JOURNAL OF GEOPHYSICAL RESEARCH, VOL. ???, XXXX, DOI:10.1029/,

What is the effect of unresolved internal climate variability on climate sensitivity estimates?

R. Olson[*,1], R. Sriver[2], M. Haran[3], W. Chang[3], N. M. Urban[4], and K. Keller[1,5]

[1] Department of Geosciences, Penn State University, University Park, PA, USA.

[2] Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

[3] Department of Statistics, Penn State University, University Park, PA, USA.

[4] Energy Security Center, Los Alamos National Laboratory, Los Alamos, NM, USA.

[5] Earth and Environmental Systems Institute, Penn State University, University Park, PA, USA.

*Corresponding author email: rzt2-wrk@psu.edu

Abstract. Current climate sensitivity (CS) estimates are highly uncertain. Quantifying the sources of this uncertainty is relevant to the design of climate policies. Here we isolate and evaluate the role of internal climate variability in driving the climate sensitivity uncertainty using observation system simulation experiments. We use ensemble runs of the University of Victoria Earth System Climate Model (UVic ESCM) spanning the last two centuries. We first construct pseudo-observations of global mean temperature and ocean heat content from the model output at a specified 'true' CS, and then re-estimate the CS using an inverse method. Our results suggest that unresolved internal climate variability is a key driver of current CS uncertainty (as measured by the 68% credible interval). We demonstrate that the internal variability can result in a large discrepancy between the best CS estimate and the truth. Since current best CS estimates based on the observed warming all rely on the same variability, they may be considerably higher or lower than the true value. The estimation uncertainties increase at higher climate sensitivities, suggesting that a high CS might be difficult to detect due to the effects of observational errors and internal climate variability.

1. Introduction

Future climate projections strongly depend on climate sensitivity (CS) [*Matthews and Caldeira*, 2007; *Knutti and Hegerl*, 2008]. CS is the equilibrium global mean near-surface temperature change for a doubling of atmospheric CO_2 concentrations [*Andronova et al.*, 2007; *Knutti and Hegerl*, 2008]. Many recent studies attempted to estimate climate sensitivity [*Forest et al.*, 2002, 2006; *Knutti et al.*, 2003; *Tomassini et al.*, 2007; *Drignei et al.*, 2008; *Holden et al.*, 2010; *Olson et al.*, 2012; *Urban and Keller*, 2010, and others], yet this quantity remains highly uncertain [*Hegerl et al.*, 2007; *Edwards et al.*, 2007].

Several sources contribute to this uncetainty. They include (i) climate model error, (ii) unresolved internal climate variability, and (iii) observational error. We refer to the sum of these processes as 'unresolved climate noise'. Quantifying the relative contribution of these sources of uncertainty is of considerable policy relevance. Here we focus on the role of the unresolved internal climate variability. The unresolved internal climate variability is the part of the observed internal climate variability record that a climate model can not reproduce.

We use observation system simulation experiments (OSSEs) to analyze the role of internal climate variability. OSSEs are a common tool in physical and environmental sciences to evaluate observation system designs [*e.g., Huang et al.*, 2010a, b; *Serra et al.*, 2011; *Zakamska et al.*, 2011; *Urban and Keller*, 2009]. In OSSEs, synthetic observations ('pseudo-observations') are first generated from a model with known 'true' parameter setting by adding noise representing observational error. Then the parameters are re-estimated using the pseudo-observations.

Our starting point is an ensemble of Earth System Model runs spanning the last two centuries where climate sensitivity is systematically varied. The ensemble also accounts for the uncertainty in ocean mixing and radiative effects of anthropogenic sulfates [*Olson et al.*, 2012]. We develop a statistical approximator ('emulator') of our climate model and use it to estimate model output at the parameter values where the model was not evaluated. In a suite of OSSEs, we construct pseudo-observations of surface temperature (T) and upper ocean heat content (0-700 m, OHC) by contaminating the model output at a set 'true' CS with unresolved climate noise. We then re-estimate CS using the pseudo-observations, and an inverse parameter estimation method. We use this approach to address three main questions: (i) How well can we constrain CS using observations of temperature and upper ocean heat content? (ii) Do the estimation uncertainties depend on the input CS? and (iii) What is the contribution of the unresolved internal climate variability to the CS uncertainty? We give further details on the Earth System model, the parameter estimation methodology, and the experimental design in the following sections.

2. Methods

2.1. Earth System Model Simulations

We use the University of Victoria Earth System model (UVic ESCM) version 2.8 [*Weaver et al.*, 2001]. Our modified version of the model includes an updated solar radiative forcing, and implements additional greenhouse gas, volcanic, and anthropogenic sulfate aerosol forcings [*Olson et al.*, 2012]. We use an ensemble of 250 historical UVic ESCM runs spanning the years 1800-2010. The ensemble samples model parameters CS (through an additional parameter f^*), background vertical ocean diffusivity (K_{bg}) and a scaling factor for albedoes due anthropogenic sulfate aerosols (A_{sc}) [*Olson et al.*, 2012]. The ranges for the climate model parameters are given in Table 1.

2.2. Gaussian Process Emulator

Our methodology to estimate the probability density function for CS given the pseudo-observations requires orders of magnitude more UVic ESCM runs than computationally feasible to carry out with a typical computational environment (see Section 2.3). We overcome this hurdle by using the UVic ESCM emulator described in *Olson et al.* [2012]. Emulators are fast statistical approximators to climate models, and are often used in climate science [*Drignei et al.*, 2008; *Holden et al.*, 2010; *Edwards et al.*, 2011; *Bhat et al.*, 2012; *Olson et al.*, 2012]. Because of their speed, they help to better sample model parameter space. Our emulator relies on model output at the 250 parameter settings of the ensemble and interpolates the model response to any desired parameter setting. Specifically, the emulator estimates global average annual surface temperature anomalies T (years 1850-2006) and upper ocean heat content anomalies OHC (0-700 m, years 1950-2003). These times reflect the coverage of pseudo-observations (Section 2.3) and are consistent with the span of observations from *Brohan et al.* [2006] and *Domingues et al.* [2008]. The temperature anomaly is with respect to years 1850-1899, while the OHC anomaly - to years 1950-2003.

The emulator works in rescaled model parameter coordinates such that each parameter ranges from zero to unity. The emulator models the climate model output as a sum of a quadratic polynomial in the rescaled parameters, and a zero-mean Gaussian process with an isotropic covariance function (*i.e.*, the smoothness of the Gaussian Process is the same in all rescaled climate model parameter directions). We use the emulator to only interpolate the model outputs between the parameter settings. There is no extrapolation beyond the range of the ensemble. The emulator provides a reasonable approximation to UVic ESCM over the parameter ranges used [*Olson et al.*, 2012].

2.3. Observation System Simulation Experiments

We conduct several OSSE to address three questions. First, how well can pseudo-observations of temperature and upper ocean heat content constrain climate sensitivity (in terms of the width of CS probability density function (pdf), and the scatter of the estimated CS mode for repeated experiments)? Second, does the estimation skill depend on the input CS? Finally, how important is the unresolved internal climate variability for the CS uncertainty (as measured by the width and the scatter of the CS pdfs)?

The OSSEs involve two main parts: (i) Generation of pseudo-observations from the UVic ESCM given assumed 'true' CS and (ii) Re-estimating CS given the UVic ESCM model output, the pseudo-observations, and the inverse parameter estimation method. In the first stage, we answer the following question: Given a 'true' CS, and assuming that the UVic ESCM emulator correctly models climate response to historical forcings, what time series of temperature and ocean heat content can we theoretically observe? To this end, we construct pseudo-observations by superimposing unresolved climate noise on the UVic ESCM emulator output at a pre-defined 'true' climate parameter setting. The unresolved noise models the sum of the processes that result in the discrepancy between the observations and the emulator. Mathematically, the noise n is defined as:

$$n_{t,k} = y_{t,k} - \tilde{f}_{t,k}(\theta), \tag{1}$$

where y refer to the observations, $\tilde{f}$ is the emulator output, θ is the vector of model parameters (K_{bg}, CS, A_{sc}), t is the time index, and k is the diagnostic index (*i.e.* $k = 1$ for T, and $k = 2$ for OHC).

We approximate the unresolved climate noise by an AR(1) process. Exploratory data analysis shows that this is a reasonable assumptions for all OSSEs presented here. Specifically,

$$n_{t,k} = \rho n_{t-1,k} + w_{t,k}, \tag{2}$$

where ρ is first-order autocorrelation and w is an independently and identically distributed Gaussian noise with the innovation standard deviation σ_k. This AR(1) process is completely specified the by σ_k and ρ_k parameters.

The second stage of the OSSE addresses that question of what CS pdfs we expect for a given 'true' CS value and different realizations of the unresolved climate noise? Following *Olson et al.* [2012], we re-estimate CS using the following statistical model:

$$y_{t,k} = \tilde{f}_{t,k} + b_k + n_{t,k}, \tag{3}$$

where b_k is an additional time-independent bias. To be consistent with *Olson et al.* [2012] we set the bias term for OHC to 0 in this stage. Associated with each parameter value $\mathbf{\Theta}$ = (K_{bg}, CS, A_{sc}, σ_T, σ_{OHC}, ρ_T, ρ_{OHC}, b_T) there is a likelihood function which describes the probability of observations given this parameter value (please see the Appendix). The posterior probability for each parameter setting is obtained using Bayes Theorem by multiplying the likelihood function by the prior probability for the parameters. We estimate the joint posterior pdf for $\boldsymbol{\Theta}$ using Markov chain Monte Carlo (MCMC). The MCMC algorithm [*Metropolis et al.*, 1953; *Hastings*, 1970] is a standard computational approach for estimating multivariate posterior pdfs. Our implementation of the method follows *Olson et al.* [2012]. Specifically, our MCMC parameter chains are 300,000 members long for each unresolved noise realization. The actual number of required emulator runs is higher because only a subset of tested parameter settings are accepted into the chain. For each experiment, we repeat the procedure of generating pseudo-observations and estimating CS sixty times, each time relying on a different random realization of the unresolved climate noise process. Two out of sixty realizations are tested for convergence by running the estimation twice with different initial values for the final MCMC chain. We have not detected any convergence problems with our algorithm.

The OSSEs share the same general set-up, with relatively minor differences. Specifically, the experiments differ in assumed 'true' parameter values, in the priors, and in the assumptions about the unresolved noise process (Table 2).

In the first experiment, called 'Standard', we address the power of the observations to constrain CS assuming realistic knowledge of climate uncertainties. Here we use mean estimates from the base case of *Olson et al.* [2012] as 'true' climate parameters. These values are $K_{bg} = 0.19$ cm^{2}s^{-1}, CS = 3.1 °C and $A_{sc} = 1.1$. For unresolved climate noise we adopt the modes from the base case of *Olson et al.* [2012]: $\sigma_T = 0.10$ [°C], $\sigma_{OHC} = 2.6$ [$\times 10^{22}$ J], $\rho_T = 0.58$, and $\rho_{OHC} = 0.079$ ('UVic ESCM Residuals' in Figures 1 and 2). For simplicity, we do not use bias terms when generating pseudo-observations, since the 95% posterior credible intervals for these terms include zero [*Olson et al.*, 2012]. We use uniform priors for all parameters.

In the experiment 'Nat. Var.', we address the following question: What could the estimated pdfs look like if the only source of the discrepancy between the model and the observations were the unresolved internal climate variability? By the internal climate variability we mean the variations in the mean state of the climate on all spatial and temporal scales beyond that of individual weather events due to natural internal processes within the climate system (as opposed to variations in natural or anthropogenic external forcing) [*Baede*, 2007]. The only difference between 'Nat. Var.' and 'Standard' lies in the values for the unresolved noise parameters. In the 'Nat. Var.' experiment we assume that the unresolved noise models the internal climate variability only. We also assume that the UVic ESCM emulator does not include any substantial internal climate variability.

Unfortunately, estimating the internal climate variability from observations is confounded by the observational errors, particularly in the case of OHC. Thus, following *Tomassini et al.* [2007] and *Sanso and Forest* [2009] we approximate the internal variability by using the output from General Circulation Models (GCMs). We fit an AR(1) process to detrended near-surface annual atmospheric temperature and 0-700 m ocean heat content anomalies from preindustrial control runs of three climate models: BCCR-BCM2.0 [*Ottera et al.*, 2009], GFDL-CM2.1 [*Delworth et al.*, 2006; *Gnanadesikan et al.*, 2006] and UKMO-HadCM3 [*Gordon et al.*, 2000; *Pope et al.*, 2000; *Johns et al.*, 2003]. The output of these runs was obtained from the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset [*Meehl et al.*, 2007]. Specifically, we use run 1 for all three models. We discard the first 100 years for BCCR-BCM2.0 because the modeled climate appears to be out of equilibrium during this period. We detrend the anomalies using robust locally weighted regression

[*Cleveland*, 1979] with the span f of 2/3. When calculating OHC, we first obtain temperatures from potential temperatures, and salinities using the UNESCO equation of state [*UNESCO*, 1981] following *Bryden* [1973] and *Fofonoff* [1977]. For this conversion we find the ocean pressure field from latitude and depth using simplified equations [*Lovett*, 1978]. The resulting AR(1) properties, averaged across the models, are: σ_T=0.12 [°C], $\sigma_{OHC} = 0.51$ [$\times 10^{22}$ J], $\rho_T = 0.45$, and $\rho_{OHC} = 0.9$ (Table 2, Figures 1 and 2, red triangles).

The 'Higher CS' experiment explores the effects of different 'true' parameter values on the estimation. It differs from 'Standard' by using a higher 'true' input CS. Specifically, we adopt $K_{bg} = 0.19$ cm^{2}s^{-1}, CS = 4.8 °C and $A_{sc} = 1.3$. These values are selected to be consistent with the bivariate joint pdfs presented in *Olson et al.* [2012].

The 'Inf. Priors' experiment examines the role of priors. It uses informative priors for CS (Figure 3) and K_{bg} following the default case of *Olson et al.* [2012]. 'Inf. Priors' has otherwise the same settings as 'Standard' (cf. Table 2).

3. Results and Discussion

Our results suggest that the process driving unresolved internal climate variability is a key factor behind the current uncertainty in climate sensitivity estimates. Specifically, the average width of the estimated CS pdfs (as measured by the 68% posterior credible intervals) in the 'Nat. Var.' case is only modestly lower compared to the 'Standard' case (Table 2, Figure 3). This indicates that even if we had perfect models of long term mean climate, and errorless observations, our CS estimates would still remain very uncertain due to the confounding effect of the unresolved internal climate variability. The variability also appears to be a key factor in the second-order uncertainty in climate sensitivity (Figures

3 and 4). This uncertainty represents the sensitivity of estimated CS pdfs to different realizations of the unresolved climate noise, and is measured by the standard deviation of CS modes between the realizations. Specifically, while the standard deviation is 1.6 °C in the 'Standard' case, it decreases only slightly to 1.4 °C in the 'Nat. Var.' case (Table 2). Of course, the pivotal role of the internal climate variability should not prevent us from investing in better future observational systems. *Webster et al.* [2008] find, using a simplified unresolved climate noise representation, that future observations are expected to further reduce the CS uncertainty. Our results suggest that internal climate variability presents a substantial obstacle to estimating climate sensitivity. Whether alternative approaches that perform joint state and parameter estimation [*e.g., Annan et al.*, 2005; *Hill et al.*, 2012; *Evensen*, 2009] can overcome this challenge, is thus far an open question.

The CS estimation uncertainties increase at higher CS. Specifically, both pdf width and scatter increase considerably compared to the 'Standard' case (Table 2, Figure 4). This suggests that higher climate sensitivities can be difficult to detect if a particular realization of climate noise biases the result low. This is consistent with the analytical model results of *Hansen et al.* [1985] which show that the dependency of transient ocean warming on climate sensitivity weakens at high CS. Thus, at high CS, a small uncertainty in a single ocean surface warming observation implies a larger uncertainty in climate sensitivity. Our numerical model shows similar response of atmospheric surface warming to changing CS. Note that there are other complicating factors influencing the CS uncertainty, such as the aerosol effects specified by A_{sc}.

Switching from uniform to informative priors (the 'Inf. Priors' experiment) substantially reduces the CS uncertainty (Table 2, Figures 3 and 4). Under the informative priors, the mean estimated CS mode (2.9 °C) is somewhat lower than the 'true' value of 3.1 °C. This difference is statistically significant ($\alpha = 0.05$). This might be in part due to the biasing effect of the mode of the CS prior, which is lower than the 'true' value. Both of these effects (lower uncertainty but potential biases under narrower priors within the context of OSSEs) have been previously found and discussed by *Webster et al.* [2008]. Thus, while using informative priors can be a promising approach, care should be given to choosing an appropriate prior.

Finally, each realization of internal climate variability can result in a considerable discrepancy between the best CS estimate and the true value ('Nat. Var.' panels, Figures 3 and 4). The average discrepancy due to the unresolved internal variability is 1.1 °C (Table 2). One of the Nat. Var. experiments leads to an estimate of 7.5 °C which is 4.4 °C higher than the 'true' value. The distribution of the discrepancy is positively skewed, with a longer upper tail (Figure 4). Historical observational constraints on climate sensitivity (*e.g.,* upper ocean heat content, and surface temperature) are based on a single realization of internal climate variability process. Assuming that the biasing effects of the observational and model errors are low, this realization can introduce a considerable discrepancy between the best CS estimate and the true value. Given that scientific models often share similar assumptions and might not be independent (see *Pennell and Reichler* [2011] for a discussion of similarities in GCMs), it is possible that the bias due to the internal variability can be in the same direction in studies using different models. As a result, current best CS estimates from these datasets may be considerably higher or lower than the true value. One of the ways to overcome this under-sampling problem is to use independent constraints from other time periods (*e.g.,* Last Glacial Maximum, *Schmittner et al.* [2011]).

4. Caveats

Our analysis uses many simplifying assumptions that point to several caveats and open research questions. First, our Earth System model relies on a number of approximations and neglects some historic forcings (*e.g.,* indirect effects of anthropogenic sulfates; and tropospheric ozone [*Forster et al.*, 2007]). Second, we do not fully account for past forcing uncertainties. Third, we change climate sensitivity using a very simplistic approach by varying longwave radiative feedbacks, while shortwave feedbacks are also uncertain [*Bony et al.*, 2006]. Fourth, our statistical model does not include any cross-correlation among the residuals for T and OHC, and relies on a simple AR(1) structure. However, our exploratory data analysis suggests that this structure is a reasonable approximation to the underlying statistical processes. Fifth, we use a relatively small number of realizations in the OSSEs to keep the computational burden manageable. Sixths, our estimates of internal climate variability rely on three climate models. Using more models might provide a better sample. Seventh, there is a distinct possibility that climate models considerably underestimate the observed decadal OHC variability (*e.g., Levitus et al.* [2001], *Hansen et al.* [2005]; but see *AchutaRao et al.* [2007] for an alternative view). If true, we hypothesize that the CS uncertainty in the 'Nat. Var.' experiment would increase, which would strengthen our conclusion that natural variability is an important driver of the uncertainty in climate sensitivity. Last, but not least, we rely on uniform priors in most experiments. We have chosen to work with the relatively simple prior specification because it still remains an open question to find more informative priors that lead to good bias, and coverage properties. Finally, we explore only a small subset of uncertainty in unresolved climate noise and in climate model parameters.

5. Conclusions

We use Observation System Simulation Experiments (OSSEs) to analyze the effects of unresolved internal climate variability on the uncertainty in climate sensitivity. We repeatedly simulate pseudo-observations from an Earth System Model with a given climate sensitivity, and then re-estimate the sensitivity using a Bayesian inversion method.

We find that unresolved internal climate variability is a key driver of the first-order (as measured by the 68% posterior credible internal) and the second-order (as measured by standard deviation of the estimated modes) uncertainty in climate sensitivity estimates. A single realization of the statistical process driving the variability can introduce a substantial discrepancy between a CS estimate and the true value. Since recent CS estimates using instrumental temperature and upper ocean heat content observations all rely on the same realization, they may be considerably higher or lower than the true CS. The unresolved internal variability represents a critical roadblock: our research suggests that even if we at present had errorless models and observations, current estimation approaches would still result in considerable CS uncertainty. Exploring the power of combined state and parameter estimation [*e.g., Annan et al.*, 2005; *Hill et al.*, 2012; *Evensen*, 2009] to confront this challenge is the subject of future research.

Appendix

This appendix provides the likelihood function for observations if the statistical model is given by Equations 2 and 3. We define $\mathbf{y_k} = y_{1,k}, ..., y_{N_k,k}$ where N_k is the number of observations for diagnostic k, and k refers to a diagnostic (*i.e.* $k = 1$ for temperature, and $k = 2$ for ocean heat content). The likelihood function for observations $\mathbf{y_k}$ given the model and the statistical parameters is given by [*Bence*, 1995; *Olson et al.*, 2012]:

$$L(\mathbf{y_k}|K_{bg}, \mathrm{CS}, A_{sc}, \sigma_k, \rho_k, b_k) = \left(2\pi\sigma_{p,k}^2\right)^{-1/2} \exp\left(-\frac{1}{2}\frac{n_{1,k}^2}{\sigma_{p,k}^2}\right) \times \\ \times \left(2\pi\sigma_k^2\right)^{-(N_k-1)/2} \times \exp\left(-\frac{1}{2\sigma_k^2}\sum_{t=2}^{N_k} w_{t,k}^2\right).$$

Here $\sigma_{p,k}^2$ refers to the stationary process variance and is defined by $\sigma_{p,k}^2 = \sigma_k^2/(1-\rho_k^2)$, and $w_{t,k}$ are whitened bias-corrected residuals. The whitened residuals are calculated as $w_{t,k} = n_{t,k} - \rho_k n_{t-1,k}$ for $t > 1$. Assuming the independence of the residuals (between the model and the pseudo-observations) across different diagnostics, the final likelihood for all pseudo-observations $\mathbf{Y} \equiv (\mathbf{y_T}, \mathbf{y_{OHC}})$ is the product of the individual likelihoods:

$$L(\mathbf{Y}|\mathbf{\Theta}) = L(\mathbf{y_T}|K_{bg}, CS, A_{sc}, \sigma_T, \rho_T, b_T) \times L(\mathbf{y_{OHC}}|K_{bg}, CS, A_{sc}, \sigma_{OHC}, \rho_{OHC}) \quad (4)$$

Acknowledgments. This work was supported by NSF through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507, and through the Center for Climate and Energy Decision Making under the cooperative agreement SES-0949710 between the NSF and Carnegie Mellon University. We are grateful to Michael Eby and to the developers of UVic ESCM for providing the model and for discussions and advice. This study would not have been possible without the efforts of scientists who collected the observations used in this study. We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. This research uses data provided by the Bergen Climate Model (BCM) project (www.bcm.uib.no) at the Bjerknes Centre for Climate Research, largely funded by the Research Council of Norway. Furthermore, we thank the scientists at the Met Office Hadley Center, and Geophysical Fluid Dynamics Laboratory for producing the GCM output used in this study. All views, errors, and opinions are solely that of the authors.

References

AchutaRao, K. M., M. Ishii, B. D. Santer, P. J. Gleckler, K. E. Taylor, T. P. Barnett, D. W. Pierce, R. J. Stouffer, and T. M. L. Wigley (2007), Simulated and observed variability in ocean temperature and heat content, *Proc. Natl. Acad. Sci.*, *104*(26), 10,768–10,773.

Andronova, N., M. Schlesinger, S. Dessai, M. Hulme, and B. Li (2007), The concept of climate sensitivity: History and development, in *Human-induced Climate Change: An Interdisciplinary Assessment*, edited by M. Schlesinger, H. Kheshgi, J. Smith, F. de la Chesnaye, J. M. Reilly, T. Wilson, and C. Kolstad, Cambridge University Press.

Annan, J. D., J. C. Hargreaves, N. R. Edwards, and R. Marsh (2005), Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter, *Ocean Modelling*, *8*(1-2), 135–154, doi:10.1016/j.ocemod.2003.12.004.

Baede, A. P. M. (2007), Annex I: Glossary, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Bence, J. R. (1995), Analysis of short time series – Correcting for autocorrelation, *Ecology*, *76*(2), 628–639.

Bhat, K. S., M. Haran, R. Olson, and K. Keller (2012), Inferring likelihoods and climate system characteristics from climate models and multiple tracers, *Environmetrics*, *23*(4), 345–362, doi:10.1002/env.2149.

Bony, S., et al. (2006), How well do we understand and evaluate climate change feedback processes?, *J. Clim.*, *19*(15), 3445–3482.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res. [Atmos.]*, *111*(D12), doi:10.1029/2005JD006548.

Bryden, H. L. (1973), New polynomials for thermal expansion, adiabatic temperature gradient and potential temperature of sea-water, *Deep-Sea Res.*, *20*(4), 401–408, doi: 10.1016/0011-7471(73)90063-6.

Cleveland, W. S. (1979), Robust Locally Weighted Regression and Smoothing Scatterplots, *J. Am. Stat. Assoc.*, *74*(368), 829–836, doi:10.2307/2286407.

Delworth, T., et al. (2006), GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics, *J. Clim.*, *19*(5), 643–674.

Domingues, C. M., J. A. Church, N. J. White, P. J. Gleckler, S. E. Wijffels, P. M. Barker, and J. R. Dunn (2008), Improved estimates of upper-ocean warming and multi-decadal sea-level rise, *Nature*, *453*(7198), 1090–U6, doi:10.1038/nature07080.

Drignei, D., C. E. Forest, and D. Nychka (2008), Parameter estimation for computationally intensive nonlinear regression with an application to climate modeling, *Ann. Appl. Stat.*, *2*(4), 1217–1230, doi:10.1214/08-AOAS210.

Edwards, N. R., D. Cameron, and J. Rougier (2011), Precalibrating an intermediate complexity climate model, *Clim. Dyn.*, *37*(7-8), 1469–1482, doi:10.1007/s00382-010-0921-0.

Edwards, T. L., M. Crucifix, and S. P. Harrison (2007), Using the past to constrain the future: how the palaeorecord can improve estimates of global warming, *Prog. Phys.*

Geog., *31*(5), 481–500, doi:10.1177/0309133307083295.

Evensen, G. (2009), The Ensemble Kalman Filter for Combined State and Parameter Estimation. Monte Carlo Techniques for Data Assimilation in Large Systems, *IEEE Control Systems Maganize*, *29*(3), 83–104, doi:10.1109/MCS.2009.932223.

Fofonoff, N. P. (1977), Computation of potential temperature of seawater for an arbitrary reference pressure, *Deep-Sea Res.*, *24*(5), 489–491, doi:10.1016/0146-6291(77)90485-4.

Forest, C. E., P. H. Stone, A. P. Sokolov, M. R. Allen, and M. D. Webster (2002), Quantifying uncertainties in climate system properties with the use of recent climate observations, *Science*, *295*(5552), 113–117.

Forest, C. E., P. H. Stone, and A. P. Sokolov (2006), Estimated PDFs of climate system properties including natural and anthropogenic forcings, *Geophys. Res. Lett.*, *33*(1), doi:10.1029/2005GL023977.

Forster, P., et al. (2007), Changes in Atmospheric Constituents and in Radiative Forcing, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Cambridge Univ. Press, Cambridge, United Kingdom and New York, NY, USA.

Gnanadesikan, A., et al. (2006), GFDL's CM2 global coupled climate models. Part II: The baseline ocean simulation, *J. Clim.*, *19*(5), 675–697.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood (2000), The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Clim. Dyn.*, *16*(2-3), 147–168.

Hansen, J., G. Russell, A. Lacis, I. Fung, D. Rind, and P. Stone (1985), Climate Response Times - Dependence on Climate Sensitivity and Ocean Mixing, *Science*, *229*(4716), 857–859, doi:10.1126/science.229.4716.857.

Hansen, J., et al. (2005), Earth's energy imbalance: Confirmation and implications, *Science*, *308*(5727), 1431–1435, doi:10.1126/science.1110252.

Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*(1), 97–109.

Hegerl, G. C., F. W. Zwiers, P. Braconnot, N. P. Gillett, Y. Luo, J. A. Marengo Orsini, N. Nicholls, J. E. Penner, and P. A. Stott (2007), Understanding and Attributing Climate Change, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Cambridge Univ. Press, Cambridge, United Kingdom and New York, NY, USA.

Hill, T. C., E. Ryan, and M. Williams (2012), The use of CO_2 flux time series for parameter and carbon stock estimation in carbon cycle research, *Glob. Change Biol.*, *18*(1), 179–193, doi:10.1111/j.1365-2486.2011.02511.x.

Holden, P. B., N. R. Edwards, K. I. C. Oliver, T. M. Lenton, and R. D. Wilkinson (2010), A probabilistic calibration of climate sensitivity and terrestrial carbon change in GENIE-1, *Clim. Dyn.*, *35*(5), 785–806.

Huang, Y., S. Leroy, P. J. Gero, J. Dykema, and J. Anderson (2010a), Separation of longwave climate feedbacks from spectral observations, *J. Geophys. Res.-Atm.*, *115*, doi:10.1029/2009JD012766, D07104.

Huang, Y., S. S. Leroy, and J. G. Anderson (2010b), Determining Longwave Forcing and Feedback Using Infrared Spectra and GNSS Radio Occultation, *J. Clim.*, *23*(22), 6027–6035, doi:10.1175/2010JCLI3588.1.

Johns, T. C., et al. (2003), Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios, *Clim. Dyn.*, *20*(6), 583–612, doi:10.1007/s00382-002-0296-y.

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

Knutti, R., T. F. Stocker, F. Joos, and G. K. Plattner (2003), Probabilistic climate change projections using neural networks, *Clim. Dyn.*, *21*(3-4), 257–272, doi:10.1007/s00382-003-0345-1.

Levitus, S., J. I. Antonov, J. L. Wang, T. L. Delworth, K. W. Dixon, and A. J. Broccoli (2001), Anthropogenic warming of Earth's climate system, *Science*, *292*(5515), 267–270, doi:10.1126/science.1058154.

Lovett, J. R. (1978), Merged Seawater Sound-Speed Equations, *J. Ac. Soc. Am.*, *63*(6), 1713–1718, doi:10.1121/1.381909.

Matthews, H. D., and K. Caldeira (2007), Transient climate-carbon simulations of planetary geoengineering, *Proc. Natl. Acad. Sci. U. S. A.*, *104*(24), 9949–9954, doi:10.1073/pnas.0700419104.

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor (2007), The WCRP CMIP3 multi-model dataset - A new era in climate change research, *Bull. Am. Meteorol. Soc.*, *88*(9), 1383–1394, doi:10.1175/BAMS-

88-9-1383.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*(6), 1087–1092.

Olson, R., R. Sriver, M. Goes, N. M. Urban, H. D. Matthews, M. Haran, and K. Keller (2012), A climate sensitivity estimate using Bayesian fusion of instrumental observations and an Earth System model, *J. Geophys. Res.*, *117*, doi:10.1029/2011JD016620, D04103.

Ottera, O. H., M. Bentsen, I. Bethke, and N. G. Kvamsto (2009), Simulated pre-industrial climate in Bergen Climate Model (version 2): model description and large-scale circulation features, *Geosci. Mod. Dev.*, *2*(2), 197–212.

Pennell, C., and T. Reichler (2011), On the Effective Number of Climate Models, *J. Clim.*, *24*(9), 2358–2367, doi:10.1175/2010JCLI3814.1.

Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton (2000), The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3, *Clim. Dyn.*, *16*(2-3), 123–146.

Sanso, B., and C. Forest (2009), Statistical calibration of climate system properties, *J. Royal Stat. Soc. Ser. C - App. Stat.*, *58*(Part 4), 485–503.

Schmittner, A., N. M. Urban, J. D. Shakun, N. M. Mahowald, P. U. Clark, P. J. Bartlein, A. C. Mix, and A. Rosell-Mele (2011), Climate Sensitivity Estimated from Temperature Reconstructions of the Last Glacial Maximum, *Science*, *334*(6061), 1385–1388, doi: 10.1126/science.1203513.

Serra, P., A. Amblard, P. Temi, D. Burgarella, E. Giovannoli, V. Buat, S. Noll, and S. Im (2011), CIGALEMC: Galaxy Parameter Estimation Using a Markov Chain Monte Carlo

Approach with CIGALE, *Astrophys. J.*, *740*(1), doi:10.1088/0004-637X/740/1/22.

Tomassini, L., P. Reichert, R. Knutti, T. F. Stocker, and M. E. Borsuk (2007), Robust Bayesian uncertainty analysis of climate system properties using Markov chain Monte Carlo methods, *J. Clim.*, *20*(7), 1239–1254, doi:10.1175/JCLI4064.1.

UNESCO (1981), Tenth Report of the Joint Panel on Oceanographic Tables and Standards, *Tech. rep.*, UNESCO Technical Reports on Marine Science 36.

Urban, N. M., and K. Keller (2009), Complementary observational constraints on climate sensitivity, L04708, *Geophys. Res. Lett.*, *36*, doi:10.1029/2008GL036457.

Urban, N. M., and K. Keller (2010), Probabilistic hindcasts and projections of the coupled climate, carbon cycle and Atlantic meridional overturning circulation system: A Bayesian fusion of century-scale observations with a simple model, *Tellus Ser. A - Dyn. Met. Ocean.*, *62*(5), 737–750, doi:10.1111/j.1600-0870.2010.00471.x.

Weaver, A. J., et al. (2001), The UVic Earth System Climate Model: Model description, climatology, and applications to past, present and future climates, *Atmos.-Ocean*, *39*(4), 361–428.

Webster, M., L. Jakobovits, and J. Norton (2008), Learning about climate change and implications for near-term policy, *Clim. Change*, *89*(1-2), 67–85, doi:10.1007/s10584-008-9406-0.

Zakamska, N. L., M. Pan, and E. B. Ford (2011), Observational biases in determining extrasolar planet eccentricities in single-planet systems, *Mon. Not. R. Astron. Soc.*, *410*(3), 1895–1910, doi:10.1111/j.1365-2966.2010.17570.x.

Table 1: Ranges for model and statistical parameters. Subscripts T and OHC refer to surface air temperature and upper ocean heat content respectively

Parameter	Units	Lower Bound	Upper Bound
K_{bg}	cm^2 s^{-1}	0.1	0.5
CS	°C per CO_2 doubling	1.1	11.2
A_{sc}	unitless	0	3
σ_T	°C	0.01	inf
σ_{OHC}	1×10^{22} J	0.01	inf
ρ_T	unitless	0.01	0.999
ρ_{OHC}	unitless	0.01	0.999
b_T	°C	-0.51	0.50

Table 2: Summary of the design and the results of the observation system simulation experiments. 'Unif.' refers to uniform priors for climate parameters, and 'Inf.' refers to informative priors for K_{bg} and CS following the default case of *Olson et al.* [2012]. The mean 68% CI refers to the mean 68% posterior credible interval of CS estimates. The interval is calculated as the range between the 16th and the 84th percentiles of the CS chains.

Experiment details							Properties of CS estimates [°C]			
Experiment	Priors	σ_T [°C]	σ_{OHC} [$\times 10^{22}$ J]	ρ_T	ρ_{OHC}	Assumed 'true' CS [°C]	Average mode	Average absolute bias of the mode	Std. of modes	Mean 68% CI of pdfs
'Standard'	Unif.	0.10	2.6	0.58	0.079	3.1	3.3	1.1	1.6	3.5
'Nat. Var.'	Unif.	0.12	0.51	0.45	0.9	3.1	3.7	1.1	1.4	3.0
'Higher CS'	Unif.	0.10	2.6	0.58	0.079	4.8	5.8	2.0	2.6	4.5
'Inf. Priors'	Inf.	0.10	2.6	0.58	0.079	3.1[a]	2.9	0.36	0.41	1.5

[a] While 'true' input CS is 3.1 °C, the mean of the non-uniform prior is 3.25 °C, and the mode is 2.96 °C.

Figure Captions

Figure 1: Statistical properties of surface atmospheric temperature anomaly (T) time series - AR(1) innovation standard deviation σ_T, and first order autocorrelation ρ_T: GCMs (BCCR-BCM2.0, GFDL-CM2.1 and UKMO-HadCM3, red circles), mean across the three GCMs (red triangle), residuals between the UVic ESCM and the observations from *Brohan et al.* [2006] (years 1850-2006, blue triangle), and detrended observations from *Brohan et al.* [2006] (years 1850-2006, green triangle). For the residuals, we use the marginal mode for the base case of *Olson et al.* [2012]. For the detrended observations, we first demean the yearly observations, and then detrend them using a lowess fit trend. Grey contours show the process standard deviation $\sigma_{p,T}$ (cf. Appendix). We use yearly average time series for the AR(1) inference.

Figure 2: Statistical properties of ocean heat content anomaly in the 0-700 m layer (OHC) - AR(1) innovation standard deviation σ_{OHC}, and first order autocorrelation ρ_{OHC}: GCMs (BCCR-BCM2.0, GFDL-CM2.1 and UKMO-HadCM3, red circles), mean across the three GCMs (red triangle), residuals between the UVic ESCM and the observations from *Domingues et al.* [2008] (years 1950-2003, blue triangle), and detrended observations from *Domingues et al.* [2008] (years 1950-2003, green triangle). For the residuals, we use the marginal mode for the base case of *Olson et al.* [2012]. For the detrended observations, we first demean the yearly observations, and then detrend them using a lowess fit trend. Grey contours show process standard deviation $\sigma_{p,OHC}$ (cf. Appendix). We use yearly average time series for the AR(1) inference.

Figure 3: Posterior probability distributions (pdfs) for climate sensitivity from observation system simulation experiments: (top left) 'Standard', (top right) 'Nat. Var., (bottom left) 'Higher CS' and (bottom right) 'Inf. Priors'. Each grey line corresponds to one realization of unresolved climate noise. 'True' input climate sensitivities are shown by vertical dotted lines. The dashed

pdf denotes CS prior in the 'Inf. Priors' experiment. Filled (open) red circles denote the mean (median) CS mode, and the red lines extend one standard deviation around the mean mode. The limits of the y-axes are the same between panels.

Figure 4: Histograms of the modes of the estimated climate sensitivity probability density functions: (top left) 'Standard', (top right) 'Nat. Var', (bottom left) 'Higher CS', and (bottom right) 'Inf. Priors'. 'True' input climate sensitivities are shown by vertical red lines. Y-axes limits are the same between panels.

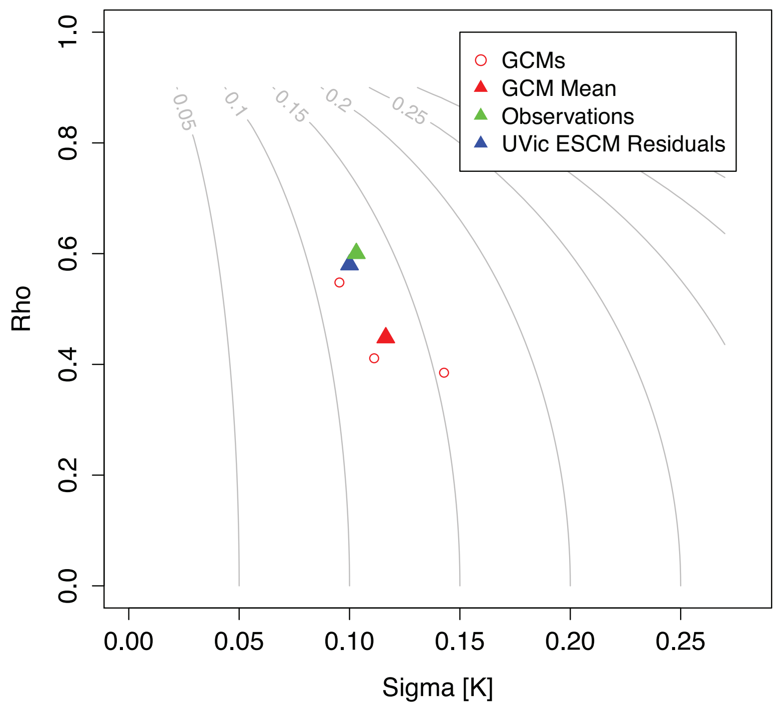

Figure 1: Statistical properties of surface atmospheric temperature anomaly (T) time series - AR(1) innovation standard deviation σ_T, and first order autocorrelation ρ_T: GCMs (BCCR-BCM2.0, GFDL-CM2.1 and UKMO-HadCM3, red circles), mean across the three GCMs (red triangle), residuals between the UVic ESCM and the observations from *Brohan et al.* [2006] (years 1850-2006, blue triangle), and detrended observations from *Brohan et al.* [2006] (years 1850-2006, green triangle). For the residuals, we use the marginal mode for the base case of *Olson et al.* [2012]. For the detrended observations, we first demean the yearly observations, and then detrend them using a lowess fit trend. Grey contours show the process standard deviation $\sigma_{p,T}$ (cf. Appendix). We use yearly average time series for the AR(1) inference.

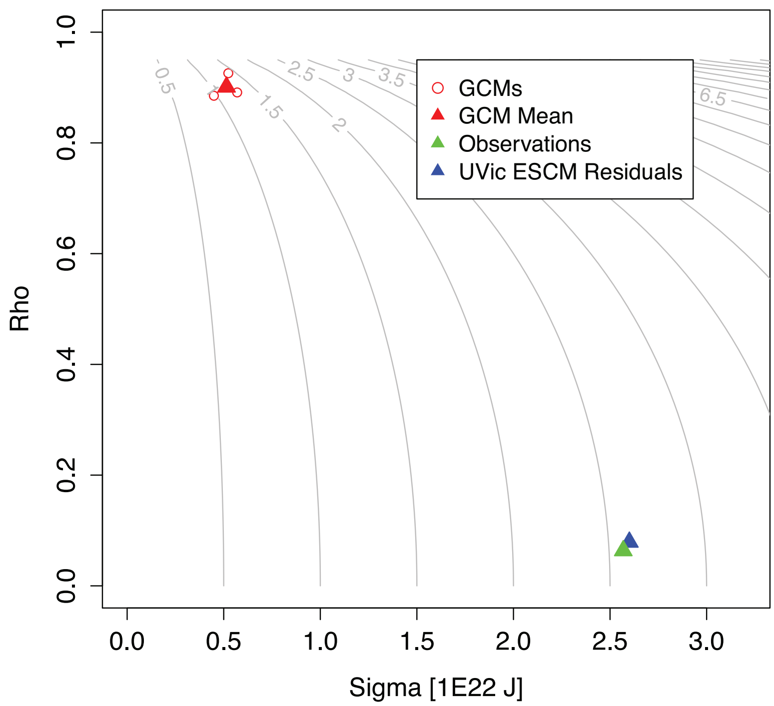

Figure 2: Statistical properties of ocean heat content anomaly in the 0-700 m layer (OHC) - AR(1) innovation standard deviation σ_{OHC}, and first order autocorrelation ρ_{OHC}: GCMs (BCCR-BCM2.0, GFDL-CM2.1 and UKMO-HadCM3, red circles), mean across the three GCMs (red triangle), residuals between the UVic ESCM and the observations from *Domingues et al.* [2008] (years 1950-2003, blue triangle), and detrended observations from *Domingues et al.* [2008] (years 1950-2003, green triangle). For the residuals, we use the marginal mode for the base case of *Olson et al.* [2012]. For the detrended observations, we first demean the yearly observations, and then detrend them using a lowess fit trend. Grey contours show process standard deviation $\sigma_{p,OHC}$ (cf. Appendix). We use yearly average time series for the AR(1) inference.

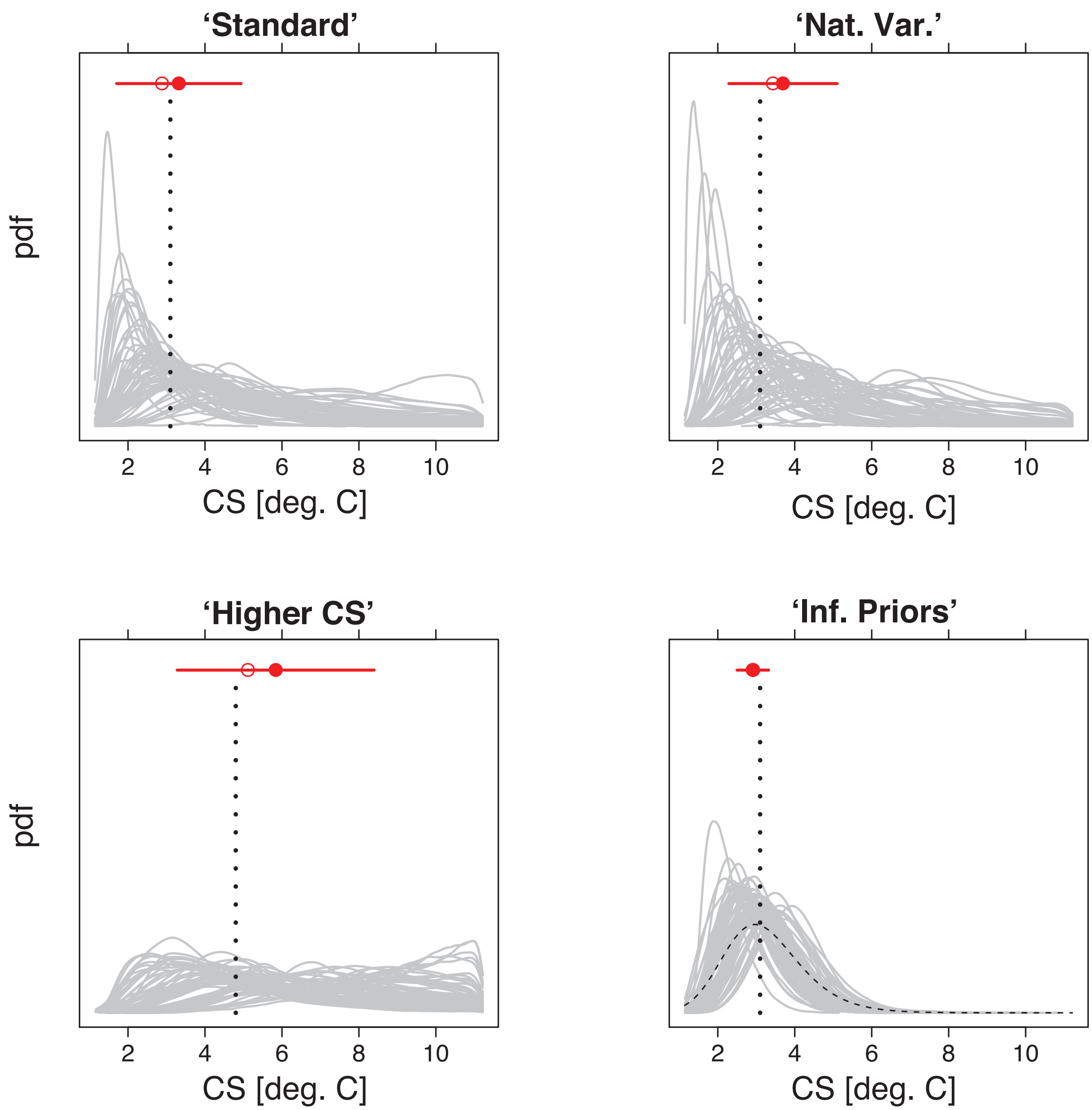

Figure 3: Posterior probability distributions (pdfs) for climate sensitivity from observation system simulation experiments: (top left) 'Standard', (top right) 'Nat. Var., (bottom left) 'Higher CS' and (bottom right) 'Inf. Priors'. Each grey line corresponds to one realization of unresolved climate noise. 'True' input climate sensitivities are shown by vertical dotted lines. The dashed pdf denotes CS prior in the 'Inf. Priors' experiment. Filled (open) red circles denote the mean (median) CS mode, and the red lines extend one standard deviation around the mean mode. The limits of the y-axes are the same between panels.

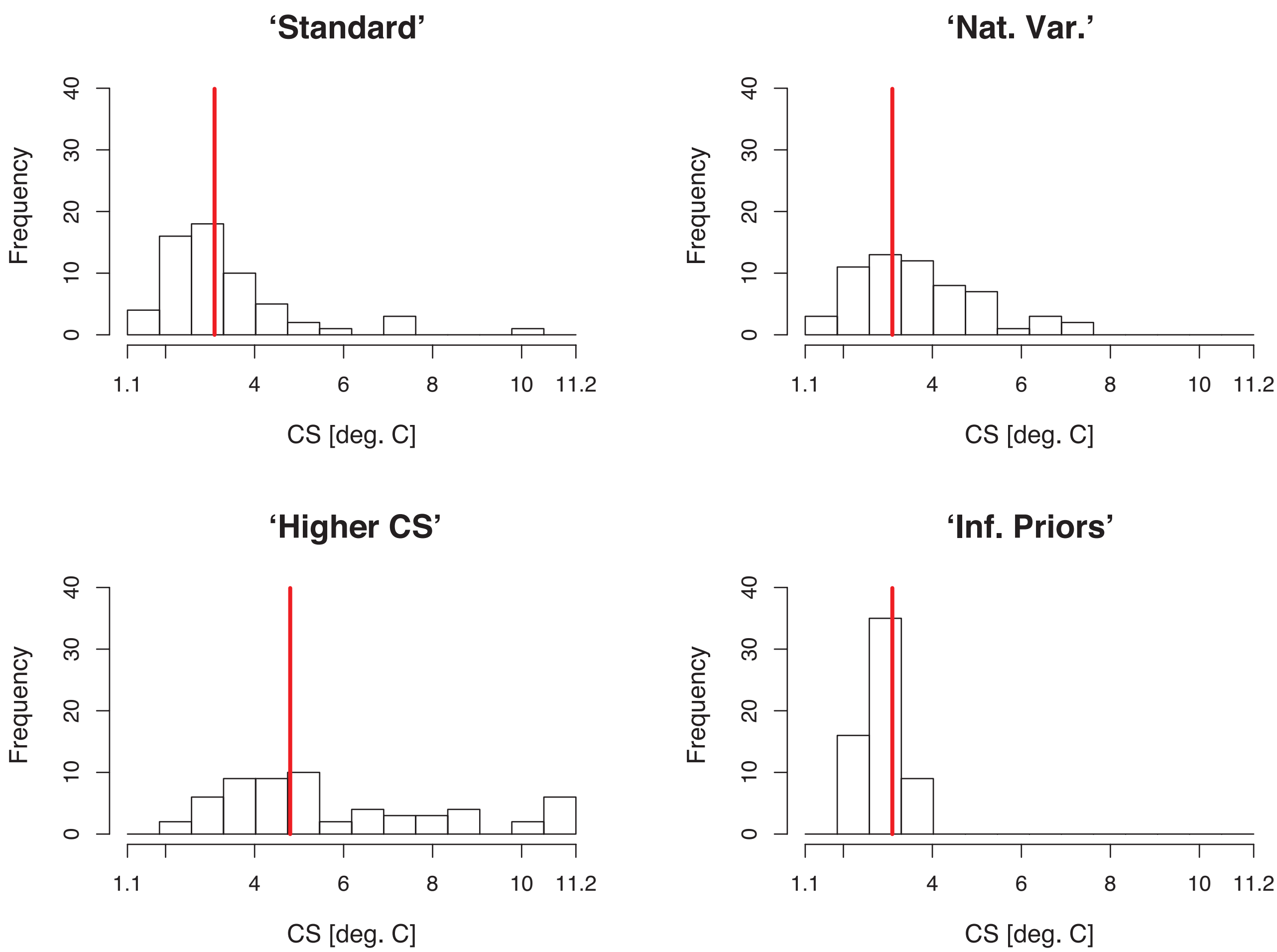

Figure 4: Histograms of the modes of the estimated climate sensitivity probability density functions: (top left) 'Standard', (top right) 'Nat. Var', (bottom left) 'Higher CS', and (bottom right) 'Inf. Priors'. 'True' input climate sensitivities are shown by vertical red lines. Y-axes limits are the same between panels.