



Objectively combining climate sensitivity evidence

Nicholas Lewis¹

Received: 17 February 2022 / Accepted: 11 August 2022 / Published online: 19 September 2022
© The Author(s) 2022

Abstract

Recent assessments of climate sensitivity per doubling of atmospheric CO₂ concentration have combined likelihoods derived from multiple lines of evidence. These assessments were very influential in the Intergovernmental Panel on Climate Change Sixth Assessment Report (AR6) assessment of equilibrium climate sensitivity, the *likely* range lower limit of which was raised to 2.5 °C (from 1.5 °C previously). This study evaluates the methodology of and results from a particularly influential assessment of climate sensitivity that combined multiple lines of evidence, Sherwood et al. (Rev Geophys 58(4):e2019RG000678, 2020). That assessment used a subjective Bayesian statistical method, with an investigator-selected prior distribution. This study estimates climate sensitivity using an Objective Bayesian method with computed, mathematical priors, since subjective Bayesian methods may produce uncertainty ranges that poorly match confidence intervals. Identical model equations and, initially, identical input values to those in Sherwood et al. are used. This study corrects Sherwood et al.'s likelihood estimation, producing estimates from three methods that agree closely with each other, but differ from those that they derived. Finally, the selection of input values is revisited, where appropriate adopting values based on more recent evidence or that otherwise appear better justified. The resulting estimates of long-term climate sensitivity are much lower and better constrained (median 2.16 °C, 17–83% range 1.75–2.7 °C, 5–95% range 1.55–3.2 °C) than in Sherwood et al. and in AR6 (central value 3 °C, *very likely* range 2.0–5.0 °C). This sensitivity to the assumptions employed implies that climate sensitivity remains difficult to ascertain, and that values between 1.5 °C and 2 °C are quite plausible.

1 Introduction

The Earth's climate sensitivity is a key measure of the longer-term climate response to external forcing. It is perhaps the most important ill-quantified climate system parameter. In principle, climate sensitivity represents the equilibrium change in mean surface temperature to a doubling of atmospheric CO₂ concentration from preindustrial levels, once the deep ocean has reached a stable state. In practice it is normally estimated using some approximate measure, often derived from disequilibrium changes. Climate sensitivity has been estimated from various types of evidence, but none of these has narrowly constrained its value. The first five Assessment Reports by the Intergovernmental Panel on Climate Change (IPCC) relied heavily on estimates of climate sensitivity from global climate model (GCM) simulations. The 1.5–4.5 K *likely* range for climate sensitivity in the 2013 IPCC Fifth Assessment Report (AR5) was

identical to the range presented in the landmark Charney (1979) report, with the great increase in GCM sophistication since 1979 not having led to any narrowing of the climate sensitivity range.

GCMs use semi-empirical approximations (parameterizations) to represent subgrid-scale cloud and convection processes that are known to be critical to determining the model's climate sensitivity, which varies by up to a factor of three among GCMs. In one well regarded GCM, a simple change to how convective precipitation was parameterized¹ varied its climate sensitivity by a factor of two, with no obvious change in how well the model otherwise performed (Zhao et al. 2016). Changing the order in which the various parameterized atmospheric modules were updated in each time step was found to vary another GCM's climate sensitivity by a factor of up to two, with ambiguity existing regarding the optimum ordering (Donahue and Caldwell 2018). Moreover, the universal use in GCMs of deterministic parameterizations may bias their climate sensitivity upwards

✉ Nicholas Lewis
nhlewis@btinternet.com

¹ Bath, UK

¹ Changing between a threshold convective scheme in which all convective condensate exceeding a threshold value is converted to precipitation, and a fractional removal scheme, in which only a fraction of such condensate is removed as precipitation.

(Strommen et al. 2019). Such issues make the reliability of GCM-derived estimates of climate sensitivity questionable.

In the light of such issues, and the further widening of the range of GCM climate sensitivities in the latest (CMIP6) generation of GCMs (Zelinka et al. 2020), the IPCC Sixth Assessment Report (AR6) abandoned the previous reliance on GCM climate sensitivities. Instead, evaluation of climate sensitivity was approached by combining estimates based on different lines of evidence, such as process understanding (feedback analysis), the historical instrumental record, and paleoclimate data.

Combining different lines of evidence should, to the extent that they are independent, enable climate sensitivity to be estimated more precisely than from any single line of evidence (Stevens et al. 2016). A comprehensive attempt to do so was made by Sherwood et al. (2020, henceforth S20), a 92-page study. S20 was conducted under the auspices of the World Climate Research Programme's Grand Science Challenge on Clouds, Circulation and Climate Sensitivity and provides a very detailed investigation of climate sensitivity. As the most influential recent assessment, S20 was cited over twenty times in the relevant AR6 chapter, which approached climate sensitivity estimation on very similar lines to S20, albeit not using its formal probabilistic methods. There are in principle considerable strengths in S20's scientific approach. Its main results were derived by combining understanding from feedback analysis (Process evidence) with evidence from changes since circa 1850 (Historical evidence), and from cold and warm past periods (Paleoclimate evidence)—three lines of evidence that S20 judged to be largely independent.

The contribution the present study makes to estimation of climate sensitivity is three-fold. First, it identifies statistical problems in S20. The main methodological argument is that, when Bayesian methods are used, an Objective rather than a Subjective Bayesian approach should be taken. This means that rather than the investigator choosing the prior distribution, the prior distribution should be mathematically computed, based on the assumed statistical model relating to all the evidence to be analyzed (Bernardo 2009). S20 used a Subjective Bayesian statistical method, with an investigator-selected prior distribution, that has been shown may produce unrealistic climate sensitivity estimation when used to combine differing types of evidence (Lewis 2018), and S20 provided no evidence that it did not do so in this case. Moreover, for all except Process evidence, S20 used a method of estimating likelihoods that turns out to be unsound. This study validates its likelihood estimates by using multiple methods and cross-checking their results. S20's method is shown to often result in serious likelihood underestimation at higher climate sensitivity levels.

The second contribution of this study is that it develops and applies an Objective Bayesian approach to combining

differing climate sensitivity evidence, using a mathematically computed prior distribution. The results using the methodology developed and the same input assumptions as S20 are then used to assess what effect the statistical problems identified in S20 have on its results. It is found that they bias S20's estimation of climate sensitivity downwards, although only to a minor extent even at the upper uncertainty bound when all three lines of evidence are combined.

This study's third contribution is to review and where appropriate revise the input assumptions used by S20, paying particular regard to more recent evidence, and to investigate the effect of the revised input assumptions on estimates of climate sensitivity using the developed Objective Bayesian methodology. Some of the revisions to input assumptions relate to the treatment in certain cases of CO₂ forcing and/or the warming it causes. This study differs from S20 regarding the appropriate scaling of CO₂ forcing, and comparison of warming, where different changes in CO₂ atmospheric concentrations are involved, and regarding scaling CO₂ forcing where its use requires a different estimation basis from that on which the forcing estimate was derived. The combined effects of the revisions to S20's CO₂ related estimates and to other input assumptions result in a major reduction in estimated climate sensitivity.

The paper is structured as follows. Climate sensitivity measures are discussed in Sect. 2. Section 3 deals with statistical methods, while Sect. 4 reviews S20's input assumptions and proposes certain revisions to them. Section 5 sets out results based on S20's original input assumptions but using the corrected likelihood estimates, using alternatively S20's chosen Subjective Bayesian prior distribution or the Objective Bayesian (mathematically computed) prior distributions. Section 6 presents results using the revised input assumptions and Objective Bayesian prior distributions. Section 7 discusses the statistical issues and both their effects and the effects of using different input assumptions.

2 Climate sensitivity measures

The traditional measure of climate sensitivity is the equilibrium change in global mean surface temperature (GMST) following a doubling of the atmospheric CO₂ concentration (equilibrium climate sensitivity, henceforth ECS). While the equilibrium involved allows for the deep ocean to reach a steady state, it excludes changes in slow components (e.g., ice sheets). Such an equilibrium is achievable in a GCM but not in the real climate system. The corresponding equilibrium change over timescales that allow for feedbacks from changes in the slow components to occur is called Earth system sensitivity (ESS).

Under the standard linear forcing-feedback framework, the excess of a change ΔF in effective radiative forcing (ERF) over the change in top-of-atmosphere (TOA) planetary radiative imbalance, ΔN , is equal and opposite to the climate system's radiative response ΔR , measuring all radiation downwards. ERF is a measure of the increase in TOA radiative imbalance resulting from a change in atmospheric composition, such as an increase in CO_2 concentration, with surface temperature held constant but the atmosphere allowed to adjust to the change. ΔR is taken to be the product of the change ΔT in global mean near-surface air temperature (GMAT), or in GMST, and a fixed climate feedback parameter λ^{fixed} . Accordingly:

$$(\Delta F - \Delta N) = -\Delta R = -\lambda^{\text{fixed}} \Delta T \quad (1)$$

Under this framework, it follows that for $\Delta N = 0$, representing equilibrium:

$$\Delta T = -\Delta F / \lambda^{\text{fixed}} \quad (2)$$

and hence for the ERF from a doubling of CO_2 concentration, $F_{2\times\text{CO}_2}$:

$$\text{ECS} = -F_{2\times\text{CO}_2} / \lambda^{\text{fixed}} \quad (3)$$

However, in GCMs (and the real climate system) the climate feedback parameter may not in fact be fixed, in which case a linear projection to $\Delta N = 0$ will not provide an accurate estimate of ECS.

Rather than seeking to estimate ECS, S20 instead estimate an effective climate sensitivity S , that corresponds to the effective sensitivity in GCMs derived from feedbacks occurring during the first 150 simulation years after an abrupt quadrupling of CO_2 concentration from its preindustrial level (abrupt4x CO_2), treating climate feedbacks as being fixed.

For GCMs, S is normally derived by linearly regressing, usually using annual average values, changes (ΔN) in TOA radiative imbalance on changes (ΔT) in GMAT over abrupt4x CO_2 simulations, with changes being measured relative to values during an unforced preindustrial control simulation by the GCM. The ΔT and ΔN values are rescaled to reflect the ratio of $F_{2\times\text{CO}_2}$ to the ERF from a quadrupling of CO_2 , $F_{4\times\text{CO}_2}$. The slope of the regression line, λ , is a measure of the effective climate feedback parameter operating over the regression period. The regression line is continued forwards to $\Delta N = 0$, indicating radiative equilibrium, with S being defined as the ΔT value at that point, and backwards to $\Delta T = 0$, the rescaled ΔN value at that point providing an estimate, $F_{2\times\text{CO}_2}^{\text{regress}}$, of $F_{2\times\text{CO}_2}$. Hence:

$$S = -F_{2\times\text{CO}_2}^{\text{regress}} / \lambda \quad (4)$$

If the climate feedback parameter is not fixed, in general λ differs from $-F_{2\times\text{CO}_2}/\text{ECS}$, S differs from ECS, and $F_{2\times\text{CO}_2}^{\text{regress}}$ differs from $F_{2\times\text{CO}_2}$.

In the vast majority of GCMs, the local slope of the relationship between ΔN and ΔT weakens over the course of 150-year abrupt4x CO_2 simulations, strongly suggesting that the model ECS exceeds S . Since feedbacks activated only on a long timescale affect the climate extremely slowly, S is more relevant than ECS (or ESS) to climate change over the next few centuries. However, deriving S from paleoclimate evidence, which reflects equilibrium changes, requires an estimate of the ECS to S ratio, with its excess over one being defined as $\zeta = \text{ECS}/S - 1$.

To obtain a valid estimate for climate sensitivity to doubled CO_2 concentration from data involving a different change in CO_2 concentration, it is necessary to scale the temperature change involved by the ratio of $\Delta F_{2\times\text{CO}_2}$ to the ERF associated with the particular change in CO_2 concentration, even assuming that climate sensitivity is unaffected by the effect of the difference in CO_2 concentration change on the climate state. S20 define their S in GCMs as the linearly regressed warming over years 1–150 after a quadrupling of CO_2 concentration, extrapolated to zero ΔN and then divided by two. This scaling factor, while popular, is difficult to justify when the actual ratio of the ERF change involved ($F_{4\times\text{CO}_2}$) to $F_{2\times\text{CO}_2}$ has been estimated with reasonable precision to be 2.10, 5% greater than twice that from doubled CO_2 concentration (Byrne and Goldblatt 2014; Etminan et al. 2016; Meinshausen et al. 2020).

S20 defend their division of abrupt4x CO_2 temperature changes by 2 (rather than 2.10) on the basis that it brings S estimated on their basis closer to estimated ECS in models with very long abrupt2x CO_2 simulations, which they estimate as 6% higher than S derived by halving temperature changes in those models' abrupt4x CO_2 simulations, implying $\zeta = 0.06$. However, that argument conflicts with their valid desire for a measure that is as closely related as possible to scenarios of practical relevance. Moreover, the increase in S that S20 introduce by basing it on a biased scaling of $F_{4\times\text{CO}_2}$ to $F_{2\times\text{CO}_2}$ results in inconsistent estimation of S between their three lines of evidence, a serious flaw. The biased scaling only affects (via the resulting ζ estimate) their estimation of S from Paleoclimate evidence, since its estimation from both Process and Historical evidence is based on directly estimated $F_{2\times\text{CO}_2}$, and is independent of scaled $F_{4\times\text{CO}_2}$.

Here, S20's one-half scaling factor and the resulting 0.06 central ζ estimate for $\text{ECS}/S - 1$ is retained when investigating the effect of the objective Bayesian statistical method on their results. However, in Sect. 4 it is revised to 0.135, the mean ζ estimate in both abrupt2x CO_2 and abrupt4x CO_2 long-run simulations (16 in all) by eleven GCMs (Rugenstein et al. 2020). No scaling from $F_{4\times\text{CO}_2}$ to $F_{2\times\text{CO}_2}$ is

required for these calculations, since temperature changes are only compared within abrupt4xCO₂ and abrupt2xCO₂ simulations, not between them.

3 Methods

Scientific knowledge regarding the properties of any real-world system, or of a simplified conceptual model used to represent it, emanates from observing aspects of the system behavior. The results of such observations or assessments based thereon ('data-variables') are typically numerical and somewhat uncertain, and are regarded as subject to random errors. Conceptual models of the system usually relate the data-variables used as input assumptions to system properties of interest that are regarded as fixed but unknown ('parameters'), assumed here to be represented by continuously-valued variables. A key role of statistical inference is then to draw valid conclusions from data-variables regarding such parameters, as regards their values and associated uncertainty.

It is essential for scientific inference that the statistical methods used are *calibrated*, in the sense that the uncertainty ranges they generate closely approximate confidence intervals. That is, over the long run (over many applications of the method to different data sets) the true parameter value will lie below a properly derived ($x\%$ – $y\%$) confidence interval in about $x\%$ of cases, and above it in about $(1-y)\%$ of cases.

The data likelihood, a joint function of data and parameter values, plays a central role in statistical parameter inference. It represents the joint probability density of the observed data as a function of the parameter value(s). Provided errors in the observed data are independent, their joint probability density is the product of that for each data-variable. Usually, only the ratio of the data likelihood to its highest value matters.

An important property of likelihood functions is that, where two likelihood functions concerning the same system and parameters but derived from independent data exist, the information they jointly contain about the parameter is representable by their product (Birnbaum 1962; Pawitan 2001 Sects. 2.3 and 7.2). This property is used when combining the three different lines of evidence.

3.1 Bayesian parameter inference

There are two main statistical paradigms, Frequentist and Bayesian (Bernardo and Smith 1994). In both, parameter inference revolves around likelihood functions. However, Bayesians treat fixed but uncertain parameters as having distributions representing degrees of belief, in effect as if random variables with a probability distribution, while

Frequentists do not, notwithstanding Frequentist confidence distributions (Schweder and Hjort 2002, 2016).

Both Stevens et al. (2016) and S20 employ Bayesian methods for combining climate sensitivity evidence; Frequentist methods appear less suitable for this task. Bayesian methods provide a means of coherently updating personal beliefs about an unknown parameter with external evidence as to its value. However, for continuously-valued parameters they do not in general provide calibrated estimates that properly reflect such evidence. By contrast, Frequentist confidence measures are derived from randomness in the data values and are intrinsically calibrated.

In the continuous case, from Bayes' theorem (Bayes 1763) the posterior probability density function (PDF), $p_{\theta}(\theta|y)$, for a parameter (vector) θ on which observed data y depend, is proportional to the data likelihood $p_y(y|\theta)$ (the probability density of the data treated as a function of θ , for fixed y) multiplied by the density of a 'prior distribution' (prior) for θ , $p_{\theta}(\theta)$:

$$p_{\theta}(\theta|y) = c p_y(y|\theta) p_{\theta}(\theta) \quad (5)$$

(the subscripts indicating the variable each density is for). The constant c is such that $p_{\theta}(\theta|y)$ integrates to unit probability; it is the reciprocal of $\int p_y(y|\theta) p_{\theta}(\theta) d\theta$. If the parameter being estimated were a random variable with actual probability distribution $p_{\theta}(\theta)$ then (5) would follow from the conditional probability lemma. However, this is not the case here (Fraser and Reid 2011).

The Bayesian equivalent of a confidence interval, a credible interval, reflects probability implied by the posterior PDF. Whether a credible interval is calibrated or not will among other things depend on the choice of prior, which the investigator is free to select. S20 select a prior that is uniform in λ -space, and therefore proportional to $F_{2 \times \text{CO}_2}/S^2$ in S -space. In the common 'Subjective Bayesian' view adopted by S20, the prior is a probability distribution representing the investigator's degrees of belief about parameter values before incorporating information from the current data. There is no requirement that the posterior PDF be calibrated, and the resulting credible intervals may be far from actual confidence intervals (Fraser 2011). However, to avoid Bayesian inference providing misleading results, it is necessary to use a prior that provides correct calibration of posterior probabilities to frequencies and hence confidence intervals (Fraser et al. 2010; Lewis and Grünwald 2018).

In the alternative 'Objective Bayesian' view, in the absence of existing evidence regarding parameter values the prior should consist of a mathematical weighting function intended to have minimal influence on inference relative to the data, so that it is as *noninformative* as possible.

Under this approach, any existing probabilistic evidence concerning the parameter being estimated may appropriately

be represented by a likelihood function for a notional observation, from which a posterior density has been calculated using a noninformative prior (Hartigan 1965), rather than by a posterior density.

Noninformative priors are mathematical priors that are generally intended to result in *probability matching* posterior PDFs, which produce credible intervals that are (at least approximately) true confidence intervals, and they are often judged on that basis (Berger and Bernardo 1992, p. 36; Kass and Wasserman 1996). Typically, a noninformative prior primarily reflects how the expected informativeness of the data about the parameter value(s) varies with the parameter value(s). How informative data are expected to be about the parameter value(s) is represented by the (expected) Fisher information,² which has a key role in likelihood-based inference.

The gold-standard method for producing noninformative priors is reference analysis (Bernardo 1979; Berger and Bernardo 1992), which results in Bayesian inference that is objective in the sense that it only depends on observed data and model assumptions, as is the case for Frequentist inference (Bernardo 2011). In Objective Bayesianism as well as Subjective Bayesianism, however, subjective choices will still be made by the investigator in relation to the data and model used.

Under Subjective Bayesianism, independent new evidence about a parameter is incorporated by updating an existing posterior PDF (treated as the prior) by multiplying it by the data likelihood function for the new evidence and renormalizing to unit total probability, as S20 does. Such updating would be valid mathematically were the parameter a random variable, but it is not. Bayesian updating satisfies Subjective Bayesian axioms (Bernardo and Smith 1994) and results in coherent personal beliefs, but that does not imply that the resulting inference will exhibit satisfactory probability matching even if the existing posterior PDF was well calibrated.

Bayesian updating is in any event unsupportable for Objective Bayesianism, since the noninformative prior for the original likelihood will in general differ from that for the new likelihood. Bayesian updating would therefore produce inference that varied with the order in which different evidence was incorporated (Kass and Wasserman 1996). It can also result in quite poor frequentist calibration (Lewis 2018).

The order-dependence problem does not arise if Bayes theorem is applied once only, to the joint likelihood function for the combined evidence, with a single noninformative prior being computed that reflects the nature of the combined evidence Lewis (2013a, b). This is the method employed in the present study. As shown in Lewis (2018) and Lewis and Grunwald (2018), using such a single-step method when combining climate sensitivity evidence results in more realistic inference than using Bayesian updating, even when a noninformative prior is used to incorporate the first line of evidence, although the magnitude of the improvement will vary.

Where a univariate parameter, such as climate sensitivity, is the only parameter being estimated, a Jeffreys' prior (Jeffreys 1946), which in that case is normally also the reference prior, gives credible intervals that match confidence intervals more closely than any other prior (Welch and Peers 1963; Hartigan 1965), and is therefore the most appropriate prior to use for weighting the combined evidence-providing data likelihood functions. A Jeffreys' prior is proportional (with arbitrary scaling) to the square root of the Fisher information (of its determinant for a multivariate parameter).

Fisher information for different likelihood functions combines additively, provided the likelihood functions are derived from independent data (Pawitan 2001, Sect. 8.4). Therefore, Jeffreys' prior for inference from the combined likelihood function is obtainable by adding in quadrature the Jeffreys' priors for the separate likelihood functions, after scaling each to equal the square root of the Fisher information. The probability matching of posterior PDFs derived by this method has been tested and found to be accurate in cases involving various probability distributions (Lewis 2013a, b), including in the context of combining evidence regarding climate sensitivity (Lewis 2018).

3.2 Statistical models

I use the same statistical models as S20. These derive from simple forcing-feedback physical models, for the various lines of evidence, as follows. Terms on the right-hand sides of (6) to (13) each represent the 'true' value of each variable, the best observational estimate of which is taken to include an additive error (ϵ) term, which has been omitted for clarity. These variables are termed "data-variables", as generally their estimated values are ultimately derived from observational data. Their estimated error characteristics are inputs to the statistical models.

For Process evidence, total climate feedback λ is taken as the sum of component feedbacks:

$$\lambda = \lambda_{\text{Planck}} + \lambda_{\text{WV}} + \lambda_{\text{LR}} + \lambda_{\text{sfcAlbedo}} + \lambda_{\text{clouds}} + \lambda_{\text{stratospheric}} + \lambda_{\text{atmosComp}} \quad (6)$$

² Fisher information (I), in the case of a univariate parameter θ and assuming regularity conditions, is the expected value of minus the second derivative with respect to θ of the log-likelihood function ($\log p(x | \theta)$) for the data (x), at a given value of θ : $I(\theta) = \int p(x|\theta) \left(-\frac{\partial^2 \log p(x|\theta)}{\partial \theta^2} \right) dx$; in the multivariate parameter case it is a matrix, sometimes just called the (expected) information matrix: $(I(\theta))_{ij} = \int p(x|\theta) \left(-\frac{\partial^2 \log p(x|\theta)}{\partial \theta_i \partial \theta_j} \right) dx$ (Bernardo and Smith, p. 288).

where λ_{Planck} is feedback from extra emission to space from vertically uniform warming, the anti-correlated water vapor (WV) and lapse rate (LR) feedbacks are combined into $\lambda_{\text{WV+LR}}$, $\lambda_{\text{sfcAlbedo}}$ is surface albedo feedback, λ_{clouds} is cloud feedback, $\lambda_{\text{stratospheric}}$ is feedback from changes in stratospheric water vapor and temperature, and $\lambda_{\text{atmosComp}}$ is feedback from changes in atmospheric composition. The error/uncertainty term (ϵ_λ) for total feedback λ represents the sum of independent error terms for its components, with λ_{clouds} being likewise the sum of components, each subject to independent errors. Since errors are independent and are assumed to be normally distributed, ϵ_λ is also normally distributed, with variance equal to the sum of the variances of its components.

Provided that the estimates of the climate feedback components are of the values they take over 150 year abrupt4xCO₂ GCM simulations, which they all are, the resulting climate feedback estimate will be on a basis consistent with its derivation over such simulations, and hence be of λ . As a result, S can be derived from this feedback estimate, using (4), without any adjustment being made to it. S20 do so by dividing the feedback estimate into (unadjusted) $F_{2\times\text{CO}_2}$, but (4) requires use of $F_{2\times\text{CO}_2}^{\text{regress}}$. $F_{2\times\text{CO}_2}$ should accordingly be multiplied by a scaling factor, γ , to convert it to $F_{2\times\text{CO}_2}^{\text{regress}}$ (Sect. 4.1), so that:

$$S = -\gamma F_{2\times\text{CO}_2} / \lambda \quad (7)$$

For Historical evidence, differences in sea surface temperature (SST) change patterns may cause feedback estimated using (1), denoted $\lambda_{\text{hist}} = -[\Delta F_{\text{Hist}} - \Delta N_{\text{Hist}}] / \Delta T_{\text{Hist}}$, to differ from that over 150-year GCM abrupt4xCO₂ simulations, λ . An estimate of the effect, denoted $\Delta\lambda = \lambda - \lambda_{\text{hist}}$, of such differences in SST change patterns (the historical pattern effect) is allowed for when computing S , but not when calculating an alternative measure, S_{hist} . Since the appropriate λ is thereby being used to estimate S , $F_{2\times\text{CO}_2}$ should be scaled by γ , as for Process evidence. Using (7):

$$\begin{aligned} S &= -\gamma F_{2\times\text{CO}_2} / (\lambda_{\text{hist}} + \Delta\lambda) \\ &= \gamma F_{2\times\text{CO}_2} / \{[\Delta F_{\text{Hist}} - \Delta N_{\text{Hist}}] / \Delta T_{\text{Hist}} - \Delta\lambda\} \end{aligned} \quad (8)$$

whereas

$$S_{\text{hist}} = -F_{2\times\text{CO}_2} / \lambda_{\text{hist}} = F_{2\times\text{CO}_2} / \{[\Delta F_{\text{Hist}} - \Delta N_{\text{Hist}}] / \Delta T_{\text{Hist}}\} \quad (9)$$

ΔF_{Hist} is the sum of forcing component ERFs, which (apart from $F_{2\times\text{CO}_2}$) have uncertainties that are independent of each other and of the likewise independent uncertainties in the other right hand side terms in (8) and (9):

$$\begin{aligned} \Delta F_{\text{Hist}} &= \Delta F_{\text{Hist}}^{\text{CO}_2/2\times} F_{2\times\text{CO}_2} + \Delta F_{\text{Hist}}^{\text{othGHG}} + \Delta F_{\text{Hist}}^{\text{O}_3} + \Delta F_{\text{Hist}}^{\text{aerosol}} + \Delta F_{\text{Hist}}^{\text{LandUse}} \\ &+ \Delta F_{\text{Hist}}^{\text{vapor}} + \Delta F_{\text{Hist}}^{\text{BCsnow}} + \Delta F_{\text{Hist}}^{\text{contrails}} + \Delta F_{\text{Hist}}^{\text{solar}} + \Delta F_{\text{Hist}}^{\text{volcanic}} \end{aligned} \quad (10)$$

where $\Delta F_{\text{Hist}}^{\text{othGHG}}$ includes well-mixed non-CO₂ greenhouse gases, $\Delta F_{\text{Hist}}^{\text{O}_3}$ includes tropospheric and stratospheric ozone, $\Delta F_{\text{Hist}}^{\text{vapor}}$ represents stratospheric water vapor and $\Delta F_{\text{Hist}}^{\text{BCsnow}}$ represents black carbon on snow and ice.

When S is inferred from Process or Historical evidence using S20's assumptions, γ is set to one, as no adjustment to $F_{2\times\text{CO}_2}$ was made in S20.

For Paleoclimate evidence, there are separate models for changes related to the Last Glacial Maximum (LGM), mid-Pliocene Warm Period (mPWP) and Paleocene-Eocene Thermal Maximum (PETM), all of which involve a division by $(1 + \zeta)$ to convert an estimate for ECS into one for S .

3.2.1 For the LGM

$$S = \Delta T_{\text{LGM}} / \{(1 + \zeta)[\Delta F_{\text{LGM}}^{\text{CO}_2/2\times} + (\Delta F_{\text{LGM}}^{\text{exCO}_2} + 0.5\alpha\Delta T_{\text{LGM}}^2) / F_{2\times\text{CO}_2}]\} \quad (11)$$

where $\Delta F_{\text{LGM}}^{\text{CO}_2/2\times}$ represents the LGM – preindustrial CO₂ forcing change as a fraction of that from CO₂ doubling; $\Delta F_{\text{LGM}}^{\text{exCO}_2}$ represents the corresponding change in non-CO₂ forcing; and α is a coefficient for state dependence in climate feedback.

3.2.2 For the mPWP

$$\begin{aligned} S &= \Delta T_{\text{mPWP}} / \{[\log(1 + \Delta\text{CO}_{2\text{mPWP}}) / \log(2)] \\ &\quad (1 + f_{\text{mPWP}}^{\text{CH}_4})(1 + \zeta)(1 + f_{\text{mPWP}}^{\text{ESS}})\} \end{aligned} \quad (12)$$

where $(1 + f_{\text{mPWP}}^{\text{ESS}})$ represents the ratio of ESS to ECS for the mid-Pliocene, $f_{\text{mPWP}}^{\text{CH}_4}$ the forcing from methane relative to that from CO₂, and $\Delta\text{CO}_{2\text{mPWP}}$ the fractional increase in mid-Pliocene CO₂ concentration over an assumed 284-ppm equivalent preindustrial well-mixed greenhouse gases state. A logarithmic CO₂ forcing–concentration relationship holds over this range.

3.2.3 For the PETM

$$\begin{aligned} S &= [1 / (1 + \zeta)] / \{[(1 + f_{\text{PETM}}^{\text{CO}_2\text{nonLog}}) \log(1 + \Delta\text{CO}_{2\text{PETM}}) / \\ &\quad \log(2)](1 + f_{\text{PETM}}^{\text{CH}_4}) / \Delta T_{\text{PETM}} - \beta_{\text{PETM}} / F_{2\times\text{CO}_2}\} \end{aligned} \quad (13)$$

where β_{PETM} allows for possible state-dependence of climate feedback and any slow feedbacks affecting ESS but not ECS, and $f_{\text{PETM}}^{\text{CO}_2\text{nonLog}}$ scales CO₂ forcing from a logarithmic relationship with concentration, to correct for deviations therefrom at high concentration. S20 omits $f_{\text{PETM}}^{\text{CO}_2\text{nonLog}}$, so it is set to zero when S is inferred using S20's assumptions. Uncertainty in $f_{\text{PETM}}^{\text{CO}_2\text{nonLog}}$ is minute relative to that in $\Delta\text{CO}_{2\text{PETM}}$ and is therefore ignored.

Significant correlation is assumed to exist between errors in ΔT_{PETM} and ΔT_{mPWP} and between errors in $\Delta \text{CO}_2_{\text{PETM}}$ and $\Delta \text{CO}_2_{\text{mPWP}}$, with errors in all other variables being independent.

The foregoing equations are rearrangements of S20 Eqs. (4), (18), (20), (21), (22), (23) and (24), with the additional γ and $\left(1 + f_{\text{PETM}}^{\text{CO}_2 \text{nonLog}}\right)$ terms.

S20 also considered, but neither created a statistical model for nor used in their main results, 'emergent constraints' on climate sensitivity. These depend on relationships between selected observationally-constrainable variables and climate sensitivity in GCMs. In almost all cases, strong relationships found in one generation of GCMs have been statistically insignificant and/or substantially different in another GCM generation (Caldwell et al. 2014, 2018; Schlund et al. 2020), casting substantial doubt on their reliability. Biases common to all or most models are another concern.

3.3 Likelihood estimation for S

Likelihood estimation for a parameter is not straightforward when its value depends on multiple data-variables, even apart from the question of how to combine different lines of evidence. For each line of evidence separately, each set of data-variable value realizations corresponds to a unique value of the parameter and data-variable errors are independent, so the joint probability density for any set of data-variable values can be derived as the product of their PDFs at the values concerned, and assigned to the parameter value that they imply.

However, each value of the parameter, S , will correspond to an infinite set of combinations of differing data-variable values. Accordingly, producing a single likelihood corresponding to each S value requires some method of weighting probability densities for the different data-variable value combinations.

S20's likelihood estimation method involves sampling S uniformly and $F_{2 \times \text{CO}_2}$ *pro rata* to its PDF, the sample ratios providing λ samples, and for each line of evidence except Process (where the likelihood is analytically calculable) also sampling *pro rata* to its PDF each remaining data-variable involved other than ΔT . They take the likelihood of each resulting multivariate sample set as equal to the PDF of ΔT at the value implied by the sample set's λ and data-variable values. They bin the multivariate sample sets by their S values, and compute the S likelihood for each bin as the average of the likelihoods of the sample sets it contains.

While S20's likelihood estimation method may satisfactorily estimate the actual likelihood for S in simple cases, it is not clear why it would provide a realistic estimate of the likelihood where, for example, a data-variable has

substantial asymmetrical uncertainty or S is related to it non-linearly. Such circumstances arise with Historical evidence, due to the asymmetrical PDF for aerosol ERF, and with PETM (and to a lesser extent mPWP) evidence, due to the logarithmic relationship between CO_2 concentration ratio and ERF. Investigation confirms that S20's method of likelihood estimation is indeed unsound (supplemental material S2). Their method causes substantial misestimation of Historical and (worsened by a coding error resulting in the $\Delta \text{CO}_2_{\text{PETM}}$ standard deviation used being one tenth of its correct value—supplemental material S2) of PETM likelihood, and non-negligible misestimation of mPWP likelihood. Therefore, I do not use S20's likelihood estimation method in this study.

Rather than relying on a single likelihood estimation method, I employ three alternative methods, with the resulting likelihoods cross-checked. Each method involves setting up S value bins on a fine (0.01 K) grid spanning 0–20 K.

The first and third likelihood estimation methods involve first randomly sampling all the data-variables involved in estimating S from the line of evidence concerned that are not treated as fixed, weighting the sampling *pro rata* to their PDFs. The S value that each resulting multivariate sample set's data-variable values implies is computed, and each sample set is allocated to the appropriate S bin. The number of sample sets in each bin then provides an estimated (posterior) PDF for S . Between 10^7 and 10^8 sample sets are drawn, depending on the case. Since this procedure requires a unique S value for each sample-set, these likelihood estimation methods can only be used for single lines of evidence (for a single period in the case of paleoclimate evidence). This sampling method of deriving a posterior PDF for S has been widely used (Gregory et al. 2002; Otto et al. 2013; Lewis and Curry 2015, 2018). The method is prior-free, in the sense that no explicit prior selection is required, however it is equivalent to Bayesian estimation using a noninformative prior.

The first likelihood estimation method effectively involves estimating, at each S -bin value, the probability-weighted likelihood integrated over data-variable space; it gives the highest weight to those combinations of data-variable values most likely to arise. This 'integrated likelihood' method is implemented by taking the sample sets generated and used to derive a PDF for S as set out in the preceding paragraph, and computing the likelihood for each sample set (as the product of the data-variable PDFs at their sampled values). The likelihood for the S value at each bin center is then derived as the simple average of the likelihoods of all sample sets in the bin.

The second method uses the profile likelihood (Pawitan 2001), a widely used measure that typically provides a close

approximation to likelihood derived using more sophisticated methods. This method applies the entire weight to that combination of data-variable values which, at the S value concerned, maximizes the likelihood. The profile likelihood is derived by using an optimization algorithm to find, for each fine-grid value of S , the data values combination that maximizes the product of the data-variable PDFs for the line or lines of evidence concerned, allowing for any uncertainties that are common or correlated between data-variables.

Finally, a likelihood is calculated using the data-doubling method (Efron 1993; Lewis 2018). This involves the supposition that the evidence involved represented an observational data set, and that an identical but independent data set had also been observed. A posterior PDF for S based on emulating the stronger evidence provided by such doubled data is computed. This can often be effected by halving the actual variance of each data-variable and sampling pro rata to the reduced variance data-variable PDFs. An implied likelihood is then computed by dividing that PDF by the posterior PDF corresponding to the actual data set, derived by sampling pro rata to the actual data-variable PDFs. The validity of this 'doubled data likelihood' method follows directly from (5) if a noninformative prior is involved, since the same prior is noninformative for repeated observations from the same experiment, and the doubled data likelihood will equal the product of the original data likelihood with itself:

$$\begin{aligned} p_{\theta}(\theta|yy)/p_{\theta}(\theta|y) \\ &= [c_1 p_y(y|\theta) p_y(y|\theta) p_{\theta}(\theta)] \\ &\quad / [c_2 p_y(y|\theta) p_{\theta}(\theta)] \\ &= c_3 p_y(y|\theta) \end{aligned} \quad (14)$$

However, this method will not work satisfactorily where evidence is represented by a distribution for which emulating a doubled data version is problematic. That is the case for S20's Historical evidence, where the aerosol ERF distribution is highly asymmetrical and has no analytical form.

Before each derived PDF and likelihood is used it is generally smoothed with a spline-based method, retaining sufficient degrees of freedom to very closely match the shape of the unsmoothed original.

3.4 Noninformative prior and posterior PDF estimation

When either an integrated or doubled-data likelihood is used, a prior for S is derived by dividing the associated sampling-derived PDF by the estimated likelihood. This is an exact probability-matching prior by construction, and, since a Jeffreys' prior provides the closest probability matching,

the derived prior is necessarily a noninformative Jeffreys' prior (provided the likelihood is valid).

The profile likelihood method only produces a likelihood, so it is necessary to separately derive a Jeffreys' prior, $\pi_{JP}(S)$, to use therewith. The 'data-space movement' method used is based on a direct measure of the local informativeness of the data about the parameter. Details of this method, and of the calibration of all the Jeffreys' priors, are given in the supplemental material (S3).

A PDF for S can then be derived as the product of the profile likelihood and the related data-space movement prior. This PDF cannot account for probability outside the range of S values used, so it is normalized to unit probability over that range. References to profile likelihood method PDFs are to such PDFs derived directly, or after combining likelihoods and priors from different lines of evidence.

3.5 Applying the Objective Bayesian statistical methods to combining S20's evidence

The statistical models employed in S20 to link S to data-based evidence necessitate a more general approach to combining lines of evidence using a single combination-based noninformative prior than was employed in Lewis (2018) and Lewis and Grunwald (2018). The mechanics involved are detailed in the supplemental material (S4).

A key motivation for using several different likelihood and noninformative prior estimation methods is that comparison of their performance when combining lines of evidence can provide confidence that both those methods, and the combination methods used, are valid. For S20's Historical evidence, to which the doubled data method cannot successfully be applied, likelihoods from the integrated likelihood and profile likelihood methods are almost identical. However, the profile likelihood data-space movement prior from Historical evidence is poor, due to imperfect optimization and difficulty representing the informativeness of the aerosol forcing distribution used by S20, and profile likelihood based estimates of S from all lines of evidence combined cannot reliably be derived by a single optimization. I therefore examine the combined-evidence likelihoods, priors and posterior PDFs for S that the three methods produce when combining S20's Process evidence with LGM and mPWP Paleoclimate evidence. For the two sampling-based methods, doing so involves combining separate estimates from Process, LGM and mPWP evidence. For the profile likelihood method, these are one-step estimates from simultaneous inference using data from all three individual lines of evidence. For all methods, the combined-evidence posterior PDF for S is normalized to unit probability over 0–20 K.

Figure 1a compares the combined Process, LGM and mPWP evidence likelihood estimates from the aforementioned three methods. They are indistinguishable.

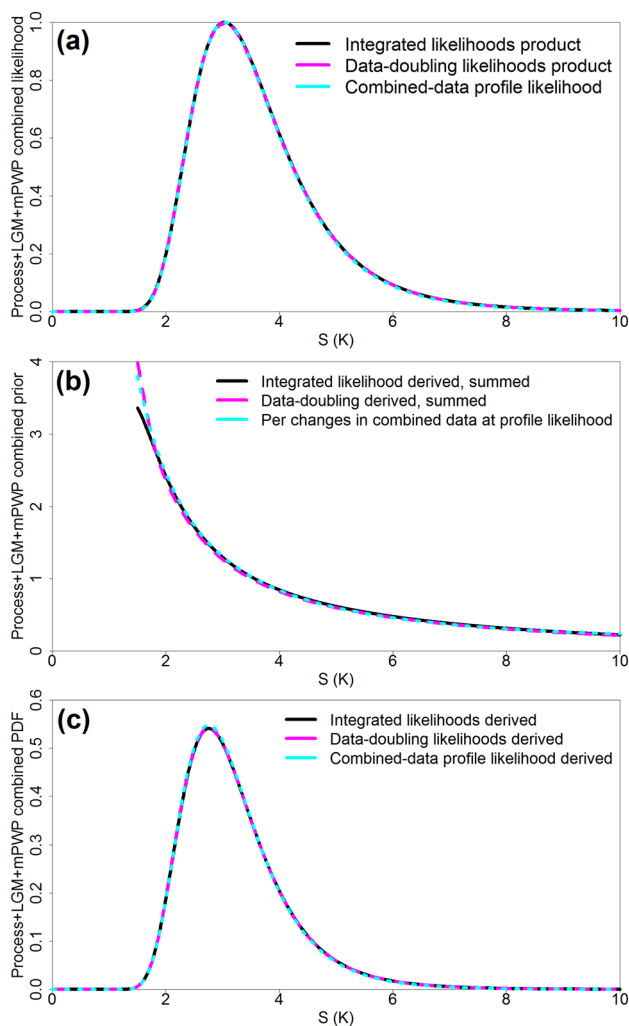


Fig. 1 Results from combining S20's Process evidence with LGM and mPWP Paleoclimate evidence using the integrated likelihood, doubled data and profile likelihood based methods. **a** Combined evidence likelihoods; **b** Combined evidence priors; **c** Combined evidence posterior probability density functions. In each case the solid black line is from the integrated likelihood method, the dashed magenta line is from the doubled data method and the dotted cyan line is from the profile likelihood method and/or the related data-space movement prior

Figure 1b compares the three related Jeffreys' priors from $S = 1.5$ K up. Below that level the integrated likelihood and doubled data sampling-derived priors are artefacted, due to the paucity of samples for Process evidence, and behave erratically. However, the likelihood is almost zero below $S = 1.5$ K, and the total probability in that region is only 0.05%, so the effect on inference for S is negligible. The resulting posterior PDFs for S using the three methods (Fig. 1c) are indistinguishable. Their medians are all within 0.01 K of each other, their 5th percentiles are all 2.07 K and their 95% percentiles are all within ± 0.03 K.

The very close agreement among combined-evidence inference for S using three different methods of deriving and combining likelihoods, noninformative Jeffreys' priors and PDFs provides strong support for the general validity of all the methods for this application.

The main results presented in Sects. 5 and 6 use the robust sampling-based integrated likelihood method—which in all cases produces a satisfactory likelihood and prior—for all lines of evidence.

4 Review and revision of S20 data-variable assumptions

I now make some revisions to S20's data-variable assumptions for various lines of evidence, which are justified on the basis of more recent evidence, by a preferable alternative interpretation of the same evidence, or because they remedy an error or omission. The scaling factor γ for $F_{2\times\text{CO}_2}$ is included in these revisions. The original and revised estimates for all data-variables are set out in Tables 1, 2 and 3, with the reasons for changes. The evidence justifying each revision is reviewed in detail in the supplemental material (S5); evidence relating to a number of the unrevised data-variable estimates is also reviewed there. Results from applying the Objective Bayesian approach to inference using S20's assumptions and the revised assumptions are given in Sects. 5 and 6 respectively.

The units of all stated feedback values are $\text{Wm}^{-2} \text{K}^{-1}$. Uncertainties indicated by \pm represent one standard deviation, with a Normal distribution, denoted $N(\text{mean, standard deviation})$, assumed.

4.1 $F_{2\times\text{CO}_2}$ and its scaling when using Eq. (4)

S20 use the estimate of stratospherically-adjusted forcing from doubled CO_2 of 3.80 Wm^{-2} per the simplified formula in Etminan et al. (2016), and add 5% for tropospheric adjustments, arriving at an ERF estimate for $F_{2\times\text{CO}_2} = 4.0 \pm 0.3 \text{ Wm}^{-2}$. Meinshausen et al. (2020) fitted Etminan et al.'s results more precisely, obtaining a $F_{2\times\text{CO}_2}$ value 1.5% lower. Based on their more accurate formula, and using the same 5% tropospheric adjustment, $F_{2\times\text{CO}_2}$ ERF was assessed at $3.93 \pm 0.3 \text{ Wm}^{-2}$ in AR6 (Forster et al. 2021: 7.3.2.1). The ratio of $F_{4\times\text{CO}_2}$ to $F_{2\times\text{CO}_2}$ per the Meinshausen et al. (2020) formula is $2.10\times$, 5.0% higher than under a $\log(\text{concentration})$ relationship. I adopt these numbers in Sect. 6 when estimating S .

Care must be taken to use the appropriate $F_{2\times\text{CO}_2}$ value when applying (4). S20 use their estimate of the actual ERF from a doubling of CO_2 concentration. However, as stated in Sect. 3.2, when feedback is estimated using a linear model on a basis consistent with behavior during years 1–150

Table 1 Parameters of the data-variable distributions used to estimate S from Process evidence

Description	Symbol	Sherwood et al		As revised		Reason for any change
Normal distribution parameters		Median	Std Dev	Median	Std Dev	
Feedback components:						
Tropical marine low cloud		0.25	0.16			
Mid-latitude marine low cloud amount		0.12	0.12			
Tropical & mid-latitude marine low cloud		0.37	0.20	0.19	0.20	Newer evidence ^a
High-latitude low cloud		0.00	0.10	0.00	0.10	
High cloud altitude		0.20	0.10	0.20	0.10	
Tropical anvil cloud area		−0.20	0.20	−0.20	0.20	
Land clouds		0.08	0.08	0.08	0.08	
Total cloud	λ_{clouds}	0.45	0.33	0.27	0.33	
Planck	λ_{Planck}	−3.20	0.10	−3.25	0.10	Later data ^b
Water vapor + Lapse rate	$\lambda_{\text{WV+LR}}$	1.15	0.15	1.15	0.15	
Surface albedo	$\lambda_{\text{sfcAlbedo}}$	0.30	0.15	0.30	0.15	
Stratospheric	$\lambda_{\text{stratospheric}}$	0.00	0.10	0.00	0.10	
Atmospheric composition	$\lambda_{\text{atmosComp}}$	0.00	0.15	0.00	0.15	
Total feedback	λ	−1.30	0.44	−1.53	0.44	
ERF from doubled CO ₂ (Wm ^{−2})	$F_{2\times\text{CO2}}$	4.00	0.30	3.93	0.30	AR6 value/ Later data ^c
$F_{2\times}$ scaling factor ($F_{2\times\text{CO2}}^{\text{regress}}/F_{2\times\text{CO2}}$)	γ	1.00	–	0.86	0.09	Omission ^d
S at data-variable median values (K)	S	3.08		2.21		

^aMyers et al. (2021), Cessana and Del Genio (2021), Mülmenstädt et al. (2021); supplemental material S5.1.3

^bZelinka et al. (2020); supplemental material S5.1.2

^cMeinshausen et al. (2020); Sect. 4.1

^dSection 4.1; supplemental material S5.1.5

of abrupt4xCO₂ simulations, normally by ordinary least squares regression, then $F_{2\times\text{CO}_2}$ should be converted into an estimate of $F_{2\times\text{CO}_2}^{\text{regress}}$, the ERF implied by the y-axis regression line intercept, by multiplying it by $\gamma = F_{2\times\text{CO}_2}^{\text{regress}}/F_{2\times\text{CO}_2}$. That is because S is defined in terms of the climate feedback λ arising from system behavior over years 1–150 after a (hypothetical or actual) quadrupling of CO₂ concentration, and as per (4) equals $-F_{2\times\text{CO}_2}^{\text{regress}}/\lambda$ (Sect. 2). When climate feedback weakens during the course of 150-year abrupt CO₂ forcing simulations, as it does in the vast majority of GCMs, $F_{2\times\text{CO}_2}^{\text{regress}}$ will underestimate the GCM's actual $F_{2\times\text{CO}_2}$. Therefore, dividing the actual $F_{2\times\text{CO}_2}$ by climate feedback estimated over the whole simulation period is bound to overestimate S (supplemental material S1; compare red and black lines in Figure S1.1). For an ensemble of CMIP5 and CMIP6 GCMs, the $F_{2\times\text{CO}_2}^{\text{regress}}/F_{2\times\text{CO}_2}$ ratio is 0.86 ± 0.09 (supplemental material Table S1, rounding up the standard deviation, S5.1.5). I adopt this estimate.

S20 recognize this issue, conceding a similar overestimation of S , but neglect it, asserting incorrectly that it only affects feedback estimates from GCMs. This misconception results in S20's estimates of S from Process and Historical evidence being biased high. The issue is intrinsic to the use of (4) and the S20 definitions of S and λ , which

involve a single λ value, estimated on a basis consistent with that obtained by regressing over 150-year abrupt4xCO₂ simulations.

For Process evidence, the bias is self-evident, since almost all λ components are estimated by or on a basis consistent with regressing changes over abrupt4xCO₂ simulations. In particular, low cloud feedback, the dominant cause of weakening feedback over abrupt4xCO₂ simulations, and hence of $\gamma < 1$, is so estimated (supplemental material S5.1.1, S5.1.3).

For Historical evidence, S20 estimate λ_{hist} as $-(\Delta F_{\text{Hist}} - \Delta N_{\text{Hist}})/\Delta T_{\text{Hist}}$ and divide it into $-F_{2\times\text{CO}_2}$ to estimate S_{hist} (supplemental material Figure S1.1, blue line). They then adjust λ_{hist} by their estimate of the difference between λ in abrupt4xCO₂ simulations and λ_{hist} , and estimate S by dividing the resulting λ estimate into $-F_{2\times\text{CO}_2}$, rather than (as should be done) into $-F_{2\times\text{CO}_2}^{\text{regress}}$, thereby overestimating S (supplemental material Figure S1.1, red line).

Accordingly, in the statistical models S20 uses to estimate S from both Process and Historical evidence, $F_{2\times\text{CO}_2}$ needs to be scaled by $F_{2\times\text{CO}_2}^{\text{regress}}/F_{2\times\text{CO}_2}$. I do so in Sect. 6. Paleoclimate estimation of S is unaffected, since that in effect estimates climate feedback from equilibrium changes, derives

Table 2 Parameters of the data-variable distributions used to estimate S from Historical evidence. All changes are in global and time means between the 1861–1880 and 2006–2018 periods

Description	Symbol	Sherwood et al		As revised		Reason for any change
Normal distribution parameters		Median	Std Dev	Median	Std Dev	
Change in (blended) GMST (K)		0.94	0.07	0.94	0.075	
Adjustment from GMST to GMAT (K)		0.08	0.04	0.00	0.057	AR6 value ^a
Change in GMAT (K)	ΔT	1.03	0.085	0.94	0.095	
<i>Changes in forcing, as ERF (Wm^{−2})</i>						
Carbon dioxide	ΔF_{CO_2}	1.731	$\propto F_{2\times\text{CO}_2}$	1.724	$\propto F_{2\times\text{CO}_2}$	AR6 value ^b
CO ₂ ERF as fraction of $F_{2\times\text{CO}_2}$	$\Delta F_{\text{CO}_2/2\times}$	0.433	Zero	0.439	Zero	Arithmetic
Other well-mixed greenhouse gases	ΔF_{othGHG}	0.969		1.015		AR6 value ^b
Ozone (tropospheric + stratospheric)	ΔF_{O_3}	0.298		0.400		AR6 value ^b
Land use	$\Delta F_{\text{LandUse}}$	−0.106		−0.150		AR6 value ^b
Stratospheric water vapor	ΔF_{vapor}	0.064		0.041		AR6 value ^b
Black carbon on snow and ice	ΔF_{BCsnow}	0.020		0.109 ^c		AR6 value ^b
Contrails and induced cirrus	$\Delta F_{\text{contrails}}$	0.048		0.048		AR6 value ^b
Solar	ΔF_{solar}	0.017		0.019		AR6 value ^b
Volcanic	$\Delta F_{\text{volcanic}}$	−0.113		0.044		AR6 value ^b
Forcing excluding that from CO ₂ and aerosol	ΔF_{other}	1.196	0.184	1.526	0.173	
Aerosol	$\Delta F_{\text{aerosol}}$	−1.104	Non-normal	−0.860 ^d	0.498 ^d	Other/new evidence ^e
Forcing excluding that from CO ₂	ΔF_{exCO_2}	0.093	n/a	0.666	0.527	
Change in radiative imbalance (Wm ^{−2})	ΔN	0.600	0.183	0.600	0.183	
Radiative response at median values (Wm ^{−2})	ΔR	1.224		1.800		
Feedback for S_{hist} at median values	λ_{hist}	−1.188		−1.915		
Historical pattern effect feedback adjustment	$\Delta \lambda$	0.500	0.305	0.350	0.305	Newer evidence ^f
Feedback for S at data-variable medians	λ	−0.687		−1.565		
ERF from doubled CO ₂ (Wm ^{−2})	$F_{2\times\text{CO}_2}$	4.00	0.30	3.93	0.30	AR6 value/ Later data ^g
$F_{2\times}$ scaling factor ($F_{2\times\text{CO}_2}^{\text{regress}}/F_{2\times\text{CO}_2}$)	γ	1.00	Zero	0.86	0.09	Omission ^h
S at data-variable median values (K)	S	5.82		2.16		

^aGulev et al. (2021); supplemental material S5.2.1. GMST is derived as a blend of land near-surface air temperature and ocean sea surface temperature

^bSmith et al. (2021); supplemental material S5.2.3

^cThe AR6 ERF has been multiplied by a 1.5 efficacy factor; supplemental material S5.2.3

^dUsing the AR6 aerosol ERF best estimate time series to scale the adopted $-0.95 \pm 0.55 \text{ Wm}^{-2}$ estimated 1850 to 2005–2015 ERF to a 1861–1880 to 2006–18 change

^eMultiple studies, e.g. Hamilton et al. (2018); Gryspeerdt et al. (2019); Paulot et al. (2020); Possner et al. (2020); Glassmeier et al. (2021); Lee et al. (2021); Liu et al. (2021); supplemental material S5.2.3

^fLewis and Mauritsen (2021); Zhou et al. (2021); Fueglistaler and Silvers (2021); supplemental material S5.2.4

^gMeinshausen et al. (2020); Sect. 4.1

^hSect. 4.1; supplemental material S5.2.5

ECS by dividing it into $-F_{2\times\text{CO}_2}$, and then converts ECS into S (see (11), (12) and (13)).

4.2 Process evidence

The data-variable distributions adopted by S20, and as revised here, for estimating S from Process evidence are summarised in Table 1. The main changes made are to low cloud feedback, reflecting strong recent evidence that it is

weaker than S20's assessment, and scaling $F_{2\times\text{CO}_2}$ to $F_{2\times\text{CO}_2}^{\text{regress}}$ (Sect. 4.1).

The significant revision in estimated tropical and mid-latitude (60°S–60°N) marine low cloud feedback is discussed in detail in the supplemental material (S5.1.3). In brief, both the S20 and the revised estimates depend primarily on observational estimates of low cloud response to cloud-controlling factors (CCF). S20's assessment of tropical marine low cloud feedback was based primarily on, and equals, an estimate

Table 3 Parameters of the data-variable distributions used to estimate S from Paleoclimate evidence

Description	Symbol	Sherwood et al		As revised		Reason for any change
Normal distribution parameters		Median	Std Dev	Median	Std Dev	
ERF from doubled CO ₂ (Wm ⁻²)	$F_{2\times\text{CO}_2}$	4.00	0.30	3.93	0.30	AR6 value/ Later data ^a
ECS/ $S-1$	ζ	0.06	0.20	0.135	0.10	Consistency ^b
LGM						
Change in GMAT (K)	ΔT	-5.0	1.0	-4.5	1.25	S20 sources ^c
<i>Changes in forcing, as ERF (Wm⁻²)</i>						
CO ₂ ERF as fraction of $F_{2\times\text{CO}_2}$	$\Delta F_{\text{CO}_2/2\times}$	-0.57	Zero	-0.57	Zero	
Methane (CH ₄)		-0.57		-0.57		
Nitrous oxide (N ₂ O)		-0.28		-0.28		
Land ice and sea level		-3.20		-3.72		Other evidence ^d
Vegetation		-1.10		-1.10		
Dust (aerosol)		-1.00		-1.00		
Forcing excluding that from CO ₂	ΔF_{exCO_2}	-6.15	2.00	-6.67	2.00	
Dependence of feedback on ΔT (Wm ⁻² K ⁻²)	α	0.10	0.10	0.10	0.10	
S at median values for all variables (K)	S	2.63		1.97		
mPWP						
Change in GMAT (K)	ΔT	3.00	1.00	2.48	1.25	Later data ^e
Fractional change in CO ₂ concentration	ΔCO_2	0.32	0.09	0.32	0.09	
CH ₄ forcing as a fraction of that from CO ₂	f_{CH_4}	0.40	0.10	0.40	0.10	
Earth system sensitivity / ECS-1	f_{ESS}	0.50	0.25	0.67	0.40	Later data ^e
S at median values for all variables (K)	S	3.36		2.33		
PETM						
Change in GMAT (K)	ΔT	5.00	2.00	5.00	2.00	
Fractional change in CO ₂ concentration	ΔCO_2	1.667	0.778	1.667	0.778	
CO ₂ ERF relative to with log(concentration)	$f_{\text{CO}_2\text{nonLog}}$	1	–	1.117	~Zero	Omission ^f
CH ₄ forcing as a fraction of that from CO ₂	f_{CH_4}	0.40	0.20	0.40	0.20	
Different climate state: feedback adjustment	β	0.00	0.50	0.00	0.50	
S at median values for all variables (K)	S	2.38		1.99		

^aMeinshausen et al. (2020); Sect. 4.1^bConsistency with estimation from Process and Historical evidence. Rugenstein et al. (2020); Sect. 2; supplemental material S5.3.1^cSame studies as cited by S20; supplemental material S5.3.2^dKohler et al. (2010); Zhu and Poulsen (2021); supplemental material S5.3.2^eTierney et al. (2019); McClymont et al. (2020); Haywood et al. (2020); supplemental material S5.3.3^fMeinshausen et al. (2020); supplemental material S5.3.4

from the Klein et al. (2017) review, which only took two CCF into consideration. For the 30–60° mid-latitude bands S20 also used GCM-derived evidence. The revised median estimate is the observationally-constrained 60°S–60°N value from Myers et al. (2021). They use a more comprehensive set of CCF and argue that their feedback estimate is more realistic than Klein et al.'s. Cessana and Del Genio (2021) likewise find the Klein et al. feedback estimate to be too high.

The revisions to the S20 median data-variable values change the central λ estimate from Process evidence alone from -1.30 ± 0.44 in S20 to -1.53 ± 0.44 , while the

maximum likelihood estimate of S changes from 3.08 K to 2.21 K.

4.3 Historical evidence

The data-variable distributions adopted by S20, and as revised here, for estimating S from Historical evidence are summarised in Table 2. The changes made to S20's data-variable estimates are discussed in detail in the supplemental material (S5.2). The main changes made are to aerosol forcing, reflecting the most quantitative of the recent evidence that it is weaker than S20 assumed, to other forcings and the ΔGMAT – ΔGMST difference, reflecting AR6 assessments,

to the historical pattern effect, reflecting evidence that most SST datasets indicate little unforced element, and scaling $F_{2\times\text{CO}_2}$ to $F_{2\times\text{CO}_2}^{\text{regress}}$.

The significant revision in estimated aerosol ERF is discussed in considerable detail in the supplemental material (S5.2.3). Briefly, S20 use the unconstrained aerosol forcing distribution from Bellouin et al. (2020; hereafter B20). That distribution is based on a complex theoretical formula that depends on a number of factors, all of which are estimated separately. There is considerable evidence suggesting that B20 overestimate aerosol forcing strength. The revised estimate median uses recent evidence (Gryspeerd et al. 2019; Possner et al. 2020; Glassmeier et al. 2021) regarding just one of the factors involved in B20's calculations: the cloud liquid water path sensitivity factor used when adjusting the radiative forcing from aerosol-cloud interactions to an ERF basis. The revised median aerosol ERF estimate is derived by carrying out the same computation as in B20 except for changing the estimate of that sensitivity factor. The revised ERF estimate adopts a Gaussian uncertainty distribution with the same 5% bound as assessed in AR6; B20's theoretically-derived distribution assigns significant probability to extremely negative aerosol ERF values, but these appear inconsistent with observational evidence that is independent of the global temperature record.

Another significant revision is that to the Historical pattern effect feedback adjustment, discussed in detail in the supplemental material (S5.2.4). Briefly, S20's median estimate is based on that reported by Andrews et al. (2018), which was derived by comparing, in six GCMs, estimated feedback in abrupt4×CO₂ simulations with that in fixed SST simulations with evolving observational historical SST patterns. S20 reduced the Andrews et al. (2018) estimate by 0.1 Wm⁻² K⁻¹ to allow for the possibility that the pattern effect may be smaller than reported in that study. However, substantial evidence now exists that the observational SST dataset used for all the fixed SST simulations assessed by Andrews et al. (2018) is an outlier in terms of the magnitude of the pattern effect that it gives rise to (Lewis and Mauritsen 2021; Zhou et al. 2021; Fueglistaler and Silvers 2021). The revised pattern effect adjustment reflects that evidence.

The revisions to the S20 data-variable median values reduce the estimates of S and S_{hist} that they imply from 5.82 to 2.16 K, and from 3.37 to 2.05 K, respectively.

4.4 Paleoclimate evidence

Paleoclimate evidence has the advantage of being largely independent of, and sometimes involving a much larger signal than, the historical period, but suffers from relating to different states of the Earth. The evidence is also derived from imprecise, geographically limited, and potentially

biased proxies that provide estimates for only some relevant variables.

S20 evaluate evidence from climate transitions during the LGM, mPWP and PETM, but their main results exclude PETM evidence. For all three periods, an estimate of ECS is converted into one for S by dividing it by $(1 + \zeta)$, the distribution of which is revised (Sect. 2; supplemental material S5.3.1), as is the median $F_{2\times\text{CO}_2}$ estimate (Sect. 4.1).

4.4.1 LGM

The best studied paleoclimate transition, and that most used for estimating climate sensitivity, is the change from the LGM, the coldest phase in the last ice age, some twenty thousand years ago to the preindustrial Holocene. A significant advantage of the LGM transition is that, unlike more distant periods, there is proxy evidence not only of changes in temperature and CO₂ concentration but also of non-CO₂ forcings, and that enables estimation of the effects on radiative balance of slow (ice sheet, etc.) feedbacks, which need to be treated as forcings in order to estimate ECS (and hence S) rather than ESS. Moreover, the temperature proxy evidence is sufficient to enable spatially-weighted global means to be estimated (Annan and Hargreaves 2013).

The two changes made to S20's data-variable estimates are discussed in detail in the supplemental material (S5.3.2). The revision to land ice and sea-level forcing adds an estimate of the omitted albedo change caused by sea-level fall exposing more land. The revision to ΔT brings it closer to the average of estimates from S20's cited sources. The revisions I make to S20's median LGM data-variable values, along with the revised ζ estimate, reduce the estimate of S that they imply from 2.63 K to 1.97 K.

4.4.2 mPWP

The mid-Pliocene warm period, approximately 3 Ma ago, was moderately warmer than preindustrial times, and in that respect a closer analogue than the LGM of conditions expected during this century. However, the temperature change involved was smaller than for the LGM, there is more uncertainty about CO₂ levels, temperature proxies are more limited, and usable proxy-based estimates of non-CO₂ forcing are unavailable.

The changes made to S20's data-variable estimates are discussed in detail in the supplemental material (S5.3.3). They relate to ΔT and the ESS–ECS ratio, in both cases reflecting ratios per the more recent PlioMIP2 project (Haywood et al. 2020). The revisions I make to S20's mPWP data-variable median values, along with the revised ζ estimate, reduce the estimate of S that they imply from 3.36 to 2.33 K, more in line with estimates from the LGM and PETM.

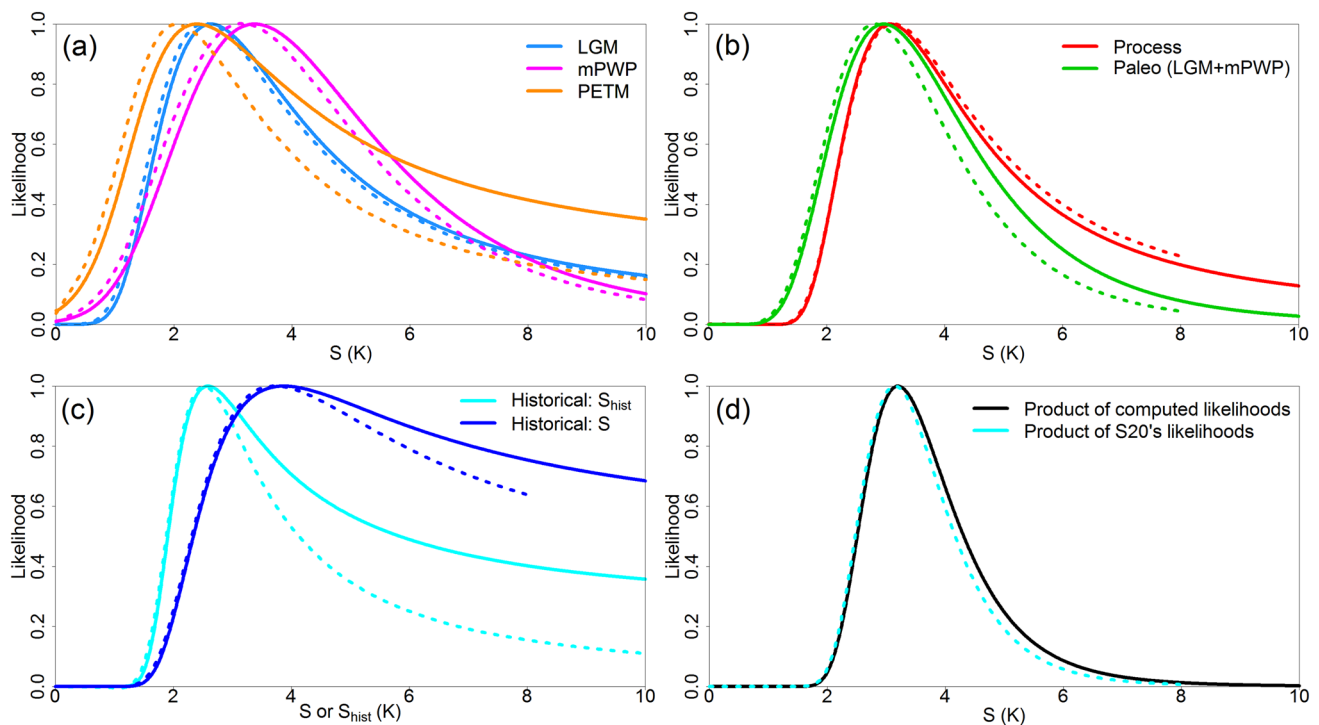


Fig. 2 Likelihoods for S based on S20's data-variable assumptions as derived in this study (solid lines) and, for comparison, those shown in S20 (dotted lines). **a** Likelihoods from evidence for the three paleoclimate periods. **b** Likelihoods from Process evidence and from combining Paleoclimate evidence for the LGM and mPWP; S20 did not use evidence for the PETM for their estimation of S . **c** Likelihoods from Historical evidence for both S and S_{hist} . **d** Likelihoods from com-

bined Process, Paleoclimate (LGM and mPWP combined), and Historical evidence, computed as the product of the likelihoods from the three lines of evidence. The likelihoods used in this study have been derived using the integrated likelihood method and normalized to a maximum of one. The S20 likelihoods were accurately digitized over the full S ranges of the relevant figures in S20

4.4.3 PETM

The PETM temperature excursion period some 56 Ma ago was much warmer than the present, and differed geographically and orographically. S20 state that the PETM is arguably the best pre-Pliocene warm interval for estimating ECS; it has been fairly well studied and involves a large signal. Nevertheless, they excluded PETM evidence in their Base-line estimates "due to the large uncertainties and the danger of over-constraining the likelihood should these be underestimated". While the PETM uncertainties are substantial, S20 makes generous allowance for them, and any underestimation of uncertainties appears much more likely to cause over-estimation than underestimation of S (supplemental material S5.3.4). In view of that and of the large signal involved in the PETM, I use it, but as an alternative to the mPWP rather than combining their evidence, since doing so provides very little benefit when the estimated data-variable error correlations between the two periods are allowed for (supplemental material S5.3.5).

The one revision I make to S20's PETM data-variable median values allows for the CO₂ ERF-concentration not

being exactly logarithmic, which in combination with the revised ζ estimate reduces the estimate of S that they imply from 2.38 K to 1.99 K. This change is discussed in detail in the supplemental material (S5.3.4), where the evidence relating to several of S20's other PETM data-variable estimates is also discussed.

The data-variable distributions adopted by S20, and as revised here, for estimating S from Paleoclimate evidence for each period are summarised in Table 3.

5 Results using S20 data-variable assumptions: comparison using different methods

I now compare S20's results with those derived here using the same input assumptions and either a computed noninformative Jeffreys' prior or the same prior as used in S20. I start by comparing likelihoods, as these are the foundation of parameter inference and are unaffected by the prior used, and then discuss the computed Jeffreys' priors. Finally, the posterior PDFs and numerical percentile values for S produced in this study are presented and compared with those in S20.

5.1 Likelihoods

The likelihoods derived using the profile likelihood method, the sampling-based integrated likelihood and the data doubling methods described in Sect. 3.3 agree very closely with each other (supplemental material Figures S1 to S3). They should also agree with the likelihoods shown in S20, as they are based on the same statistical models and data-variable assumptions. The integrated likelihoods derived in this study are shown by solid lines in Fig. 2; likelihoods from S20 are the same color but dotted. The overall Paleoclimate likelihood is that from combining evidence from the LGM and mPWP.

Figure 2a shows that mPWP paleoclimate evidence discriminates more strongly against very high S values than does LGM or, particularly, PETM evidence, despite its median value being the highest of the three. This is primarily because fractional uncertainty in forcing, and in those terms that effectively modify forcing, is lowest for mPWP evidence and highest for PETM evidence.

Figure 2b and c show that Paleoclimate evidence likelihoods downweight the possibility of very high S values most strongly, while Historical evidence does so very weakly. The latter primarily reflects the use of the unconstrained Bellouin et al. (2020) aerosol forcing distribution, which assigns substantial probability to strongly negative values.

S20's own likelihoods generally peak marginally earlier than those computed here, and thereafter decline faster. The difference is barely noticeable for the LGM, but rather larger for the mPWP and for the overall Paleoclimate (LGM and mPWP combined) evidence. The difference is major for the likelihoods from PETM and Historical evidence. This is particularly the case, in ratio terms, for S_{hist} (cyan lines in Fig. 2c). The differences arise because S20 employ an invalid method for deriving likelihoods (supplemental material S2). The virtual identity of the present study's estimated likelihoods using three different methods (two for Historical evidence) provides further confirmation that the S20 likelihoods are incorrect, and quantifies their inaccuracy.

When likelihoods from all lines of evidence used in S20 are combined multiplicatively (Fig. 2d), the resulting likelihood drops far more sharply than those from any individual line of evidence. The absolute difference between the combined S20 likelihoods and those computed here is relatively small in this case. This is because the S20 likelihoods are reasonably accurate below the likelihood maxima, and the combined likelihood drops to a low level before the errors in the S20 likelihoods (which partially cancel) grow very significant. By $S = 5$ K, the S20 combined likelihood is ~25%

lower than that calculated here, but by that point the 95th probability percentile has been reached.

The likelihood difference for Process evidence is small, but of the opposite sign to the other cases; S20's likelihood peaks marginally later than that computed here and declines more slowly after the peak. The Process likelihood can also be derived using a formula that accurately approximates the distribution of the ratio of two normally-distributed variables, here λ and $F_{2 \times \text{CO}_2}$ (Raftery and Schweder 1993; Lewis 2018). The likelihood per that formula almost exactly matches the likelihoods derived using this study's three methods, but not S20's likelihood. Since S20's Process likelihood computation is not subject to the defects identified in their other likelihood computations, that suggests some other problem may exist in S20's statistical computations.

5.2 Computed Jeffreys' priors

Figure 3 shows the computed and calibrated integrated likelihood based Jeffreys' priors for each of the three main lines of evidence, before and after transformation to λ -space (which aids comparing them), and so-transformed priors for individual Paleoclimate periods. Since these priors are very similar to those estimated using the revised assumptions, comments on them are deferred until the latter are discussed in Sect. 6.

5.3 Posterior PDFs and percentiles for S

Figure 4a shows (solid lines) sampling-derived primary posterior PDFs for each main line of evidence, representing in each case the product of the estimated integrated likelihood and a Jeffreys' prior, normalizing to unit probability over the 0–20 K S range used. Posterior PDFs from using a uniform-in- λ prior, as in S20, with the same likelihoods are also shown (dashed lines). For Process evidence, for which Jeffreys' prior is uniform in λ , the two PDFs coincide. For Historical and particularly Paleoclimate evidence, use of a uniform-in- λ prior biases the posterior PDF towards lower S values and, in the Paleoclimate case, excessively constrains high S values.

The primary combined evidence posterior PDF (Fig. 4b, solid red line) represents the product of the estimated Process, Historical and Paleoclimate likelihoods, being the combined likelihood, and Jeffreys' prior for the combined evidence. The PDF using a uniform-in- λ prior is also shown (solid blue line). The difference between these is much smaller than in the case of separate Historical or

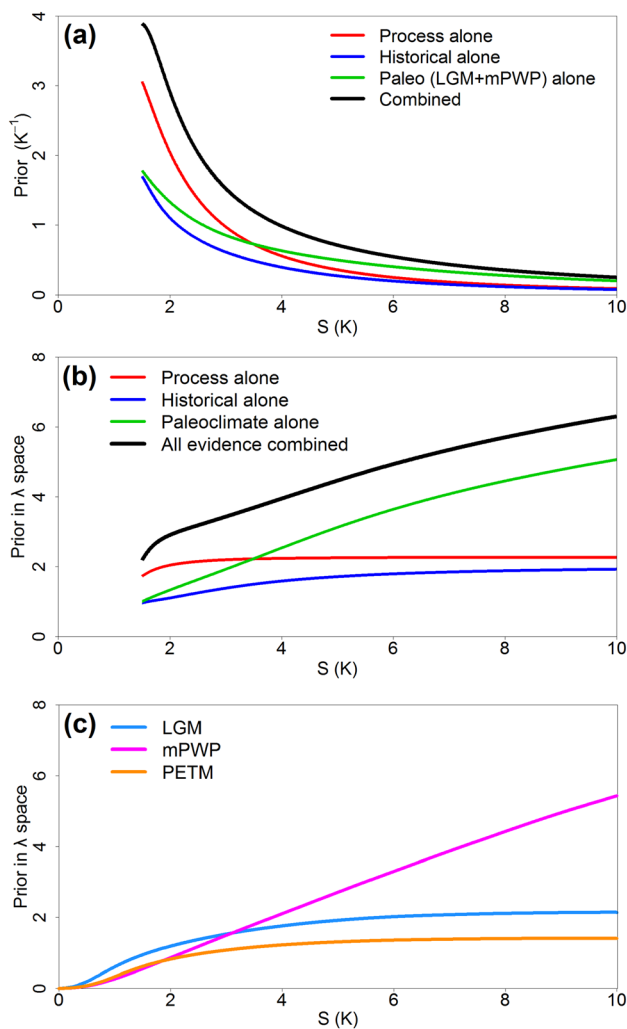


Fig. 3 Jeffreys' priors based on S20's data-variable assumptions, computed using the integrated likelihoods. **a** For each main line of evidence and for the combined evidence from all three lines. Paleoclimate evidence from the LGM and mPWP is included but, as in S20, not that from the PETM. **b** As (a) but with the prior, although still plotted against S values, having been transformed into λ -space. **c** As for (b), but for evidence from individual paleoclimate periods. Transformation of the prior to λ -space is effected in the standard way, by multiplying the original, S -space, prior by the absolute Jacobian of the inverse transformation, being $S^2/F_{2\times\text{CO}_2}$. The relatively small uncertainty in $F_{2\times\text{CO}_2}$ is ignored when effecting this transformation, which accounts for the decline in the transformed Process evidence prior as S reduces to a low level. Note that the separate priors shown cannot validly be added in quadrature to obtain a Jeffreys' prior for the All combined case, or for Paleoclimate evidence using the priors in panel (c), because such summing would multiply count the influence of uncertain data-variables in common. The panel (a) and (b) plots start at 1.5 K because the priors become artefacted below that level

Paleoclimate evidence, reflecting the combined likelihood (Fig. 2d) being much narrower. The dotted cyan line shows S20's Baseline, uniform-in- λ prior based, PDF. It is very close to the uniform-in- λ prior based PDF derived here,

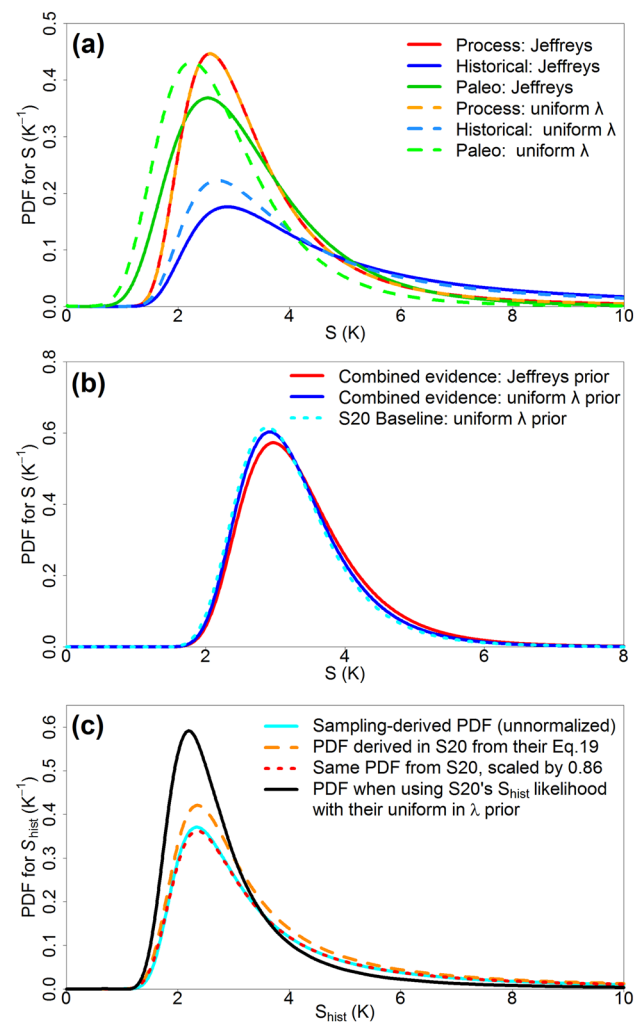


Fig. 4 Posterior PDFs based on S20's data-variable assumptions. **a** PDFs for S from separate Process, Historical and Paleoclimate evidence, using computed Jeffreys' priors (solid lines) and uniform-in- λ priors (dashed lines). **b** Combined evidence PDFs for S derived in this study (solid lines) using both Jeffreys' prior and a uniform-in- λ prior and, for comparison, the Baseline PDF in S20 using a uniform-in- λ prior (dotted line), accurately digitized from the relevant figure in S20. In panels (a) and (b) this study's PDFs for S have been normalized to unit probability over 0–20 K. Normalization has a negligible effect on the comparison with the S20 Baseline PDF. **c** Unnormalized PDF for S_{hist} , that from this study (solid cyan line) and the "non-Bayesian S_{hist} PDF" from S20's Fig. 11(b), accurately digitized (dashed orange line). The dotted red line shows the S20 PDF scaled by a factor of 0.86. The black line shows the PDF implied by use of S20's S_{hist} likelihood with their uniform-in- λ prior

which follows from the closeness of its combined evidence likelihood to that derived here, notwithstanding the substantial differences in the Paleoclimate and, particularly, Historical likelihoods at high S .

Figure 4c shows (solid cyan line) this study's posterior PDF for S_{hist} , derived by a sampling-based method, without normalization to unit probability over 0–20 K. The dashed

Table 4 Primary results using S20's data-variable assumptions, sampling-derived integrated likelihoods and Jeffreys' priors. PDFs for S have been normalized to unit probability over 0–20 K except where stated. All values are in K and except for the medians (50th percentiles) are rounded to the nearest 0.05 K

Percentile of posterior PDF for S (as %)	2.5%	5%	17%	50%	83%	95%	97.5%
All combined	2.2	2.3	2.65	3.23	4.1	5.05	5.6
<i>S20 Baseline (UL) results for comparison</i>		2.3	2.6	3.1	3.9	4.7	
Process alone	1.8	1.95	2.3	3.07	4.55	6.75	8.45
Historical alone	1.95	2.15	2.7	4.21	8.15	13.65	16.15
<i>Historical: no restriction to 0–20 K</i>	2.05	2.3	3.0	6.09	None	None	None
Paleoclimate alone	1.45	1.6	2.1	3.02	4.5	6.1	7.05
No Process	2.05	2.2	2.6	3.40	4.65	6.1	7.0
<i>S20 UL No Process results</i>		2.0	2.4	3.1	4.1	5.2	
No Historical	1.95	2.05	2.4	3.04	4.0	5.0	5.6
<i>S20 UL No Historical results</i>		2.0	2.3	2.9	3.7	4.6	
No Paleoclimate	2.1	2.25	2.65	3.39	4.8	6.8	8.35
<i>S20 UL No Paleo results</i>		2.2	2.6	3.3	4.6	6.4	
Percentile of posterior PDF for $S_{\text{hist}}/\text{TCR}$	2.5%	5%	17%	50%	83%	95%	97.5%
Historical S_{hist} : no restriction to 0–20 K	1.75	1.9	2.25	3.38	13.75	None	None
Historical TCR: no restriction to 0–20 K	1.4	1.5	1.7	2.26	4.45	None	None

Note: PDFs for Process alone and Historical alone are directly derived by sampling but are equivalent to their integrated likelihoods multiplied by Jeffreys' priors

orange line shows S20's corresponding "non-Bayesian S_{hist} PDF", derived directly (by sampling) from their (19), the equivalent of (9) here. When S20's PDF is scaled by a factor of 0.86^3 (dotted red line) it closely matches this study's sampling-based S_{hist} PDF (which equates to a Bayesian posterior PDF derived using a noninformative prior).

Although all S20's main probabilistic estimates are based on Bayesian analysis with a uniform-in- λ prior, no such results are given for S_{hist} . However, the solid black line in Fig. 4c shows a uniform-in- λ prior based PDF, using the accurate emulation of S20's S_{hist} likelihood (supplemental material Figure S2.1(a)). This PDF is much better constrained at high S_{hist} than the non-Bayesian sampling-derived PDFs, primarily reflecting S20's misestimation of the S_{hist} likelihood.

Table 4 presents results in the form of medians and 66%, 90% and 95% uncertainty ranges for posterior PDFs for S and S_{hist} on S20's data-variable assumptions, using this study's methods, with the comparative S20 results where available. It is evident from the high percentile S values that Paleoclimate evidence gives the strongest constraints on upper uncertainty bounds, with Historical evidence constraining them least. That is consistent with the relative shapes of the likelihood functions for the three lines of evidence.

Notwithstanding the difficulty the optimization-based profile likelihood method has in deriving a satisfactory data-space movement prior for S20's Historical evidence, the 'All combined' 5% to 95% percentile values from combining that method's results for each line of evidence are all within

0.05 K of those using the sampling-based integrated likelihood method and related Jeffreys' priors, as given in Table 4.

Compared to S20's Baseline combined evidence results, the Table 4 median estimate for S is approximately 0.13 K higher, and the 95% bound 0.35 K higher. The likely reasons for these differences being small are that (i) neither of S20's most seriously inaccurate likelihood estimates were used for its Baseline estimate, resulting in the combined likelihood that they used for that purpose deviating only modestly from that given by this study's methods until both have fallen to a moderate level, as well as being nearly identical to it up to the likelihood maximum; (ii) both S20's uniform-in- λ prior and this study's combined-evidence Jeffreys' prior fall sharply over the high likelihood ($> 0.5 \times$ its maximum) region, by about two-thirds in S20's case, which means that differences in the likelihood and (still declining) prior used beyond that region have a minor effect on S estimates; and (iii) over the high likelihood region the Jeffreys' prior only increases by $\sim 25\%$ relative to a uniform-in- λ prior, which difference is only sufficient to produce small upward shifts in the median and higher percentile S estimates.

PDFs for S used in Table 4 were normalized to unit probability over 0–20 K, except in one case. As discussed in Sect. 3, for combined evidence virtually all probability lies within the 0–20 K range over which computations are performed and over which total probability is normalized to unity. Likewise, almost no probability for S lies outside 0–20 K when combining two or more lines of evidence, or using Paleoclimate evidence alone, and under 1% does for Process evidence alone. However, when using Historical evidence alone 30% of samples produce S values that are above

³ So that over the range plotted in S20 the cumulative probability of the two PDFs is then equal.

Table 5 Results if using S20's uniform-in- λ prior as well as their data-variable assumptions, with PDFs normalized to unit probability over 0–20 K, and comparisons with similarly based results in S20's Table 10. All values are in K and except for the medians (50th percentiles) are rounded to the nearest 0.05 K

Percentile of posterior PDF for S (as %)	2.5%	5%	17%	50%	83%	95%	97.5%
All combined	2.15	2.25	2.6	3.16	4.0	4.85	5.35
<i>S20 Baseline results for comparison</i>		2.3	2.6	3.1	3.9	4.7	
Process alone	1.8	1.95	2.3	3.07	4.55	6.75	8.45
Historical alone	1.9	2.05	2.55	3.80	7.2	12.55	15.3
Paleoclimate alone	1.25	1.4	1.8	2.62	3.85	5.15	5.95
No Process	1.95	2.1	2.45	3.17	4.3	5.5	6.25
<i>S20 UL No Process results</i>		2.0	2.4	3.1	4.1	5.2	
No Historical	1.9	2.05	2.35	2.97	3.85	4.75	5.3
<i>S20 UL No Historical results</i>		2.0	2.3	2.9	3.7	4.6	
No Paleoclimate	2.1	2.25	2.6	3.35	4.7	6.65	8.1
<i>S20 UL No Paleo results</i>		2.2	2.6	3.3	4.6	6.4	

20 K or negative due to $\Delta R > 0$, implying an unstable climate system, and 15% of samples do so for S_{hist} . The substantial proportion of sampled S and S_{hist} values exceeding 20 K primarily reflects the significant probability assigned by S20's data-variable assumptions to highly negative aerosol ERF values: $\Delta F_{\text{Hist}}^{\text{aerosol}} < -2 \text{ W m}^{-2}$ in 17% of samples. Unnormalized results for S and S_{hist} from Historical evidence are therefore also given, without probability being restricted to any range of S values. This unrestricted basis, which correctly reflects the implications of the data-variable uncertainty distributions, is usual for sampling-based energy budget studies that derive S_{hist} (Gregory et al. 2002; Otto et al. 2013; Lewis and Curry 2015, 2018).

S20 did not provide any estimate of the transient climate response (TCR), a shorter-term climate sensitivity measure, however Table 4 does do so. TCR is estimated as for S_{hist} but omitting the deduction for ΔN in (9), a common method (Otto et al. 2013; Lewis and Curry 2015, 2018; Forster et al. 2021 7.5.2.1), so that:

$$\text{TCR} = F_{2\times\text{CO}_2} / (\Delta F_{\text{Hist}} / \Delta T_{\text{Hist}}) \quad (15)$$

The resulting median TCR estimate of 2.26 K exceeds the AR6 likely range, and there is a 7% probability that TCR exceeds 20 K. Moreover, if S20's estimates of the historical pattern effect are accurate then over half of it is unforced (supplemental material S5.2.4) and will have depressed ΔT_{Hist} . Amending (15) to correct for that would increase the implied S20 TCR estimate.

Table 5 presents equivalent results from posterior PDFs based on uniform-in- λ priors, with the comparative S20 results where available. Where Process evidence is not used, the entire posterior probability will in theory be located immediately above zero S , resulting in all S percentiles being almost zero (supplemental material S7). This study's Table 5 results reflect imposing a restriction to $S \geq 0.01 \text{ K}$, which avoids the uniform-in- λ prior producing non-negligible probability at very low S values.

The S values are lower, particularly at high percentiles, than when using Jeffreys' priors, except for Process alone where the two types of prior and hence the S values are identical. This behavior reflects the fact that, while all the Jeffreys' priors decrease with S , they decline less rapidly than the uniform-in- λ prior. Equivalently, when transformed to λ -space, all the non-Process Jeffreys' priors increase with S , whereas the uniform-in- λ prior does not. The S values at higher percentiles derived here using a uniform-in- λ prior differ from those per S20's Baseline results, as would be expected given the identified differences between their likelihoods. However, the differences in S values are small, only reaching 0.1 K beyond the median. That is consistent with the differences in likelihoods only becoming significant beyond the medians; since the uniform-in- λ prior varies with S^{-2} the effect on the posterior PDFs of sizeable likelihood differences is muted at higher S values.

Comparing the present study's results using the two types of prior, the differences between their 95th percentile S values are quite significant when Historical and/or Paleoclimate evidence are used, either alone (differences of $\sim 1 \text{ K}$) or in combination (a difference of 0.6 K, or 0.9 K using S20's results). When Process evidence is used in combination with Historical and/or Paleoclimate evidence the differences are smaller. When combining all three lines of evidence, the 95% bound is only 0.2 K lower when using a uniform-in- λ prior, and the difference in medians is under 0.1 K. That is mainly due to the combined evidence being much more informative about S , and constraining it more tightly, than any of the separate lines of evidence, and partly due to contributions to the combined Jeffreys' prior from Process and Historical evidence respectively being uniform-in- λ , and increasing only gently relative to a uniform-in- λ prior.

S20 also present various posterior PDFs based on a uniform-in- S prior. That prior is unsuitable for S estimation unless ΔT uncertainty dominates, which it almost never does, and will often result in uncertainty ranges that are far from being true confidence intervals. For Historical

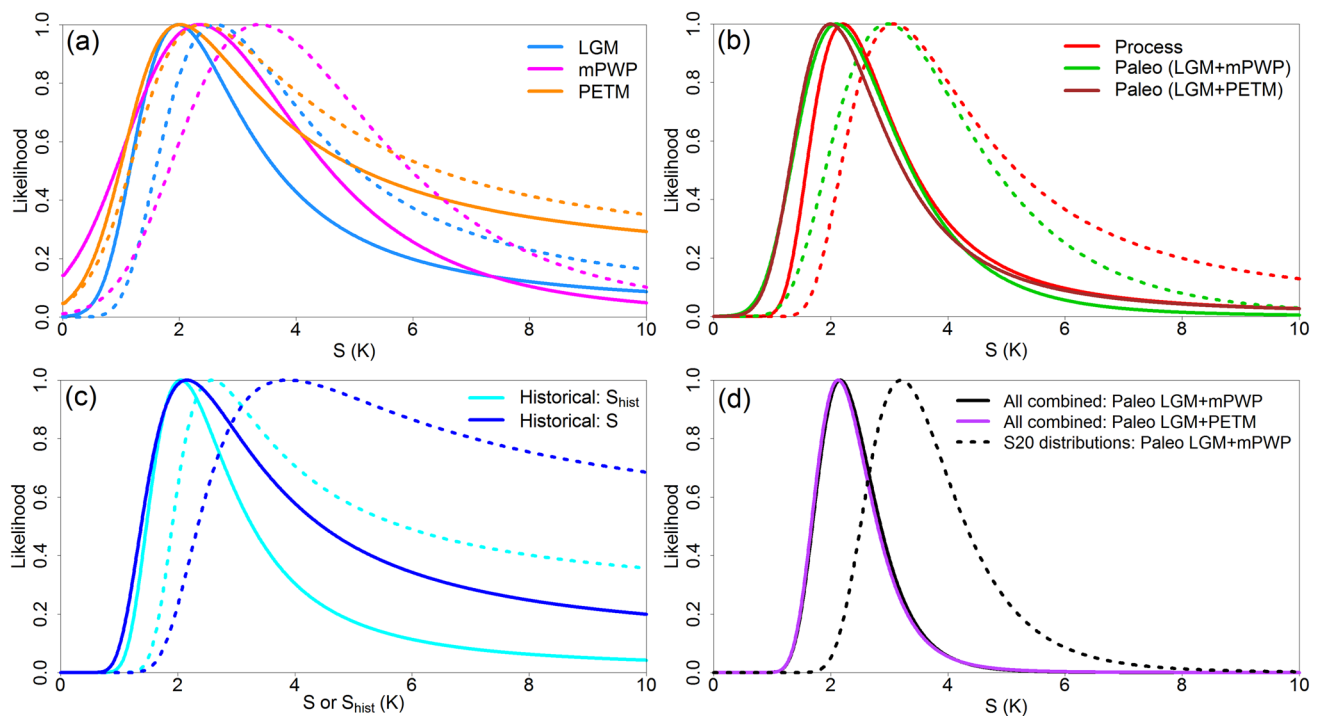


Fig. 5 Likelihoods for S based on this study's revised data-variable assumptions (solid lines) and, for comparison, those derived in this study using S20's data-variable assumptions (same color, dotted, lines). Likelihoods, on both sets of data-variable assumptions, have been derived using the integrated likelihood method and normalized to a maximum of one. **a** Likelihoods from evidence for the three Paleoclimate periods: LGM, mPWP and PETM. **b** Likelihoods from

Process evidence and from combining Paleoclimate evidence from the LGM and either the mPWP or the PETM. S20 did not combine LGM and PETM evidence so no comparison dotted line is shown for this case. **c** Likelihoods from Historical evidence for both S and S_{hist} . **d** Likelihoods from combining Process and Historical evidence with Paleoclimate evidence from the LGM combined with that from either the mPWP or the PETM

evidence, use of a uniform-in- S prior has been shown to be unsuitable and to result in seriously biased estimation (Annan and Hargreaves 2011; Lewis 2014). For Process evidence, the likelihood for λ is normal in λ space, a case for which there is general agreement that use of a uniform prior is appropriate. When transformed to S space, the resulting appropriate prior for S for Process evidence will (applying the Jacobian factor) be proportional to S^{-2} , very far from uniform.⁴ Most damningly, use of a uniform-in- S prior would result, even for 'All combined' evidence, in the S values for all percentiles being an unbounded function of the upper bound of the S range over which normalization to unit probability occurs, and to increase without limit as that bound tends to infinity.⁵

6 Results using the revised data-variable assumptions

Results are now presented using the data-variable assumptions as revised per Sect. 4, including rectification of the omission of the necessary scaling of $F_{2\times\text{CO}_2}$ when using Process and Historical evidence ("the revised assumptions"), employing the integrated likelihood method with computed Jeffreys' priors.

6.1 Likelihoods

Figure 5 shows likelihoods derived using the revised assumptions (solid lines), with similarly calculated likelihoods using S20's data-variable assumptions in the same colors but dotted. The revised assumptions produce likelihoods that peak at lower S values, and are lower at high S values, than do S20's assumptions.

The likelihoods for both separate and combined evidence derived using the integrated likelihood method, the data doubling method, and the profile likelihood method, are all almost identical when employing the revised assumptions

⁴ Ignoring uncertainty in $F_{2\times\text{CO}_2}$, which has a negligible effect on the transformation over the 2.5%–97.5% uncertainty range.

⁵ Because use of a uniform-in- S prior means the PDF for S is a scaled version of its likelihood, which remains above zero as $\lambda \rightarrow 0$ and hence $S \rightarrow \infty$.

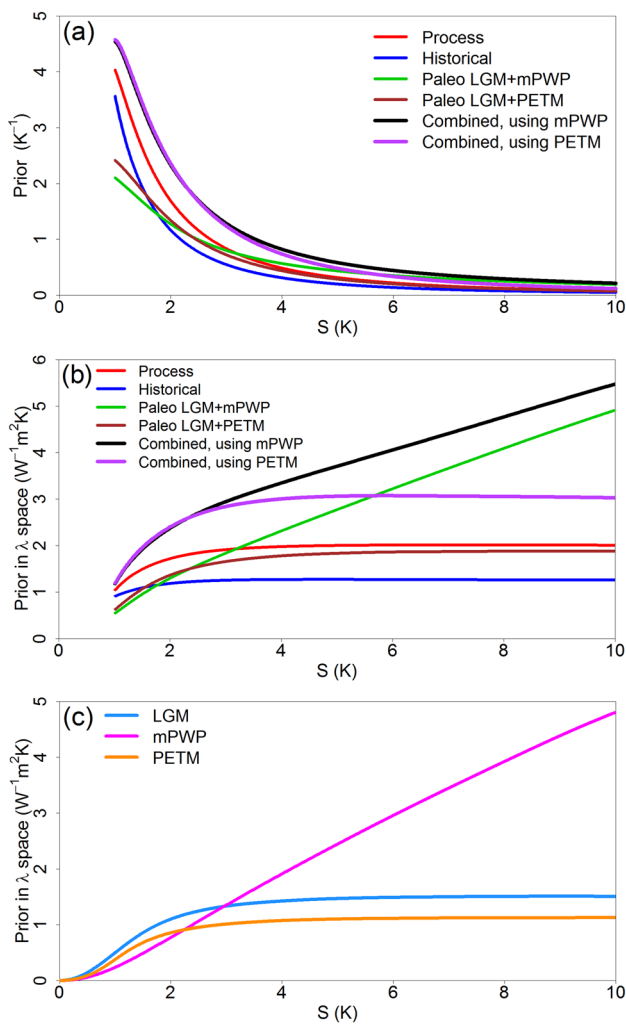


Fig. 6 Computed Jeffreys' priors based on this study's revised data-variable assumptions. **a** Prior in S -space for each main line of evidence and for the combined evidence from all three types. Paleoclimate evidence from the LGM and PETM combined is included as well that from the LGM and mPWP combined. **b** As for (a) but with the S priors transformed into priors in λ -space (that is, priors for λ , but plotted at the corresponding S values). **c** Priors transformed into λ -space for evidence from individual paleoclimate periods. The transformations from S -space to λ -space are effected by multiplying them by $S^2/F_{2\times CO_2}$, for simplicity ignoring the minor uncertainty in $F_{2\times CO_2}$ and hence not being exact

(supplemental material Figures S4 to S7). This includes, unlike on S20's data-variable assumptions, likelihoods from the doubled-data method applied to Historical evidence, confirming that it is difficult incorporating the unusual historical aerosol forcing distribution employed by S20 that prevents data-doubling working satisfactorily in that case.

6.2 Computed Jeffreys' priors

Figure 6a shows the computed and calibrated Jeffreys' priors for S for each main line of evidence, with Paleoclimate

represented by LGM evidence combined with either mPWP or PETM evidence, and for all main lines of evidence combined, from $S = 1$ K upwards. Below that level, they start to become artefacted—as they do below ~ 1.5 K when using S20's data-variable assumptions—due to the paucity of samples for Process and Historical evidence. The likelihood and the probability are both almost zero below those levels, so the effect on inference for S is negligible.

In all cases the priors decline rapidly with increasing S , reflecting declining informativeness of the evidence about S , making it difficult to compare their behavior at high S levels. Figure 6b shows these priors transformed into λ -space—that is converted into priors for λ —but plotted against the corresponding values of S . Figure 6c shows the transformed priors for LGM, mPWP and PETM Paleoclimate evidence separately. Strictly, because for computational reasons the relatively small uncertainty in $F_{2\times CO_2}$ is ignored when effecting the transformation, it is into priors for $\overline{F_{2\times CO_2}}/S$, whereas $\lambda = F_{2\times CO_2}/S$. The transformed priors show how informative the evidence is about λ at each S value, which varies comparatively little. Their shapes are all similar when using the revised data-variable assumptions to those when using S20's assumptions (Fig. 3b, c), except that the revised Historical evidence prior flattens out much earlier with rising S . For Process evidence, the prior should in both cases be uniform (constant) when transformed to λ -space—the same prior as used for all lines of evidence in S20—since the λ likelihood is for the sum of normally distributed variables. The drop in the transformed priors that occurs at low S is due to uncertainty in $F_{2\times CO_2}$ not being removed by the transformation. At low S values, which arise where ΔR is large and ΔT is small, fractional uncertainty in ΔT translates to significant absolute uncertainty in λ , hence Paleoclimate evidence (which has large ΔT uncertainty) is relatively uninformative.

At low S values, Process evidence is most informative, and dominates the combined evidence prior. As S increases, mPWP and hence Paleoclimate evidence becomes increasingly more informative, due mainly to fractional uncertainty in ΔR being low for the mPWP, and at high S dominates the combined evidence prior. Historical evidence is less informative than Process evidence at all S values, although only modestly so at high S when using S20's data-variable assumptions. The priors generally have slightly lower values when using the revised rather than S20's original data-variable assumptions. The reasons for this are discussed in the supplemental material (S6).

In all cases the priors derived using the doubled data and profile likelihood data-space movement methods are close to those derived using the primary integrated likelihood method over S of 1–10 K, outside which virtually zero probability lies (supplemental material Figures S8, S9). In some Paleoclimate cases there is minor divergence at $S > 4$ K when using the data-space movement method,

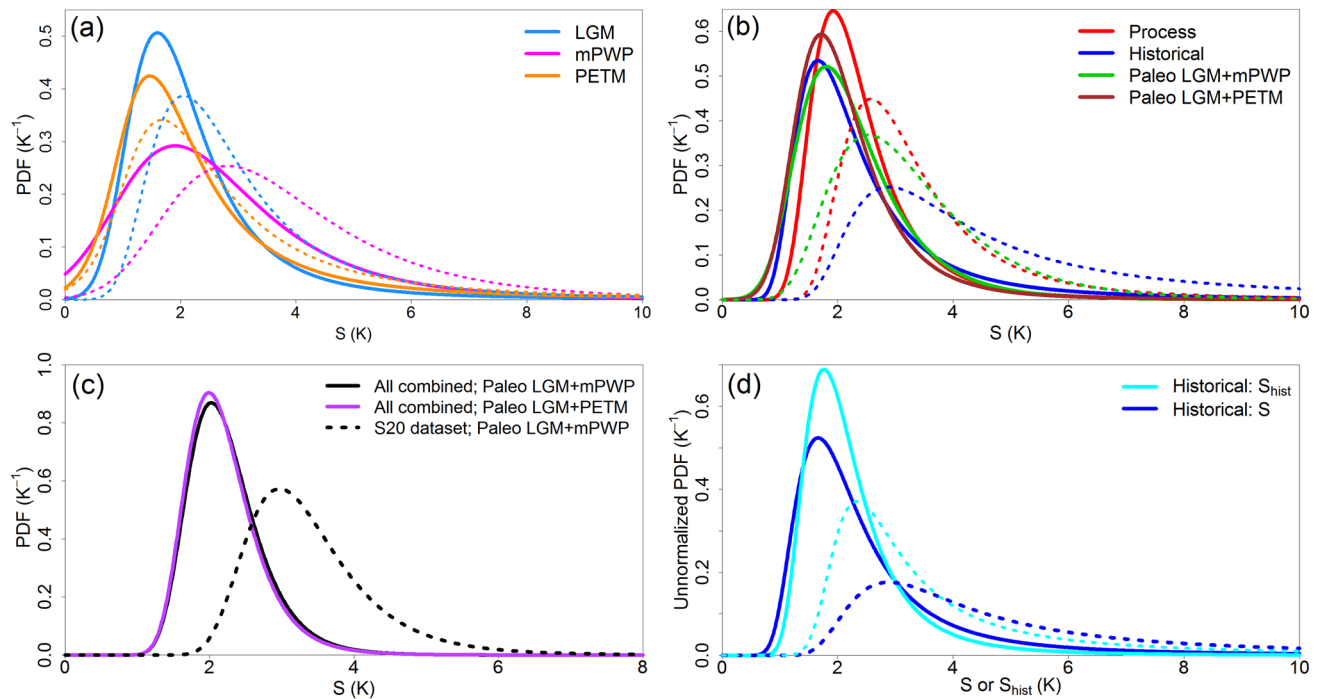


Fig. 7 Posterior PDFs for S based on the revised data-variable assumptions (solid lines) and, for comparison, on S20's original assumptions (dotted lines), in both cases using the sampling-based integrated likelihood method. The PDFs for individual separate lines of evidence, are obtained directly by sampling, but equate to the product of the estimated integrated likelihood and derived Jeffreys' prior. Save in panel (a), comparative PDFs involving the PETM are not given, since S20 did not provide any such PDFs. Except in panel (d), the PDFs have all been normalized to unit probability over

0–20 K. **a** PDFs for S from Paleoclimate LGM, mPWP and PETM evidence separately. **b** PDFs for S from separate Process, Historical and Paleoclimate (LGM combined with mPWP or PETM) evidence. **c** Combined evidence PDFs derived using, as Paleoclimate evidence, that from the LGM combined with that from either the mPWP (black lines) or the PETM (purple line). **d** Unnormalized PDFs for S and S_{hist} . These account for probability outside 0–20 K, which is substantial when using S20's original assumptions

but for combined Paleoclimate evidence little probability exists there. Although all the priors have been calibrated to equal (the square root of) Fisher information, those derived using the profile likelihood data-space movement method are almost identical with and without calibration, confirming the soundness of that method. These comments apply also (at $S > 1.5$ K) to the priors derived when using S20's data-variable assumptions (Fig. 3), save that only the integrated likelihood method produces a usable prior for S20's Historical evidence (see Sects. 3.3 and 3.5), and hence for the combined evidence.

6.3 Posterior PDFs and percentiles for S

The estimated posterior PDFs based on the revised assumptions (Fig. 7, solid lines) all peak at lower S values, and are better constrained beyond their peaks, than those based on S20's assumptions (shown dotted).

Figure 7a shows the PDFs for separate lines of Paleoclimate evidence, derived using the revised data-variable assumptions and, for comparison, using S20's assumptions.

Figure 7b shows PDFs from combining Paleoclimate LGM evidence with that from the mPWP or PETM, and for separate Process and Historical evidence. Figure 7c shows the final PDFs after combining Process, Historical and Paleoclimate evidence. The PDFs incorporating PETM evidence are almost identical to those incorporating mPWP evidence; they peak marginally earlier, and at a slightly higher level.

Figure 7d shows unnormalized posterior PDFs for S and S_{hist} . The PDFs for S are the same as in panel (b) except for the lack of normalization. When using the revised assumptions, there is little difference between the normalized and unnormalized PDFs, as only 2.0% of the probability lies beyond $S = 20$ K, down from 30% when using S20's assumptions.

Posterior PDFs computed using the doubled data method, or the profile likelihood method and its data-space movement prior, are visually identical to those using the primary integrated likelihood method shown in Fig. 7, save for a marginal difference in the peak PDF level for separate LGM and PETM evidence (supplemental material Figures S10 to S13). For separate evidence, the sampling based integrated likelihood and doubled data method both derive their PDFs

Table 6 Primary results using the revised data-variable assumptions and Jeffreys' priors. All values are in K and except for the medians (50th percentiles) are rounded to the nearest 0.05 K

Percentile of posterior PDF for S (as %)	2.5%	5%	17%	50%	83%	95%	97.5%
All combined: Paleo LGM + mPWP	1.45	1.55	1.75	2.16	2.7	3.2	3.5
<i>Same but using original assumptions</i>	2.2	2.3	2.65	3.23	4.1	5.05	5.6
All combined: Paleo LGM + PETM	1.45	1.5	1.75	2.12	2.65	3.15	3.45
No Process: Paleo LGM + mPWP	1.25	1.35	1.6	2.12	2.85	3.65	4.15
No Process: Paleo LGM + PETM	1.25	1.35	1.6	2.05	2.75	3.55	4.1
No Historical: Paleo LGM + mPWP	1.4	1.5	1.75	2.16	2.8	3.4	3.75
No Historical: Paleo LGM + PETM	1.35	1.45	1.7	2.12	2.7	3.35	3.7
No Paleoclimate	1.35	1.5	1.7	2.19	2.9	3.7	4.25

Table 7 Sensitivity of the primary combined evidence results (using LGM + mPWP for Paleo) to the various revisions of data-variable assumptions, using the integrated likelihood method and Jeffreys' priors. All values are in K and except for the medians (50th percentiles) are rounded to the nearest 0.05 K

Percentile of posterior PDF for S (as %)	2.5%	5%	17%	50%	83%	95%	97.5%
Using S20's original data-variable assumptions	2.2	2.3	2.65	3.23	4.1	5.05	5.6
<i>After adopting revisions relating to:</i>							
Use of CO ₂ ERF & abrupt2/4× simulations (ζ & γ)	1.85	1.95	2.3	2.82	3.6	4.35	4.85
Also AR6 Historical ex-aerosol ERF & ΔT basis ^a	1.7	1.85	2.15	2.64	3.35	4.1	4.55
Also, per later data/newer evidence (non-cloud)	1.6	1.7	1.95	2.39	3.0	3.65	4.05
Also cloud feedback, per newer evidence	1.5	1.6	1.85	2.25	2.8	3.4	3.75
Also LGM ΔT and ΔF_{exCO_2} (preferred evaluation)	1.45	1.55	1.75	2.15	2.7	3.25	3.55
All revisions, including to Historical aerosol ERF	1.45	1.55	1.75	2.16	2.7	3.2	3.5

^aAR6 Historical ERF time series are used to estimate ΔF_{other} , but only to scale the main 1850 to 2005–2015 $\Delta F_{\text{aerosol}}$ estimate to a 1861–80 to 2006–18 change. The AR6 revision to the ΔT basis relates to changing S20's estimated GMAT—GMST adjustment to match the AR6 zero estimate of their difference

directly by sampling, so they are bound to be identical in these cases. However, PDFs computed using the non-sampling profile likelihood method and its very differently derived data-space movement prior are completely independent of those from the integrated likelihood and doubled data methods.

Table 6 presents this study's primary results, in the form of medians and 66%, 90% and 95% uncertainty ranges from posterior PDFs, on the revised data-variable assumptions, derived using the integrated likelihood method and Jeffreys' priors. Results based on combining different pairs of lines of evidence, as well as all of them, are given. PDFs for S used in Table 6 were normalized to unit probability over 0–20 K; almost no probability ($\leq 0.1\%$) lies outside that range.

As when using S20's assumptions, the S values at high percentiles confirm that Historical evidence is least important for constraining the upper uncertainty bounds, but Process evidence now constrains them almost as strongly as Paleoclimate evidence.

The limited revisions made to S20's assumptions reduce by one-third, from 3.23 to 2.16 K, the median estimate of S given by the combined evidence, using Jeffreys' priors and warm Paleoclimate evidence from the mPWP in both cases. The 83% and 95% uncertainty bounds reduce respectively from 4.1 to 2.7 K and from 5.05 to 3.2 K. If warm Paleoclimate evidence is instead taken from the PETM when the

revised assumptions are used, the PDF percentiles from the median upwards reduce further, by ~ 0.05 K.

All the profile likelihood method (using the associated data-space movement prior) derived percentile values are within ± 0.02 K of those in Table 6, with medians identical, when combining evidence in two stages (supplemental material S4). Moreover, when using the revised assumptions, the profile likelihood optimization process can simultaneously combine Process, Historical, LGM, and either mPWP or PETM evidence, and hence produce, in a single step, a posterior PDF for all lines of evidence combined. The 1% to 99% percentile points of those two PDFs match those from the sampling-based integrated likelihood method within ± 0.02 K.

The median S values when omitting evidence from each of the three main lines in turn, with Paleoclimate evidence combining LGM evidence with that from either the mPWP or PETM, are all within 0.1 K of the average of the two 'All combined' values.

It is useful to establish how sensitive the combined-evidence results are to the various categories of revisions to data-variable assumptions. A few of the revisions might be regarded as more questionable since they are based wholly or partly on reevaluation of existing evidence. That category includes LGM cooling and non-CO₂ forcing, and also the revision to Historical aerosol forcing, which although largely

Table 8 Percentile results for individual paleoclimate periods, and for each main line of evidence, separately, including for S_{hist} and TCR, using the revised assumptions and, in italics in the following row,

using the same methods but S20's assumptions. All values are in K and except for the medians (50th percentiles) are rounded to the nearest 0.05 K

Percentile of posterior PDF for S (as %)	2.5%	5%	17%	50%	83%	95%	97.5%
Paleoclimate: LGM alone	0.75	0.9	1.25	1.95	3.15	4.9	6.35
<i>Using original data-variable assumptions</i>	<i>1.15</i>	<i>1.3</i>	<i>1.75</i>	<i>2.64</i>	<i>4.35</i>	<i>7.05</i>	<i>9.1</i>
LGM: without restriction to 0–20 K	0.75	0.9	1.25	1.95	3.15	5.1	7.0
Paleoclimate: mPWP alone	0.35	0.6	1.25	2.42	4.2	6.3	7.7
<i>Using original data-variable assumptions</i>	<i>1.05</i>	<i>1.35</i>	<i>2.1</i>	<i>3.46</i>	<i>5.55</i>	<i>8.0</i>	<i>9.55</i>
mPWP: without restriction to 0–20 K	0.05	0.4	1.15	2.38	4.2	6.3	7.75
Paleoclimate: PETM alone	0.45	0.65	1.1	1.94	3.65	6.85	9.55
<i>Using original data-variable assumptions</i>	<i>0.55</i>	<i>0.8</i>	<i>1.3</i>	<i>2.33</i>	<i>4.55</i>	<i>8.4</i>	<i>11.25</i>
PETM: without restriction to 0–20 K	0.4	0.65	1.15	2.01	4.2	17.0	None
Paleoclimate: LGM + mPWP	0.9	1.05	1.45	2.10	3.1	4.2	4.9
<i>Using original data-variable assumptions</i>	<i>1.45</i>	<i>1.6</i>	<i>2.1</i>	<i>3.02</i>	<i>4.5</i>	<i>6.1</i>	<i>7.05</i>
Paleoclimate: LGM + PETM	0.95	1.05	1.4	1.98	2.9	4.05	4.95
<i>Using original data-variable assumptions</i>	<i>1.25</i>	<i>1.4</i>	<i>1.8</i>	<i>2.59</i>	<i>3.95</i>	<i>5.75</i>	<i>7.05</i>
Process	1.3	1.4	1.65	2.21	3.1	4.3	5.15
<i>Using original data-variable assumptions</i>	<i>1.8</i>	<i>1.95</i>	<i>2.3</i>	<i>3.07</i>	<i>4.55</i>	<i>6.75</i>	<i>8.45</i>
Historical	1.05	1.15	1.45	2.13	3.55	6.1	8.3
<i>Using original data-variable assumptions</i>	<i>1.95</i>	<i>2.15</i>	<i>2.7</i>	<i>4.21</i>	<i>8.15</i>	<i>13.65</i>	<i>16.15</i>
Historical: without restriction to 0–20 K	1.05	1.2	1.5	2.15	3.7	7.6	14.75
<i>Using original data-variable assumptions</i>	<i>2.05</i>	<i>2.3</i>	<i>3.0</i>	<i>6.09</i>	<i>None</i>	<i>None</i>	<i>None</i>
Percentile of posterior PDF for S_{hist} /TCR	2.5%	5%	17%	50%	83%	95%	97.5%
Historical S_{hist} : no restriction to 0–20 K	1.2	1.3	1.55	2.05	2.95	4.25	5.35
<i>Using original data-variable assumptions</i>	<i>1.75</i>	<i>1.9</i>	<i>2.25</i>	<i>3.38</i>	<i>13.75</i>	<i>None</i>	<i>None</i>
Historical TCR: no restriction to 0–20 K	1.0	1.05	1.25	1.54	2.0	2.45	2.75
<i>Using original data-variable assumptions</i>	<i>1.4</i>	<i>1.5</i>	<i>1.7</i>	<i>2.26</i>	<i>4.45</i>	<i>None</i>	<i>None</i>

based on newer evidence concerns a very poorly constrained forcing.

Table 7 divides the revisions made into six categories, starting with those that appear least debatable, being that to $F_{2\times\text{CO}_2}$ together with appropriate adjustments (omitted in S20) to the calculation of the CO_2 ERF estimates used, or which arise from alignment of the CO_2 concentrations used for estimating of the ECS to S ratio. Without these changes, the bases of S estimation are biased and are not consistent between lines of evidence.

The next set of changes relate primarily to the substitution of the AR6 Historical non-aerosol ERF time series estimates, and of the AR6 zero-mean estimate of the difference between Historical GMST and GMAT warming, for the estimates used by S20. Together, these two categories of revision reduce the median S estimate by almost 0.6 K, and the 95% uncertainty bound by almost 1 K.

The third category comprises all other revisions that are based entirely on newer evidence or later data, other than cloud feedback, the least well constrained feedback. The main changes involved are to scaling factors used for

estimation of S from mPWP evidence, with those factors being derived from a more recent model intercomparison project than previously, and to the estimate of the Historical pattern effect. These revisions reduce the median S estimate by a further 0.25 K. Next, cloud feedback is revised, which reduces the median S estimate by almost 0.15 K more, to 2.25 K, with the 95% uncertainty bound now down to 3.4 K. Including the penultimate category, of revisions to LGM cooling and non- CO_2 forcing, brings the median S estimate down by another 0.1 K, to 2.15 K.

Finally, and somewhat counter intuitively, revising the Historical aerosol ERF distribution, with the resulting S estimation basis now being the same as that in Table 6, does not further reduce the median S estimate. That estimate remains unchanged within computational uncertainty, although the 95% (and 97.5%) uncertainty bounds reduce by a further 0.05 K. Investigation suggests the principal cause is likely that, although the aerosol ERF distribution used by S20 has a median that (after scaling between periods using the AR6 time series) is some 0.2 Wm^{-2} more negative than for the adopted revised distribution, its mode

is actually the less negative of the two. The shapes of the two distributions are such that the Historical S likelihoods resulting from their use, with other revisions to Historical data-variables having been made, are very similar below approximately 2 K. While highly negative aerosol ERF values, which correspond to high S values, are much more probable when using the S20 aerosol distribution, resulting in a larger Historical S likelihood, those high S values are almost ruled out by their low likelihood from Process and Paleoclimate evidence.

Table 8 presents sampling-derived medians and 66%, 90% and 95% uncertainty range bounds for S , for the three main lines of evidence and for each Paleoclimate period separately, using the revised and (in italics) the original S20 assumptions. S20 did not provide any similar estimates. Posterior PDFs for S were normalized to unit probability over 0–20 K except where stated; results are also given without excluding probability outside that range, where it is non-negligible.

Consistency between median S values from different lines of evidence is much improved when based on the revised assumptions; they span (without restriction to 0–20 K) only 1.9–2.4 K, compared to 2.4–6.1 K using S20's assumptions.

The median TCR is 1.54 K, within the AR6 likely range. The median TCR and S_{hist} estimates are respectively almost one-third lower, and almost 40% lower, than when using S20's assumptions.

7 Discussion

This study first identifies statistical problems in S20. Using a Subjective Bayesian statistical method involving an investigator-selected prior distribution, as S20 does, may produce unrealistic climate sensitivity estimation when used to combine differing types of evidence (Lewis 2018), even assuming that the data likelihood functions are correct. In this case, I found that the method S20 used for estimating likelihoods for all but Process evidence was in fact unsound, and that it underestimated likelihood at high S levels, substantially so in some cases. I also found that S20 used an uncertainty estimate for PETM CO_2 forcing that was a factor of ten too low, due to an apparent coding error, further biasing their likelihood estimate (although not affecting their main results).

This study then develops an Objective Bayesian approach to combining differing climate sensitivity evidence that, unlike the method used in Lewis and Grünwald (2018), is not restricted to dealing with a particular simple statistical model. The approach involves computationally deriving Jeffreys' prior distributions that are designed to maximize the influence of the data on the results and to produce probabilistic estimates that are as close as possible to being

confidence intervals, and thus are well calibrated. Three different inferential methods employed for this purpose each provide nearly identical estimated likelihoods and Jeffreys' priors, and final results. This result is very supportive of the validity of the methods used and of the results they produce.

The robustness of S20's results to the use of properly calibrated statistical methods and validly calculated likelihood estimates is then examined, using the Objective Bayesian methods developed in this study. It is shown that while S20's choice of prior and its likelihood misestimation lead to over-constraining of high S levels, based on S20's data-variable assumptions the downwards bias in S20's Baseline combined evidence results is modest: the median S estimate is approximately 0.13 K low, and the 95% uncertainty bound 0.35 K low. However, the bias in S20's No Process results is over twice as large.

The other main contribution of this study is to assess the impact of revising various input data-variable distributions used by S20, by:

- (i) adjusting the $F_{2\times\text{CO}_2}$ value used for inferring S from Process and Historical evidence to reflect the effect of climate feedback changing over GCM abrupt4x CO_2 simulations, as should undoubtedly be done;
- (ii) allowing for the CO_2 concentration-ERF relationship being slightly non-logarithmic, and estimating the ECS to S ratio in a way that is unaffected by that relationship;
- (iii) changing some of S20's other data-variable estimates to reflect more recent information; and
- (iv) using arguably better justified (albeit not based purely on more recent information), alternative estimates for a few other data-variables.

I find that doing so results in substantially lower and better constrained estimates for S . The median S estimate when combining all lines of evidence, using the Objective Bayesian method and the LGM and mPWP for Paleoclimate evidence, reduces from 3.23 to 2.16 K.

All the revised data-variable estimates are not only defensible but, given the evidence now available, in my view are better justified than S20's original estimates. Moreover, omitting the only revisions dependent, to a greater or lesser extent, on reevaluation of existing evidence only very modestly changes the combined evidence results, with the omission just of the revision of the Historical aerosol forcing having almost no effect on the results.

It therefore currently remains quite plausible that S is below 2 K. The truncation in S20's results of the lower bound for S does not appear justified given the range of data-variable estimates supported by relevant, mainly more recent, studies. There is 36% probability of S being under 2 K, considerably greater than the 26% probability of S

exceeding 2.5 K, according to the revised data-variable assumptions 'All combined: Paleo LGM + mPWP' results; they also imply that it is extremely unlikely that S is below 1.5 K, and extremely unlikely that S is above 3.2 K.

The revised data-variable median Historical evidence estimates of S_{hist} and TCR are somewhat higher than the comparable estimates in Lewis and Curry (2018), of 1.66 K and 1.33 K respectively. The excess is mainly due to a stronger aerosol ERF change, even after revising S20's assumptions. Further revising S20's median aerosol ERF to match the change per the AR5 time-series, extended post-2011 using AR6's annual changes, would reduce the Table 8 median S_{hist} and TCR to respectively 1.82 K and 1.40 K. Changing the base period to 1869–1882 to match Lewis and Curry (2018), avoiding the poorly observed 1861–1868 period, would further reduce those estimates, to 1.79 K and 1.37 K. The methane shortwave ERF adjustment, and greater estimated change in radiative imbalance, in AR6 can account for the small remaining differences.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00382-022-06468-x>.

Acknowledgements I thank Judith Curry, Ross McKittrick, and Frank Bosse for commenting on drafts, and Mark Zelinka for contributing data. I thank Tamas Bodai and another (anonymous) reviewer for constructive comments that helped improve the submitted manuscript. I acknowledge the World Climate Research Programme's (WCRP) Working Group on Coupled Modelling and the climate modeling groups for producing and making their output available.

Funding No funding has been received for this work.

Declarations

Conflict of interest The author has no relevant interests to disclose.

Data availability All data generated or analysed during this study are included in this published article or are available from publicly accessible archive sites, from its supplementary information files or from its cited references. Raw CMIP5 and CMIP6 abrupt4xCO2 simulation data are available at <https://esgf-node.llnl.gov/projects/esgf-llnl/>.

Code availability Code to generate this paper's results is available from the author on reasonable request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Andrews T, Gregory JM, Paynter D, Silvers LG, Zhou C, Mauritsen T, Webb MJ, Armour KC, Forster PM, Titchner H (2018) Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophys Res Lett* 45(16):8490–8499
- Annan JD, Hargreaves JC (2011) On the generation and interpretation of probabilistic estimates of climate sensitivity. *Clim Change* 104(3):423–436. <https://doi.org/10.1007/s10584-009-9715-y>
- Annan JD, Hargreaves JC (2013) A new global reconstruction of temperature changes at the Last Glacial Maximum. *Clim Past* 9(1):367–376. <https://doi.org/10.5194/cp-9-367-2013>
- Bayes T (1763) An essay towards solving a problem in the doctrine of chances. *Philos Trans R Soc Lond* 53 (1763) 370–418; 54 (1764) 269–325. Reprinted in *Biometrika* 45 (1958), 293–315.
- Bellouin N, Quaas J, Gryspeerdt E, Kinne S, Stier P, Watson-Parris D, Boucher O, Carslaw K.S., Christensen M., Daniau A.L. and Dufresne J.L. (2020) Bounding global aerosol radiative forcing of climate change. *Reviews of Geophysics*, 58(1), p.e2019RG000660. <https://doi.org/10.1029/2019RG000660>
- Berger JO, Bernardo JM (1992) On the development of reference priors (with discussion). In: Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) *Bayesian Statistics 4*. Oxford University Press, pp 35–60
- Bernardo JM (1979) Reference posterior distributions for Bayesian inference (with discussion). *J Roy Stat Soc Ser B* 41:113–147
- Bernardo JM (2009) Modern Bayesian inference: foundations and objective methods. In: Bandyopadhyay P, Forster M (eds) *Philosophy of statistics*. North Holland, Oxford, pp 263–306
- Bernardo JM, Smith AFM (1994) *Bayesian theory*. Wiley, 608pp
- Bernardo JM (2011) Modern Bayesian Inference: Foundations and Objective Methods, 263–306. In *Philosophy of Statistics*, P. Bandyopadhyay and M. Forster, eds. North Holland, 1253 pp
- Birnbaum A (1962) On the foundations of statistical inference (with discussion). *J Am Stat Assoc* 57:269–332
- Byrne B, Goldblatt C (2014) Radiative forcing at high concentrations of well-mixed greenhouse gases. *Geophys Res Lett* 41:152–160. <https://doi.org/10.1002/2013gl058456>
- Caldwell PM, Bretherton CS, Zelinka MD, Klein SA, Santer BD, Sanderson BM (2014) Statistical significance of climate sensitivity predictors obtained by data mining. *Geophys Res Lett* 41:1803–1808. <https://doi.org/10.1002/2014GL059205>
- Caldwell PM, Zelinka MD, Klein S, A.: Evaluating Emergent Constraints on Equilibrium Climate Sensitivity, (2018) *J Climate* 31:3921–3942
- Cesana GV, Del Genio AD (2021) Observational constraint on cloud feedbacks suggests moderate climate sensitivity. *Nat Clim Chang* 11(3):213–218
- Charney JG (1979) *Carbon dioxide and climate: a scientific assessment*. National Academies of Science Press, Washington, DC, p 22
- Donahue AS, Caldwell PM (2018) Impact of physics parameterization ordering in a global atmosphere model. *Journal of Advances in Modeling Earth Systems* 10(2):481–499. <https://doi.org/10.1002/2017MS001067>
- Efron B (1993) Bayes and likelihood calculations from confidence intervals. *Biometrika* 80:3–26
- Etminan M, Myhre G, Highwood EJ, Shine KP (2016) Radiative forcing of carbon dioxide, methane, and nitrous oxide: a significant revision of the methane radiative forcing. *Geophys Res Lett* 43:12614–12623. <https://doi.org/10.1002/2016GL071930>
- Forster P et al (2021) The Earth's energy budget, climate feedbacks, and climate sensitivity. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate*

- Change [Masson-Delmotte, V. et al. (eds.)]. Cambridge University Press
- Fraser DA (2011) Is Bayes posterior just quick and dirty confidence? *Stat Sci* 26(3):299–316
- Fraser DAS, Reid N (2011) On default priors and approximate location models. *Br J Probab Stat* 25(3):353–361
- Fraser DAS, Reid N, Marras E, Yi GY (2010) Default priors for Bayesian and frequentist inference. *J R Stat Soc B* 72(5):631–654
- Fueglistaler S, Silvers LG (2021) The peculiar trajectory of global warming. *J Geophys Res Atmos* 26:e2020JD033629. <https://doi.org/10.1029/2020JD033629>
- Glassmeier F, Hoffmann F, Johnson JS, Yamaguchi T, Carslaw KS, Feingold G (2021) Aerosol-cloud-climate cooling overestimated by ship-track data. *Science* 371(6528):485–489
- Gregory JM, Stouffer RJ, Raper SCB, Stott PA, Rayner NA (2002) An observationally based estimate of the climate sensitivity. *J Clim* 15:3117–3121
- Gryspeerd E et al (2019) Constraining the aerosol influence on cloud liquid water path. *Atmos Chem Phys* 19:5331–5347
- Gulev SK et al (2021) Changing state of the climate system. In: *Climate Change 2021: The Physical Science Basis*. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V. et al. (eds.)]. Cambridge University Press
- Hamilton DS et al (2018) Reassessment of pre-industrial fire emissions strongly affects anthropogenic aerosol forcing. *Nat Commun* 9(1):1–12. <https://doi.org/10.1038/s41467-018-05592-9>
- Hartigan JA (1965) The asymptotically unbiased prior distribution. *Ann Math Statist* 36(4):1137–1152
- Haywood AM et al (2020) The pliocene model intercomparison project phase 2: large-scale climate features and climate sensitivity. *Clim past* 16(6):2095–2123. <https://doi.org/10.5194/cp-16-2095-2020>
- Jeffreys H (1946) An invariant form for the prior probability in estimation problems. *Proc Roy Soc A* 186:453–461
- Kass RE, Wasserman L (1996) The selection of prior distributions by formal rules. *J Am Stat Assoc* 91(435):1343–1370
- Klein SA, Hall A, Norris JR, Pincus R (2017) Low-cloud feedbacks from cloud-controlling factors: a review. Shallow clouds, water vapor, circulation, and climate sensitivity, pp 135–157
- Kohler P, Bintanja R, Fischer H, Joos F, Knutti R, Lohmann G, Masson-Delmotte V (2010) What caused Earth's temperature variations during the last 800,000 years? Data-based evidence on radiative forcing and constraints on climate sensitivity. *Q Sci Rev* 29(1.2):129–145. <https://doi.org/10.1016/j.quascirev.2009.09.026>
- Lee SS, Chu JE, Timmermann A, Chung ES, Lee JY (2021) East Asian climate response to COVID-19 lockdown measures in China. *Sci Rep* 11(1):1–9. <https://doi.org/10.1038/s41598-021-96007-1>
- Lewis N (2013b) An objective Bayesian improved approach for applying optimal fingerprint techniques to estimate climate sensitivity. *J Clim* 26:7414–7429
- Lewis N (2014) Objective inference for climate parameters: Bayesian, transformation of variables and profile likelihood approaches. *J Clim* 27:7270–7284. <https://doi.org/10.1175/JCLI-D-13-00584.1>
- Lewis N (2018) Combining independent Bayesian posteriors into a confidence distribution, with application to estimating climate sensitivity. *J Stat Plan Inference* 195:80–92. <https://doi.org/10.1016/j.jspi.2017.09.013>
- Lewis N, Curry JA (2015) The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Clim Dyn* 45:1009–1023. <https://doi.org/10.1007/s00382-014-2342-y>
- Lewis N, Curry JA (2018) The impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity. *J Clim* 31(6051):6071
- Lewis N, Grünwald P (2018) Objectively combining AR5 instrumental period and paleoclimate climate sensitivity evidence. *Clim Dyn* 50(5):2199–2216
- Lewis N, Mauritsen T (2021) Negligible unforced historical pattern effect on climate feedback strength found in HadISST-based AMIP simulations. *J Clim* 34(1):39–55
- Lewis N (2013a) Modification of Bayesian updating where continuous parameters have differing relationships with new and existing data. [arXiv:1308.2791](https://arxiv.org/abs/1308.2791) [stat.ME]
- Liu P, Kaplan JO, Mickley LJ, Li Y, Chellman NJ, Arienzo MM, Kodros JK, Pierce JR, Sigl M, Freitag J, Mulvaney R (2021) Improved estimates of preindustrial biomass burning reduce the magnitude of aerosol climate forcing in the Southern Hemisphere. *Sci Adv* 7(22), p.eabc1379. <https://doi.org/10.1126/sciadv.abc1379>
- McClymont EL et al (2020) Lessons from a high CO2 world: an ocean view from ~ 3 million years ago. *Clim past* 16(4):1599–1615. <https://doi.org/10.5194/cp-2019-161>
- Meinshausen M, Nicholls ZRJ, Lewis J, Gidden MJ, Vogel E, Freund M et al (2020) The shared socioeconomic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geosci Model Dev* 13:3571–3605. <https://doi.org/10.5194/gmd-13-3571-2020>
- Mülmenstädt J, Salzmann M, Kay JE, Zelinka MD, Ma PL, Nam C, Kretzschmar J, Hörnig S, Quaas J (2021) An underestimated negative cloud feedback from cloud lifetime changes. *Nat Clim Chang* 11(6):508–513
- Myers TA, Scott RC, Zelinka MD, Klein SA, Norris JR, Caldwell PM (2021) Observational constraints on low cloud feedback reduce uncertainty of climate sensitivity. *Nat Clim Chang* 11(6):501–507
- Otto A, Coauthors, (2013) Energy budget constraints on climate response. *Nat Geosci* 6:415–416. <https://doi.org/10.1038/ngeo1836>
- Paulot F, Paynter D, Winton M, Ginoux P, Zhao M, Horowitz LW (2020) Revisiting the impact of sea salt on climate sensitivity. *Geophys Res Lett* 47:e2019GL085601. <https://doi.org/10.1029/2019GL085601>
- Pawitan Y (2001) In all Likelihood: Statistical Modeling and Inference Using Likelihood Ch. 3.4. Oxford Univ. Press, 514 pp
- Possner A, Eastman R, Bender F, Glassmeier F (2020) Deconvolution of boundary layer depth and aerosol constraints on cloud water path in subtropical stratocumulus decks. *Atmos Chem Phys* 20:3609–3621
- Raftery AE, Schweder T (1993) Inference about the ratio of two parameters, with application to whale censusing. *Amer Stat* 47(4):259–264
- Rugenstein M, Bloch-Johnson J, Gregory J, Andrews T, Mauritsen T, Li C et al (2020) Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophys Res Lett* 47:e2019GL083898. <https://doi.org/10.1029/2019GL083898>
- Schlund M, Lauer A, Gentine P, Sherwood SC, Eyring V (2020) Emergent constraints on equilibrium climate sensitivity in CMIP5: do they hold for CMIP6? *Earth Syst Dyn* 11(4):1233–1258. <https://doi.org/10.5194/esd-11-1233-2020>
- Schweder T, Hjort NL (2002) Confidence and likelihood. *Scand J Stat* 29:309–332
- Schweder T, Hjort NL (2016) Confidence, likelihood. Cambridge University Press, Probability, p 500
- Sherwood SC, Webb MJ, Annan JD, Armour KC, Forster PM, Hargreaves JC, Hegerl G, Klein SA, Marvel KD, Rohling EJ, Watanabe M, Andrews T, Braconnot P, Bretherton CS, Foster GL, Hausfather Z, von der Heydt AS, Knutti R, Mauritsen T, Norris JR, Proistosescu C, Rugenstein M, Schmidt GA, Tokarska KB, Zelinka MD (2020) An assessment of Earth's climate sensitivity using multiple lines of evidence. *Rev Geophys* 58(4):e2019RG000678
- Smith CJ et al (2021) Figure and data generation for Chapter 7 of the IPCC's Sixth Assessment Report, Working Group I (plus assorted other contributions). Version 1.0. <https://doi.org/10.5281/zenodo.5211357>; Accessed 15 Sept 2021

- Stevens B, Sherwood SC, Bony S, Webb MJ (2016) Prospects for narrowing bounds on Earth's equilibrium climate sensitivity. *Earth's Fut* 4(11):512–522. <https://doi.org/10.1002/2016EF000376>
- Strommen K, Watson PA, Palmer TN (2019) The impact of a stochastic parameterization scheme on climate sensitivity in EC-Earth. *J Geophys Res Atmos* 124(23):12726–12740
- Tierney JE, Haywood AM, Feng R, Bhattacharya T, Otto-Bliesner BL (2019) Pliocene warmth consistent with greenhouse gas forcing. *Geophys Res Lett* 46:9136–9144
- Welch BL, Peers HW (1963) On formulae for confidence points based on integrals of weighted likelihoods. *J R Soc Ser B* 25:318–329
- Zelinka MD, Myers TA, McCoy DT, Po-Chedley S, Caldwell PM, Ceppi P et al (2020) Causes of higher climate sensitivity in CMIP6 models. *Geophys Res Lett* 47:e2019GL085782. <https://doi.org/10.1029/2019GL085782>
- Zhao M, Coauthors, (2016) Uncertainty in model climate sensitivity traced to representations of cumulus precipitation microphysics. *J Clim* 29:543–560. <https://doi.org/10.1175/JCLI-D-15-0191.1>
- Zhou C, Zelinka MD, Dessler AE, Wang M (2021) Greater committed warming after accounting for the pattern effect. *Nat Clim Chang* 11(2):132–136. <https://doi.org/10.1038/s41558-020-00955-x>
- Zhu J, Poulsen CJ (2021) Last Glacial Maximum (LGM) climate forcing and ocean dynamical feedback and their implications for estimating climate sensitivity. *Clim past* 17(1):253–267

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.