large, resulting in the prior having the form

$$\mathrm{ECS} = -\frac{F + \sigma_f}{\lambda + \sigma_\lambda} \tag{S3}$$

where σ_f is a gaussian distributed random variable with standard deviation of $0.2F$. Likewise σ_λ is also a gaussian distributed random variable with standard deviation of 0.5λ. For F and λ we adopt $3.7\,\mathrm{W\,m^{-2}}$ and $-1.6\,\mathrm{W\,m^{-2}\,K^{-1}}$ ($\lambda_{\mathrm{clr}} + \lambda_{\mathrm{FAT}} = -1.8 + 0.2\,\mathrm{W\,m^{-2}\,K^{-1}}$) respectively. The prior is constrained to sample only positive values of ECS by renormalizing the resultant conditional distribution $p(\chi|\chi > 0)$

There is an extensive literature on the choice of priors, and the issues are reviewed by *Annan* [2015]. We believe our choice of prior is reasonable, and not unduly influenced by the lines of evidence used to construct the posteriori distribution. Assuming a larger uncertainty in the forcing and the feedback, whereby $\sigma_\lambda = 0.75\lambda$ and $\sigma_F = 0.25F$ produces a similar result as compared to Fig. 2 of the manuscript, but for different values of σ (Fig. S1). It illustrates, that inferences drawn from the likelihood of the evidence are not especially sensitive to the details of the prior.

The exact form for the error function that was adopted in creating the likelihoods is:

$$P(e_j|\chi) = \frac{(1 - 2\varepsilon_j)\ \mathrm{erf}\,(2\chi - 2\chi_j) + 1}{2}, \quad \text{where} \quad 0 \le \varepsilon_j \le 0.5 \tag{S4}$$

where erf denotes the standardized error function. Physically one can think of this as a way to modify the null-hypothesis, as $\varepsilon_j = 0.5$ leads to all values of ECS being equally likely and thus does not modify the prior. The parameters for Eq. (S4) are provided in Table S2. The parameter ε_j is equal to the likelihood of misleading evidence, α_j, for the conditions related to the low ECS. For the high ECS, the likelihood of misleading evidence is given by $1 - \varepsilon_j$.

The parameter ε_j weights the marginal value of the evidence in refuting an argument for a value of the ECS exceeding a given bound. Because the form we adopted for the likelihood is a symmetric around χ_j, a 25 % likelihood of evidence being misleading for values of $\chi \gg \chi_j$ is balanced by a 75 % likelihood that the evidence is compatible with a value of $\chi \ll \chi_j$. And because the weight of the evidence in shaping the the expectation depends on the differential likelihood, this formulation ends up being a weaker statement about the strength of the evidence than might otherwise intuited. The use of the error function provides a smooth transition between the likelihoods shaped by the marginal value of the evidence. Experiments with piecewise linear functions give similar results.

Different elements of the story line for a very high or very low ECS were quantitatively explored in §1 through §4 above and these were used to construct values of χ_j and ε_j in Table S2. Certainly these choices are subject to discussion, with "pattern effects" arising as a common thread adding uncertainty in the interpretation of many of the lines of evidence, especially for the very high ECS story line. To show how our framework can address this uncertainty we explore how alternative interpretations of the evidence would effect the posterior distribution that arises from this analysis.

In one sensitivity study we made the transition region about the threshold ECS χ_j twice as large, and increasing the value of χ_j for each of the lines of evidence for conditions for the high ECS storyline by 0.5 K, for instance to reflect concerns that pattern effects are not sufficiently accounted for when interpreting past records [*Gregory and Andrews*, 2016; *Andrews*, 2014]. This slightly changes the 5th and 95th percentiles of the distribution, to 1.5 K and 4.5 K respectively, but the median remains unchanged (Fig. S2, also Table S3). To estimate how a better understanding of the sign of cloud feedbacks might influence the posterior distribution we reduced ε_j for this condition of the very low ECS story-line to 0.15 and increased χ_j to 2.0 K. This increased the value of the 5th percentile from 1.6 K to 1.8 K. Without this process constraint on the low-end ECS whatsoever the 5th percentile value remains quite low, at 1.4 K. Finally, we explored the influence of our interpretation of the instrumental record on the likelihood of high-end sensitivities. Here we modified $\varepsilon_{(i)}$ to 0.6, giving it as little weight as we give to understanding of feedback processes. This increases the value of the 95th percentile by 15 %, from 4.1 K to 4.7 K. Given that the 95th percentile value of the prior is 6.1 K, this suggests that even weak statements about our understanding of the past record and its relation to the present can provide useful constraints on the high-end sensitivities.

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*(3), 253–271.

Alexander, K., K. J. Meissner, and T. J. Bralower (2015), Sudden spreading of corrosive bottom water during the Palaeocene-Eocene Thermal Maximum, *Nature Geoscience*, *8*(6), 458–U57.

Allen, M. R., and D. J. Frame (2007), Call off the quest, *Science*, *318*(5850), 582–583.

Andrews, T. (2014), Using an AGCM to Diagnose Historical Effective Radiative Forcing and Mechanisms of Recent Decadal Climate Change, *J. Climate*, *27*(3), 1193–1209.

Andrews, T., J. M. Gregory, and M. J. Webb (2015), The Dependence of Radiative Forcing and Feedback on Evolving Patterns of Surface Temperature Change in Climate Models, *J. Climate*, *28*(4), 1630–1648.

Annan, J. D. (2015), Recent Developments in Bayesian Estimation of Climate Sensitivity, *Curr Clim Change Rep*, *1*(4), 263–267.

Annan, J. D., and J. C. Hargreaves (2006), Using multiple observationally-based constraints to estimate climate sensitivity, *Geophys. Res. Lett.*, *33*, (L06,704).

Annan, J. D., and J. C. Hargreaves (2015), A perspective on model-data surface temperature comparison at the Last Glacial Maximum, *Quaternary Science Reviews*, *107*, 1–10.

Armour, K. C., C. M. Bitz, and G. H. Roe (2013), Time-Varying Climate Sensitivity from Regional Feedbacks, *J. Climate*, *26*(13), 4518–4534.

Arrhenius, S. (1896), On the influence of carbonic acid in the air upon the temperature of the ground , *Philosophical Magazine Series 5*, *41*(251), 237–276.

Berner, R. A., and Z. Kothavala (2001), Geocarb III: A Revised Model of Atmospheric CO2 over Phanerozoic Time, *American Journal of Science*, *301*(2), 182–204.

Betts, A. K., and Harshvardhan (1987), Thermodynamic Constraint on the Cloud Liquid Water Feedback in Climate Models, *J Geophys Res-Atmos*, *92*(D7), 8483–8485.

Bijl, P. K., A. J. P. Houben, S. Schouten, S. M. Bohaty, A. Sluijs, G.-J. Reichart, J. S. S. Damste, and H. Brinkhuis (2010), Transient Middle Eocene Atmospheric CO_2 and Temperature Variations, *Science*, *330*(6005), 819–821.

Bony, S., B. Stevens, I. H. Held, J. F. Mitchell, J.-L. Dufresne, K. A. Emanuel, P. Friedlingstein, S. Griffies, and C. Senior (2013), Carbon Dioxide and Climate: Perspectives on a Scientific Assessment, in *Climate Science for Serving Society*, edited by J. W. Hurrell and G. Asrar, pp. 391–413, Springer Netherlands, Dordrecht.

Bony, S., B. Stevens, D. M. W. Frierson, C. Jakob, M. Kageyama, R. Pincus, T. G. Shepherd, S. C. Sherwood, A. P. Siebesma, A. H. Sobel, M. Watanabe, and M. J. Webb (2015), Clouds, circulation and climate sensitivity, *Nature Geoscience*, *8*(4), 261–268.

Bony, S., B. Stevens, D. Coppin, T. Becker, K. A. Reed, A. Voigt, and B. Medeiros (2016), Thermodynamic control of anvil cloud amount, *Proc Natl Acad Sci USA*, p. 201601472.

Boucher, O., D. Randall, P. Artaxo, C. Bretherton, G. Feingold, P. Forster, V.-M. Kerminen, Y. Kondo, H. Liao, U. Lohmann, P. Rasch, S. K. Satheesh, S. Sherwood, B. Stevens, and X. Y. Zhang (2013), Clouds and Aerosols, 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*, edited by T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, pp. 571–657, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Bretherton, C. S. (2015), Insights into low-latitude cloud feedbacks from high-resolution models, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2054), 20140,415.

Bretherton, C. S., P. N. Blossey, and C. R. Jones (2013), Mechanisms of marine low cloud sensitivity to idealized climate perturbations: A single-LES exploration extending the CGILS cases, *J. Adv. Model. Earth Syst.*, *5*(2), 316–337.

Brient, F., and S. Bony (2013), Interpretation of the positive low-cloud feedback predicted by a climate model under global warming, *Climate Dynamics*, *40*(9-10), 2415–2431.

Brient, F., T. Schneider, Z. Tan, S. Bony, X. Qu, and A. Hall (2015), Shallowness of tropical low clouds as a predictor of climate models' response to warming, *Climate Dynamics*, *47*(1-2), 433–449.

Charney, J. G., A. Arakawa, D. J. Baker, B. Bolin, R. E. Dickinson, R. M. Goody, C. E. Leith, H. M. Stommel, and C. I. Wunsch (1979), Carbon Dioxide and Climate: A Scientific Assessment, *Tech. rep.*, National Academy of Sciences, Washington DC.

Clement, A. C., R. Burgman, and J. R. Norris (2009), Observational and Model Evidence for Positive Low-Level Cloud Feedback, *Science*, *325*(5939), 460–464.

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, Commitments 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*, edited by T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, pp. 1029–1136,

Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Cooke, R., B. A. Wielicki, D. F. Young, and M. G. Mlynczak (2013), Value of information for climate observing systems, *Environ Syst Decis*, *34*(1), 98–109.

Covey, C., L. C. Sloan, and M. I. Hoffert (1996), Paleoclimate data constraints on climate sensitivity: The paleocalibration method, *Climatic Change*, *32*(2), 165–184.

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. (2013), Observations of Climate Feedbacks over 2000-10 and Comparisons to Climate Models, *J. Climate*, *26*(1), 333–342.

Flato, G., J. Marotzke, B. Abiodun, P. Braconnot, S. C. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov, C. Reason, and M. Rummukainen (2013), Evaluation of Climate Models, 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*, edited by T. F. Stocker, D. Qin, G. K. Plattner, M. Tingor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, pp. 741–866, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., Cambridge.

Forster, P. M. (2016), Inference of Climate Sensitivity from Analysis of Earth's Energy Budget, *Annu. Rev. Earth Planet. Sci.*, *44*(1), 85–106.

Geoffroy, O., D. Saint-Martin, D. J. L. Olivié, A. Voldoire, G. Bellon, and S. Tyteca (2013), Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution and Parameter Calibration Using CMIP5 AOGCM Experiments, *J. Climate*, *26*(6), 1841–1857.

Gettelman, A., J. E. Kay, and K. M. Shell (2012), The Evolution of Climate Sensitivity and Climate Feedbacks in the Community Atmosphere Model, *J. Climate*, *25*(5), 1453–1469.

Gordon, N. D., and S. A. Klein (2014), Low-cloud optical depth feedback in climate models, *J Geophys Res-Atmos*, *119*(10), 6052–6065.

Gregory, J., and M. Webb (2008), Tropospheric Adjustment Induces a Cloud Component in CO2 Forcing, *J. Climate*, *21*(1), 58–71.

Gregory, J. M., T. Andrews, and P. Good (2015), The inconstancy of the transient climate response parameter under increasing CO_2, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, pp. 1–25.

Gregory, J. M., T. Andrews (2016), Variation in climate sensitivity and feedback parametersduring the historical period, *Geophys. Res. Lett.*, *43*, 39113920

Hargreaves, J. C., J. D. Annan, M. Yoshimori, and A. Abe-Ouchi (2012), Can the Last Glacial Maximum constrain climate sensitivity?, *Geophys. Res. Lett.*, *39*(24), n/a–n/a.

Hartmann, D. L., and K. Larson (2002), An important constraint on tropical cloud - climate feedback, *Geophys. Res. Lett.*, *29*(20), 1951.

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.

Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis (2010), Probing the Fast and Slow Components of Global Warming by Returning Abruptly to Preindustrial Forcing, *J. Climate*, *23*(9), 2418–2427.

Hopcroft, P. O., and P. J. Valdes (2015), How well do simulated last glacial maximum tropical temperatures constrain equilibrium climate sensitivity?, *Geophys. Res. Lett.*, *42*(13), 5533–5539.

Huber, M., U. Beyerle, and R. Knutti (2014), Estimating climate sensitivity and future temperature in the presence of natural climate variability, *Geophys. Res. Lett.*, *41*(6), 2086–2092.

Kang, S. M., I. M. Held, and S.-P. Xie (2013), Contrasting the tropical responses to zonally asymmetric extratropical and tropical thermal forcing, *Climate Dynamics*, *42*(7-8), 2033–2043.

Kay, J. E., B. Medeiros, Y. T. Hwang, A. Gettelman, J. Perket, and M. G. Flanner (2014), Processes controlling Southern Ocean shortwave climate feedbacks in CESM, *Geophys. Res. Lett.*, *41*(2), 616–622.

Knutti, R., and G. C. Hegerl (2008), The equilibrium sensitivity of the Earth's temperature to radiation changes, *Nature Geoscience*, *1*(11), 735–743.

Knutti, R., and M. A. A. Rugenstein (2015), Feedbacks, climate sensitivity and the limits of linear models, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2054), 20150,146.

Kuang, Z., and D. L. Hartmann (2007), Testing the Fixed Anvil Temperature Hypothesis in a Cloud-Resolving Model, *J. Climate*, *20*(10), 2051–2057.

Lewis, N., and J. A. Curry (2014), The implications for climate sensitivity of AR5 forcing and heat uptake estimates, *Climate Dynamics*, pp. 1–15.

Li, C., J.-S. Storch, and J. Marotzke (2012), Deep-ocean heat uptake and equilibrium climate response, *Climate Dynamics*.

Lindzen, R., M. D. Chou, and A. Y. Hou (2001), Does the earth have an adaptive infrared iris?, *Bulletin of the American Meteorological Society*, *83*(2), 417–432.

Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong (2009), Toward Optimal Closure of the Earth's Top-of-Atmosphere Radiation Budget, *J. Climate*, *22*(3), 748–766.

Lüthi, D., M. Le Floch, B. Bereiter, T. Blunier, J.-M. Barnola, U. Siegenthaler, D. Raynaud, J. Jouzel, H. Fischer, K. Kawamura, and T. F. Stocker (2008), High-resolution carbon dioxide concentration record 650,000–800,000?years before present, *Nature*, *453*(7193), 379–382.

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*.

Meraner, K., T. Mauritsen, and A. Voigt (2013), Robust increase in equilibrium climate sensitivity under global warming, *Geophys. Res. Lett.*, *40*(22), 5944–5948.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, *Journal of Geophysical Research: Atmospheres (1984–2012)*, *117*(D8), D08,101.

Neubersch, D., H. Held, and A. Otto (2014), Operationalizing climate targets under learning: An application of cost-risk analysis, *Climatic Change*, *126*(3-4), 305–318.

Nuijens, L., B. Medeiros, I. Sandu, and M. Ahlgrimm (2015), The behavior of trade-wind cloudiness in observations and models: The major cloud components and their variability, *J. Adv. Model. Earth Syst.*, *7*(2), 600–616.

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.

Pagani, M., K. Caldeira, D. Archer, and J. C. Zachos (2006), An Ancient Carbon Mystery, *Science*, *314*(5805), 1556–1557.

Pagani, M., M. Huber, and B. Sageman (2014), Greenhouse Climates, in *Treatise on Geochemistry*, pp. 281–304, Elsevier.

Pierrehumbert, R. T., D. S. Abbot, A. Voigt, and D. Koll (2011), Climate of the Neoproterozoic, *Annu. Rev. Earth Planet. Sci.*, *39*(1), 417–460.

Qu, X., A. Hall, S. A. Klein, and P. M. Caldwell (2013), On the spread of changes in marine low cloud cover in climate model simulations of the 21st century, *Climate Dynamics*, *42*(9-10), 2603–2626.

Rieck, M., L. Nuijens, and B. Stevens (2012), Marine Boundary Layer Cloud Feedbacks in a Constant Relative Humidity Atmosphere, *J. Atmos. Sci*, *69*(8), 2538–2550.

Rohling, E. J., A. Sluijs, H. A. Dijkstra, P. Köhler, R. S. W. van de Wal, A. S. von der Heydt, D. J. Beerling, A. Berger, P. K. Bijl, M. Crucifix, R. DeConto, S. S. Drijfhout, A. Fedorov, G. L. Foster, A. Ganopolski, J. Hansen, B. Hönisch, H. Hooghiemstra, M. Huber, P. Huybers, R. Knutti, D. W. Lea, L. J. Lourens, D. Lunt, V. Masson-Demotte, M. Medina-Elizalde, B. Otto-Bliesner, M. Pagani, H. Pälike, H. Renssen, D. L. Royer, M. Siddall, P. Valdes, J. C. Zachos, and R. E. Zeebe (2012), Making sense of palaeoclimate sensitivity, *Nature*, *491*(7426), 683–691.

Royer, D. L., R. A. Berner, and J. Park (2007), Climate sensitivity constrained by CO_2 concentrations from the last 420 million years, *Nature*, *446*, 530–532.

Schmidt, G. A., J. D. Annan, P. J. Bartlein, B. I. Cook, E. Guilyardi, J. C. Hargreaves, S. P. Harrison, M. Kageyama, A. N. LeGrande, B. Konecky, S. Lovejoy, M. E. Mann, V. Masson-Delmotte, C. Risi, D. Thompson, A. Timmermann, L. B. Tremblay, and P. Yiou (2014), Using palaeo-climate comparisons to constrain future projections in CMIP5, *Clim. Past*, *10*(1), 221–250.

Schmittner, A., N. M. Urban, J. D. Shakun, N. M. Mahowald, P. U. Clark, P. J. Bartlein, A. C. Mix, and A. Rosell-Mele (2011), Climate Sensitivity Estimated from Temperature Reconstructions of the Last Glacial Maximum, *Science*, *334*(6061), 1385–1388.

Seneviratne, S. I., M. G. Donat, A. J. Pitman, R. Knutti, and R. L. Wilby (2016), Allowable CO2 emissions based on regional and impact-related climate targets, *Nature*, *529*(7587), 477–483.

Senior, C. A., and J. F. B. Mitchell (2000), The time-dependence of climate sensitivity, *Geophys. Res. Lett.*, *27*(17), 2685–2688.

Sherwood, S. C., W. Ingram, Y. Tsushima, M. Satoh, M. Roberts, P. L. Vidale, and P. A. O'Gorman (2010), Relative humidity changes in a warmer climate, *J Geophys Res-Atmos*, *115*(D9), D09,104.

Sherwood, S. C., S. Bony, and J.-L. Dufresne (2014), Spread in model climate sensitivity traced to atmospheric convective mixing, *Nature*, *505*(7481), 37–42.

Sherwood, S. C., S. Bony, O. Boucher, C. Bretherton, P. M. Forster, J. M. Gregory, and B. Stevens (2015), Adjustments in the Forcing-Feedback Framework for Understanding Climate Change, *Bulletin of the American Meteorological Society*, *96*(2), 217–228.

Stevens, B. (2015), Rethinking the Lower Bound on Aerosol Radiative Forcing, *J. Climate*, *28*(12), 4794–4819.

Stevens, B., and S. Bony (2013), Water in the atmosphere, *Physics Today*, *66*(6), 29.

Stevens, B., and J.-L. Brenguier (2009), Cloud-controlling Factors: Low Clouds, in *Clouds in the Perturbed Climate System*, edited by J. Heintzenberg and R. J. Charlson, pp. 1–24, MIT Press.

Stevens, B., A. Abe-Ouchi, S. Bony, G. C. Hegerl, G. A. Schmidt, S. C. Sherwood, and M. J. Webb (2015), Ringberg15: Earth's Climate Sensitivities, *Tech. Rep. 11*, World Climate Research Programme, Geneva.

Stocker, T. F., D. Qin, G. K. Plattner, M. Tingor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (2013), *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*, vol. 413, Cambrdige Univ Press, Berlin, Heidelberg.

Su, H., J. H. Jiang, C. Zhai, T. J. Shen, J. D. Neelin, G. L. Stephens, and Y. L. Yung (2014), Weakening and strengthening structures in the Hadley Circulation change under global warming and implications for cloud response and climate sensitivity, *Journal of Geophysical Research: Atmospheres*, *119*(10), 5787–5805.

Tebaldi, C., and J. M. Arblaster (2014), Pattern scaling: Its strengths and limitations, and an update on the latest model simulations, *Climatic Change*, *122*(3), 459–471.

Tripati, A. K., S. Sahany, D. Pittman, R. A. Eagle, J. D. Neelin, J. L. Mitchell, and L. Beaufort (2014), Modern and glacial tropical snowlines controlled by sea surface temperature and atmospheric mixing, *Nature Geoscience*, *7*(3), 205–209.

Vial, J., J.-L. Dufresne, and S. Bony (2013), On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates, *Climate Dynamics*, *41*(11-12), 3339–3362.

Waelbroeck, C., A. Paul, M. Kucera, A. Rosell-Mele, M. Weinelt, R. Schneider, A. C. Mix, A. Abelmann, L. Armand, E. Bard, S. Barker, T. T. Barrows, H. Benway, I. Cacho, M. T. Chen, E. Cortijo, X. Crosta, A. de Vernal, T. Dokken, J. Duprat, H. Elderfield, F. Eynaud, R. Gersonde, A. Hayes, M. Henry, C. Hillaire-Marcel, C. C. Huang, E. Jansen, S. Juggins, N. Kallel, T. Kiefer, M. Kienast, L. Labeyrie, H. Leclaire, L. Londeix, S. Mangin, J. Matthiessen, F. Marret, M. Meland, A. E. Morey, S. Mulitza, U. Pflaumann, N. G. Pisias, T. Radi, A. Rochon, E. J. Rohling, L. Sbaffi, C. Schäfer-Neth, S. Solignac, H. Spero, K. Tachikawa, and J. L. Turon (2009), Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum, *Nature Geoscience*, *2*(2), 127–132.

Webb, M. J., and A. P. Lock (2012), Coupling between subtropical cloud feedback and the local hydrological cycle in a climate model, *Climate Dynamics*, *41*(7-8), 1923–1939.

Webb, M. J., A. P. Lock, C. S. Bretherton, S. Bony, J. N. S. Cole, A. Idelkadi, S. M. Kang, T. Koshiro, H. Kawai, T. Ogura, R. Roehrig, Y. Shin, T. Mauritsen, S. C. Sherwood, J. Vial, M. Watanabe, M. D. Woelfle, and M. Zhao (2015), The impact of parametrized convection on cloud feedback, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2054), 20140,414.

Winton, M., K. Takahashi, and I. M. Held (2010), Importance of Ocean Heat Uptake Efficacy to Transient Climate Change, *J. Climate*, *23*(9), 2333–2344.

Zelinka, M. D., and D. L. Hartmann (2011), The observed sensitivity of high clouds to mean surface temperature anomalies in the tropics, *J Geophys Res-Atmos*, *116*(D23).

Zelinka, M. D., S. A. Klein, K. E. Taylor, T. Andrews, M. J. Webb, J. M. Gregory, and P. M. Forster (2013), Contributions of Different Cloud Types to Feedbacks and Rapid Adjustments in CMIP5*, *J. Climate*, *26*(14), 5007–5027.

Zelinka, M. D., C. Zhou, and S. A. Klein (2016), Insights from a refined decomposition of cloud feedbacks, *Geophys. Res. Lett.*, pp. 1–11.

Zhang, M., C. S. Bretherton, P. N. Blossey, P. H. Austin, J. T. Bacmeister, S. Bony, F. Brient, S. K. Cheedela, A. Cheng, A. D. Genio, S. R. Roode, S. Endo, C. N. Franklin, J.-C. Golaz, C. Hannay, T. Heus, F. A. Isotta, J.-L. Dufresne, I. S. Kang, H. Kawai, M. Köhler, V. E. Larson, Y. Liu, A. P. Lock, U. Lohmann, M. F. Khairoutdinov, A. M. Molod, R. A. J. Neggers, P. Rasch, I. Sandu, R. Senkbeil, A. P. Siebesma, C. Siegenthaler Le Drian, B. Stevens, M. J. Suarez, K. M. Xu, K. Salzen, M. J. Webb, A. Wolf, and M. Zhao (2013), CGILS: Results from the first phase of an international project to understand the physical mechanisms of low cloud feedbacks in single column models, *J. Adv. Model. Earth Syst.*, *5*(4), 826–842.

Zhao, M. (2014), An Investigation of the Connections among Convection, Clouds, and Climate Sensitivity in a Global Climate Model, *J. Climate*, *27*(5), 1845–1862.

Zhou, C., M. D. Zelinka, and A. E. Dessler (2015), The relationship between interannual and long-term cloud feedbacks, *Geophys. Res. Lett.*, *42*.

Study	ECS 17-83 % CI (K)	5-95 % CI (K)
Aldrin et al. [2012]	1.2–2.0	1.1–2.6
Otto et al. [2013]	1.5–2.8	1.2–3.9
Lewis and Curry [2014]	1.3–2.5	1.1–4.1
Lewis and Curry [2014], revised	1.2–1.8	1.1–2.2

Table S1. ECS estimates inferred by simple models constrained by the instrumental record. The revised version of the Lewis and Curry estimate employs the aerosol-forcing bounds of *Stevens* [2015].

	Low ECS			High ECS		
Condition	(ii)	(iii)	(iv)	(i)	(ii)	(iii)
ε_j	0.25	0.35	0.20	0.75	0.65	0.60
χ_j [K]	1.50	1.50	2.00	4.00	4.50	4.50

Table S2. Parameters for likelihood functions for applicable conditions of a low ($< 1.5\,\mathrm{K}$) and high ($> 4.5\,\mathrm{K}$) ECS.

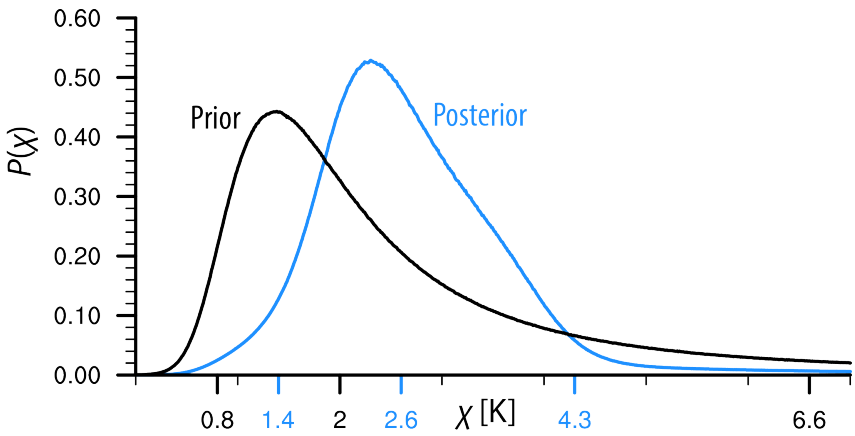

Figure S1. The prior (black) and posterior (blue) distributions with 5, 50 and 95 percentiles of the respective distributions indicated by major tick marks for the case of a broader prior ($\sigma_\lambda = 0.75\lambda$ and $\sigma_F = 0.25F$.)

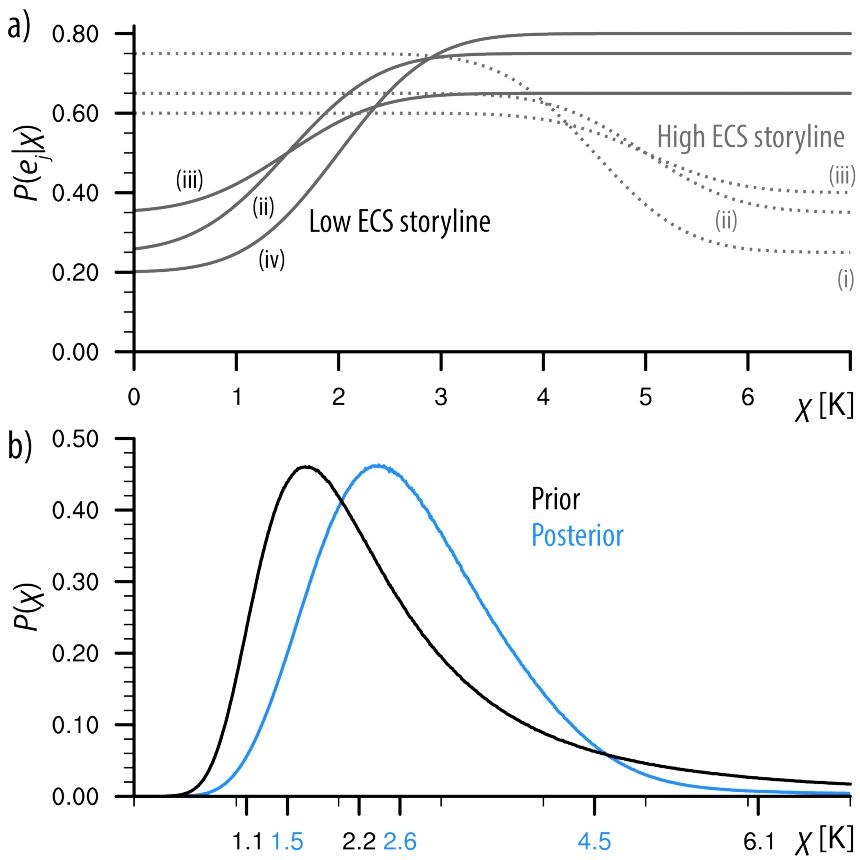

Figure S2. For comparison with Fig. 1. A case where to account for a larger than anticipated pattern effect, the values of χ_j are increased by 0.5 K for all the lines of evidence associated with the storyline for a very high ECS and the uncertainty in all values of χ_j is increased by a factor of 2.

Experiment	$P(\chi < 1.5)$	$P(\chi < 2.0)$	$P(\chi > 4.0)$	$P(\chi > 4.5)$	$P(\chi > 6.0)$
Prior	0.19	0.41	0.15	0.11	0.05
Broad Prior	0.31	0.50	0.16	0.13	0.07
Posterior P	0.03	0.18	0.06	0.03	0.01
Posterior P1	0.06	0.21	0.11	0.05	0.01
Posterior P2	0.21	0.46	0.06	0.04	0.02
Posterior P3	0.03	0.17	0.10	0.06	0.02
Posterior BP	0.06	0.21	0.08	0.04	0.02

Table S3. Probability for the ECS, χ, to lie in a specific interval. Shown are the distributions for the priors and posterior for four experiments: Example in manuscript (P); case for χ_j increased by 0.5 K for all lines of evidence in the high-ECS storyline and uncertainty in χ_j increased by a factor of two for all lines of evidence (P1); case where only the line of evidence from the instrumental record is adopted (P2); case where the historical evidence is assumed to be much weaker, so that $\alpha_j = 0.4$ instead of 0.25, (P3); same as case P but for the broader prior (BP).